Optimal Taxation with Time-Inconsistent Agents *

Pei Cheng Yu †

November 20, 2015

Job Market Paper
Please check http://pcyu.weebly.com/research.html for latest version

Abstract

People with time-inconsistent preferences tend to make intertemporal choices different from their original intentions. In particular, time-inconsistent agents make sub-optimal saving decisions. This paper studies the optimal savings and insurance policy in an economy with time-inconsistent agents who privately observe their skill. I introduce a mechanism that can elicit private information at zero cost by threatening sophisticated time-inconsistent agents with off-equilibrium path policies that can undo commitments, and fooling naïvely time-inconsistent agents with empty promises. The government can implement the full information efficient optimum, which is better than the constrained optimum obtained with traditional proposals to increase savings, like linear savings subsidies or mandatory savings rules. I introduce regressive savings subsidies for naïve agents and government loans with multiple repayment plans for sophisticated agents as policy instruments for decentralization. I show that welfare increases monotonically with the population of time-inconsistent agents. In essence, the presence of time-inconsistent agents improves the government's ability to provide insurance.

Keywords: Optimal taxation, Time-inconsistency, Retirement savings, Dynamic mechanism design, Non-common priors

JEL Classification Numbers: D03, D62, D82, D84, D86, D91, H21

*I thank Aldo Rustichini, Jan Werner, Manuel Amador, Martin Szydlowski, Erzo Luttmer, Kim-Sau Chung, Anmol Bhandari, V.V. Chari, Emmanuel Farhi, Xavier Gabaix, Simone Galperti, Mike Golosov, Radek Paluszynski and Hsin-Jung Yu for many helpful comments and suggestions. I am especially grateful to David Rahman for his guidance and endless encouragements. I also wish to acknowledge the support of the University of Minnesota Doctoral Dissertation Fellowship.

†University of Minnesota (e-mail: *yuxxx680@umn.edu*)

1 Introduction

Empirical evidence shows that time-inconsistent behaviors exhibited in the real world have substantial impact and are pervasive.[1] Skeptics argue that such behavioral bias may be attenuated once more important issues like retirement savings or investment portfolios are considered. However, evidence shows the contrary.[2] For example, for retirement savings, the National Research Council [(2012)] finds that up to $\frac{2}{3}$ of the US population is saving inadequately for retirement.[3] It is argued that people want to save more, but controlling ones' impulses is difficult.[4] Furthermore, research implies that models with time-inconsistent preference can explain the consumption and savings patterns observed in the data.[5] As a result, the literature argues for the government to implement policies that could offset the time-inconsistent behavior, for example, savings subsidies and mandatory savings policies have been suggested.[6] However, the recommended policies for time-inconsistent agents have so far been analyzed in isolation from other government objectives.

In this paper, I introduce new policy tools to increase savings when the government also wishes to insure agents against their productivity realizations. More specifically, this paper considers an environment with present-biased agents who privately observe their productivity. As a result, in addition to the efficiency and equity trade-off in a traditional Mirrlees environment, the government has the additional motive to increase retirement savings. I show that regressive savings subsidies and government provided loans can perform better than traditional policy proposals. The main contribution of this paper is to describe how the insurance problem changes with the introduction of time-inconsistent preferences, and

[1] DellaVigna and Malmendier [(2006)] study gym membership data and show that 80% of monthly gym members would have been better off had they chosen to pay per visit. Ausubel [(1999)] and Shui and Ausubel [(2005)] find similar biases in the credit card market. Gottlieb and Smetters [(2013)] also find similar evidence in the life insurance market. DellaVigna [(2009)] provides an overview of the empirical evidence for behavioral economics.

[2] Benartzi and Thaler [(2001)] find that individuals do poorly when it comes to investment diversification. Madrian and Shea [(2001)] study participation in the 401(k) plan and find evidence of strong default effects on the participation decision of individuals. O'Donoghue and Rabin [(2001)] show that a model of time-inconsistent individuals with naïveté can help explain the strong influence of the default option on retirement savings.

[3] Even 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.

[4] Choi, Laibson, Madrian and Metrick [(2002)] randomly surveyed employees of a large US corporation and found that 67.7% of the respondents felt that their saving rate is too low relative to their ideal saving rate, but only 4% of them took action to increase their saving rate over the next few months.

[5] Angeletos et al. [(2001)] and Laibson, Repetto and Tobacman [(2007)] examine the implications of quasi-hyperbolic discounting on life-cycle intertemporal decisions. They find that introducing quasi-hyperbolic discounting helps explain several patterns that cannot be explained by traditional exponential discounting models, like the co-existence of high credit card debt with a high demand for illiquid assets.

[6] Theoretical models with time-inconsistent preferences have been used to justify the introduction of such policies. See Section 1.1 for more details.

to characterize the optimal policy.

This paper also analyzes the problem for a full spectrum of sophistication levels. Agents who are fully aware of their time-inconsistency are *sophisticated*, while those who are completely unaware are *naïve*. Agents who are aware of their time-inconsistency but are wrong about the severity are called *partially naïve*. The distinction among different sophistication levels is important because the recommended policy varies with the sophistication level of the agents.

Traditional policy suggestions, such as linear savings subsidies or mandatory savings, can mitigate the agents' present bias. Such policy proposals offset the bias independent of the surrounding economic environment. Figure 1 shows how varying the linear savings subsidy rate can help the government attain the constrained efficient optimum when time-inconsistent agents are present. The government can set the savings subsidy rate to exactly offset the agents' present bias and proceed to implement the constrained efficient allocations. Hence, using traditional policy proposals, the government is able to guarantee the constrained efficient welfare level, but it cannot do better.

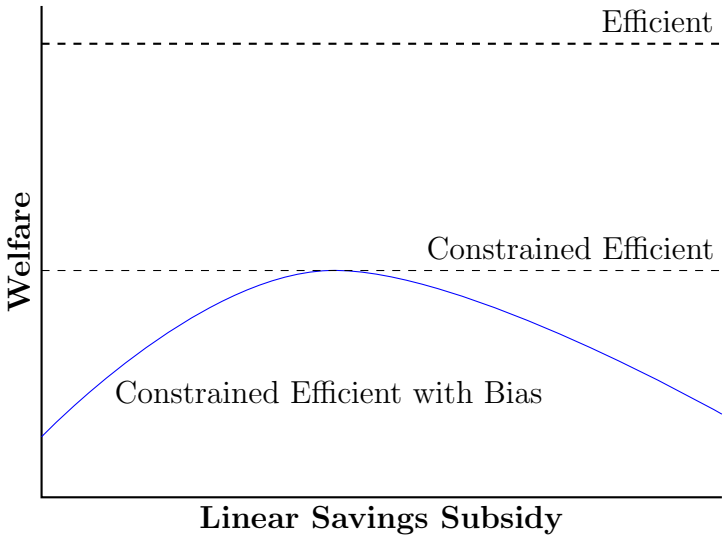

Figure 1: Welfare Comparison with Traditional Policy Instruments

I consider a set of more sophisticated policy instruments and find that the optimal set of allocations differs from the constrained efficient allocations. In particular, the government is able to fully insure all agents and avoid any labor distortions. More specifically, the main result shows that, with time-inconsistent agents, the government can implement the full information efficient allocation (Henceforth, referred to as efficient allocation) despite the presence of information asymmetry. Therefore, the government is able to do better than the constrained efficient optimum.

This surprising result is due to the fact that with naïve agents, the government induces efficient labor provision by promising a savings subsidy that increases with income. However, after preference reversal, the agents no longer value savings as much and would instead prefer the proportion of consumption and savings that corresponds to the efficient allocation. In other words, the government loads the information rent on savings to induce efficient output, which the agents do not value after preference reversal. The government *fools* the naïve agents into revealing their private information and does not need to deliver on the promise. Thus, the government can implement the full insurance policy without paying any information rent. Since partially naïve agents hold incorrect beliefs about their bias, deception is effective in implementing the efficient allocation as long as agents are non-sophisticated.

For sophisticated agents, the government exploits the fact that they know about their time-inconsistency and thus have a demand for commitment. The government will provide commitment only if the agents report truthfully. To deter misreports, the government designs a threat allocation that unravels the commitment such that a misreporting agent would be tempted to choose it, while an honest agent would not. This is possible due to differing marginal rates of substitution in labor and consumption for agents of different productivity types. Since credible threats can be constructed and the sophisticated agents can foresee the future consequences of lying, they report truthfully. In other words, efficiency is obtained using credible off-equilibrium path threats. Since partially naïve agents also have a demand for commitment, screening with credible threats also works for partially naïve agents.

In an environment where both sophisticated and naïve agents coexist, the government needs to screen the agents' productivity and their sophistication level. I show that mechanisms designed to fool more naïve agents and threaten more sophisticated ones can be combined without any efficiency loss. Therefore, the multidimensional screening problem does not alter the main result. The same method can be extended to an environment where time-inconsistent agents differ in the degree of present bias.

I also consider an economy where the government is uncertain whether a preference reversal would occur, in essence, some agents could be time-consistent. The environment with time-consistent agents limits the ability of the government to fool or threaten the agents. I find that welfare increases with the population of time-inconsistent agents. Figure 2 demonstrates this. This suggests that incentive compatibility causes distortions only in environments with time-consistent agents. This is because total output increases with the proportion of time-inconsistent agents. Hence, the government can provide more information rent per time-consistent agent without using more resources. This relaxes the incentive compatibility constraints and improves welfare. This also suggests that the estimate in Farhi

and Werning [(2012)] for the welfare benefits of adopting sophisticated intertemporal policies is a lower bound. The potential gains for adopting sophisticated intertemporal policies could be much larger depending on the population size of time-inconsistent individuals.

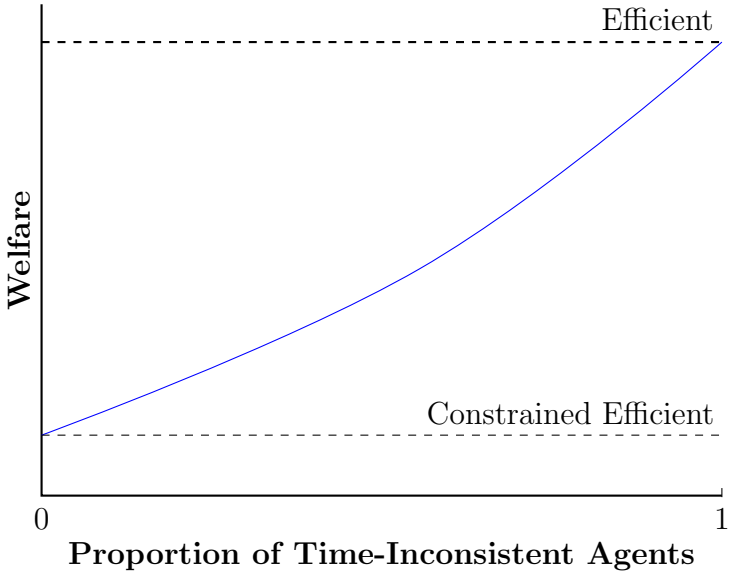

Figure 2: Welfare and Proportion of Time-Inconsistent Agents

The model with time-consistent agents also sheds light on whether the government should help the agents become more sophisticated. I compare the welfare of different sophistication levels and find that a higher sophistication level weakly improves welfare. In particular, the government would prefer fully naïve agents to be at least partially aware of their temptation. Meanwhile, the government is indifferent between other levels of sophistication.

For implementation, as was already mentioned, the government can use an income regressive non-linear savings subsidy to fool the non-sophisticated agents. For sophisticated agents, the government can implement the efficient allocation using loans with progressive repayment schemes. The loan acts as a commitment device and the repayment schemes are progressive so that a misreporting agent who earns more than initially planned would be punished with a reduction in retirement savings.

The results of this paper apply to a very general setting with agents who experience preference changes and may or may not be accurate in their predictions of these changes.[7] This demonstrates that the results may be applied more generally to different environments with dynamic inconsistencies and private information.[8]

[7] Appendix A provides a general model.

[8] Yu [(2015)] studies the optimal pricing of a monopoly firm facing time-inconsistent consumers.

1.1 Related Literature

This paper tries to combine two strands of literature: the optimal taxation literature and the behavioral contracting literature. Some papers have already attempted to do so.

Farhi and Gabaix [(2015)] studied optimal taxation with agents who suffer from a wide array of behavioral biases. They analyzed how the optimal tax formulas would be altered under Ramsey, Pigou and Mirrlees taxation. In contrast to their work, this paper attempts to focus on a certain behavioral bias, time-inconsistency, and to analyze the optimal policies from a mechanism design perspective. In particular, this paper adopts a different timing: agents are allowed to commit before making allocation decisions. In Farhi and Gabaix [(2015)], this ex-ante stage is absent.

Krusell, Kuruscu and Smith [(2010)] study the optimal taxation of consumers who suffer from temptation. They find that the government should subsidize savings to correct the agent's impatience and tendency to save too little. My work differs from Krusell, Kuruscu and Smith [(2010)] in two aspects. Firstly, their environment is a complete information one, while I introduce asymmetric information in productivity. Secondly, I also consider non-sophisticated agents, while their agents are sophisticated.

Amador, Werning and Angeletos [(2006)] examine government policies that could help agents with temptation. They study agents who suffer from temptation and are subject to future taste shocks. Hence, a preference for commitment and flexibility coexists, which causes a trade-off. They find that a minimum savings rule is optimal in this environment. Their work considers a sophisticated agent, and the adverse selection problem is in the agent's taste shock. The main difference is that my work seeks to analyze how government policies aimed at increasing savings of a time-inconsistent agent could affect labor decisions. Therefore, I have a production economy and the government has insurance motives, while Amador, Werning and Angeletos [(2006)] focus on an endowment economy with a government only trying to smooth consumption.

A few papers have studied the optimal taxation problem with hidden productivity and time-inconsistent agents. Bassi [(2010)] considers an environment where the quasi-hyperbolic discount factor is also non-observable, which creates a two-dimensional screening problem for the government. Guo and Krause [(2015)] study an environment where the government does not have full commitment. These two papers share a common goal with this one, but they employ traditional policy instruments to offset the present bias. In essence, they do not exploit the time-inconsistency of the agents and are unable to achieve maximum welfare.

Several papers have examined taxation models where individuals are differentiated along

two or more dimensions.[9] Most closely related to this paper in terms of the policy issue is Diamond and Spinnewijn [(2011)]. Their paper discusses a model with heterogeneity in both productivity and time preference (agents are time-consistent with different discount factors). This paper is also concerned with the policy on savings, but in addition to heterogeneous productivity, the agents also differ in their awareness of the underlying present bias and in their time-inconsistency. This type of multidimensional screening has not yet been analyzed in public policy or the economic literature.

The paper is also related to several behavioral contracting papers. In particular, Esteban and Miyagawa [(2005)] examine optimal pricing schemes with time-inconsistent agents and also find that distortions caused by information asymmetry can be averted for a given type of temptation: when agents are tempted to over-consume. Yu [(2015)] generalizes Esteban and Miyagawa [(2005)] to include non-sophistication and introduces a pricing scheme that achieves full rent extraction for all types of temptation, even when agents are tempted to under-consume. 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 also find that relatively sophisticated agents exert no informational externality on the relatively naïve agents. More recently, Galperti [(2015)] extends Amador, Werning and Angeletos [(2006)] to a sequential screening model where a mechanism designer first screens an agent's time-consistency and then the realized taste shock. This paper will consider an economy where the government screens both the agents' productivity and their time-consistency simultaneously.

The paper is organized as follows. Section 2 outlines the setup of a three-period model with present bias. Section 3 and Section 4 work out the results for the three-period model for non-sophisticated and sophisticated agents respectively. Section 5 examines the environment with hidden sophistication level and hidden time-inconsistency. Section 6 presents an analysis of a model where the government is not certain whether an agent is time-consistent or time-inconsistent. Section 7 provides methods for decentralizing the allocations. Section 8 considers extensions with dynamic stochastic productivity shocks and commitment versus flexibility preferences, and discusses some of the impediments to the implementation of the proposed mechanism. Section 9 concludes the paper. All proofs of results before Section 6 can be found in Appendix A, where I present a general three-period model. Proofs of results contained in Section 6 and beyond can be found in Appendix B.

[9] More notably, Cremer, Pestieau and Rochet [(2001)] examine a model where both the productivity level and endowments are not observed by the government. Cremer, Pestieau and Rochet [(2003)] extend the model to an overlapping generations setting and endogenize individual endowments as inherited wealth. Beaudry, Blackorby and Szalay [(2009)] examines an economy where agents could participate in both market and non-market production and have different unobservable productivity levels for both sectors.

2 The Model

This section describes the environment, the welfare criterion, and sets up the mechanism.

2.1 Setup of the Model

A continuum of agents of measure one live for three periods. In the first period, the agents realize their productivity and do not consume.[10] They produce and make consumption and savings decisions in the second period, and consume their savings in the third and final period. Agents are differentiated by their production efficiency θ. There are $|M| \geq 2$ types of agents denoted by the set $\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$, with $\theta_{m+1} > \theta_m$. The types are distributed according to $\Pr(\theta = \theta_m) = \pi_m > 0$, for all $\theta_m \in \Theta$ with $\sum_{m=1}^{M} \pi_m = 1$.

The production technology is linear and depends only on labor input l and the productivity of the agent: $y = \theta l$. In a competitive equilibrium, the wages are equated to the marginal productivity of labor. There is also a storage technology that transfers one unit of good in the first period to one unit of second period good. (Alternatively, they have access to a bond with interest rate 0.)

As is standard in Mirrlees taxation, both the production efficiency θ of each agent and their labor input l are not observed by the government. The government can only observe output y.

2.1.1 Agent Utility

The agents have the following utility before consumption

$$U(c, k, y; \theta) = u(c) - h\left(\frac{y}{\theta}\right) + w(k),$$

where c is consumption in the second period and k is savings for the final period. I will refer to U as the *ex-ante utility*. I also refer the incarnation of the agent with the ex-ante utility as the *planner*. This is the utility agents use to evaluate their consumption plans. The agents' utility changes when they are consuming to

$$V(c, k, y; \theta) = u(c) - h\left(\frac{y}{\theta}\right) + \beta w(k).$$

I will refer to V as the *ex-post utility*. I also refer the incarnation of the agent with the ex-post utility as the *doer*.[11] This utility models the tendency of a time-inconsistent agent

[10] The main results do not change if the agents require consumption in the first period.

[11] The terminologies 'planner' and 'doer' is derived from Thaler and Shefrin [(1981)], and has subsequently been widely used in the behavioral literature. In this paper, the word 'planner' would refer to the farsighted

to deviate from plans when confronted with an allocation decision.

The period utilities $u : \mathbb{R}_+ \mapsto \mathbb{R}$ and $w : \mathbb{R}_+ \mapsto \mathbb{R}$ are continuously differentiable, unbounded below and satisfy the usual strictly increasing and concavity assumptions: $u', -u'' > 0$ and $w', -w'' > 0$, while the dis-utility from labor $h : \mathbb{R}_+ \mapsto \mathbb{R}$ satisfies $h', h'' > 0$. Furthermore, $u', w' > \epsilon$ for some $\epsilon > 0$. This rules out the case where $u(c)$ or $w(k)$ asymptotically approaches a finite number when c or k approaches infinity, so $u(c)$ and $w(k)$ are unbounded above.

The Spence-Mirrlees single crossing property is automatically satisfied under the assumptions on the utility functions and the production function for both U and V. In other words, the marginal rates of substitution between consumption and output and between savings and output are smaller for more efficient agents: $\frac{\partial}{\partial\theta}\left(-\frac{\partial U}{\partial y}/\frac{\partial U}{\partial c}\right), \frac{\partial}{\partial\theta}\left(-\frac{\partial U}{\partial y}/\frac{\partial U}{\partial k}\right) < 0$ and $\frac{\partial}{\partial\theta}\left(-\frac{\partial V}{\partial y}/\frac{\partial V}{\partial c}\right), \frac{\partial}{\partial\theta}\left(-\frac{\partial V}{\partial y}/\frac{\partial V}{\partial k}\right) < 0$.

I will focus on the case with present bias, where $\beta < 1$.[12] A smaller β represents a stronger bias for present consumption. I will refer to β as measuring the degree of temptation the agents sufferer from. With $\beta \neq 1$, the marginal rate of substitution between consumption and savings ($MRS_{c,k}$) is different for U and V : $\frac{\partial U}{\partial c}/\frac{\partial U}{\partial k} \neq \frac{\partial V}{\partial c}/\frac{\partial V}{\partial k}$. Since both utilities are separable in consumption and labor, $MRS_{c,k}$ is independent of the agents' labor choice. Along with strictly increasing and concavity assumptions, the difference in $MRS_{c,k}$ between U and V implies a single crossing condition on the indifference curves for the ex-ante and ex-post utilities of c and k. I will refer to preferences that display this difference between ex-ante and ex-post utility as exhibiting *preference reversal*.[13] Preference reversal is key for allowing the government to implement policies that would seem attractive to the planner while remaining undesirable for the doer, or vice versa.

incarnation of the agent, and should not be confused with the government.

[12] The main results do not change if $\beta > 1$.

[13] In the current setup, the present bias is similar to a temptation shock that triggers immediate gratification and under-weighs the virtues of saving. This formulation has an equivalent quasi-hyperbolic discounting representation. The three period quasi-hyperbolic representation shown in Figure 3 is

$$\begin{aligned}
U_1(c,k,y;\theta) &= \hat{\beta}\delta\left[u(c) - h\left(\frac{y}{\theta}\right) + \delta w(k)\right],\\
U_2(c,k,y;\theta) &= u(c) - h\left(\frac{y}{\theta}\right) + \hat{\beta}\delta w(k),\\
U_3(k) &= w(k),
\end{aligned}$$

with $\delta = 1$. In the first period, the agents enroll in an income-specific savings plan (with several options available in the plan). In the second period, agents work and make consumption and savings decision according to the plan. Finally, in the third period, agents consume their retirement savings. If $\hat{\beta} = \beta$, then the agents are sophisticated and the transformed model is similar to Laibson (1997), and if not, then it is similar to the model with cognitive limitations as presented in O'Donoghue and Rabin (2001).

2.1.2 Timing

At date 0, the government designs the tax system. By the law of large numbers, the government knows the measure of each type of agent. At date 1, the agents' types are realized. At date 2, the agents report their types according to reporting strategy $\sigma : \Theta \mapsto \Theta$. They make decisions according to the ex-ante utility from date 0 to date 2. At date 3, just when the agents are making their consumption decisions, their preferences switch, and they make their consumption and labor decisions based on the ex-post utility. The timing of the model is shown in Figure 3.

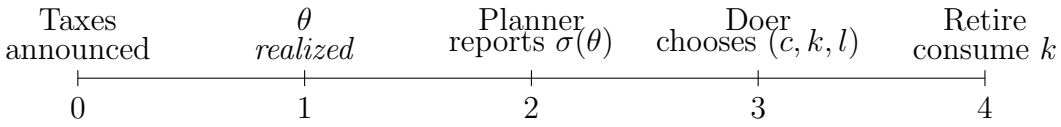

Figure 3: Timing of Events

I assume that the government has full commitment. Hence, the revelation principle can be applied. For ease of analysis, the current timing focuses on a direct mechanism, where the agents report their productivity types and the government assigns allocations according to the reports. For policy implementation, the timing can be dependent on the sophistication level of the agents.

2.1.3 Modeling Non-Sophistication

Following O'Donoghue and Rabin [(2001)], partially naïve agents perceive their degree of present bias to be $\hat{\beta} \in (\beta, 1]$. Let $W(c, k, y; \theta)$ denote the non-sophisticated agents' perceived ex-post utility:

$$W(c, k, y; \theta) = u(c) - h\left(\frac{y}{\theta}\right) + \hat{\beta} w(k).$$

Notice if $\hat{\beta} = 1$, then the planner is fully naïve and unaware of the doer's present bias. If $\hat{\beta} = \beta$, then the planner is sophisticated and aware of the doer's bias. Partially naïve agents know they have present bias, $\hat{\beta} < 1$, but their perceived present bias is always strictly greater than the actual present bias, $\hat{\beta} > \beta$. In other words, the planner underestimates the severity of the doer's temptation problem.[14] I will refer to the perception or belief $\hat{\beta}$ as describing the *sophistication level* of an agent.[15]

[14] The main results do not change if $\hat{\beta} < \beta$.

[15] There are two common ways to model partial naïveté. Loewenstein, O'Donoghue and Rabin [(2003)] and Heidhues and Koszegi [(2010)] have interpreted partial naïveté as the underestimation of the *magnitude* of temptation. Eliaz and Spiegler [(2006)] have interpreted partial naïveté as the underestimation of the *likelihood* of temptation occurring. Following Spiegler [(2011)], I will refer to the former as *magnitude*

2.1.4 Welfare Criterion

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

The government evaluates allocations at date 0 according to the following welfare criterion

$$\sum_{m=1}^{M} \pi_m \left[\kappa U\left(c_m, k_m, y_m; \theta_m\right) + (1-\kappa) V\left(c_m, k_m, y_m; \theta_m\right)\right], \tag{1}$$

where (c_m, k_m, y_m) denotes the allocation type θ_m agent consumes and $\kappa \in (0, 1]$ represents the welfare weight the government places on the planner. Since both U and V are strictly increasing and concave in (c, k), the government has a desire to insure agents against the realization of θ.

Much of the literature on dynamically inconsistent preferences have evaluated welfare solely with the ex-ante utility, or $\kappa = 1$. I adopt a Pareto criterion and allow the government to place positive welfare weight on the doer as well, as long as $\kappa \in (0, 1]$. Therefore, the government has a desire to smooth the agents' consumption across periods. Regardless of the perspective on the appropriate welfare weights, the main results of the paper are robust to changes in κ.[16]

2.1.5 The Availability of Insurance and Commitment

I assume that no private markets exist to insure against productivity shocks so the government has a role in insurance provision.[17] I also assume that there are no markets for illiquid assets or other commitment devices. If such a market exists, then sophisticated agents can use it for commitment. The government would have a limited role in helping the sophisticated agents smooth consumption. However, for non-sophisticated agents, the

naïveté and the latter as *frequency naïveté*. The paper focuses on magnitude naïveté and address *frequency naïveté* along with the general model in Appendix A. The main results go though for both types of non-sophistication.

[16] The welfare weight on the planner is strictly positive because I adopt the perspective that the ex-ante utility reflects the agents' long-term planning, while the ex-post utility reflects the agents' short-term temptations. In other words, the ex-post utility is not immune to regret and a benevolent government would consider the adverse implications if the agents give in to their urges. As a result, the choice of the welfare criteria is non-arbitrary, since the actions undertaken by the doer can be regarded as a systematic mistake the agents make, as in Bernheim and Rangel [(2004)].

[17] Prescott and Townsend [(1984)] show that the presence of an efficient market that allows agents to insure against future productivity shocks can make distortionary taxes redundant.

presence of such markets does not preclude the need for government intervention.[18] In Section 8, I discuss how the presence of a market for commitment can affect the results of this paper.

2.2 The Benchmark

2.2.1 No Private Information

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

$$\sum_{m=1}^{M} \pi_m \left(\theta_m l_m - c_m - k_m\right) - G = 0, \tag{2}$$

where G is government's external revenue needs. For the remainder of the paper, $G = 0$, so the sole role of the government is to provide insurance and smooth consumption.

With complete information, the government can achieve full insurance regardless of the agents' time-inconsistency or their degree of naïveté. This is because with complete information, the agents work according to their productivity type. The government then chooses an appropriate linear tax to correct the distortion caused by the present bias. The optimal allocation of the government's problem without private information is referred to as the efficient allocation.

Let (c_m^*, k_m^*, y_m^*) denote the efficient allocation for type θ_m. The efficient consumption satisfies $\forall \theta_m \in \Theta$,

$$\frac{\partial \left[\kappa U + (1-\kappa)V\right]}{\partial c_m^*} = \frac{\partial \left[\kappa U + (1-\kappa)V\right]}{\partial k_m^*}.$$

Since $U \neq V$ and $\kappa > 0$, the government can choose linear taxes or subsidies τ^c and τ^k on c and k respectively such that $\frac{1}{1+\tau^c}\frac{\partial V}{\partial c_m^*} = \frac{1}{1+\tau^k}\frac{\partial V}{\partial k_m^*}$, which implements the efficient consumption. Therefore, if the agents save too little, the government can subsidize savings k or tax c. The following proposition characterizes the properties of the efficient allocation.

Proposition 1 *The efficient allocation* $\{(c_m^*, k_m^*, l_m^*)\}_{\theta_m \in \Theta}$ *satisfies the following:*

1. *Full insurance: For any* $\theta_m, \theta_{m'} \in \Theta$, $c_m^* = c_{m'}^*$ *and* $k_m^* = k_{m'}^*$.
2. *Consumption smoothing: For any* $\theta_m \in \Theta$, $u'(c_m^*) = \left[\kappa + (1-\kappa)\beta\right] w'(k_m^*)$.
3. *Efficient output: For any* $\theta_m \in \Theta$, $u'(c_m^*) = \frac{1}{\theta_m} h'(l_m^*)$.

[18] Heidhues and Koszegi (2009) demonstrate how commitment devices can do more harm than good for non-sophisticated agents, since they pay the cost of commitment but still suffer from self-control problems.

Consumption smoothing is achieved since the government puts a strictly positive weight κ on the ex-ante utility. As a special case, if $\kappa = 1$, then I will refer to the efficient allocation as attaining *perfect consumption smoothing*: $u'(c_m^*) = w'(k_m^*)$.

There are three important features of the policy instrument in this environment. First, the taxes or subsidies can be linear. Second, these linear instruments are the same across all productivity types. This is because all productivity types suffer from the same degree of temptation. Finally, the linear taxes or subsidies also work independently of sophistication levels. This is because regardless of the sophistication level, the policy enacted to correct for the taste change does not distort the incentives to work. The government can always 'force' the productive agents to work more than the less productive agents.

2.2.2 Private Information without Time-Inconsistency

For the case 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, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$u\left(c_m\right) - h\left(\frac{y_m}{\theta_m}\right) + w\left(k_m\right) \geq u\left(c_{\hat{m}}\right) - h\left(\frac{y_{\hat{m}}}{\theta_m}\right) + w\left(k_{\hat{m}}\right). \tag{3}$$

The efficient allocations are not incentive compatible, so with private information and time-consistent agents, the government can only implement the constrained efficient allocation. The following proposition characterizes the properties of the constrained efficient allocation.

Proposition 2 *The constrained efficient allocation* $\{(c_m^{**}, k_m^{**}, l_m^{**})\}_{\theta_m \in \Theta}$ *satisfies the following:*

1. *Partial insurance: For any* $\theta_m, \theta_{m'} \in \Theta$, *with* $\theta_m > \theta_{m'}$, $c_m^{**} > c_{m'}^{**}$ *and* $k_m^{**} > k_{m'}^{**}$.
2. *Consumption smoothing: For any* $\theta_m \in \Theta$, $u'(c_m^{**}) = [\kappa + (1-\kappa)\beta]\, w'(k_m^{**})$.
3. *Output distortions: For any* $\theta_m < \theta_M$, $u'(c_m^{**}) > \frac{1}{\theta_m} h'(l_m^{**})$.

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

With sophisticated time-inconsistent agents, if the government uses a linear savings subsidy to correct the present bias, the constrained efficient optimum can still be obtained, as shown in Figure 1. This is because the linear savings subsidy would correct the doer's present

bias without changing the set of incentive compatible allocations. Hence, the planner would face the same incentive compatibility constraints and the government can implement the constrained efficient allocations. In the following subsection, I will discuss a more general mechanism which would allow for a richer set of policy instruments.

2.3 Incentive Compatibility

By Yu [(2015)], the revelation principle applies to this setting, so the analysis will focus on a truth-telling direct mechanism. The government presents a menu of allocations C_m for type θ_m agents defined as

$$C_m = \{(c_m, k_m, y_m), (c'_m, k'_m, y'_m), \ldots\}.$$

An agent could be assigned a menu C_m after a report $\sigma(\theta) = \theta_m$ as opposed to an allocation. Previous literature has also highlighted the benefits of using enlarged menus in mechanism design problems with time-inconsistent agents.[19] Let $\left(c_m^R, k_m^R, y_m^R\right) \in C_m$ denote the *real allocation*, which is the optimal allocation the government can implement. Following the timing in Figure 3, the government posts $C = \{C_1, \ldots, C_M\}$. The agents then proceed to choose a menu from C after learning θ. After temptation sets in, the agents choose an allocation from the menu they initially selected.

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

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

Hence, $C_m^{\hat{\beta}}$ denotes the set of allocations a truthful θ_m agent with sophistication level $\hat{\beta}$ predicts the doer would choose. Let

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

Hence, $C_{\hat{m}|m}^{\hat{\beta}}$ denotes the set of allocations a type θ_m agent with sophistication level $\hat{\beta}$ predicts the doer would choose after he misreports to be a type $\theta_{\hat{m}}$ agent. Incentive compatibility is

[19] Enlarged menus have been used to exploit time-inconsistent agents in the literature. For instance, Esteban and Miyagawa [(2005)] showed how enlarging menus could help monopolists achieve perfect price discrimination for a particular type of temptation even when consumer valuations are unobservable. Yu [(2015)] extended their analysis to all temptations. Galperti [(2015)] used large menus to help separate time-inconsistent agents from the time-consistent agents.

thus expressed as, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

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

The incentive compatibility constraints (4) make sure the agents choose the menu that is intended for their productivity type given their sophistication level $\hat{\beta}$. Additional constraints are needed to make sure the real allocations are implemented. The executability constraints are, $\forall \theta_m \in \Theta$,

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

The executability constraints (5) make sure that the doer would choose the real allocations. In a model with non-common priors, the direct revelation mechanism has to take into account the agents' beliefs through the incentive compatibility constraints and the government's beliefs through the executability constraints.

All other allocations besides the real allocations are considered off-equilibrium path. However, depending on the sophistication level of the agents, the presence of other allocations in the menu can affect the reporting strategy σ.

3 Savings with Non-sophisticated Agents

For this section, I will first describe the method the government uses to elicit truth-telling and setup the optimization problem for $\hat{\beta} \in (\beta, 1]$. I will then introduce the first main result which characterizes the optimal allocation for non-sophisticated agents.

3.1 The Fooling Mechanism: Real and Imaginary Allocations

Non-sophisticated agents hold erroneous beliefs about their level of temptation, and hence make incorrect predictions about their behavior. This means that their reporting strategies reflect the expected choices of a fictitious doer, and not the real doer. The government can exploit this misspecified belief by using a *fooling mechanism*.

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

After the menu is announced, non-sophisticated agents of any type θ_m 'mentally' choose a set of allocations (c_m^I, k_m^I, y_m^I). However, the government intends the agents to 'actually'

choose allocation (c_m^R, k_m^R, y_m^R) to maximize the ex-post utility. The superscript I represents 'imaginary,' since it is never actually selected, but were the perceived choices of the planner before preference reversal. I will set $y_m^R = y_m^I = y_m$, and show that $(c_m^R, k_m^R) \neq (c_m^I, k_m^I)$ is enough to exploit non-sophistication.[20]

The planners of non-sophisticated agents underestimate the degree of temptation, so the design of imaginary allocations is important. Planners make their reporting decisions based on $U(c, l)$ while anticipating an ex-post utility of $W(c, l)$. Therefore, to fool the non-sophisticated agents, the government requires the imaginary allocations to be more appealing than the real allocations under $W(c, l)$, and $\sigma(\theta) = \theta$. However, to implement the real allocation, it is required to be more appealing than the imaginary allocation for the actual doer. The analysis in this section will focus on a truth-telling direct revelation mechanism. The following definition defines truthful implementation in a direct fooling mechanism.

Definition 2 *An allocation* $\left\{\left(c_m^R, k_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ *is truthfully implementable by a direct fooling mechanism if there exists* $\left\{\left(c_m^I, k_m^I, y_m^I\right)\right\}_{\theta_m \in \Theta}$ *such that the*

i. incentive compatibility constraints, and

ii. executability constraints

are satisfied.

The government's problem is to choose the menu of allocations, C, to maximize (1) subject to the feasibility constraint (2) evaluated at the real allocations and the incentive compatibility constraints (4), executability constraints (5) and the fooling constraints. The imaginary allocations are not required to satisfy the feasibility constraint. This is because the government only cares about the real allocations, and views the imaginary allocations as an empty promise. The government is certain about the degree of the naïveté and present bias of the agents, so it places no weight on a future where it may need to actually honor the delivery of imaginary allocations. Another concern is that the agents may realize that the aggregate imaginary allocation violates the feasibility constraint and doubt the validity of the government's promise. However, each agent is infinitesimally small, and even though an agent believes he would consume the imaginary allocation, he does not consider the belief and behavior of other agents.

[20]The use of imaginary allocations for exploiting non-sophisticates has been explored in the literature before, for instance, Eliaz and Spiegler [(2006)] and Heidhues and Koszegi [(2010)]. However, it is to the best of this author's knowledge that this paper is the first to use it to elicit private information without cost.

3.2 Main Result for Non-Sophisticated Agents

It can be shown that the government achieves the efficient allocation by using a fooling mechanism. In other words, surprisingly, private information does not matter in an environment where all agents harbor some naiveté.

Theorem 1 *If $\hat{\beta} \in (\beta, 1]$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct fooling mechanism.*

Theorem 1 states that the private information problem can be alleviated if the agents are non-sophisticated. With the imaginary allocations, the government is able to provide the information rents necessary for truth-telling. However, these rents are fictional. After preference reversal, the government implements the efficient allocation without paying information rents. Indeed, it follows that it is optimal for the government to deceive the agents when they are not sophisticated.

Corollary 1 *If $\hat{\beta} \in (\beta, 1]$, it is optimal for the government to implement a fooling mechanism.*

The key to deceiving the agents is to load the rents on savings, $k^I \geq k^R$ and $c^R \geq c^I$, which the planner values relatively more during the reporting stage, but the doer would not value as much. The government can promise a higher return on savings for the imaginary allocations to elicit truthful reports as long as the agents hold the wrong beliefs on their temptation level. In essence, non-sophisticated agents are willing to trade information rents for empty promises.

It is interesting to note that there is a discontinuity in the optimal welfare with respect to the sophistication level of the agents. The government achieves the efficient welfare level for any sophistication level $\hat{\beta} \in (\beta, 1]$. However, with fully sophisticated agents, the best the fooling mechanism can implement is the constrained efficient allocation which requires information rent for the productive types. This is because the sophisticated agents would perfectly foresee their doers' present bias at the consumption stage and would thus be immune to any deceptions at the reporting stage. As a result, there is a discontinuity in welfare which is similar to the discontinuity in Heidhues and Koszegi [(2010)].

I will show that the discontinuity is only present if the government restricts attention to a fooling mechanism. For sophisticated agents, the government can also take advantage of their time-inconsistency and implement the efficient optimum. In the next section, I will present such a mechanism.

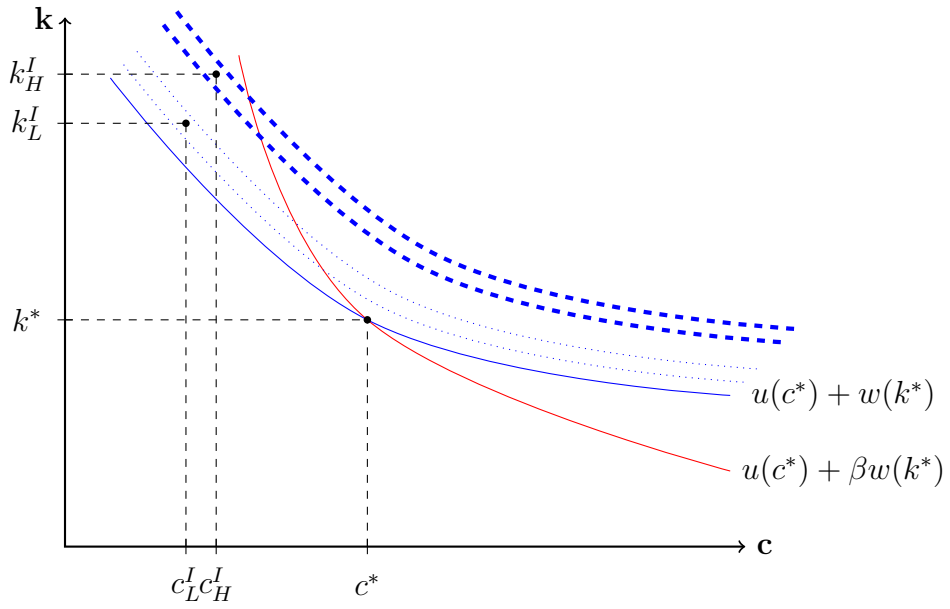

Figure 4: Finding the Imaginary Allocations

3.2.1 An Illustration of the Fooling Mechanism

To demonstrate how the fooling mechanism works, consider an economy with two productivity types $\Theta = \{\theta_L, \theta_H\}$, where $\theta_H > \theta_L$, and let $\kappa = 1$, so the government maximizes the sum of ex-ante utility and wants to achieve perfect consumption smoothing. By Theorem 1, $\left\{\left(c_m^R, k_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ will be the efficient allocation, where $c_H^R = c_L^R = c^*$ and $k_H^R = k_L^R = k^*$ and $y_m^R = y_m^*$ for all $\theta_m \in \Theta$. The efficient allocation satisfies Proposition 1, so $y_H^* > y_L^*$. For simplicity, I will examine the fully naïve case: $\hat{\beta} = 1$.

In Figure 4, the flatter solid (blue) curve represents the indifference curve of the ex-ante utility at allocation (c^*, k^*). The steeper solid (red) curve represents the indifference curve of the ex-post utility at allocation (c^*, k^*). The imaginary allocations have to be in the area bounded by the solid line indifference curves in the north-west region.[21] Any allocation within this area satisfies the fooling constraint and executability constraint. Furthermore, the incentive compatibility constraints provide an upper and lower bounds to the difference in ex-ante utility of the two types of agents. In essence,

$$h\left(\frac{y_H^*}{\theta_L}\right) - h\left(\frac{y_L^*}{\theta_L}\right) \geq \left[u(c_H^I) + w(k_H^I)\right] - \left[u(c_L^I) + w(k_L^I)\right] \geq h\left(\frac{y_H^*}{\theta_H}\right) - h\left(\frac{y_L^*}{\theta_H}\right).$$

Therefore, given y_m^*, the imaginary allocations have to be within the dashed indifference

[21] If $\beta > 1$, then the imaginary allocations would be bounded by the solid line indifference curves in the south-east region.

curves, where the high productivity type's imaginary allocation is within the bold dashed area, and the low productivity type's is within the light dotted area.

4 Savings with Sophisticated Agents

The fooling mechanism introduced for non-sophisticated agents no longer work for sophisticated agents. For sophisticated agents, the government can instead design an off-equilibrium path threat that will be chosen only if an agent misreports. This threat helps attain the efficient allocation.[22]

4.1 The Threat Mechanism: Real and Threat Allocations

When sophisticated agents choose reporting strategy σ, they know that their preferences will change and their doer will be tempted to consume a set of allocations that is inconsistent with the ex-ante utility U. The government can take advantage of the agents' time-inconsistency and their awareness by using a *threat mechanism*.

Definition 3 *A direct threat mechanism for sophisticated agents* ($\hat{\beta} = \beta$) *has* $C = \{C_m\}_{\theta_m \in \Theta}$ *with* $C_m = \left\{\left(c_m^R, k_m^R, y_m^R\right), \left(c_m^T, k_m^T, y_m^T\right)\right\}$ *satisfying the threat constraints:* $\left(c_{\hat{m}}^T, k_{\hat{m}}^T, y_{\hat{m}}^T\right) \in C_{\hat{m}|m}^{\beta}, \forall \theta_m, \theta_{\hat{m}} \in \Theta$.

I will refer to $\left(c_m^T, k_m^T, y_m^T\right)$ as the *threat allocation*. The analysis will focus on a truth-telling direct mechanism. Sophisticated agents have the correct belief about their temptation, so the executability constraints imply that truth-telling is evaluated at the real allocation and the threat constraints imply that misreports are evaluated at the threat allocations. The following definition defines truthful implementation in a direct threat mechanism.

Definition 4 *An allocation* $\left\{\left(c_m^R, k_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ *is truthfully implementable by a direct threat mechanism if there exists* $\left\{\left(c_m^T, k_m^T, y_m^T\right)\right\}_{\theta_m \in \Theta}$ *such that the*

i. incentive compatibility constraints, and

ii. executability constraints

are satisfied

[22] Adverse selection models with sophisticated agents have been explored in Esteban and Miyagawa [(2005)], Galperti [(2015)] and Chung [(2015)].

The threat allocations, $\left(c_m^T, k_m^T, y_m^T\right)$, for type θ_m are designed such that, after preference reversal, a type θ_m planner who reports truthfully would never choose it (by the executability constraint), but a planner who misreports as type θ_m would (by the threat constraint). The superscript T represents 'threat,' because the misreporting planner would consider this allocation to be inferior according to his ex-ante preferences. Adding the threat allocation to the menu deters the agents from misreporting. When agents are sophisticated, a threat allocation that satisfies the threat and executability constraints will be referred to as satisfying the *credible threat constraints*.[23]

It is important to note that threats are non-credible if $y_m^R = y_m^T$. This is because all agents share the same preference for goods consumption, so a misreporting agent would make the same consumption choice as a truthful agent if $y_m^R = y_m^T$. Hence, agents will not be deterred from misreporting if $y_m^R = y_m^T$. As a result, unlike a fooling mechanism, the real allocations and threat allocations have to be different in both consumption and output.

The government's problem is to choose the menu of allocations, C, to maximize (1) subject to the feasibility constraint (2) evaluated at the real allocations and the incentive compatibility constraints (4) and the credible threat constraints.

4.2 Main Result for Sophisticated Agents

The following theorem shows that private information does not matter in an environment with sophisticated agents as well. When the government uses a threat mechanism, the efficient allocation is achievable.

Theorem 2 *If $\hat{\beta} = \beta$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct threat mechanism.*

Theorem 2 holds because sophisticated agents are aware that their doers would distort their savings plan, and would desire a commitment device that deters them from doing so. The government provides this commitment device only if the sophisticated agents report truthfully. If not, the threat allocation caters to the temptations of the doers. In essence, the government holds the agents' doers hostage and threatens to distort the savings plan unless agents report truthfully. This threat also helps screening because the threat allocations can be designed such that it separates the productivity of the agents by using the Spence-Mirrlees single crossing property. In essence, sophisticated agents trade information rents for commitment.

[23] The government has full commitment, so the credible threat constraints are not to ensure subgame perfection. They merely address the need for the government to find threats that deter misreports and to avoid truthful agents from choosing the threats, so the government is credibly benevolent.

4.2.1 An Illustration of the Threat Mechanism

To demonstrate Theorem 2, consider the setting with two productivity types and $\kappa = 1$ investigated previously. In a Mirrlees setting, the productive agent has an incentive to pretend to be a less productive agent to decrease labor supply while enjoying the gains of insurance. In a relaxed problem, the threat is targeted at the productive types to discourage them from pretending to be the less productive. Hence, the government designs the threat allocation $\left(c_L^T, k_L^T, y_L^T\right) \in C_L$. I will focus on the downward incentive compatibility constraint.

By Theorem 2, $\left\{\left(c_m^R, k_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ will be the efficient allocation, so $c_H^R = c_L^R = c^*$ and $k_H^R = k_L^R = k^*$ and $y_m^R = y_m^*$ for all $\theta_m \in \Theta$. Let $\Phi_{j,k}^i = u(c_j^i) - h\left(y_j^i/\theta_k\right)$ and $\Phi_{k,k}^i = \Phi_k^i$, where $i \in \{R, T\}$ and $j, k \in \{L, H\}$. From incentive compatibility and credible threat constraints and from the fact that $\theta_H > \theta_L$ and $\beta < 1$, the efficient and threat allocations have to satisfy

$$\Phi_{L,H}^T > \Phi_{L,H}^R > \Phi_L^R > \Phi_H^R, \text{and } k^* > k_L^T.$$

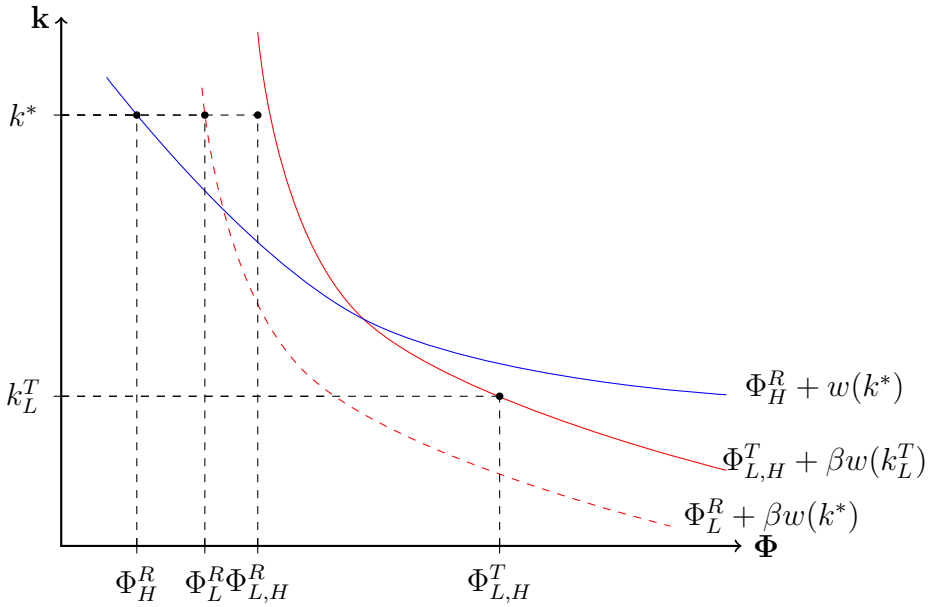

Figure 5: Finding the Threat Allocation: Part I

Figure 5 shows how the incentive compatibility constraint restricts the set of threat allocations. The steeper solid (red) curve represents the indifference curve of the ex-post utility for the θ_H agent who pretends to be θ_L. The flatter solid (blue) curve represents the indifference curve of the ex-ante utility for the θ_H agent who reports truthfully. The dashed (red) curve represents the indifference curve of the ex-post utility for the truthful θ_L agent. Figure 5 shows that the government can choose $\left(c_L^T, k_L^T, y_L^T\right)$ such that the incentive

compatibility constraint is satisfied by decreasing k_L^T and increasing $\Phi_{L,H}^T$.

By the Spence-Mirrlees single crossing property, the government can increase $\Phi_{L,H}^T$ without violating the credible threat constraints. To see this, fix the choice of k_L^T and $\Phi_{L,H}^T$ at the level shown in Figure 5. If the threat satisfies incentive compatibility and credible threat constraints, it implies

$$\Delta w \geq \Phi_{L,H}^T - \Phi_H^R > \Phi_{L,H}^T - \Phi_{L,H}^R \geq \beta \Delta w \geq \Phi_L^T - \Phi_L^R,$$

where $\Delta w \equiv \left[w(k^*) - w(k_L^T)\right]$. The problem now is to find c_L^T and y_L^T such that $u(c_L^T) - h\left(\frac{y_L^T}{\theta_H}\right) = \Phi_{L,H}^T$ and satisfies the executability constraint, $\beta \Delta w \geq \Phi_L^T - \Phi_L^R$. Figure 6 shows how this could be done.

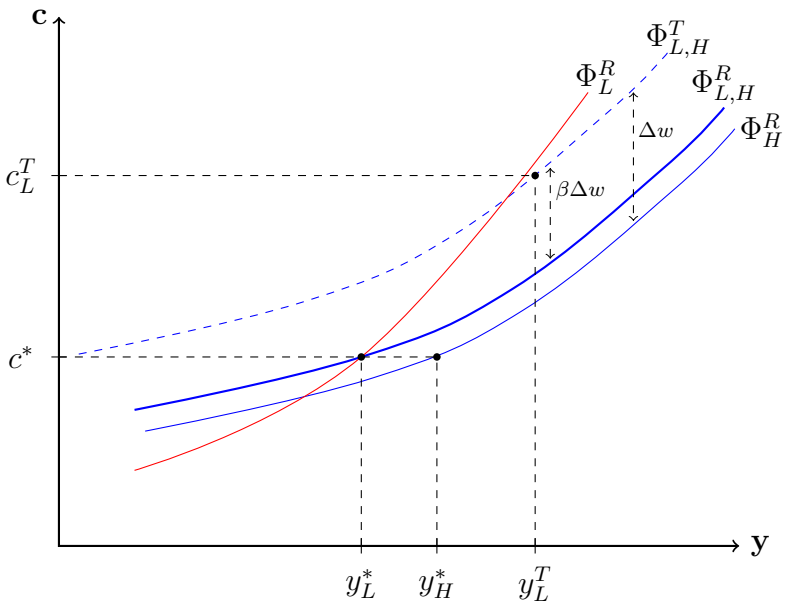

Figure 6: Finding the Threat Allocation: Part II

In Figure 6, the flatter thick solid (blue) curve represents the indifference curve of Φ for the θ_H agents at allocation (c^*, y_L^*). The steeper solid (red) curve represents the indifference curve of Φ for the θ_L agents at allocation (c^*, y_L^*). The dashed (blue) curve represents the indifference curve of Φ for the θ_H agent at allocation $\left(c_L^T, y_L^T\right)$, chosen so that $u(c_L^T) - h\left(\frac{y_L^T}{\theta_H}\right) = \Phi_{L,H}^T$ and the executability constraint holds.

The key to making a threat credible and incentive compatible is to decrease savings and increase present utility from choosing the threat allocation. The doer has a present bias, so a threat allocation would be more preferred than the real allocation. To make sure that only a misreporting agent is punished, the government exploits the fact that the more productive agents have a lower marginal rate of substitution in consumption and labor than

less productive agents. Hence, the threat allocation can always be constructed such that the compensation in consumption for the output is too low for the low productivity agents, but sufficient for the high productivity agents.

4.3 The Threat Mechanism for Partially Naïve Agents

Fully naïve agents have to be fooled, since they do not respond to threats. While sophisticated agents have to be threatened, because they can never be deceived. Since partially naïve agents have demand for commitment devices too, they are also susceptible to threats. The following definition defines a direct threat mechanism for agents with sophistication level $\hat{\beta} \in (\beta, 1)$ and truthful implementation in this environment.

Definition 5 *A direct threat mechanism has a menu* $C = \{C_m\}_{\theta_m \in \Theta}$ *with* $C_m = \left\{\left(c_m^R, k_m^R, y_m^R\right), \left(c_m^T, k_m^T, y_m^T\right)\right\}$ *satisfying credible threat constraints:* $\left(c_m^R, k_m^R, y_m^R\right) \in C_m^{\hat{\beta}}$ *and* $\left(c_{\hat{m}}^T, k_{\hat{m}}^T, y_{\hat{m}}^T\right) \in C_{\hat{m}|m}^{\hat{\beta}}$, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$. *The real allocation* $\left\{\left(c_m^R, k_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ *is truthfully implementable by a direct threat mechanism if there exists* $\left\{\left(c_m^T, k_m^T, y_m^T\right)\right\}_{\theta_m \in \Theta}$ *such that the incentive compatibility constraints (4) and the executability constraints (5) are satisfied.*

For partial naïveté with sophistication level $\hat{\beta} \in (\beta, 1)$, the threats are evaluated using the erroneous $\hat{\beta}$. Therefore, the credible threat constraints are defined to make sure the planner perceives his doer choosing the real allocation when report is truthful and the threat allocation if he misreported.

The government also needs to make sure that after preference reversal, the truthful agents will indeed choose the real allocation, so the executability constraint is needed. In contrast to Definition 3, the constraint, $\left(c_m^R, k_m^R, y_m^R\right) \in C_m^{\hat{\beta}}$, is redundant when agents are sophisticated. The following corollary shows partially naïve agents can be screened using a threat mechanism and Table 1 summarizes the applicability of the mechanisms to each sophistication level.

Corollary 2 *It is optimal to threaten the agents when* $\hat{\beta} \in [\beta, 1)$.

	Fully Naïve	Partially Naïve	Sophisticated
Fooling	✓	✓	✗
Threat	✗	✓	✓

Table 1: Summary of Mechanism for Different Sophistication Levels

5 Diversely Time-Inconsistent Agents

The previous model had agents differing in their productivity while sharing the same degree of temptation and sophistication level. In this section, I consider a setting where the government faces time-inconsistent agents who differ in their degree of temptation and sophistication level, which are both hidden from the government. I will first examine a model with hidden levels of sophistication, and then introduce hidden degrees of temptation.

5.1 Hidden Sophistication

In this economy, agents share the same temptation β but have varying sophistication levels. The hidden type of each agent is indexed by $(\theta_m, \hat{\beta})$, so the optimal policy has to solve a multidimensional screening problem.

The sophistication level of agents is distributed within the bounded support of $[\beta, 1]$. Let $\Pi(\theta_m, \hat{\beta})$ denote the joint distribution of productivity and sophistication level. I will refer to mechanisms that attain the efficient allocation as *effective*.

Lemma 1 *A fooling mechanism that is effective for agents of sophistication level $\acute{\beta} \in (\beta, 1)$ is also effective for all agents with $\hat{\beta} \geq \acute{\beta}$. A threat mechanism that is effective for agents of sophistication level $\grave{\beta} \in (\beta, 1)$ is also effective for all agents with $\hat{\beta} \leq \grave{\beta}$.*

Lemma 1 shows that a fooling mechanism designed for sophistication level $\acute{\beta}$ agents can also fool agents who are more naïve. While a threat mechanism for sophistication level $\grave{\beta}$ agents, can also credibly threaten agents who are more sophisticated. This is because providing incentives for the least naïve agents for truth-telling is the most difficult, so any incentives that could separate the productivity of the least naïve agents will also induce truth-telling for more naïve agents. Similarly, a threat mechanism for $\grave{\beta}$ will work for any $\hat{\beta} < \grave{\beta}$, since the less sophisticated agents need a stronger threat to be willing to divulge the truth, so the threat would also work for more sophisticated agents.

With Lemma 1, the government can choose an arbitrary *target sophistication level*, $\tilde{\beta}$, such that all agents who are more sophisticated than $\tilde{\beta}$ are threatened by using the same threat mechanism and those who are more naïve than $\tilde{\beta}$ are fooled by using the same fooling mechanism. I will henceforth refer to the agents with sophistication level $\hat{\beta} \in (\tilde{\beta}, 1]$ as *relatively naïve* and agents with sophistication level $\hat{\beta} \in [\beta, \tilde{\beta})$ as *relatively sophisticated*. Hence, a fooling mechanism designed for agents with sophistication level $\tilde{\beta}$ is applied to relatively naïve agents and a threat mechanism designed for agents with sophistication level $\tilde{\beta}$ is applied to relatively sophisticated agents.

The only concern is whether implementing a fooling mechanism for the relatively naïve would affect the effectiveness of the threat mechanism for the relatively sophisticated. The government can choose a fixed target sophistication level at $\tilde{\beta} \in (\beta, 1)$ and introduce the menu $C = \{C_m\}_{\theta_m \in \Theta}$, with

$$C_m = \left\{ \left(c_m^R, k_m^R, y_m^R\right), \left(c_m^I, k_m^I, y_m^I\right), \left(c_m^T, k_m^T, y_m^T\right) \right\}.$$

The imaginary and threat allocations are chosen such that agents with sophistication level $\tilde{\beta}$ are fooled and threatened with effective mechanisms. I will refer to this mechanism as a *hybrid* mechanism. The following theorem shows that the efficient allocation is attainable in this environment.

Theorem 3 *For the environment with hidden sophistication, the efficient allocation* $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ *is truthfully implementable with a hybrid mechanism.*

Theorem 3 follows immediately from the fact that relatively naïve and relatively sophisticated agents can be costlessly separated using a hybrid mechanism. This is because the relatively naïve agents focus on the imaginary allocations for truth-telling as long as the imaginary benefits from truth-telling are sufficiently appealing, while the relatively sophisticated agents focus on the threat allocations for misreporting. Therefore, the fooling and threat mechanisms do not interact when they are integrated, so the government can consider the relatively naïve agents independently of the relatively sophisticated agents. Along with Lemma 1, the government can choose any arbitrary target sophistication level $\tilde{\beta}$, so the population that is being fooled and threatened can be arbitrary.

Another feature of Lemma 1 and Theorem 3 is that it relies on very little information about the economic environment. The government does not need to know the joint distribution $\Pi(\theta_m, \hat{\beta})$, which is an integral information in usual multi-dimensional screening mechanisms.[24] The only information the government needs is the distribution of productivity to determine the efficient allocation.

5.2 Hidden Temptation

Here, I will consider an environment where all agents are time-inconsistent, but vary in the degree of their temptation β and sophistication level $\hat{\beta}$. Let $\mathcal{B} = \left[\underline{\beta}, \overline{\beta}\right] \subset [0, 1]$ be the set of possible levels of temptation, with $\overline{\beta} > \underline{\beta}$ and $\overline{\beta} < 1$. In this economy, an agent's type is represented by $(\theta_m, \beta, \hat{\beta})$, where $\hat{\beta} \in [\beta, 1]$. Let $\Pi(\theta_m, \beta, \hat{\beta})$ denote the joint distribution of productivity, temptation level and sophistication level.

[24] For an introduction to the multi-dimensional screening model, see Armstrong and Rochet [(1999)]

Notice the government is unable to ascertain whether an agent with reported belief $\hat{\beta}$ is sophisticated or non-sophisticated. It is thus impossible to simultaneously screen for the agents' degree of temptation and sophistication levels. With an additional hidden dimension, a mechanism with an arbitrary target level of sophistication, like above, may no longer work. For example, if the government chose a cutoff $\tilde{\beta} \in \left(\underline{\beta}, \overline{\beta}\right)$ and attempted to threaten all agents more sophisticated and fool all agents more naïve, then agents with temptation $\beta > \tilde{\beta}$ and sophistication level $\hat{\beta} \geq \beta$ will not be fooled.

However, the government can choose $\tilde{\beta} \in \left[\overline{\beta}, 1\right)$ and design both a threat mechanism and fooling mechanism for sophistication level $\tilde{\beta}$. All agents with $\hat{\beta} \leq \tilde{\beta}$ would be threatened, and all agents with $\hat{\beta} \geq \tilde{\beta}$ would be fooled. This fact immediately follows from Lemma 1. Lemma 1 can be applied with this cutoff, because the temptation level of the agents can be ignored, since all agents have $\beta \leq \tilde{\beta}$. The subtlety lies in how the threats and empty promises are constructed to satisfy the executability constraints. For the imaginary allocations, the government needs to ensure the agents with the least temptation $\beta = \overline{\beta}$ would prefer the real allocations, then all agents with stronger temptation $\beta < \overline{\beta}$ would strictly prefer the real allocation. For the threat allocations, the government needs to ensure the agents with the most temptation $\beta = \underline{\beta}$ would prefer the real allocations, then all agents with less temptation $\beta > \underline{\beta}$ would strictly prefer the real allocation. This gives us the following theorem.

Theorem 4 *For the environment with hidden temptation and sophistication, the efficient allocation* $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ *is truthfully implementable with a hybrid mechanism.*

Theorem 4 is also robust to changes in the joint distribution $\Pi\left(\theta_m, \beta, \hat{\beta}\right)$. In addition to the primitives introduced in Section 2, the government does not need to know more than the support of $\mathcal{B}$. Theorem 4 also demonstrates how the efficient allocation is implementable as long as all agents are time-inconsistent, regardless of the sophistication or temptation. The next section will show how this result no longer holds when time-consistent agents are present.

6 Model with Time-Consistent Agents

In the previous sections, all agents in the economy were time-inconsistent (henceforth, TI). In this section, I will explore an economy where some agents are time-consistent (henceforth, TC). The presence of TC agents causes distortions, because without preference reversal, TC agents can follow through with their consumption plan and choose the imaginary allocation and avoid the threat allocation. This restricts the ability of the government to provide insurance.

By Theorem 4, it is without loss of generality to focus on TI agents with the same temptation level. The government is uncertain whether the agents are TC ($\beta = 1$) or TI ($\beta < 1$), with probability $\text{PR}(TI) = \phi$. The TC agents know their time-consistency (sophisticated), while TI agents could be non-sophisticated.[25] I assume the distribution of productivity is independent of the agents' consistency level. For this section, I will focus on the case with $\kappa = 1$, so the government puts all of the welfare weights on the planner. Thus, the welfare of both TC and TI agents are measured using the same utility.

6.1 The Threat Mechanism with Time-Consistent Agents

Suppose TI agents have a fixed sophistication level of $\hat{\beta} \in [\beta, 1)$, the government can implement a threat mechanism. The government introduces the following menu for TC agents of productivity θ_m:

$$C_m^{TC} = \left\{ \left(c_m^P, k_m^P, y_m^P\right); \left(c_m^D, k_m^D, y_m^D\right) \right\}$$

and the following menu for TI agents of productivity θ_m:

$$C_m^{TI} = \left\{ \left(c_m^R, k_m^R, y_m^R\right); \left(c_m^T, k_m^T, y_m^T\right) \right\}.$$

The allocation $\left(c_m^P, k_m^P, y_m^P\right)$ will be referred to as the *persistent allocation*, and it is the allocation the government implements for TC agents with θ_m productivity. The allocation $\left(c_m^D, k_m^D, y_m^D\right)$ is referred to as the *deterrent allocation*, and it is meant to deter the TI agents from misreporting as TC agents of θ_m productivity.

The deterrent allocation has $k_m^D < k_m^P$ and $c_m^D > c_m^P$, which appeals to the doer's present bias. This deters the TI agent's planner from misreporting the consistency level. The deterrent allocation can also be constructed so that the TC agents would never prefer it over the persistent allocation. The construction of a deterrent allocation will be shown.

The government offers the following menu at date $t = 0$, $C = \left\{C_m^{TC}, C_m^{TI}\right\}_{\theta_m \in \Theta}$. The planner first chooses a menu C_m^j from C, and then an allocation from C_m^j is chosen by the doer. The mechanism is meant to separate agents along two dimensions: productivity and level of consistency. Let

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

[25] I will relax this assumption and discuss the case with paranoid TC agents, TC agents believing they may be TI agents, in the last subsection of this section.

Hence, $C^{1|\hat{\beta}}_{\hat{m}|m}$ denotes the set of allocations a TI agent with productivity θ_m and sophistication level $\hat{\beta}$ predicts his doer will choose after he misreports to be a TC agent with productivity $\theta_{\hat{m}}$. Let

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

Hence, C^1_m denotes the set of allocations a truth-telling TC agent would choose. Let $C^{\hat{\beta}}_m$ and $C^{\hat{\beta}}_{\hat{m}|m}$ be defined as before. The following definition defines a direct threat mechanism with TC agents.

Definition 6 *A direct threat mechanism with TC agents has* $C = \left\{C^{TC}_m, C^{TI}_m\right\}_{\theta_m \in \Theta}$, *with* $C^{TC}_m = \left\{\left(c^P_m, k^P_m, y^P_m\right); \left(c^D_m, k^D_m, y^D_m\right)\right\}$ *and* $C^{TI}_m = \left\{\left(c^R_m, k^R_m, y^R_m\right); \left(c^T_m, k^T_m, y^T_m\right)\right\}$ *satisfying:*

i. credible threat constraints: $\left(c^R_m, k^R_m, y^R_m\right) \in C^{\hat{\beta}}_m$ *and* $\left(c^T_{\hat{m}}, k^T_{\hat{m}}, y^T_{\hat{m}}\right) \in C^{\hat{\beta}}_{\hat{m}|m}$, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

ii. deterrent constraints: $\left(c^P_m, k^P_m, y^P_m\right) \in C^1_m$, *and* $\left(c^D_{\hat{m}}, k^D_{\hat{m}}, y^D_{\hat{m}}\right) \in C^{1|\hat{\beta}}_{\hat{m}|m}$, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$.

To define the set of truthfully implementable allocations that can be implemented by a direct threat mechanism with TC agents, note that the incentive compatibility constraints for the TI agents are, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$\max_{(c_m, k_m, y_m) \in C^{\hat{\beta}}_m} U\left(c_m, k_m, y_m; \theta_m\right) \geq \max \left\{ \max_{(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C^{\hat{\beta}}_{\hat{m}|m}} U\left(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}; \theta_m\right), \max_{(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C^{1|\hat{\beta}}_{\hat{m}|m}} U\left(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}; \theta_m\right) \right\}. \tag{6}$$

By the incentive compatibility constraints (6), it is optimal for the TI agents to report truthfully about their productivity and consistency level. Let

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

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

Hence, $C^{\hat{\beta}|1}_{\hat{m}|m}$ denotes the set of allocations a TC agent with productivity θ_m would select from a menu meant for TI agents with productivity $\theta_{\hat{m}}$, while $C^1_{\hat{m}|m}$ denotes the set of allocations a TC agent would pick if he misreports his productivity as $\theta_{\hat{m}}$ and is truthful about his

consistency. The incentive compatibility constraints for the TC agents are, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$\max_{(c_m,k_m,y_m)\in C_m^1} U\left(c_m,k_m,y_m;\theta_m\right) \geq \max\left\{ \max_{(c_{\hat{m}},k_{\hat{m}},y_{\hat{m}})\in C_{\hat{m}|m}^1} U\left(c_{\hat{m}},k_{\hat{m}},y_{\hat{m}};\theta_m\right), \right. \tag{7}$$

$$\left. \max_{(c_{\hat{m}},k_{\hat{m}},y_{\hat{m}})\in C_{\hat{m}|m}^{\hat{\beta}|1}} U\left(c_{\hat{m}},k_{\hat{m}},y_{\hat{m}};\theta_m\right) \right\}.$$

Incentive compatibility constraints (7) discourage the TC agents from misreporting about their productivity and level of consistency. The executability constraints for the real allocation are defined by (5).

Definition 7 *The allocation* $\left\{\left(c_m^P,k_m^P,y_m^P\right),\left(c_m^R,k_m^R,y_m^R\right)\right\}_{\theta_m\in\Theta}$ *is truthfully implementable by a direct threat mechanism if there exists* $\left\{\left(c_m^D,k_m^D,y_m^D\right),\left(c_m^T,k_m^T,y_m^T\right)\right\}_{\theta_m\in\Theta}$ *such that the*

i. incentive compatibility constraints, and

ii. executability constraints

are satisfied.

The government maximizes welfare

$$\sum_{\theta_m\in\Theta} \pi_m \left[\phi U\left(c_m^R,k_m^R,y_m^R;\theta_m\right)+\left(1-\phi\right)U\left(c_m^P,k_m^P,y_m^P;\theta_m\right)\right], \tag{8}$$

subject to the incentive compatibility constraints ((6) and (7)), executability constraints (5), credible threat constraints, deterrent constraints and the feasibility constraint

$$\sum_{\theta_m\in\Theta} \left\{\phi\pi_m\left[y_m^R-c_m^R-k_m^R\right]+\left(1-\phi\right)\pi_m\left[y_m^P-c_m^P-k_m^P\right]\right\}=0. \tag{9}$$

A suitably chosen set of threat and deterrent allocations can help relax (6). As a result, 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 5 *There exists* $\bar{\theta}\in\Theta\setminus\theta_1$ *such that*

i. for productivity types $\theta_m\geq\bar{\theta}$, $\left(c_m^P,k_m^P\right)\gg\left(c_m^R,k_m^R\right)$ *with* $y_m^P<y_m^R$,

ii. for productivity types $\theta_1<\theta_m<\bar{\theta}$, $\left(c_m^P,k_m^P,y_m^P\right)\ll\left(c_m^R,k_m^R,y_m^R\right)$,

iii. for θ_1, $\left(c_1^P,k_1^P,y_1^P\right)=\left(c_1^R,k_1^R,y_1^R\right)$.

By Theorem 5, 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. Since TI agents are manipulable, 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. 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 they are already receiving the least utility and 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.

To construct the deterrent allocation, let $\left(c^D, k^D, y_m^D\right)$ with $y_m^D = y_m^P$ be the deterrent allocation satisfying:

$$\min_{\theta_m, \theta_{\hat{m}} \in \Theta} \left[u\left(c^D\right) - h\left(\frac{y_{\hat{m}}^D}{\theta_m}\right) + \hat{\beta} w\left(k^D\right) \right] \geq \max_{\theta_m, \theta_{\hat{m}} \in \Theta} \left[u\left(c_{\hat{m}}^P\right) - h\left(\frac{y_{\hat{m}}^P}{\theta_m}\right) + \hat{\beta} w\left(k_{\hat{m}}^P\right) \right], \quad (10)$$

and

$$\min_{\theta_m \in \Theta} \left[u\left(c_m^R\right) - h\left(\frac{y_m^R}{\theta_m}\right) + w\left(k_m^R\right) \right] \geq \max_{\theta_m, \theta_{\hat{m}} \in \Theta} \left[u\left(c^D\right) - h\left(\frac{y_{\hat{m}}^D}{\theta_m}\right) + w\left(k^D\right) \right]. \quad (11)$$

By inequality (10), any TI agent who misreports his consistency level would select the deterrent allocation over the persistent allocation. Inequality (11) guarantees the TI agents would prefer to report their time consistency truthfully. If (11) is satisfied, the TC agents would never choose the deterrent allocations over the persistent allocations. If (10) and (11) hold, then TI agents of all productivity types would never misreport to be TC agents, and it is possible to construct the deterrent allocations such that (10) and (11) are satisfied.[26]

Let the intertemporal wedge be defined as $\tau_K = 1 - \frac{U_c\left(c,k,\frac{y}{\theta}\right)}{U_k\left(c,k,\frac{y}{\theta}\right)}$, and the labor wedge as $\tau_L = 1 + \frac{U_y\left(c,k,\frac{y}{\theta}\right)}{U_c\left(c,k,\frac{y}{\theta}\right)}$. The following theorem helps characterize the optimal allocation and wedges in an environment with TC agents.

[26] To see how, first choose $\left(c^D, k^D\right)$ so (11) holds. Next, increase c^D and decrease k^D such that $u\left(c^D\right) + w\left(k^D\right)$ remains unchanged, and since $1 > \hat{\beta}$, then it is possible to find $\left(c^D, k^D\right)$ such that (10) holds.

Theorem 6 *For a threat mechanism, if $\phi \in (0,1)$ and $\hat{\beta} \in (\beta, 1)$, the optimal allocation has the following properties*

i. $\tau_K = 0$ *for all agents.*

ii. $\tau_L = 0$ *for all TI agents with* $\theta_m \geq \bar{\theta}$ *and TC agents of productivity* θ_M.

iii. $\tau_L \geq 0$ *for all TI agents with* $\theta_m < \bar{\theta}$ *and* $\tau_L > 0$ *for all TC agents with* $\theta_m \in \Theta \setminus \theta_M$.

The usual trade-off between insurance and output efficiency is present in this economy. Theorem 6 demonstrates how the the output of the less productive agents is distorted downwards. This result is standard in Mirrlees taxation. The government is able to provide consumption smoothing for all agents.

The next corollary shows that as long as TI agents not fully naïve, then the threat mechanism can implement the same optimal allocation for all levels of sophistication. This follows from the fact that the optimal allocation in a threat mechanism does not depend on the sophistication level of the TI agents.

Corollary 3 *In a threat mechanism, the optimal allocation is the same for any sophistication level* $\hat{\beta} \in [\beta, 1)$.

The usual Mirrlees taxation with TC agents occurs when $\phi = 0$, and the constrained efficient optimum is achieved. This paper has shown that the full information efficient optimum is attainable when the economy is populated by TI agents ($\phi = 1$). Let $W^T(\phi, \hat{\beta})$ denote the welfare under a threat mechanism with measure ϕ of TI agents with sophistication level $\hat{\beta} \in [\beta, 1)$. It is natural to presume that social welfare increases as the proportion of TI agents increases, as shown in Figure 2. Theorem 7 confirms this intuition.

Theorem 7 *In a threat mechanism, $W^T(\phi, \hat{\beta})$ increases continuously 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 5. 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 (7). Hence, the government is able to provide better insurance when more agents are time-inconsistent.

Also of note, Theorem 7 shows that continuity in welfare can be achieved with respect to the proportion of TC agents. However, in an economy with homogeneous temptation β,

there is a discontinuity in welfare as $\beta \to 1$. This discontinuity is due to the assumption on utility $u(\cdot)$ and $w(\cdot)$. Since $u(\cdot)$ and $w(\cdot)$ are unbounded below and above, the government is always able to construct possibly extreme threats (very small k and very large c) that deter TI agents. In Section 8, I will discuss the effects of relaxing this assumption on utility.

6.2 The Fooling Mechanism with Time-Consistent Agents

The government can also implement a fooling mechanism when $\hat{\beta} \in (\beta, 1)$. The government introduces the following menu for agents of productivity θ_m : $C_m = \left\{C_m^{TC}, C_m^{TI}\right\}$, where C_m^{TC} consists of the persistent and deterrent allocations and C_m^{TI} consists of the real and imaginary allocations.

Definition 8 *A direct fooling mechanism with TC agents has* $C = \left\{C_m^{TC}, C_m^{TI}\right\}_{\theta_m \in \Theta}$, *with* $C_m^{TC} = \left\{\left(c_m^P, k_m^P, y_m^P\right); \left(c_m^D, k_m^D, y_m^D\right)\right\}$ *and* $C_m^{TI} = \left\{\left(c_m^R, k_m^R, y_m^R\right); \left(c_m^I, k_m^I, y_m^I\right)\right\}$ *satisfying:*

i. fooling constraints: $\left(c_m^I, k_m^I, y_m^I\right) \in C_m^{\hat{\beta}}$ *and* $\left(c_{\hat{m}}^I, k_{\hat{m}}^I, y_{\hat{m}}^I\right) \in C_{\hat{m}|m}^{\hat{\beta}}$, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

ii. deterrent constraints: $\left(c_m^P, k_m^P, y_m^P\right) \in C_m^1$, *and* $\left(c_m^D, k_m^D, y_m^D\right) \in C_{\hat{m}|m}^{1|\hat{\beta}}$, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$.

The allocations that are truthfully implementable by a direct fooling mechanism with TC agents is bounded by the incentive compatibility constraints (6) and (7) and the executability constraints (5). Thus, the definition of truthfully implementable allocations in this mechanism is similar to Definition 7.

Definition 9 *The allocation* $\left\{\left(c_m^P, k_m^P, y_m^P\right), \left(c_m^R, k_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ *is truthfully implementable by a direct fooling mechanism if there exists* $\left\{\left(c_m^D, k_m^D, y_m^D\right), \left(c_m^I, k_m^I, y_m^I\right)\right\}_{\theta_m \in \Theta}$ *such that the*

i. incentive compatibility constraints, and

ii. executability constraints

are satisfied.

The following theorem shows that, for a given $(\phi, \hat{\beta})$, the government is indifferent between implementing a fooling or a threat mechanism when agents are partially naïve.

Theorem 8 *For* $\hat{\beta} \in (\beta, 1)$, *the optimal allocation in a direct fooling mechanism is equivalent to the optimal allocation in a direct threat mechanism.*

Theorem 8 implies that the properties derived for the optimal allocation in a threat mechanism with TC agents also applies to the fooling mechanism. Combined with Corollary 3, for a given ϕ, the optimal allocation and welfare are the same for any sophistication level $\hat{\beta} \in (\beta, 1)$ regardless of whether a threat or a fooling mechanism is implemented. The caveat here is that this result is only true when TI agents are partially naïve, in essence, $\hat{\beta} \in (\beta, 1)$. However, when agents are completely naïve, the optimal allocation and maximal welfare is different.

6.2.1 Fully Naïve Time-Inconsistent Agents

When $\hat{\beta} \in (\beta, 1)$, the government was able to use deterrent allocations to prevent the TI agents from pretending to be TC agents. However, separation in consistency is no longer possible when $\hat{\beta} = 1$.[27] When $\hat{\beta} = 1$, the government introduces the following menu at date $t = 0$, $C = \{C_m\}_{\theta_m \in \Theta}$, with

$$C_m = \left\{ \left(c_m^R, k_m^R, y_m^R\right), \left(c_m^P, k_m^P, y_m^P\right) \right\}.$$

The government is unable to separate the TC agents from the TI agents at date $t = 0$, so it expects agents of the same productivity to pick the same menu. After preference reversal, the government expects the TI agents to choose the real allocations and the TC agents to choose the persistent allocations. Here, the persistent allocations serve a similar purpose as the imaginary allocations to the TI agents. The TI agents evaluate their reporting strategy using the persistent allocations, but consume the real allocations after preference reversal.

The definition of a direct fooling mechanism for full naiveté with TC agents and the corresponding definition of truthful implementation is presented below.

Definition 10 *A direct fooling mechanism with TC agents and $\hat{\beta} = 1$ has $C = \{C_m\}_{\theta_m \in \Theta}$, with $C_m = \left\{ \left(c_m^R, k_m^R, y_m^R\right), \left(c_m^P, k_m^P, y_m^P\right) \right\}$ satisfying the fooling constraints: $\left(c_m^P, k_m^P, y_m^P\right) \in C_m^{\hat{\beta}}$ and $\left(c_{\hat{m}}^P, k_{\hat{m}}^P, y_{\hat{m}}^P\right) \in C_{\hat{m}|m}^{\hat{\beta}}$, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$. The allocation $\left\{ \left(c_m^R, k_m^R, y_m^R\right), \left(c_m^P, k_m^P, y_m^P\right) \right\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct fooling mechanism if the allocation satisfies the incentive compatibility constraints (4) and the executability constraints (5).*

Note that since $\hat{\beta} = 1$, the incentive compatibility constraints (4) are the same for the TC and TI agents. The government maximizes the utilitarian welfare (8) subject to the fooling constraints, the incentive compatibility constraints (4), executability constraints (5), and feasibility (9). The following theorem characterizes the intertemporal wedge in a fully naïve environment and the allocations for the top and bottom of the productivity distribution.

[27] This point was also made in Galperti (2015), but its implications were not fully explored.

Theorem 9 *For a fooling mechanism, if* $\phi \in (0,1)$ *and* $\hat{\beta} = 1$*, the optimal allocation has the following properties:*

i. $\tau_K < 0$ *for all TC agents with* $\theta_m > \theta_1$ *and* $\tau_K > 0$ *for all TI agents with* $\theta_m > \theta_1$*.*

ii. For θ_M*, the allocation satisfies* $c_M^P < c_M^R$*,* $k_M^P > k_M^R$ *and* $y_M^P > y_M^R$*.*

iii. For θ_1*,* $\left(c_1^P, k_1^P, y_1^P\right) = \left(c_1^R, k_1^R, y_1^R\right)$ *with* $\tau_K = 0$*.*

Theorem 9 shows how the government provides consumption smoothing for agents with $\theta_m > \theta_1$ in expectation. In essence, the best the government can do is to require the TC agents to over-save and the TI agents to under-save. This is true even for the most productive agents, so there will be distortion at the top.

The reason for this distortion is due to the binding executability constraints (5). If the government implements the allocation described in Theorem 5, then the executability constraints would be violated since the TI agents would pick the persistent allocation. This stems from the government's inability to separate the TI agents from the TC agents, so the persistent allocation acts as a proxy for the imaginary allocation when TI agents are fully naïve. The next section discusses the welfare implications of increasing the TI agents' awareness of their foibles.

6.3 Incentives for Self-Awareness

Comparing the optimal welfare for each sophistication level helps answer the question of whether the government wants to raise the TI agents' awareness of their present bias. It is clear from the setting where all agents were TI, the government is indifferent among the various sophistication levels of the agents. However, when TC agents are present, this is no longer the case. Theorem 8 and Corollary 3 show that when agents are aware of their present bias, the government is indifferent between implementing a threat mechanism or a fooling mechanism, which would both achieve the same optimal allocation and welfare regardless of the sophistication level of the TI agents. Thus, the question is whether it is in the best interest of the government to alert the fully naïve agents of their present bias.

For a given ϕ, let $W(\hat{\beta} < 1)$ denote the social welfare when agents are sophisticated or partially naïve, and $W(\hat{\beta} = 1)$ be the welfare when agents are fully naïve. The following theorem shows that the government would want the TI agents to be at least partially naïve.

Theorem 10 *If* $\phi \in (0,1)$*, then* $W(\hat{\beta} < 1) > W(\hat{\beta} = 1)$*.*

The difference in welfare in Theorem 10 is due to the inability of the government to separate the fully naïve TI agents from the TC agents, while separation is possible if TI agents are at least partially naïve. The inability to separate along consistency levels causes two problems. First, the government is unable to design a deterrent allocation to dissuade TI agents from mimicking TC agents. Second, the persistent allocation serves as a proxy for the imaginary allocation, so the allocation used for fooling the TI agents are no longer empty promises and are not off-equilibrium path when TC agents are present. As a result, when $\hat{\beta} = 1$, the optimal allocation distorts the intertemporal wedge, which is an additional distortion not present when TI agents are at least partially naïve.

Theorem 10 has new implications on the education policy for TI agents. Educating non-sophisticated TI agents is only weakly welfare increasing. This result contrasts with the policy recommendations from the literature.[28]

6.4 Paranoia

Previous analysis assumed the TC agents were sophisticated ($\hat{\beta} = \beta = 1$). Sophisticated TC agents do not respond to threats or empty promises, and require strictly positive information rents for truth-telling. Here, I will briefly discuss how paranoia affects the behavior of TC agents, and consequently how government policy can be adjusted to take advantage of this.

A paranoid agent is a non-sophisticated TC agent who believes he is time-inconsistent. Paranoid agents respond to threats, and can also be fooled. It is trivial to see that a threat mechanism can extract all information rents from paranoid agents, since the government can construct the threat allocation using similar methods employed for the TI agents. The logic for the fooling mechanism is more subtle.

To see how a fooling mechanism achieves the full information optimum in an economy with only paranoid agents, $\hat{\beta} < \beta = 1$, consider the example with two productivity types $\Theta = \{\theta_L, \theta_H\}$. Let $C_m = \left\{\left(c^*, k^*, y_m^*\right), \left(c_m^I, k_m^I, y_m^*\right)\right\}$ and the allocations satisfy the fooling constraints

$$u\left(c_m^I\right) + \hat{\beta} w\left(k_m^I\right) \geq u\left(c^*\right) + \hat{\beta} w\left(k^*\right),$$

the executability constraints

$$u\left(c^*\right) + w\left(k^*\right) \geq u\left(c_m^I\right) + w\left(k_m^I\right),$$

[28] Previous literature has focused on the pitfalls of being non-sophisticated. For example, behavioral contract theory has focused on how limited awareness of self control problems makes consumers vulnerable to exploitation by firms, see Spiegler [(2011)] or Koszegi [(2014)] for examples.

and the incentive compatibility constraints, which implies the following

$$h\left(\frac{y_H^*}{\theta_L}\right) - h\left(\frac{y_L^*}{\theta_L}\right) \geq \left[u(c_H^I) + w(k_H^I)\right] - \left[u(c_L^I) + w(k_L^I)\right] \geq h\left(\frac{y_H^*}{\theta_H}\right) - h\left(\frac{y_L^*}{\theta_H}\right).$$

The fooling and executability constraints imply $c_m^I > c^*$ and $k_m^I < k^*$. Combined with the incentive compatibility constraints, it must be that $c_L^I > c_H^I$ with $k_L^I < k_H^I$. 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 he is strictly worse off by choosing it, because he 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 he is afraid that by misreporting, he would have even less savings.

The way the fooling mechanism works for paranoid agents is in stark contrast to the logic presented in the previous sections. Non-sophisticated TI agents were fooled by empty promises, but paranoid agents were fooled by empty threats.[29]

If the economy has both paranoid and TI agents, then using a fooling mechanism could be problematic for the government. 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 temptation $\beta < \hat{\beta}$. If the government tries to fool the agents, then it is not possible for the government to separate agents along consistency level. In this particular case, depending on who the government chooses to fool, either the paranoid TC agent 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. This is analogous to the case with sophisticated TC agents and fully naïve TI agents. However, such a problem does not arise when the government uses a threat mechanism. This is because the same threats for the TI agents can also deter the paranoid agents from misreporting.[30]

7 Policy Implementation

In the previous sections, a direct revelation mechanism was used to characterize the optimal allocations. In this section, I explore a decentralized setting and the policy tools

[29] In the previous sections, it was sufficient to set $\left(c_L^I, k_L^I, y_L^*\right) = (c^*, k^*, y_m^*)$. For paranoid agents, however, it must be the case that $\left(c_L^I, k_L^I, y_L^*\right) \neq (c^*, k^*, y_L^*)$, but the government can set $\left(c_H^I, k_H^I, y_H^*\right) = (c^*, k^*, y_H^*)$.

[30] 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.

that could implement the optimum.

The government introduces a policy $\mathcal{P}$, and the agents earn before-tax income y, choose consumption c and purchase risk-free bond b with gross rate $R = 1$ in response to $\mathcal{P}$. Let $\hat{A} = \left\{\left(\hat{c}_m, \hat{k}_m, \hat{y}_m\right)\right\}_{\theta_m \in \Theta}$ denote the resulting equilibrium allocation given $\mathcal{P}$. The following definition specifies what implementation means in a decentralized environment.

Definition 11 *A government policy $\mathcal{P}$ implements an optimum $\tilde{A}$ if the equilibrium allocation given $\mathcal{P}$ is $\hat{A} = \tilde{A}$.*

The section will proceed by analyzing different policies that could implement the full information efficient optimum. For illustrative purposes, the government evaluates welfare only by the ex-ante utility, in essence, $\kappa = 1$. The outline of the policy implementation does not change with κ.

7.1 Decentralizing the Fooling Mechanism

For non-sophisticated agents ($\hat{\beta} > \beta$), the government can introduce a savings plan with a default savings subsidy set at τ^* and an income contingent savings subsidy $\tau(y)$ to incentivize work. Though an agent who earns income y is qualified for the savings subsidy $\tau(y)$, it is costly to switch from τ^*. Let the income dependent tax for switching be $T^s(y) > 0$. An agent with income y is only eligible for $\tau(y)$ and the default τ^*. The savings subsidy is combined with a standard income tax $T^e(y)$ independent of the subsidy. The gross income tax will be $T^e(y)+T^s(y)$ if the agent switches from the default subsidy, and $T^e(y)$ if no switching occurs. The government policy is $P^n = (\tau^*, \tau(y), T^e(y), T^s(y))$. I will proceed to construct a savings subsidy scheme $(\tau^*, \tau(y))$ with an income tax schedule $(T^e(y), T^s(y))$ that implements the full information efficient allocation.

The agents face the following budget constraints:

$$c + (1-\tau)b \leq y - T^e(y) - \mathbf{1}_{\tau \neq \tau^*} T^s(y), \text{ and } k \leq b,$$

where $\mathbf{1}_{\tau \neq \tau^*}$ is an indicator function that equals to one if $\tau \neq \tau^*$. At the optimum, $k = b$, so the sequential budget constraints can be rewritten as

$$c + (1-\tau)k \leq y - T^e(y) - \mathbf{1}_{\tau \neq \tau^*} T^s(y). \tag{12}$$

An agent solves the following problem:

$$\max_y U\left(c(y), k(y), y; \theta\right)$$

subject to

$$(c(y), k(y)) \in \arg\max_{(c,k)} W(c, k, y; \theta) \text{ s.t. } (12).$$

The agents are required to work before making consumption and savings decisions. They choose work effort by deducing future consumption and savings given policy P^n and belief $\hat{\beta}$. However, due to the incorrect beliefs, the government expects the agents to stick with the default option of τ^*, which does not vary with income.

To see how such a policy implements the efficient allocation, first choose $\tau^* = 1 - \beta$ so, on the equilibrium path, consumption is smooth: $u'(c(y)) = w'(k(y))$. Also, set the income tax to be

$$T^e(y) = \begin{cases} y - c^* - \beta k^* & \text{if } y \in [y_1^*, \infty) \\ y & \text{if } y \in [0, y_1^*) \end{cases}$$

where (c^*, k^*) is the efficient consumption and savings and y_1^* is the efficient output for the least productive agents. By construction, no agent will ever produce $y < y_1^*$.

Let $Y = \{y_1^*, \ldots, y_M^*\}$ denote the set of efficient outputs. Next, define $\tau(y)$ as an increasing step function, so $\tau(y) = \tau(y_m^*)$ if $y \in \left[y_m^*, y_{m+1}^*\right)$ with $\tau(y_1^*) = \tau^*$ and $\tau(y_m^*) < \tau\left(y_{m+1}^*\right)$. Also, define $T^s(y)$ as an increasing step function, so $T^s(y) = T^s(y_m^*)$ if $y \in \left(y_{m-1}^*, y_m^*\right]$ with $T^s(y_1^*) = 0$ and $T^s(y_m^*) > T^s\left(y_{m-1}^*\right)$. By the construction of $\tau(y)$, $T^e(y)$ and $T^s(y)$, agents would only choose $y \in Y$. If the planner expects the doer to switch from the default savings subsidy τ^*, for a given y, the doer is perceived to choose consumption and savings by solving

$$\max_{c,k} u(c) + \hat{\beta} w(k) \text{ s.t. } c + (1 - \tau(y)) k = y - T^e(y) - T^s(y).$$

Hence, the perceived optimal consumption and savings choices are functions of $\tau(y)$ and $T^s(y)$: $c(\tau(y), T^s(y))$ and $k(\tau(y), T^s(y))$.

An agent with productivity θ_m selects y by solving the following problem:

$$\max_{y \in Y} u\left[c(\tau(y), T^s(y))\right] - h\left(\frac{y}{\theta_m}\right) + w\left[k(\tau(y), T^s(y))\right].$$

By the setup, it is sufficient to choose policies that deter adjacent downward deviations. For type θ_m agents, the government can construct $\tau(y_m^*)$ for given $T^s(y_m^*) > T^s(y_{m-1}^*)$ and $\tau\left(y_{m-1}^*\right)$ to satisfy

$$\begin{aligned} &u\left[c\left(\tau\left(y_m^*\right), T^s\left(y_m^*\right)\right)\right] + w\left[k\left(\tau\left(y_m^*\right), T^s\left(y_m^*\right)\right)\right] \\ &\qquad - \left[u\left[c\left(\tau\left(y_{m-1}^*\right), T^s\left(y_{m-1}^*\right)\right)\right] + w\left[k\left(\tau\left(y_{m-1}^*\right), T^s\left(y_{m-1}^*\right)\right)\right]\right] = \Delta_m, \end{aligned}$$

where $\Delta_m = h\left(\frac{y_m^*}{\theta_m}\right) - h\left(\frac{y_{m-1}^*}{\theta_m}\right)$. Note that the least productive agent would choose the efficient allocation for θ_1. Therefore, by the incentive compatibility constraints, $\tau(y_m^*)$ is a function of $(T^s(y_2^*), \ldots, T^s(y_m^*))$.[31]

The government can choose $T^s(y_2^*)$ so the following holds

$$u\left[c\left(\tau\left(y_2^*\right), T^s\left(y_2^*\right)\right)\right] + \hat{\beta} w\left[k\left(\tau\left(y_2^*\right), T^s\left(y_2^*\right)\right)\right] = u\left(c^*\right) + \hat{\beta} w\left(k^*\right).$$

Since $\tau(y_2^*) > \tau^*$, then it must be that $T^s(y_2^*) > T^s(y_1^*)$ for the above fooling constraint to hold with equality. By iteration, the government chooses $T^s = (T^s(y_2^*), \ldots, T^s(y_M^*))$ so that the fooling constraints bind and both $T^s(y)$ and $\tau(y)$ are increasing step functions.

To check that the construction above implements the efficient allocation, notice that the executability constraints are trivially satisfied, because $\tau(y) \geq \tau^*$ for any y and the fooling constraints bind. As a result, the agents would choose the efficient allocation if they produced the efficient output. Also, notice that both the fooling and incentive compatibility constraints hold by construction, so the agents would produce efficiently. Finally, the government budget constraint holds, since $\sum \pi_m T^e(y_m^*) + \tau^* k^* = 0$.

Proposition 3 *If $\hat{\beta} \in (\beta, 1]$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ can be implemented by $P^n = (\tau^*, \tau(y), T^e(y), T^s(y))$, where savings subsidy $\tau(y)$ is an increasing step function and the gross income tax is strictly increasing ($T^e(y)$ is strictly increasing and $T^s(y)$ is an increasing step function).*

The savings subsidy is regressive in income, while the gross income tax increases with income. This is because the government increases the savings subsidy for higher incomes to encourage efficient output. With a regressive savings subsidy, it is necessary for the gross income tax to increase with income, so the income effect can offset the increase in utility caused by the substitution effect from the higher subsidies $\tau(y) > \tau^*$. Consequently, given an increasing $\tau(y)$, the increasing $T^s(y)$ guarantees the implementation of the efficient allocation after preference reversal.

O'Donoghue and Rabin [(2001)] have shown that naiveté and present bias can cause status quo bias. If agents suffer from status quo bias then with τ^* as the default savings subsidy, their doers may find switching to be inherently costly. For reasons outside of the model, the government may want to utilize the status quo bias to soften the need for an increasing $T^s(y)$.[32]

[31] This is due to the fact that from the incentive compatibility constraint for type θ_2 agents, $\tau(y_2^*)$ is a function of $T^s(y_2^*)$. By induction, $\tau(y_m^*)$ is a function of $(T^s(y_2^*), \ldots, T^s(y_m^*))$.

[32] Other policy proposals have suggested utilizing the status quo bias to help increase retirement savings. For instance, Thaler and Benartzi [(2004)] utilizes present bias and status quo bias to increase 401(k) enrollment and raise contribution rates.

There are other policies that could implement the efficient outcome for non-sophisticated agents, especially since the timing of when the agents work is flexible in the implementation for the relatively naïve. Other policies that require agents to commit to a menu of subsidy plans before working can also be conceived.

7.2 Decentralizing the Threat Mechanism

In this section, I propose a policy implementation for partially naïve and sophisticated agents ($\hat{\beta} < 1$). The government provides agents with *commitment loans*, where its associated repayment plans help smooth consumption. Let $(L(y^e), R(b, y^e, y))$ be the commitment loans and repayment pairs in the government plan P^s, where each loan is a function of projected income y^e and the repayment depends on savings b, realized income y and the original projection.

Agents take out commitment loans $L(y^e)$ after productivity is realized and before allocation decisions are made, with associated repayment plan $R(b, y^e, y)$. Notice the repayment plan depends on the savings and income of the agent, so the planner needs to predict what the doer would choose before committing to a plan. The government policy is comprised of the commitment loans and repayment plan with a standard income tax: $P^s = (L(y^e), R(b, y^e, y), T(y))$. It is assumed that there will be no interest earned from savings b.

The agents face the following sequential budget constraints after choosing $L(y^e)$

$$c + b \leq y + L(y^e) - T(y), \tag{13}$$

$$k \leq b - R(b, y^e, y). \tag{14}$$

Since selecting a commitment loan is equivalent to choosing a projected income y^e, a planner solves the following problem:

$$\max_{y^e} U(c(y^e), k(y^e), y(y^e); \theta)$$

subject to

$$(c(y^e), k(y^e), y(y^e)) \in \arg\max_{(c,k,y)} W(c, k, y; \theta) \text{ s.t. } (13) \text{ and } (14).$$

The loan is referred to as a commitment loan, because the agents do not use the loan to invest or consume. In contrast to usual loans, the purpose of a commitment loan is to facilitate commitment to increasing savings.

The agents obtain commitment from the structure of the repayment plan, which is constructed to incentivize savings. Another essential component of the repayment plan is its dependence on realized income and projected income. The repayment plan can be engineered so that if the realized income does not match the original projection, then the government is able to punish the agent by exacerbating his present bias. It also awards the agent with a consumption smoothing plan if the projection matches with the actual realization.

Proposition 4 *If $\hat{\beta} \in [\beta, 1)$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ can be implemented through commitment loans $L(y^e)$ with associated repayment $R_m(b, y^e, y)$ contingent on savings, expected income and realized income.*

Since the repayment plan depends on realized income, the government is always able to verify whether an agent chose the appropriate loans plan by observing an agent's income. If an agent projects an income of y^e and has an actual income not significantly higher than y^e, then the repayment plan provides a subsidy on bond savings that would offset the present bias. However, a more productive agent who selects a commitment loan meant for less productive agents would be tempted to work significantly more than y^e in exchange for a consumption subsidy (savings tax). This would lower the ex-ante utility since the consumption subsidy aggravates the present bias problem, so each agent chooses the commitment loan designed for their productivity.

The commitment loans and the associated repayment plan proposed here is not unusual. The agents have to take out the loan before earning income but after realizing their future possible productivity, much like student loans for medical or law school. Current government funded student loans programs also have repayment plans that depend on income. For instance, it is possible to select an income-driven repayment plan for federal student loans. Therefore, only minor adjustments, such as adding savings incentives into the repayment plan, need to be made to implement the type of policy suggested above.

7.3 Decentralization with Time-Consistent Agents

In this subsection, I will present a policy for the TC agents which can be combined with both fooling and threat mechanisms in an environment with partially naïve agents.

The persistent allocation can be implemented with a standard income tax. Let $Y^P = \left\{y_1^P, \ldots, y_M^P\right\}$ be the set of constrained efficient output for the TC agents. Since the constrained efficient consumption $\left\{\left(c_m^P, k_m^P\right)\right\}_{\theta_m \in \Theta}$ depends on the productivity θ_m only through y_m^P, it is possible to define the following consumption mapping $c^P : \left[y_1^P, \infty\right) \mapsto \mathbb{R}$ such that $c_m^P = c^P\left(y_m^P\right)$ and $c^P(y) = c^P\left(y_m^P\right)$ for any $y \in \left(y_m^P, y_{m+1}^P\right)$. The mapping for savings $k^P(y)$

is defined in a similar fashion. Since the constrained efficient consumption is monotonically increasing, the functions $c^P(y)$ and $k^P(y)$ are increasing step functions. The government can define an income tax schedule $T^P(y)$ for the TC agents as a function of y :

$$T^P(y) = \begin{cases} y - \left(c^P(y) + k^P(y)\right) & \text{if } y \geq y_1^P \\ y & \text{if } y < y_1^P. \end{cases}$$

Given the income tax schedule $T^P(y)$, TC agents of productivity θ_m would solve the following problem

$$\max_{c,k,b} U(c,k,y;\theta_m) \text{ s.t. } c + b \leq y - T^P(y) \text{ and } k \leq b.$$

Note the TC agents are allowed to save freely (uninhibited use of the storage technology). It is easy to show that the income tax can implement the constrained efficient allocation for the TC agents, provided that they do not pretend to be TI agents. This is because given the income tax $T^P(y)$, agents would always at least produce y_1^P, and combined with the consumption and savings functions, $c^P(y)$ and $k^P(y)$, the agents would never produce $y \notin Y^P$. Also, since the allocations satisfy incentive compatibility, a TC agent with productivity θ_m would choose to produce output y_m^P. Finally, since the agents are time-consistent, they would consume and save the constrained efficient amount given the after-tax income.

The deterrent allocation is a penalty for TI agents pretending to be TC agents. It can be constructed as a loan with a high interest rate combined with a bankruptcy insurance χ. Let $\left(c^D, k^D\right)$ satisfy (11) and (10). Agents take out a loan $L(y) = c^D - \left[c^P(y) + k^P(y)\right]$ and are required to repay $R(y) = c^P(y) + k^P(y) + L$. The bankruptcy insurance is set at $\chi = k^D$. An agent taking out this loan would solve

$$\max_{c,k,y,b} U(c,k,y;\theta_m) \text{ s.t. } c + b \leq y - T^P(y) + L(y) \text{ and } k \leq \max\{b - R(y), \chi\}.$$

Agents taking out this loan would find it optimal to work y_1^P, save $b = 0$, and proceed to file for bankruptcy to receive χ. Since $\left(c^D, k^D\right)$ satisfies (11) and (10), the TC agents would never choose this loans plan, and only the TI agents would choose this sub-optimal plan if they do not take out a commitment loan presented in Section 7.2, or if they do not apply for the savings subsidy introduced in Section 7.1.

If an agent takes out a commitment loan, then he chooses $L(y^e)$ with the associated repayment plan $R(b, y^e, y)$ to maximize the ex-ante utility. The income tax facing the agents who take out a commitment loan will be different from the previous subsection, and

is defined as

$$T^R(y) = \begin{cases} y - \left(c^R(y) + \beta k^R(y)\right) & \text{if } y \geq y_1^R \\ y & \text{if } y < y_1^R. \end{cases}$$

where the consumption functions depends on y_m^R. The government defines the following consumption mapping $c^R : \left[y_1^R, \infty\right) \mapsto \mathbb{R}$ such that $c_m^R = c^R\left(y_m^R\right)$ and $c^R(y) = c^R\left(y_m^R\right)$ for any $y \in \left(y_m^R, y_{m+1}^R\right)$. The mapping for savings $k^R(y)$ is defined in a similar way. Following the blueprint for constructing commitment loans in Section 7.2, the policy can be constructed so that TI agents who choose the commitment loan meant for their productivity will work, consume and save the constrained efficient amount, while the agents who do not choose the appropriate commitment loan would be worse off. Similarly, by using the construction oulined in Section 7.1, it is also possible to use regressive savings subsidy to implement the constrained optimum.

8 Discussions

In this section, I address some immediate extensions, which include environments with dynamic stochastic productivity shocks and commitment versus flexibility preferences. I also discuss some limitations of the main results.

8.1 Dynamic Stochastic Productivity Shocks

The model abstracts from concerns with capital taxation (work does not occur in the retirement stage). However, even with dynamic productivity, the efficient optimum is still implementable.

Consider a three period model, similar to Section 3 of Kocherlakota [(2005)], where work occurs in the first and second period, while the agent retires in the third period. To illustrate, let the set of possible second period productivity be $\Theta_2 = \{\theta_L, \theta_H\}$, with distribution $\Pr(\theta = \theta_m) = \pi_m \in (0,1)$. The set of possible first period productivity is a singleton $\Theta_1 = \{\theta_1\}$. In each period, the agents have access to a one-period bond with interest rate zero. The expected lifetime utility of the agents at $t = 0$ is

$$u_1(c_1) - h\left(\frac{y_1}{\theta_1}\right) + \delta \sum_{\theta_m \in \Theta_2} \pi_m \left[u_2(c_m) - h\left(\frac{y_m}{\theta_m}\right) + \delta u_3(k_m)\right].$$

This is also the ex-ante utility for the agents' planners at $t = 0$. The agents suffer from time-inconsistency, and they make decisions in each period according to a quasi-hyperbolic

discounting model. The agents' doers have the following ex-post utility at each period

$$U_1 = u_1(c_1) - h\left(\frac{y_1}{\theta_1}\right) + \beta\delta \sum_{\theta_m \in \Theta_2} \pi_m \left[u_2(c_m) - h\left(\frac{y_m}{\theta_m}\right) + \delta u_3(k_m)\right],$$

$$U_{2,m} = u_2(c_m) - h\left(\frac{y_m}{\theta_m}\right) + \beta\delta u_3(k_m),$$

$$U_{3,m} = u_3(k_m).$$

I will assume that $\delta = 1$ and $\beta < 1$. Similar to previous sections, in each period, the doer is triggered when consumption and savings decisions are made.

If all agents were time-consistent, then by the first order condition, the constrained efficient optimal allocation has to satisfy the inverse Euler equation.[33] This implies the agents are savings constrained at the optimum. In essence, if agents are allowed to save freely, they would want to save more for the second period than what the optimum prescribes.

If all agents are time-inconsistent and not fully naïve, in the second period, the government can either adopt a fooling mechanism or a threat mechanism. In a threat mechanism, the planner in the first period knows that regardless of the productivity realization in the second period, the planner in the second period would reveal the true productivity. Since all agents have the same productivity in the first period, for this specific model, a linear savings subsidy set at $1 - \beta$ for first period savings along with a threat mechanism in the second period can implement the full information efficient optimum. In essence, the agents can save freely at the savings subsidy augmented interest rate.

The problem is more subtle for a fooling mechanism. In the first period, the agents believe they will consume the at imaginary consumption levels in the second and third periods. As a result, if the agents are allowed to save freely, they would save according to an incorrect projection of their future consumption. In this case, the linear savings subsidy in the first period would also have to take this mis-specification into account.

In a model with multiple periods and stochastic productivity shocks, it is still possible to implement both the fooling and threat mechanisms. However, the implementation of a fooling mechanism has certain subtleties. More specifically, at the beginning of each period, the non-sophisticated agents learn their productivity and make decisions based on the imaginary allocation and an imaginary promised utility. When the agents make consumption and savings decisions, another imaginary promised utility is used by the agents to evaluate the merits of choosing the real allocations. In essence, within each period, the planners use a different imaginary promised utility than the doers to make intertemporal decisions:

[33] From the first order conditions, it is possible to show that $1/u'(c_1^{**}) = \pi_L/u'(c_L^{**}) + \pi_H/u'(c_H^{**})$. See Kocherlakota (2005) for details.

the promised utility for the planners is higher than the one for the doers. As a result, the fooling mechanism is dynamic, since the planners for the next period would use the imaginary allocations contained in the promised utility for the previous period's doers.

The threat mechanism does not have the complications described above. With a threat mechanism, the dynamic taxation problem only needs to keep track of the history of reports. In fact, if the productivity shocks are independent across periods, the government can implement the same threat mechanism every period. In this case, the dynamic taxation problem is essentially a static problem for each period.

In a multi-period model, other concerns may arise, chief among them is the agents' ability to learn. In a fooling mechanism, the doers consume differently from what the planners expected to consume. This could trigger learning and the agents could become more sophisticated.[34] Therefore, in a dynamic fooling mechanism, the government may also need to account for the evolving sophistication level of the agents. The threat mechanism does not have such concerns. This is because a threat mechanism confirms the planners' initial beliefs about their doers' bias, regardless of whether the beliefs were correct or not. In other words, partially naïve agents never have an opportunity to learn about their true degree of temptation in a threat mechanism, and the government does not need to worry about the effects of learning in a threat mechanism.

8.2 Demand for Flexibility

In the previous sections, the agents only have a demand for commitment. Recent developments in the literature have highlighted the trade-off between commitment and flexibility.[35] In these models, agents suffer from temptation but would also like to accommodate their intertemporal taste shock. The demand for commitment comes from the temptation problem. The demand for flexibility comes from the taste shock, which is an unresolved uncertainty.

I examine a two period model as in Section 2. I include a taste shock and augment the timing so that the agents receive additional private information over time. To be explicit, the taste shock γ is distributed according to probability distribution $f(\cdot|\theta)$ (CDF $F(\cdot|\theta)$) within a bounded support of $\left[\underline{\gamma}, \overline{\gamma}\right]$, with $0 < \underline{\gamma} < \overline{\gamma} < \infty$. The taste shock is realized when the agent makes consumption and savings decision. The agent does not know the extent of the taste shock when productivity is learned, but I allow the distribution of the taste shock

[34] For instance, Ali (2011) has the naïve planner learn about the doer's present bias through Bayesian updating by observing the outcome of an intertemporal task with noise.

[35] Amador, Werning, Angeletos (2006), Ambrus and Egorov (2013), Bond and Sigurdsson (2015) and Galperti (2015) are some of the papers that have explored the trade-off in the context of time-inconsistent preferences.

to vary with productivity. Hence, both the government and the agents do not know the true taste shock, but the agent is better informed than the government.

The ex-ante utility of the agent is

$$U(c,k,y;\theta) = \mathrm{E}_{\gamma}\left[\gamma u\left(c\right) - h\left(\frac{y}{\theta}\right) + w(k)\right].$$

The ex-post utility is

$$V(c,k,y;\theta) = \gamma u\left(c\right) - h\left(\frac{y}{\theta}\right) + \beta w(k).$$

The tension between commitment and flexibility arises from the fact that the agents would like to save according to the realization of γ, but the act of saving triggers the present bias.

This is a sequential screening problem. The government asks agents to report productivity first, and then the taste shock. As a benchmark, if the taste shock is public and all agents are time-inconsistent to the same degree, then by using the appropriate fooling or threat mechanism, the government can learn about the productivity of the agents and prescribe the optimal savings plan for each agent according to their observed taste shock. The government can implement the efficient allocation if the taste shock was public.

As was shown in the previous sections, by implementing a fooling mechanism, it is possible to use imaginary allocations to fool the non-sophisticated agents and elicit their actual productivity. For sophisticated agents, a threat mechanism can costlessly reveal the productivity type. If all agents are time-inconsistent, then after learning the productivity, the government can implement a savings dependent transfer $T(k)$. By Proposition 1 in Galperti [(2015)], an optimally chosen $T(k)$, with $T(k)$ strictly increasing and $T(k) > 0$ above some threshold $\bar{k}$ and $T(k) \leq 0$ below it, can help the government achieve the efficient allocation. This is because the transfer aligns the incentives of the planner and the doer. More importantly, since a fooling or threat mechanism elicits all of the private information from the agent before consumption and savings in a costless manner, the government can use the savings dependent transfer to ensure the doer saves according to the realized γ.

However, if the consistency of the agents is unobservable by the government (for example, if time-consistent agents exist in the economy), then the optimal allocations may not only be distorted in output, as in Section 6, but also intertemporally, as in Galperti [(2015)]. It should be interesting future work to explore the optimal sequential mechanism for screening both consistency and productivity in the first period and then the taste shock in the second.

8.3 Impediments to Implementation

This paper has focused on a setting where the government can screen time-inconsistent agents without impunity. In Section 6, I showed how the presence of time-consistent agents would impede the screening process and make distortions necessary. Here, I discuss other situations where full efficiency may not be achievable.

8.3.1 Limited Promises and Punishments

With non-sophisticated agents, if the government is limited in the amount of empty promise it can make, then full efficiency might not be achievable. With sophisticated agents, if the government is limited in the amount of punishment it can dole out for misreporting agents, then the efficient allocation might not be implementable. This can arise from the preferences of the agents.

In the previous sections, in addition to the usual assumptions, I assumed the period utility functions were unbounded below and above. This assumption provides a non-empty set of consumption bundles that could deceive non-sophisticated agents for any amount of information rent. Consider the case where the utility function is bounded below. Figure 7 illustrates how fooling can be limited for fully naïve agents in the two types example. The flatter solid (blue) curve represents the indifference curve of the ex-ante utility and the steeper solid (red) curve represents the indifference curve of the ex-post utility, both evaluated at allocation (c^*, k^*). 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 7, which does not provide enough rents to the productive agents to elicit their true type without distorting the allocations for the low productivity type agents. Hence, there will be distortions caused by asymmetric information in the optimal allocation to provide these rents.

Similarly, for sophisticated agents, if the period utility functions are bounded below, then the punishment for misreporting will be limited. Figure 8 illustrates how credible threats are limited in a fully sophisticated case with two productivity types. The flatter solid (blue) curve represents the indifference curve of the ex-ante utility for the productive truthful agents who reports truthfully at efficient allocation. The productive agents have an incentive to misreport and receive $\left(\Phi^R_{L,H}, k^*\right)$ instead. However, the harshest threat the government can issue is to choose the threat allocation indicated in Figure 8. The steeper solid (red) curve represents the indifference curve of the ex-post utility for the efficient agents at the threat allocation. This is clearly not enough to deter the agents from misreporting. Hence, the

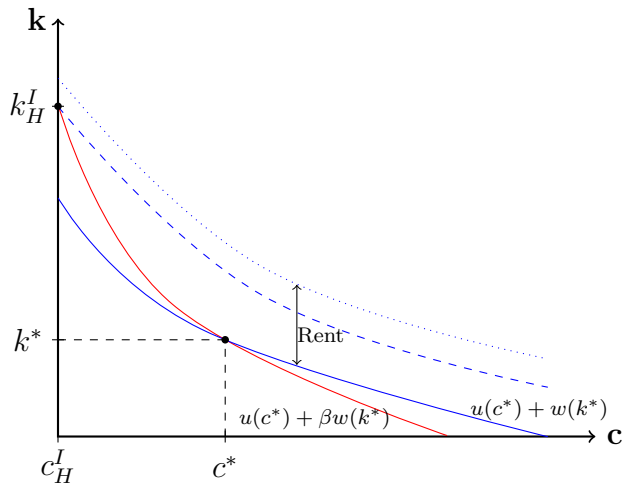

Figure 7: Limited Fooling

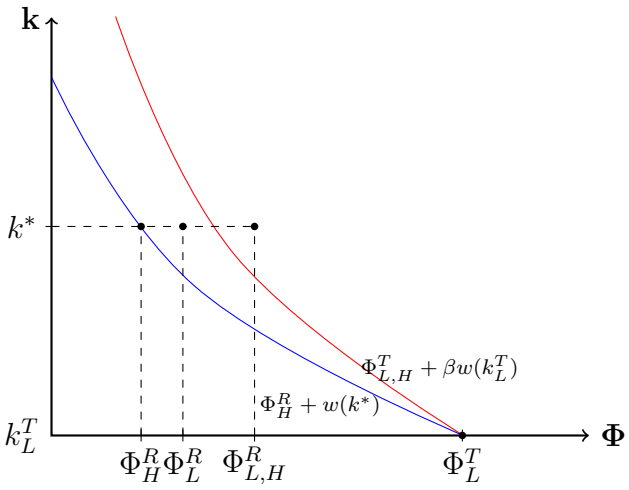

Figure 8: Limited Threats

efficient allocation is not achievable and the optimal allocation will involve distortions.[36]

However, this does not detract from the main message of the paper, which is threat or fooling mechanisms can improve welfare. Even with limitations on the extent of threats or fooling, the government is still able to improve welfare above the constrained efficient optimum by using the mechanisms introduced in this paper.

8.3.2 Outside Commitment Devices

In reality, people with self-control problems can choose from a wide array of commitment devices available in the market, for example, illiquid assets.[37] In the case of sophisticated agents, if commitment devices are available and its usage is unobservable, then threats become much less potent. This is because an agent can purchase illiquid assets and bind himself to an intertemporal allocation. This reduces the effectiveness of a threat, because the ex-post utility is maximized over a smaller consumption set. Therefore, private information could matter when commitment devices for sophisticated or partially naïve agents are available.

However, fooling mechanisms for non-sophisticated agents could override the demand for commitment devices. The government can always choose imaginary allocations that make buying an outside commitment device undesirable.

[36] Similar arguments can be made for fooling and threat mechanisms if the period utility functions were bounded above.

[37] There is also a growing market for commitment devices. For example, StickK, Pact and Beeminder are some recent websites that offer contracts contingent on the completion of stated goals.

8.3.3 Political Economy

In an economy with political constraints, the incentives to be re-elected may affect the set of implementable policies. Even for benevolent political candidates, if the primary goal is to win the election, political incentives would distort the choice of policies.

The intertemporal rate of consumption could be distorted. This is especially true when elections are held after preference reversal. Suppose an election occurs after the agents' preferences change, then the political candidates have an incentive to undo policies that encourage savings. The competition for votes could force the candidates to pander to the voters' desire for present consumption and undermine the implementation of optimal savings policies. The timing of elections has shown to be of crucial importance in models with time-inconsistent voters. Bisin, Lizzeri and Yariv [(2015)] showed how political candidates would exploit the voters' present bias and undo the incentives for private commitment when elections are held in tandem with the intertemporal decisions of the agents. This fact is true regardless of the agents' sophistication level.

9 Summary and Conclusion

In this paper, I demonstrate how government policies can harness the present bias of agents and improve welfare in a Mirrlees taxation model. Contrary to traditional policy proposals, where the primary goal was to mitigate the present bias, I provide methods on utilizing the agents' time-inconsistency to the government's advantage. The methods I developed are also appealing because they can be implemented with familiar policy instruments. For sophisticated time-inconsistent agents, the government can use loans and progressive income repayment plans to increase welfare from the constrained efficient optimum. For naïve agents, the government can use regressive savings subsidy.

The results presented in this paper could be applied to other settings (for example, industrial organization) and to other biases (for example, overconfidence). The concept of fooling and issuing threats could potentially be used in a wider array of mechanism design problems with dynamically inconsistent agents or other biases.

It may seem that this paper is against the enactment of commitment devices that could help agents ameliorate their time-inconsistency, since giving agents access to commitment devices would undermine the government's ability to exploit the bias. However, this is not meant to be the message of the paper. This paper highlights the policy concerns and welfare opportunities of the government when the optimal savings policy for time-inconsistent agents is derived in a richer setting. The focus on savings has obscured other costs associated

with being time-inconsistent, such as inadequate human capital development. There are substantial costs of having time-inconsistent preferences not included in this model.

References

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

Ambrus, Attila and Georgy Egorov, "A Comment on 'Commitment vs. Flexibility'," *Econometrica*, 2013, 81 (5), 2113-2124.

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

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

Armstrong, Mark and Jean-Charles Rochet, "Multi-dimensional screening: A user's guide," *European Economic Review*, 1999, 43, 959-979.

Ausubel, Lawrence, "Adverse Selection in the Credit Card Market," *Unpublished*, 1999.

Bassi, Matteo, "Mirrlees Meets Laibson: Optimal Income Taxation with Bounded Rationality," *CSEF Working Paper*, 2010, No. 266.

Beaudry, Paul, Charles Blackorby and Dezso Szalay, "Taxes and Employment Subsidies in Optimal Redistribution Programs," *American Economic Review*, 2009, 99(1), 216-242.

Benartzi, Shlomo and Richard Thaler, "Naïve Diversification Strategies in Defined Contribution Saving Plans," *American Economic Review*, 2001, 91(1), 79-98.

Bernheim, Douglas and Antonio Rangel, "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, 2004, 94(5), 1558-1590.

Bisin, Alberto, Alessandro Lizzeri and Leeat Yariv, "Government Policy with Time Inconsistent Voters," *American Economic Review*, 2015, 105(6), 1711-1737 .

Bond, Philip and Gustav Sigurdsson, "Commitment Contracts," *Working Paper*, 2015.

Cremer, Helmuth, Pierre Pestieau and Jean-Charles Rochet, "Direct Versus Indirect Taxation: The Design of the Tax Structure Revisited," *International Economic Review*, 2001, 42, 781-799.

Cremer, Helmuth, Pierre Pestieau and Jean-Charles Rochet, "Capital Income Taxation When Inherited Wealth is Not Observable," *Journal of Public Economics*, 2003, 87, 2475-2490.

Choi, James, David Laibson, Brigitte Madrian and Andrew Metrick, "Defined Contribution Pensions: Plan Rules, Participant Decisions, and the Path of Least Resistance," In *Tax Policy and the Economy,* ed. James Poterba, 2002, 67-113.

Chung, Kim-Sau, "How I Learned to Love Being Dynamically Inconsistent," *Working Paper*, 2015.

Dellavigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 2009, 47(2), 315-372.

DellaVigna, Stefano and Ulrike Malmendier, "Paying Not to Go to the Gym," *American Economic Review*, 2006, 96(3), 694-719.

Diamond, Peter and Johannes Spinnewijn, "Capital Income Taxes with Heterogeneous Discount Rates," *American Economic Journal: Economic Policy*, 2011, 3, 52-76.

Eliaz, Kfir and Ran Spiegler, "Contracting with Diversely Naïve Agents," *Review of Economic Studies*, 2006, 73, 689-714.

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

Farhi, Emmanuel and Iván Werning, "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," *Journal of Political Economy*, 2012, 120(3), 398-445.

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

Galperti, Simone, "Commitment, Flexibility, and Optimal Screening of Time Inconsistency," *Econometrica*, 2015, 83(4), 1425-1465.

Gottlieb, Daniel and Kent Smetters, "Lapse-Based Insurance", *Working Paper*, 2013.

Guo, Jang-Ting and Alan Krause, "Dynamic Nonlinear Income Taxation with Quasi-Hyperbolic Discounting and No Commitment," *Journal of Economic Behavior and Organization*, 2015, 109, 101-119.

Heidhues, Paul and Botond Koszegi, "Exploiting Naïveté about Self-Control in the Credit Market," *American Economic Review*, 2010, 100(5), 2279-2303.

Heidhues, Paul and Botond Koszegi, "Futile Attempts at Self-Control," *Journal of the European Economic Association*, 2009, 7(2-3), 423-434.

Hellwig, Martin F., "A Contribution to the Theory of Optimal Utilitarian Income Taxation," *Journal of Public Economics*, 2007, 91, 1449-1477.

Kocherlakota, Narayana, "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 2005, 73(5), 1587-1621.

Koszegi, Botond, "Behavioral Contract Theory," *Journal of Economic Literature*, 2014, 52(4), 1075-1118.

Krusell, Per, Burhanettin Kuruscu and Anthony Smith Jr., "Temptation and Taxation," *Econometrica*, 2010, 78(6), 2063-2084.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 1997, 112(2), 443-477.

Laibson, David, Andrea Repetto and Jeremy Tobacman, "Estimating Discount Functions with Consumption Choices over the Lifecycle," *NBER Working Paper*, 2007, No. 13314.

Loewenstein, George, Ted O'Donoghue and Matthew Rabin, "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics*, 2003, 118, 1209-1248.

Madrian, Brigitte and Dennis Shea, "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 2001, 116(4), 1149-1187.

Mirrlees, James, "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 1971, 38, 175-208.

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

O'Donoghue, Ted and Matthew Rabin, "Choice and Procrastination," *Quarterly Journal of Economics*, 2001, 116, 121-160.

O'Donoghue, Ted and Matthew Rabin, "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes," *American Economic Review Papers and Proceedings*, 2003, 93(2), 186-191.

Prescott, Edward and Robert Townsend, "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard," *Econometrica*, 1984, 52(1), 21-46.

Scholz, John Karl, Ananth Seshadri and Surachai Khitatrakun, "Are Americans Saving 'Optimally' for Retirement?" *Journal of Political Economy*, 2006, 114(4), 607-643.

Shui, Haiyan and Lawrence Ausubel, "Time Inconsistency in the Credit Market," *Working Paper*, 2005.

Spiegler, Rani, "Bounded Rationality and Industrial Organization," *Oxford University Press*, 2011.

Thaler, Richard and Shlomo Benartzi, "Save More Tomorrow: Using Behavioral Economics to Increase Employee Savings," *Journal of Political Economy*, 2004, 112, S164-S187.

Thaler, Richard and H.M. Shefrin, "An Economic Theory of Self-Control," *Journal of Political Economy*, 1981, 89(2), 392-406.

Yu, Pei Cheng, "Non-Linear Pricing with Preference Reversal," *Working Paper*, 2015.

A General Model

The model in this section will outline a general form of dynamic inconsistency. I will examine a less restrictive utility function and discuss both types of partial naïveté. The results presented on the savings problem with present biased agents can then be treated as corollaries to the results presented in this section. The proofs for the results in the paper before Section 5 are included here.

A.1 Setup of Model

Consider an economy with $|N| \geq 2$ goods produced with labor or other goods and a continuum of agents of measure one. There are $|M| \geq 2$ types of agents denoted by the set $\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$. The types are distributed according to $\Pr(\theta = \theta_m) = \pi_m > 0$, for all $\theta_m \in \Theta$ with $\sum_{m=1}^{M} \pi_m = 1$.

The production of good n depends on the labor input l_n and the vector amount of other inputs $\mathbf{x}_n \in \mathbb{R}_+^N$. Let $y_n = F_n(\mathbf{x}_n, l_n; \theta_m) \in \mathbb{R}_+$ denote a continuous and differentiable production process strictly increasing in l_n and $\mathbf{x}_n$ of a type θ_m agent for good n. Let $l_n = G_n(y_n, \mathbf{x}_n; \theta_j)$ denote the inverse of $F_n(\mathbf{x}_n, l_n; \theta_j)$ with fixed input $\mathbf{x}_n$. Each type of agent differs in their labor production efficiency: $F_n(\mathbf{x}_n, l_n; \theta_j) > F_n(\mathbf{x}_n, l_n; \theta_k)$, for any labor input $l_n > 0$ and $\mathbf{x}_n$ and good n with $\theta_j > \theta_k$. If a good does not depend on labor input, then it does not depend on θ. The production of at least one of the goods requires labor. Both the production efficiency of each agent and their labor input $l = (l_1, l_2, \ldots, l_N)$ are not

observable by the government. The government can only observe output $y = (y_1, y_2, \ldots, y_N)$ and input $\mathbf{x}$. With a slight abuse of notation, I will define $G\left(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m\right) \in \mathbb{R}_+^N$ as the labor input vector for a type θ_m agent with reporting strategy $\sigma(\theta_m) = \theta_{\hat{m}}$.

A.1.1 Consumer Utility

The agents' have the following ex-ante utility before consumption

$$U(c, l),$$

where $c = (c_1, c_2, \ldots, c_N)$. The agents' utility changes when they are consuming to the ex-post utility

$$V(c, l).$$

Let $U : \mathbb{R}_+^{2N} \mapsto \mathbb{R}$ and $V : \mathbb{R}_+^{2N} \mapsto \mathbb{R}$ be continuously differentiable and let them be strictly increasing and concave in consumption: $\frac{\partial U}{\partial c_n} > 0, \frac{\partial^2 U}{\partial c_n^2} < 0$ and $\frac{\partial V}{\partial c_n} > 0, \frac{\partial^2 V}{\partial c_n^2} < 0$. Let U and V be strictly decreasing and convex in labor: $\frac{\partial U}{\partial l_n} < 0, \frac{\partial^2 U}{\partial l_n^2} < 0$. Also, there exists $\epsilon > 0$ such that $\frac{\partial U}{\partial c_n}, \frac{\partial V}{\partial c_n} > \epsilon$ for all goods. Finally, for any good $n \in N$, the utility from consumption for both ex-ante and ex-post utility is unbounded below, so $\lim_{c_n \to 0} \frac{\partial U}{\partial c_n} = +\infty$ and $\lim_{c_n \to 0} \frac{\partial V}{\partial c_n} = +\infty$. Hence, an interior solution for consumption is ensured. Also, for any good $n \in N$, let $\lim_{l_n \to 0} \frac{\partial U}{\partial l_n} = 0$ and $\lim_{l_n \to +\infty} \frac{\partial U}{\partial l_n} = +\infty$ so the labor supply is always strictly positive and finite. Finally, I assume additive separability of consumption and leisure, so that the labor decision does not affect the marginal utility or marginal rates of substitution in goods consumption.

I will assume that the utility from consumption is different. More precisely, I assume the marginal rate of substitution for some consumption goods is different for U than for V.

Assumption 1 *There exist $j, k \in N$ such that $\frac{\frac{\partial U}{\partial c_k}}{\frac{\partial U}{\partial c_j}} \neq \frac{\frac{\partial V}{\partial c_k}}{\frac{\partial V}{\partial c_j}}$,*

First notice that since the utility is separable in consumption and labor, Assumption 1 is independent of the agents' labor choice. Assumption 1 along with strictly increasing and concave utility implies a *single crossing condition* on the indifference curves for the ex-ante and ex-post utilities of the two goods, j and k. If the ex-post and ex-ante utilities satisfy Assumption 1, then agents exhibit *preference reversal*. I will impose an additional standard assumption on the agents' preferences.

Assumption 2 *For any good $n \in N$ that depends on labor for production, the ex-ante preferences satisfy the single crossing property: $\frac{\partial}{\partial \theta}\left(-\frac{\frac{\partial U}{\partial y_n}}{\frac{\partial U}{\partial c_n}}\right) < 0$ and $\frac{\partial}{\partial \theta}\left(-\frac{\frac{\partial V}{\partial y_n}}{\frac{\partial V}{\partial c_n}}\right) < 0$*

A.1.2 Types of Non-sophistication

Following Spiegler [(2011)], I will analyze the optimal allocation under two types of partial naïveté: magnitude naïveté and frequency naïveté.

Definition 12 *For some $\hat{\alpha} \in (0,1]$, agents are partially naïve in magnitude if, with probability one, they perceive their ex-post utility to be*

$$W(c,l) = \hat{\alpha}U(c,l) + (1-\hat{\alpha})V(c,l).$$

Definition 12 defines magnitude naïveté. If $\hat{\alpha} < 1$, then agents are certain that their preferences will change. However, since $\hat{\alpha}$ is bounded away from 0, agents underestimate the degree of their preference reversal.

Definition 13 *Agents are partially naïve in frequency if they believe their ex-post utility to be $V(c,l)$ with probability $1-\hat{\alpha}$, where $\hat{\alpha} \in (0,1]$. In other words, let $W(c,l)$ denote the expected ex-post utility of the agent:*

$$W(c,l) = \hat{\alpha}U(c,l) + (1-\hat{\alpha})V(c,l).$$

Definition 13 defines frequency naïveté. In essence, if $\hat{\alpha} < 1$, the agents attach a positive probability to the likelihood of a change in the preference. However, since $\hat{\alpha}$ is bounded away from 0, they underestimate the probability of their preferences changing.

Under both definitions, if $\hat{\alpha} = 1$, the agents are fully naïve, and if $\hat{\alpha} = 0$, the agents are sophisticated. I will refer to $\hat{\alpha}$ as describing the *sophistication level* of an agent.

A.1.3 Timing, Welfare Criterion and Other Assumptions

The timing of the model is shown in Figure 9 and is the same as in the previous sections.

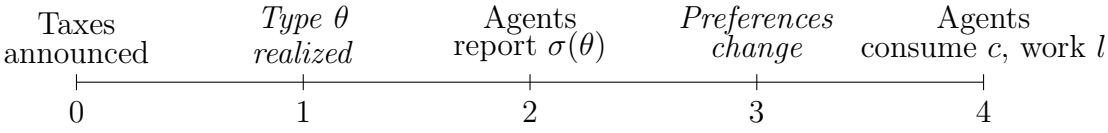

Figure 9: Timing of Events

Similar to the savings problem, the government has full commitment so the revelation principle applies and the analysis proceeds by using a direct mechanism as illustrated in Yu [(2015)].

The government evaluates allocations at date 0 using the welfare criterion presented in Section 2,

$$\sum_{m=1}^{M} \pi_m \left[\kappa U\left(c(\theta_m), l(\theta_m)\right) + (1-\kappa) V\left(c(\theta_m), l(\theta_m)\right)\right], \tag{15}$$

where $(c(\theta_m), l(\theta_m))$ denotes the allocation type m agent consumes, and let $\kappa \in (0, 1]$.

I will continue to assume that the agents do not have access to an insurance market for skill realizations nor does a market for commitment device exists.

A.1.4 The No Private Information Case

Without private information, the government maximizes social welfare (15) subject to the feasibility constraint

$$\sum_{m=1}^{M} \pi_m \left[F_n\left(\mathbf{x}_n(\theta_m), l_n(\theta_m); \theta_m\right) - c_n(\theta_m)\right] = 0, \forall n \in N. \tag{16}$$

As was the case in the previous section, with complete information, the agents work according to their skill type: more productive agents work more than the less productive agents. The government is also able to achieve full insurance without distortions. For consumption smoothing, the government chooses an appropriate linear tax to correct the distortion caused by the preference reversal. The implementation does not change with regards to the agents' sophistication level.

A.2 The Effects of Non-sophistication

With non-sophisticated agents, the government issues the following menu

$$\left\{\left(c^R(\theta_m), y^R(\theta_m), \mathbf{x}^R(\theta_m)\right), \left(c^I(\theta_m), y^I(\theta_m), \mathbf{x}^I(\theta_m)\right)\right\}_{\theta_m \in \Theta}.$$

Set $y^R(\theta_m) = y^I(\theta_m) = y(\theta_m)$ and $\mathbf{x}^R(\theta_m) = \mathbf{x}^I(\theta_m) = \mathbf{x}(\theta_m)$, since $c^R(\theta) \neq c^I(\theta)$ is enough to exploit non-sophistication. Denote $l(\theta_m) = G\left(y(\theta_m), \mathbf{x}(\theta_m); \theta_m\right)$ as the labor supply under truth-telling.

For magnitude naïveté, agents make their reporting decision based on $U(c, l)$ while anticipating a taste change of $W(c, l)$. Therefore, they require $c^I(\theta)$ to be more appealing than $c^R(\theta)$ under $W(c, l)$, and the reporting strategy is evaluated using $U(c, l)$. While for frequency naïveté, agents make their reporting decision based on their expected ex-post utility $W(c, l)$. More specifically, they require $c^I(\theta)$ to be more appealing than $c^R(\theta)$ under $U(c, l)$, and the reporting strategy is evaluated at the expected utility of $U(c, l)$.

A.2.1 Planning Problem with Magnitude Naïveté

Under magnitude naïveté, the government's problem is to choose an allocation menu $\left\{c^R(\theta_m), c^I(\theta_m), y(\theta_m), \mathbf{x}(\theta_m)\right\}_{m=1}^M$ to maximize (15) subject to the feasibility constraint (16) evaluated at the real allocations and

$$U\left(c^I(\theta_m), l(\theta_m)\right) \geq U\left[c^I(\theta_{\hat{m}}), G\left(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m\right)\right], \forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}, \tag{17}$$

$$W\left(c^I(\theta_m), l(\theta_m)\right) \geq W\left(c^R(\theta_m), l(\theta_m)\right), \forall \theta_m \in \Theta, \tag{18}$$

$$V\left(c^R(\theta_m), l(\theta_m)\right) \geq V\left(c^I(\theta_m), l(\theta_m)\right), \forall \theta_m \in \Theta. \tag{19}$$

Constraint (17) is the incentive compatibility constraint. Constraints (18) and (19) are the fooling and executability constraints.

A.2.2 Planning Problem with Frequency Naïveté

For frequency naïveté, the government chooses $\left\{c^R(\theta_m), c^I(\theta_m), y(\theta_m), \mathbf{x}(\theta_m)\right\}_{m=1}^M$ to maximize (15) subject to the feasibility constraint (16) evaluated at the real allocations and the fooling constraint (19) with the following incentive compatibility constraint $\forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}$,

$$\begin{aligned} &\hat{\alpha} U\left(c^I(\theta_m), l(\theta_m)\right) + (1-\hat{\alpha}) U\left(c^R(\theta_m), l(\theta_m)\right) \\ &\geq \hat{\alpha} U\left[c^I(\theta_{\hat{m}}), G\left(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m\right)\right] + (1-\hat{\alpha}) U\left[c^R(\theta_{\hat{m}}), G\left(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m\right)\right], \end{aligned} \tag{20}$$

and the fooling constraint for the imaginary allocation

$$U\left(c^I(\theta_m), l(\theta_m)\right) \geq U\left(c^R(\theta_m), l(\theta_m)\right), \forall \theta_m \in \Theta. \tag{21}$$

The difference between frequency naïveté and magnitude naïveté lies in the beliefs of the future preference. In frequency naïveté, the agents believe with some probability $\hat{\alpha}$ that they will choose the imaginary allocation evaluated at the ex-ante preference $U(c, l)$, which is represented in (21). This is different from the fooling constraint (18) for magnitude naïveté, where the agents are certain that their preferences would change, but underestimate the extent of this shift.

A.2.3 Result for Non-sophisticated Agents

By Assumption 1 and Assumption 2, it can be shown that the government can achieve the efficient allocation.

Proposition 5 *The optimal allocation for the environment where agents have private information on productivity and are fully naïve or partially naïve in magnitude or frequency about their preference changes is the efficient allocation.*

Proof Let $\{c^*(\theta_m), y^*(\theta_m), \mathbf{x}^*(\theta_m)\}_{\theta_m \in \Theta}$ denote the set of efficient allocations, and

$$U^*(\theta_m) = U\left(c^*(\theta_m), l^*(\theta_m)\right),$$

$$V^*(\theta_m) = V\left(c^*(\theta_m), l^*(\theta_m)\right),$$

where $l^*(\theta_m)$ denotes the efficient labor supply when the agents are truthful. Let the real allocation be the efficient allocation. It suffices to show that the efficient allocation can be supported by imaginary allocations that satisfy the fooling and incentive compatibility constraints. I will first prove the case for magnitude naïveté.

The set of incentive compatibility constraints can be rewritten as, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$U\left(c^I(\theta_m), l^*(\theta_m)\right) \geq U\left(c^I(\theta_{\hat{m}}), l^*(\theta_{\hat{m}})\right) + \Delta(\theta_{\hat{m}}; \theta_m),$$

where $\Delta(\theta_{\hat{m}}; \theta_m) \geq 0$ denotes the minimum information rent that prevents a θ_m type from misreporting to be a $\theta_{\hat{m}}$ type. Since labor is additively separable from consumption in utility and output is fixed at the efficient level, $\Delta(\theta_{\hat{m}}; \theta_m)$ is only a function of types θ_m and $\theta_{\hat{m}}$. Since there exists $\epsilon > 0$ such that $\frac{\partial U}{\partial c_n} > \epsilon$ for all goods, it is possible to find $c^I(\theta_m)$ such that the incentive compatibility constraint is satisfied for any $\Delta(\theta_{\hat{m}}; \theta_m)$. By Assumption 2, since the efficient labor supply is strictly increasing with productivity, the ex-ante utility evaluated at the imaginary allocations also need to be strictly increasing, which can be constructed. This implies that it is sufficient to focus on local downward incentive compatibility constraints.

Let $c^I(\theta_m)$ be chosen so that $\forall \theta_m \in \Theta$,

$$V\left(c^I(\theta_m), l^*(\theta_m)\right) = V^*(\theta_m). \tag{22}$$

Hence, for the magnitude naïveté case, if the incentive compatibility constraints are satisfied, then the fooling constraints are immediately satisfied as well.

I will now show that for any given $c^I(\theta_{m-1})$ and for an arbitrary $\Delta(\theta_{m-1}; \theta_m) \in \mathbb{R}_+$, it is possible for the efficient allocations to satisfy the incentive compatibility constraints. This is akin to showing that there does not exist a finite solution to the following programming problem

$$\max_{c^I(\theta_m)} U\left(c^I(\theta_m), l^*(\theta_m)\right)$$

subject to (22). The proof will proceed by contradiction.

Suppose there exists a finite $\hat{c}^I(\theta_m)$ such that it solves the programming problem above. Since the utility functions are unbounded below and $\hat{c}^I(\theta_m)$ is finite, it must be the case that $\hat{c}^I(\theta_m) \in \mathbb{R}^N_{++}$. By Assumption 1 and without loss of generality, there exists $j, k \in N$ such that

$$\frac{\frac{\partial U}{\partial \hat{c}^I_k(\theta_m)}}{\frac{\partial U}{\partial \hat{c}^I_j(\theta_m)}} > \frac{\frac{\partial V}{\partial \hat{c}^I_k(\theta_m)}}{\frac{\partial V}{\partial \hat{c}^I_j(\theta_m)}}.$$

For any $\epsilon > 0$, choose a new imaginary consumption allocation $\hat{c}^{I'}(\theta_m)$, where

$$\hat{c}^{I'}_k(\theta_m) = \hat{c}^I_k(\theta_m) + \epsilon$$

and

$$\hat{c}^{I'}_j(\theta_m) = \hat{c}^I_j(\theta_m) - \frac{\frac{\partial V}{\partial \hat{c}^I_k(\theta_m)}\epsilon}{\frac{\partial V}{\partial \hat{c}^I_j(\theta_m)}}.$$

This leaves the ex-post utility unchanged: $V\left(c^{I'}(\theta_m), l^*(\theta_m)\right) = V^*(\theta_m)$. However, this new imaginary allocation strictly increases the ex-ante utility:

$$U\left(c^{I'}(\theta_m), l^*(\theta_m)\right) \approx U\left(c^I(\theta_m), l^*(\theta_m)\right) + \frac{\partial U}{\partial \hat{c}^I_k(\theta_m)}\epsilon - \frac{\partial U}{\partial \hat{c}^I_j(\theta_m)}\left(\frac{\frac{\partial V}{\partial \hat{c}^I_k(\theta_m)}\epsilon}{\frac{\partial V}{\partial \hat{c}^I_j(\theta_m)}}\right),$$

which due to preference reversal,

$$\frac{\partial U}{\partial \hat{c}^I_k(\theta_m)}\epsilon - \frac{\partial U}{\partial \hat{c}^I_j(\theta_m)}\left(\frac{\frac{\partial V}{\partial \hat{c}^I_k(\theta_m)}\epsilon}{\frac{\partial V}{\partial \hat{c}^I_j(\theta_m)}}\right) > 0.$$

This is a contradiction. Therefore, there does not exist a finite solution to the programming problem above. As a result, the efficient allocation is implementable for magnitude naïveté regardless of the amount of information rent $\Delta(\theta_{m-1}; \theta_m)$.

It is now straightforward to prove for the case of frequency naïveté. Again, choose the imaginary consumption such that $V\left(c^I(\theta_m), l^*(\theta_m)\right) = V^*(\theta_m)$. By Assumption 2, the labor supply is strictly increasing with productivity, so the ex-ante utility evaluated at the imaginary consumption can be constructed to be strictly increasing as well. Thus, it suffices to focus on local downward incentive compatibility constraints, which can be rewritten as $\forall \theta_m \in \Theta \backslash \theta_1$,

$$\hat{\alpha} U\left(c^I(\theta_m), l^*(\theta_m)\right) + (1-\hat{\alpha}) U\left(c^*(\theta_m), l^*(\theta_m)\right) \geq U\left(c^*(\theta_m), l^*(\theta_m)\right) + \Delta(\theta_{m-1}; \theta_m).$$

Notice that if it is possible to choose the imaginary consumption so that the incentive compatibility constraints are satisfied, then it immediately follows that the fooling constraints are satisfied. Hence, it suffices to show that the same programming problem shown above

does not have any finite solutions. This completes the proof. ■

Proposition 5 states that the private information problem can be alleviated if the agents are not sophisticated regardless of the type of naïveté. This is accomplished by using a fooling mechanism.

Corollary 4 *If $\hat{\alpha} > 0$, it is optimal for the government to implement a fooling mechanism.*

Proof By Proposition 5, the government can implement the efficient allocation. Let $\left\{\left(c^R(\theta_m), y(\theta_m)\right), \mathbf{x}(\theta_m)\right\}_{\theta_m \in \Theta}$ be the efficient allocation. Suppose the government does not fool the agents, then by definition, for all productivity types, agents evaluate the incentive compatibility constraints at the real allocation. This violates incentive compatibility if the real allocations are the efficient allocations. It follows that for some types, the agents must use the imaginary allocations to evaluate incentive compatibility and thus the government must implement a fooling mechanism to achieve the efficient allocation. ■

To exploit the agents' naïveté, the government loads the information rent on goods that they value during the reporting stage, but would not value as much relative to other goods after the preference change. By Assumption 1, suppose $\frac{\frac{\partial U}{\partial c_k}}{\frac{\partial U}{\partial c_j}} > \frac{\frac{\partial V}{\partial c_k}}{\frac{\partial V}{\partial c_j}}$, then the agents value good k more than good j at the reporting stage. The government can then promise more of good k than good j for the imaginary allocations to elicit truthful reports as long as the agents hold the wrong beliefs. However, after the preference change, the promise of more good k is less appealing, and the agents would no longer choose the imaginary allocations but the real allocations, with less of good k.

With magnitude naïveté, the government is able to achieve full information efficient welfare for any sophistication level $\hat{\alpha} \in (0, 1]$. However, with fully sophisticated agents ($\hat{\alpha} = 0$), the fooling mechanism can only implement the constrained efficient optimum. Similar discontinuities in welfare can also be found in analysis involving magnitude naïveté, as in Heidhues and Koszegi [(2010)]. A more surprising result for partial naïveté is that this discontinuity is present for naïveté in frequency as well. Spiegler [(2011)] has shown the optimal contract to be continuous with respect to cognitive limitations in a second-degree price discrimination setting for frequency naïveté. However, Proposition 5 shows that this continuity result does not hold in the optimal taxation setting even with naïveté in frequency.

A.3 The Effects of Sophistication

The government introduces the following menu for sophisticated agents

$$\left\{\left(c^R(\theta_m), y^R(\theta_m), \mathbf{x}^R(\theta_m)\right), \left(c^T(\theta_m), y^T(\theta_m), \mathbf{x}^T(\theta_m)\right)\right\}_{\theta_m \in \Theta}.$$

Denote $l^R(\theta_m) = G\left(y^R(\theta_m), \mathbf{x}^R(\theta_m); \theta_m\right)$ and $l^T(\theta_m) = G\left(y^T(\theta_m), \mathbf{x}^T(\theta_m); \theta_m\right)$ as the labor supply under truth-telling for the real and threat allocations respectively.

The government maximizes (15) subject to the feasibility constraint (16) evaluated at the real allocations and the following incentive compatibility constraints, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$U\left(c^R(\theta_m), l^R(\theta_m)\right) \geq U\left[c^T(\theta_{\hat{m}}), G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right], \tag{23}$$

and the credible threat constraints $\forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}$,

$$V\left(c^R(\theta_m), l^R(\theta_m)\right) \geq V\left(c^T(\theta_m), l^T(\theta_m)\right), \tag{24}$$

$$V\left[c^T(\theta_{\hat{m}}), G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right] \geq V\left[c^R(\theta_{\hat{m}}), G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right]. \tag{25}$$

A.3.1 Result for Sophisticated Agents

Using Assumption 1 and Assumption 2, the following proposition shows that private information does not matter in an environment with sophisticated agents.

Proposition 6 *The optimal allocation for the environment where agents have private information on productivity and are sophisticated is the efficient allocation.*

Proof Let $\{c^*(\theta_m), y^*(\theta_m), \mathbf{x}^*(\theta_m)\}_{\theta_m \in \Theta}$ denote the set of efficient allocations. Let the real allocations be the efficient allocations. Since the utility functions are additively separable in consumption and labor, let

$$V(c, l) = v(c) - h(l),$$

and

$$U(c, l) = u(c) - h(l).$$

I will first deter local downward misreporting for type θ_2, and then show how global downward misreports can be prevented by deterring local downward misreports. Note that type θ_1 does not need to be deterred from misreporting when the real allocation is the efficient allocation. Choose the threat allocation so that the following holds

$$v\left(c^T(\theta_1)\right) - h\left[G\left(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2\right)\right] = v\left(c^*(\theta_1)\right) - h\left[G\left(y^*(\theta_1), \mathbf{x}^*(\theta_1); \theta_2\right)\right]. \tag{26}$$

By Assumption 2, labor supply for the threat allocation can be increased with a corresponding increase in utility from threat consumption such that (26) holds and productivity type θ_1 agent would not choose the threat allocation. To see this, notice that the following problem has no finite solution (the government is choosing a utility level $v\left(c^T(\theta_1)\right)$, not the consumption bundle $c^T(\theta_1)$)

$$\min_{v(c^T(\theta_1)),l^T(\theta_1)} v\left(c^T(\theta_1)\right) - h\left(l^T(\theta_1)\right) \text{ s.t. (26)}.$$

Hence, it is possible to find $\left(v\left(c^T(\theta_1)\right), l^T(\theta_1)\right)$ so that, along with (26), the following holds

$$v\left(c^*(\theta_1)\right) - h\left(l^*(\theta_1)\right) \geq v\left(c^T(\theta_1)\right) - h\left(l^T(\theta_1)\right).$$

Thus, by construction, the credible threat constraints are satisfied with $v\left(c^T(\theta_1)\right) > v\left(c^*(\theta_1)\right)$ and $l^T(\theta_1) > l^*(\theta_1)$.

Next, fix the utility derived from the threat allocation $v\left(c^T(\theta_1)\right)$ and inputs $l^T(\theta_1)$ and $\mathbf{x}^T(\theta_1)$ such that all credible threat constraints are satisfied. Hence, all that remains is to show the threat consumption can be chosen to satisfy the incentive compatibility constraint. It is always possible to find such a threat consumption if the following programming problem yields no finite solutions for any given $\Omega(\theta_1;\theta_2) > 0$,

$$\min_{c^T(\theta_1)} u\left(c^T(\theta_1)\right)$$

subject to

$$v\left(c^T(\theta_1)\right) = v\left(c^*(\theta_1)\right) + \Omega(\theta_1;\theta_2),$$

where the constraint is a rewriting of (26), in essence, $\Omega(\theta_1;\theta_2) = h\left[G\left(y^T(\theta_1), \mathbf{x}^T(\theta_1);\theta_2\right)\right] - h\left[G\left(y^*(\theta_1), \mathbf{x}^*(\theta_1);\theta_2\right)\right]$.

By Assumption 1, it can be shown that the programming problem does not have a finite solution using the arguments established for the proof of Proposition 5. Hence, it is possible to choose a set of threat allocation in the menu for type θ_1 such that local downward incentive constraints and credible threat constraints hold.

By the convexity of $h(\cdot)$ and the fact that $l^T(\theta_1) > l^*(\theta_1)$, for any $\theta_m > \theta_2$ the following relationship holds: $\Omega(\theta_1;\theta_2) > \Omega(\theta_1,\theta_m)$. Hence, more productive agents misreporting as a θ_1 agent would also choose the threat allocation. As a result, if $c^T(\theta_1)$ is chosen so that

$$u\left(c^T(\theta_1)\right) \leq \min_{\theta_m > \theta_1} \left[u\left(c^*(\theta_m)\right) + h\left[G\left(y^T(\theta_1), \mathbf{x}^T(\theta_1);\theta_m\right)\right] - h\left(l^*(\theta_m)\right)\right],$$

then all types are deterred from misreporting as θ_1. Hence, to deter type θ_3 from local downward misreporting, a similar process can be used to find the threat allocation. Choose the threat allocation $\left(c^T(\theta_2), y^T(\theta_2), \mathbf{x}^T(\theta_2)\right)$ so that the following holds

$$\begin{aligned} &v\left(c^T(\theta_2)\right) - h\left[G\left(y^T(\theta_2), \mathbf{x}^T(\theta_2); \theta_3\right)\right] \\ &= \max\left\{v\left(c^*(\theta_2)\right) - h\left[G\left(y^*(\theta_2), \mathbf{x}^*(\theta_2); \theta_3\right)\right], v\left(c^T(\theta_1)\right) - h\left[G\left(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_3\right)\right]\right\}. \end{aligned}$$

Using a similar argument, the threat allocation can be chosen so that the credible threat constraints hold and the incentive compatibility constraints for all types are satisfied.

Finally, by induction, it is possible to achieve global incentive compatibility with credible threats. Hence, the efficient allocation is implementable. ■

The efficient allocation is achieved by using a threat mechanism for sophisticated agents. Since as long as agents are not fully naïve, they are either fully or partially aware of their time-inconsistency problem ($\hat{\alpha} < 1$) and a threat mechanism can potentially be utilized.

For magnitude naïveté, the threats are evaluated using the erroneous ex-post utility W, so the credible threat constraints $\forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}$ are

$$W\left(c^R(\theta_m), l^R(\theta_m)\right) \geq W\left(c^T(\theta_m), l^T(\theta_m)\right),$$

$$W\left[c^T(\theta_{\hat{m}}), G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right] \geq W\left[c^R(\theta_{\hat{m}}), G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right],$$

and the executability constraint

$$V\left(c^R(\theta_m), l^R(\theta_m)\right) \geq V\left(c^T(\theta_m), l^T(\theta_m)\right), \forall \theta_m \in \Theta. \tag{27}$$

For frequency naïveté, the credible threat constraints remain the same.[38] However, the incentive compatibility constraints $\forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}$ are

$$\begin{aligned} U\left(c^R(\theta_m), l^R(\theta_m)\right) \geq &\hat{\alpha} U\left[c^R(\theta_{\hat{m}}), G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right] \\ &+ (1-\hat{\alpha}) U\left[c^T(\theta_{\hat{m}}), G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right]. \end{aligned}$$

The threat allocation makes sure that the partially naïve agent does not think misreporting is worth the risk of consuming the threat allocation. The following corollary shows that a threat mechanism works on partially naïve agents.

[38] I require that a truthful agent would prefer the real allocation regardless of which utility it is evaluated at, while a misreporting agent would prefer the threat allocation regardless of the utility it is evaluated at.

Corollary 5 *It is optimal to threaten the agents when* $\hat{\alpha} < 1$.

Proof Let $\{c^*(\theta_m), y^*(\theta_m), \mathbf{x}^*(\theta_m)\}_{\theta_m \in \Theta}$ denote the set of efficient allocations, and set the real allocations to be the efficient allocations. Set both the threat and real inputs be at the efficient level. Also, let $l^*(\theta_m)$ and $l^T(\theta_m)$ denote the efficient labor supply under truth-telling for real and threat allocations respectively. I will first show the result for magnitude naïveté.

First, examine downward misreporting for type θ_2. Choose the threat allocation so that the following holds

$$\min_{\theta_m} \{u\left(c^*(\theta_m)\right) - h(l^*(\theta_m))\} = u\left(c^T(\theta_1)\right) - h\left[G\left(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2\right)\right]. \tag{28}$$

Notice that the following programming problem has no finite solution:

$$\min_{c^T(\theta_1), y^T(\theta_1)} v\left(c^T(\theta_1)\right) - h\left[G\left(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_1\right)\right]$$

subject to

$$\begin{aligned} &\hat{\alpha} u\left(c^T(\theta_1)\right) + (1-\hat{\alpha}) v\left(c^T(\theta_1)\right) - h\left[G\left(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2\right)\right] \\ &\geq \hat{\alpha} u\left(c^*(\theta_1)\right) + (1-\hat{\alpha}) v\left(c^*(\theta_1)\right) - h\left[G\left(y^*(\theta_1), \mathbf{x}^*(\theta_1); \theta_2\right)\right]. \end{aligned} \tag{29}$$

and (28). To see why this is the case, suppose there exists a finite solution $\left(\hat{c}^T(\theta_1), \hat{y}^T(\theta_1)\right)$, then (29) can be rewritten as

$$v\left(\hat{c}^T(\theta_1)\right) - h\left[G\left(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2\right)\right] = \Psi(\theta_1, \theta_2),$$

where Ψ only depends on the efficient allocation for types θ_1 and θ_2 and the sophistication level $\hat{\alpha}$. This implies that

$$\hat{y}^T(\theta_1) = \arg\min h\left[G\left(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2\right)\right] - h\left[G\left(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_1\right)\right] \text{ s.t. (28)},$$

which is a contradiction.[39] Hence, it is possible to choose the threat allocations to satisfy the executability, local downward incentive compatibility and credible threat constraints.

Once it is established that type θ_2 is deterred from choosing $\left(c^*(\theta_1), y^*(\theta_1), \mathbf{x}^*(\theta_1)\right)$, higher types can be deterred. This can be done by choosing $c^T(\theta_1)$ so that the ex-ante utility is sufficiently low while the ex-post utility is sufficiently high, and then adjusting the output so

[39] Since increasing the threat output such that $h\left[G\left(\tilde{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2\right)\right] = h\left[G\left(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2\right)\right] + \epsilon$, would further decrease $h\left[G\left(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2\right)\right] - h\left[G\left(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_1\right)\right]$. By Assumption 1, it is possible to choose $\tilde{c}^T(\theta_1)$ such that (28) holds and $v\left(\tilde{c}^T(\theta_1)\right)$ satisfies (28).

the credible threat and executability constraints hold. Using a similar argument above, by induction, the efficient allocation is implementable using a threat mechanism for magnitude naïveté.

Now, I will analyze the case for frequency naïveté. For frequency naïveté, all constraints are the same as in the magnitude naïveté case, except for the incentive compatibility constraints. However, since it is possible to arbitrarily decrease the ex-ante utility evaluated at the threat allocation, the argument for the magnitude naïveté holds for frequency naïveté as well. This completes the proof. ■

A.4 Model with Diversely Naïve Agents

When agents are heterogeneous in productivity and sophistication level, the government is still capable of achieving the efficient allocation. The following lemma demonstrates a monotonicity result for fooling and threat mechanisms similar to the one in Section 5.

Lemma 2 *A fooling mechanism that is effective for agents of sophistication level $\acute{\alpha}$ is also effective for more naïve agents. A threat mechanism that is effective for agents of sophistication level $\grave{\alpha}$ is also effective for more sophisticated agents.*

Proof First, look at fooling mechanisms for magnitude naïveté. If a fooling mechanism

$$\left\{\left(c^R(\theta_m), y(\theta_m), \mathbf{x}(\theta_m)\right), \left(c^I(\theta_m), y(\theta_m), \mathbf{x}(\theta_m)\right)\right\}_{\theta_m \in \Theta}$$

is effective for an agent with sophistication level $\acute{\alpha}$ of magnitude naïveté, then the following fooling constraint must be satisfied

$$\begin{aligned} &\acute{\alpha} U\left(c^I(\theta_m), l(\theta_m)\right) + (1-\acute{\alpha})\, V\left(c^I(\theta_m), l(\theta_m)\right) \\ &\geq \acute{\alpha} U\left(c^R(\theta_m), l(\theta_m)\right) + (1-\acute{\alpha})\, V\left(c^R(\theta_m), l(\theta_m)\right), \end{aligned}$$

where $l(\theta)$ denotes the efficient labor supply when the agent is truthful. Since the executability constraint is also satisfied, for any $\hat{\alpha} > \acute{\alpha}$, the fooling constraint is relaxed and all other constraints are unchanged. Hence, a fooling mechanism that is effective for $\acute{\alpha}$ is effective for more naïve agents.

Next, fooling mechanisms for frequency naïveté also have the same form as the mechanism above. Suppose the fooling mechanism is effective for an agent with sophistication level $\acute{\alpha}$

of frequency naïveté, then the following incentive compatibility constraint must be satisfied

$$\begin{aligned}&\acute{\alpha}U\left(c^I(\theta_m), l(\theta_m)\right) + (1-\acute{\alpha})U\left(c^R(\theta_m), l(\theta_m)\right)\\ &\geq \acute{\alpha}U\left[c^I(\theta_{\hat{m}}), G\left(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m\right)\right] + (1-\acute{\alpha})U\left[c^R(\theta_{\hat{m}}), G\left(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m\right)\right].\end{aligned}$$

Due to separability between consumption and effort, the utility function can be written as

$$U(c, l) = u(c) - h(l),$$

so the incentive compatibility constraint becomes

$$\begin{aligned}&\acute{\alpha}\left[u\left(c^I(\theta_m)\right) - u\left(c^I(\theta_{\hat{m}})\right)\right] + (1-\acute{\alpha})\left[u\left(c^R(\theta_m)\right) - u\left(c^R(\theta_{\hat{m}})\right)\right]\\ &\geq h(l(\theta_m)) - h\left[G\left(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m\right)\right].\end{aligned}$$

Since the social welfare function is strictly concave, the efficient allocation would provide full insurance: $c^R(\theta_m) = c^R(\theta_{\hat{m}})$ for all $\theta_m \neq \theta_{\hat{m}}$. As a result, for any $\theta_m, \theta_{\hat{m}} \in \Theta$ such that $\theta_m > \theta_{\hat{m}}$ it must be that $c^I(\theta_m) > c^I(\theta_{\hat{m}})$ for the incentive constraint to hold. The incentive compatibility constraint holds trivially for $\theta_m < \theta_{\hat{m}}$ of any sophistication level. It is easy to see that for any agent with $\hat{\alpha} > \acute{\alpha}$, the incentive compatibility constraint holds as well. Since all other constraints are invariant to the sophistication level, a fooling mechanism that is effective for sophistication level $\acute{\alpha}$ is effective for more naïve agents.

Finally, for threat mechanisms, the analysis will begin by analyzing frequency naïveté. If a threat mechanism

$$\left\{\left(c^R(\theta_m), y^R(\theta_m), \mathbf{x}^R(\theta_m)\right), \left(c^T(\theta_m), y^T(\theta_m), \mathbf{x}^T(\theta_m)\right)\right\}_{\theta_m \in \Theta}.$$

is effective for an agent with sophistication level $\grave{\alpha}$ of frequency naïveté, then the incentive compatibility constraint must be satisfied

$$\begin{aligned}U\left(c^R(\theta_m), l^R(\theta_m)\right) \geq &\grave{\alpha}U\left[c^R(\theta_{\hat{m}}), G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right]\\ &+ (1-\grave{\alpha})\,U\left[c^T(\theta_{\hat{m}}), G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right],\end{aligned}$$

where $l^R(\theta)$ denotes the efficient labor supply under the real allocation when the agent is truthful. Note that for $\theta_m > \theta_{\hat{m}}$ to implement the efficient allocation, it must be that

$$U\left[c^R(\theta_{\hat{m}}), G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right] > U\left[c^T(\theta_{\hat{m}}), G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right].$$

If not, then the incentive compatibility constraint would not hold.[40] As a result, for any agent with $\hat{\alpha} < \grave{\alpha}$, the incentive compatibility constraint would hold as well. The incentive compatibility constraint would hold trivially for $\theta_m < \theta_{\hat{m}}$ of any sophistication level. The other constraints do not depend on $\hat{\alpha}$, so a threat mechanism that is effective for sophistication level $\grave{\alpha}$ would also be effective for more sophisticated agents.

A threat mechanism for magnitude naïveté agents with sophistication level $\grave{\alpha}$ would also have the same policy menu as the threat mechanism for frequency naïveté. The credible threat constraints are

$$\begin{aligned}
&\grave{\alpha} U\left(c^R(\theta_m), l^R(\theta_m)\right) + (1-\grave{\alpha})\, V\left(c^R(\theta_m), l^R(\theta_m)\right) \\
&\geq \grave{\alpha} U\left(c^T(\theta_m), l^T(\theta_m)\right) + (1-\grave{\alpha})\, V\left(c^T(\theta_m), l^T(\theta_m)\right),
\end{aligned}$$

$$\begin{aligned}
&\grave{\alpha} U\left[c^T(\theta_{\hat{m}}), G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right] + (1-\grave{\alpha})\, V\left[c^T(\theta_{\hat{m}}), G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right] \\
&\geq \grave{\alpha} U\left[c^R(\theta_{\hat{m}}), G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right] + (1-\grave{\alpha})\, V\left[c^R(\theta_{\hat{m}}), G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right],
\end{aligned}$$

where $l^T(\theta)$ denotes the efficient labor supply under the threat allocation when the agent is truthful. The executability constraint ensures that for any $\hat{\alpha} < \grave{\alpha}$, a truth-telling agent would always prefer the real allocations over the threat allocations. Separability between consumption and effort helps check that the other credible threat constraint (a misreporting agent would always choose the threat allocation over the real allocation) works for more sophisticated agents as well. Let

$$V(c, l) = v(c) - h(l),$$

so the credible threat constraint can be rewritten as

$$\begin{aligned}
&\grave{\alpha}\left[u\left(c^R(\theta_{\hat{m}})\right) - u\left(c^T(\theta_{\hat{m}})\right)\right] + (1-\grave{\alpha})\left[v\left(c^R(\theta_{\hat{m}})\right) - v\left(c^T(\theta_{\hat{m}})\right)\right] \\
&\leq h\left[G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right] - h\left[G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right].
\end{aligned}$$

Notice that the incentive compatibility implies

$$u\left(c^R(\theta_m)\right) - u\left(c^T(\theta_{\hat{m}})\right) \geq h\left(l^R(\theta_m)\right) - h\left[G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right].$$

[40] This is because $U\left(c^R(\theta_m), l^R(\theta_m)\right) < U\left[c^R(\theta_{\hat{m}}), G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right]$ at the efficient allocation.

The mechanism is effective which implies that it has full insurance, $c^R(\theta_m) = c^R(\theta_{\hat{m}})$, and for any $\theta_m > \theta_{\hat{m}}$,

$$h\left(l^R(\theta_m)\right) > h\left[G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right].$$

Hence, $u\left(c^R(\theta_m)\right) - u\left(c^T(\theta_{\hat{m}})\right) > h\left[G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right] - h\left[G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right]$, so it must be the case that

$$v\left(c^R(\theta_{\hat{m}})\right) - v\left(c^T(\theta_{\hat{m}})\right) < h\left[G\left(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m\right)\right] - h\left[G\left(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m\right)\right].$$

As a result, for any agent with $\hat{\alpha} < \grave{\alpha}$, a misreporting agent would also choose the threat allocation over the real allocation. This proves that a threat mechanism that is effective for sophistication level $\grave{\alpha}$ is also effective for more sophisticated agents. ■

With Lemma 2, the government can target sophistication level, $\tilde{\alpha}$, such that all agents who are more sophisticated than $\tilde{\alpha}$ are threatened by using the same threat mechanism and those who are more naïve are fooled by using the same fooling mechanism. The government targets sophistiction level $\tilde{\alpha} \in (0, 1)$ and introduce the following hybrid mechanism

$$\left\{\left(c^R(\theta_m), y^R(\theta_m), \mathbf{x}^R(\theta_m)\right), \left(c^I(\theta_m), y^I(\theta_m), \mathbf{x}^I(\theta_m)\right), \left(\mathbf{c}^T(\theta_m), \mathbf{y}^T(\theta_m), \mathbf{x}^T(\theta_m)\right)\right\}_{\theta_m \in \Theta}.$$

The imaginary and threat allocations are chosen such that agents with sophistication level $\tilde{\alpha}$ are fooled and threatened with effective mechanisms. It can be shown that the fooling and the threat components do not interact in a hybrid mechanism and full efficiency is achievable.

Proposition 7 *The optimal allocation for the environment with diversely naïve agents where agents have private information on productivity and sophistication level is the efficient allocation.*

Proof The result follows immediately from Lemma 2. ■

A.4.1 Model with Diversely Time-Inconsistent Agents

To model agents with varying degrees of time-inconsistency, let $\alpha \in [\underline{\alpha}, \overline{\alpha}] \subset [0, 1]$ denote the degree of preference reversal (temptation level) of the agents, where $\underline{\alpha} = 0$ and $\overline{\alpha} < 1$.

Definition 14 *An agent with preference reversal $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ has ex-post utility*

$$V(c, l; \alpha) = \alpha U(c, l) + (1 - \alpha)\, \underline{V}(c, l),$$

where $\underline{V}(c,l) \neq U(c,l)$ *is the ex-post utility of the agents with the largest degree of preference reversal.*

Since $\overline{\alpha} < 1$, all agents in the economy are time-inconsistent. In this setting, the definitions of magnitude or frequency naiveté follow. In other words, a partially naïve agents has misperception $\hat{\alpha}$ in magnitude or frequency within the support $(\alpha, 1]$. If $\hat{\alpha} = \alpha$, then the the agents are sophisticated. The following proposition shows that a hybrid mechanism can implement full efficiency in an economy with diversely time-inconsistent and naïve agents.

Proposition 8 *The optimal allocation for the environment with diversely time-inconsistent and naïve agents where agents have private information on productivity and sophistication level is the efficient allocation.*

Proof I will first show how the mechanism can be constructed for magnitude naiveté. First, choose a target $\tilde{\alpha} \in (\overline{\alpha}, 1)$, and construct a fooling mechanism that implements the full information efficient optimum for $\hat{\alpha} = \tilde{\alpha}$ and $\alpha = \overline{\alpha}$. By Proposition 5, this can be done. The executability and fooling constraints are satisfied $\forall \theta_m \in \Theta$,

$$V\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right); \overline{\alpha}\right) \geq V\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right); \overline{\alpha}\right),$$

$$\begin{aligned}
&\tilde{\alpha} U\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right) + \left(1-\tilde{\alpha}\right) \underline{V}\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right) \\
&\geq \tilde{\alpha} U\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right) + \left(1-\tilde{\alpha}\right) \underline{V}\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right).
\end{aligned}$$

The executability constraint can be rewritten as

$$\begin{aligned}
&\overline{\alpha} U\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right) + \left(1-\overline{\alpha}\right) \underline{V}\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right) \\
&\geq \overline{\alpha} U\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right) + \left(1-\overline{\alpha}\right) \underline{V}\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right).
\end{aligned}$$

The executability and fooling constraints imply the following:

$$U\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right) - U\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right) > \underline{V}\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right) - \underline{V}\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right). \quad (30)$$

I will now show that the efficient allocation is implementable for all agents more naïve ($\hat{\alpha} > \tilde{\alpha}$) and with larger degrees of preference reversals ($\alpha < \overline{\alpha}$). To see this, first notice that by Lemma 2, all agents with $\alpha = \overline{\alpha}$ and $\hat{\alpha} > \tilde{\alpha}$ would be fooled into choosing the efficient

allocation in equilibrium. Notice for agents with $\alpha < \overline{\alpha}$, the executability constraint is

$$\begin{aligned}&\alpha U\left(c^{*}\left(\theta_{m}\right), l^{*}\left(\theta_{m}\right)\right)+(1-\alpha) \underline{V}\left(c^{*}\left(\theta_{m}\right), l^{*}\left(\theta_{m}\right)\right)\\&\geq \alpha U\left(c^{I}\left(\theta_{m}\right), l^{*}\left(\theta_{m}\right)\right)+(1-\alpha) \underline{V}\left(c^{I}\left(\theta_{m}\right), l^{*}\left(\theta_{m}\right)\right),\end{aligned}$$

which holds trivially by (30). Since the fooling constraints do not depend on α and the incentive compatibility constraints do not depend on α and $\hat{\alpha}$, a fooling mechanism for $\alpha = \overline{\alpha}$ and $\hat{\alpha} = \tilde{\alpha}$ is able to deceive all agents in the economy more naïve than $\tilde{\alpha}$.

Next, consider a threat mechanism that implements the full information efficient optimum for $\hat{\alpha} = \tilde{\alpha} \in (\overline{\alpha}, 1)$ and $\alpha = \underline{\alpha}$. This can be accomplished by Corollary 5. I will demonstrate that this mechanism achieves the full information optimum for more sophisticated agents of any degree of preference reversal.

By Lemma 2, the efficient allocation is implementable for agents with $\alpha = \underline{\alpha}$ and $\hat{\alpha} < \tilde{\alpha}$. Similar to the argument for the fooling mechanism, only the executability constraint depends on α. However, for magnitude naiveté, if the credible threat constraints hold for all $\hat{\alpha} \in [\underline{\alpha}, \tilde{\alpha}]$, then the executability constraint would be satisfied for any $\alpha \in [\underline{\alpha}, \overline{\alpha}]$. This is because one of the credible threat constraints is $\forall \hat{\alpha} \in [\underline{\alpha}, \tilde{\alpha}]$,

$$\begin{aligned}&\hat{\alpha} U\left(c^{*}\left(\theta_{m}\right), l^{*}\left(\theta_{m}\right)\right)+(1-\hat{\alpha}) \underline{V}\left(c^{*}\left(\theta_{m}\right), l^{*}\left(\theta_{m}\right)\right)\\&\geq \hat{\alpha} U\left(c^{T}\left(\theta_{m}\right), l^{T}\left(\theta_{m}\right)\right)+(1-\hat{\alpha}) \underline{V}\left(c^{T}\left(\theta_{m}\right), l^{T}\left(\theta_{m}\right)\right).\end{aligned}$$

Hence, the set allocations satisfying it is a superset for the set of allocations satisfying the executability constraint for any $\alpha \in [\underline{\alpha}, \overline{\alpha}]$. Therefore, the threat mechanism for agents with $(\hat{\alpha}, \alpha) = (\tilde{\alpha}, \underline{\alpha})$ would be able to implement the efficient allocation for more sophisticated agents of any degree of preference reversal. Furthermore, the relatively sophisticated agents would weakly prefer the threat mechanism over the fooling mechanism, and the relatively naïve would strictly prefer the fooling mechanism over the threat mechanism. This completes the proof for magnitude naiveté.

I will now prove the case for frequency naiveté. First, target a sophistication level $\tilde{\alpha} \in (\overline{\alpha}, 1)$ and degree of preference reversal $\alpha = \overline{\alpha}$ with an effective fooling mechanism, which can be done by Proposition 5. By Lemma 2, the efficient allocation is implemented for more naïve agents and $\alpha = \overline{\alpha}$. Hence, the incentive compatibility and fooling constraints would also hold for all agents with $\hat{\alpha} > \tilde{\alpha}$ and $\alpha \in [\underline{\alpha}, \overline{\alpha}]$. Only the executability constraint varies with α. By construction, both the executability and fooling constraints hold at $\alpha = \overline{\alpha}$, for

any $\theta_m \in \Theta$,

$$\begin{aligned}&\overline{\alpha}U\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right) + \left(1-\overline{\alpha}\right)\underline{V}\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right)\\ &\geq \overline{\alpha}U\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right) + \left(1-\overline{\alpha}\right)\underline{V}\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right),\end{aligned}$$

and $U\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right) \geq U\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right)$. This implies that $\underline{V}\left(c^*\left(\theta_m\right), l^*\left(\theta_m\right)\right) \geq \underline{V}\left(c^I\left(\theta_m\right), l^*\left(\theta_m\right)\right)$. Hence, the executability constraints for agents with $\alpha < \overline{\alpha}$ would also hold trivially. Therefore, the efficient allocation can be implemented for all agents more naïve than $\tilde{\alpha}$ regardless of their degree of preference reversal.

Finally, construct an effective threat mechanism for agents with the same sophistication level $\tilde{\alpha}$ and degree of preference reversal $\alpha = \underline{\alpha}$, which can be done by Corollary 5. By Lemma 2, agents with $\hat{\alpha} < \tilde{\alpha}$ and $\alpha = \underline{\alpha}$ would be threatened into choosing the efficient allocation. Notice the only difference between magnitude and frequency naiveté for a threat mechanism are the incentive compatibility constraints, which do not depend on α. Hence, the same argument used to show that the executability constraint holds for all degrees of preference reversal in the magnitude naiveté case also works for frequency naiveté. As a result, the efficient allocation can be implemented for all agents more sophisticated than $\tilde{\alpha}$ regardless of the degree of preference reversal. This completes the proof. ■

B Proofs for Section 6 and Beyond

Proof of Theorem 5:

Notice that the deterrent and threat allocations can deter TI agents of any productivity type from misreporting. Hence, only the incentive compatibility constraints for the TC agents need to be considered. The proof will focus on a relaxed problem, where for within consistency deviations, only the local downward incentive compatibility constraints are considered along with a monotonicity constraints on $y^P : y_m^P \leq y_{m+1}^P$ for all $\theta_m \in \Theta$.

The incentive compatibility constraints are non-smooth. The problem can be reformulated by introducing a vector of new variables $x = (x_2, \ldots, x_M)$ so the constraints are smooth.

In essence, the incentive compatibility constraints are rewritten as $\forall \theta_m, \theta_{\hat{m}} \in \Theta$

$$u\left(c_m^P\right) - h\left(\frac{y_m^P}{\theta_m}\right) + w\left(k_m^P\right) \geq x_m \tag{31}$$

$$x_m - u\left(c_{m-1}^P\right) - h\left(\frac{y_{m-1}^P}{\theta_m}\right) + w\left(k_{m-1}^P\right) \geq 0 \tag{32}$$

$$x_m - u\left(c_{\hat{m}}^R\right) - h\left(\frac{y_{\hat{m}}^R}{\theta_m}\right) + w\left(k_{\hat{m}}^R\right) \geq 0. \tag{33}$$

Let λ_m denote the Lagrange multiplier associated with inequality (31). Let λ_m^{TC} denote the Lagrange multiplier associated with inequality (32). Let $\lambda_{\hat{m}|m}^{TI}$ denote the Lagrange multiplier associated with inequality (33). Let μ be the Lagrange multiplier associated with the feasibility constraint. Finally, denote γ_m as the Lagrange multiplier on the monotonicity constraint. The Lagrangian for the problem is

$$\begin{aligned}
\mathcal{L} = & \sum_{\theta_m} \pi_m \left\{ \phi \left[u\left(c_m^R\right) - h\left(\frac{y_m^R}{\theta_m}\right) + w\left(k_m^R\right) \right] + (1-\phi) \left[u\left(c_m^P\right) - h\left(\frac{y_m^P}{\theta_m}\right) + w\left(k_m^P\right) \right] \right\} \\
& + \sum_{\theta_m} \lambda_m \left[u\left(c_m^P\right) - h\left(\frac{y_m^P}{\theta_m}\right) + w\left(k_m^P\right) - x_m \right] \\
& + \sum_{\theta_m} \lambda_m^{TC} \left[x_m - u\left(c_{m-1}^P\right) + h\left(\frac{y_{m-1}^P}{\theta_m}\right) - w\left(k_{m-1}^P\right) \right] \\
& + \sum_{\theta_m} \sum_{\theta_{\hat{m}}} \lambda_{\hat{m}|m}^{TI} \left[x_m - u\left(c_{\hat{m}}^R\right) + h\left(\frac{y_{\hat{m}}^R}{\theta_m}\right) - w\left(k_{\hat{m}}^R\right) \right] \\
& + \mu \left\{ \sum_{\theta_m} \pi_m \left[\phi\left(y_m^R - c_m^R - k_m^R\right) + (1-\phi)\left(y_m^P - c_m^P - k_m^P\right) \right] \right\}.
\end{aligned}$$

The first order conditions for the real allocations are

$$\left(\phi \pi_m - \sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} \right) u'\left(c_m^R\right) = \phi \pi_m \mu,$$

$$\left(\phi \pi_m - \sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} \right) w'\left(k_m^R\right) = \phi \pi_m \mu$$

$$\phi \pi_m \frac{1}{\theta_m} h'\left(\frac{y_m^R}{\theta_m}\right) - \sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} \frac{1}{\theta_{\hat{m}}} h'\left(\frac{y_m^R}{\theta_{\hat{m}}}\right) = \phi \pi_m \mu.$$

The first order conditions for the persistent allocations are

$$\begin{aligned}
&\left[(1-\phi)\,\pi_m + \lambda_m - \lambda^{TC}_{m+1}\right] u'\left(c^P_m\right) = (1-\phi)\,\pi_m \mu,\\
&\left[(1-\phi)\,\pi_m + \lambda_m - \lambda^{TC}_{m+1}\right] w'\left(k^P_m\right) = (1-\phi)\,\pi_m \mu,\\
&\left[(1-\phi)\,\pi_m + \lambda_m\right] \frac{1}{\theta_m} h'\left(\frac{y^P_m}{\theta_m}\right) - \lambda^{TC}_{m+1} \frac{1}{\theta_{m+1}} h'\left(\frac{y^P_m}{\theta_{m+1}}\right) = (1-\phi)\,\pi_m \mu + \gamma_m - \gamma_{m+1}.
\end{aligned}$$

The first order condition on x_m gives the following relationship for Lagrange multipliers

$$\lambda_m = \lambda^{TC}_m + \sum_{\theta_{\hat{m}}} \lambda^{TI}_{\hat{m}|m}.$$

The Kuhn-Tucker conditions for a solution are the first order conditions above and the complementary slackness conditions. By Hellwig [(2007)], at the optimum, it must be the case that at least one of the multipliers λ_m or γ_m is strictly positive. Also, notice that perfect consumption smoothing is implemented: $c^R_m = k^R_m$ and $c^P_m = k^P_m$ for all $\theta_m \in \Theta$.

Suppose for some θ_m, $\left(c^R_m, k^R_m, y^R_m\right) \neq \left(c^P_m, k^P_m, y^P_m\right)$. There are three possible cases. The first case is $c^R_m < c^P_m$ and $y^R_m < y^P_m$ with $\lambda^{TI}_{m|m} > 0$ and $\lambda^{TI}_{m|\hat{m}} = 0$ for all $\theta_{\hat{m}} \neq \theta_m$. The second case is $c^R_m > c^P_m$ and $y^P_m < y^R_m$ with some $\theta_{m+k} > \theta_m$ such that $\lambda^{TI}_{m|m+k} \geq 0$ and $\lambda^{TI}_{m|\hat{m}} = 0$ for all $\theta_{\hat{m}} \neq \theta_{m+k}$. The third case is $\lambda^{TI}_{m|\hat{m}} = 0$ for all $\theta_{\hat{m}} \in \Theta$.

For the first case, the first order conditions on consumption imply

$$\frac{\lambda_m - \lambda^{TC}_{m+1}}{1-\phi} > -\frac{\lambda^{TI}_{m|m}}{\phi}.$$

From the first order conditions on output and by the fact that $h(\cdot)$ is strictly convex with $\theta_{m+1} > \theta_m$, the following inequality must hold at the optimum

$$\left(1 - \frac{\lambda^{TI}_{m|m}}{\phi \pi_m}\right) \frac{1}{\theta_m} h'\left(\frac{y^R_m}{\theta_m}\right) \geq \left[1 + \frac{\lambda_m - \lambda^{TC}_{m+1}}{(1-\phi)\,\pi_m}\right] \frac{1}{\theta_m} h'\left(\frac{y^P_m}{\theta_m}\right) + \frac{\gamma_{m+1} - \gamma_m}{(1-\phi)\,\pi_m},$$

Thus, for $y^P_m > y^R_m$, it must be that $\gamma_m > \gamma_{m+1}$. Thus, $y_m = y_{m-1}$, which implies $c_m = c_{m-1}$ and $k_m = k_{m-1}$, or else the upward incentive compatibility constraints would be violated. However, if $c^R_m < c^P_m$, $k^R_m < k^P_m$ and $y^R_m < y^P_m$ with $\lambda^{TI}_{m|m} > 0$, then the TC agent with productivity θ_{m-1} would be strictly better off misreporting as a TI agent with productivity θ_m. The first case is not possible.

For the second case, notice that if for some $\theta_m > \theta_1$, $\left(c^R_m, k^R_m, y^R_m\right) = \left(c^P_m, k^P_m, y^P_m\right)$, then the government has a welfare improving deviation for the real allocations. If $u'\left(c^P_m\right) < \mu$, then consider the following new deviation in real allocation: $\tilde{c}^R_m = c^R_m + \alpha\epsilon$, $\tilde{y}^R_m = y^R_m + \epsilon$ and

$\tilde{k}_m^R = k_m^R$, with $\alpha = \left[u'\left(c_m^P\right)\right]^{-1} \frac{1}{\theta_{m+1}} h'\left(\frac{y_m^P}{\theta_{m+1}}\right)$. The gain in welfare from a higher net output more than offsets the loss in utility for the type θ_m TI agent. Notice that this deviation is not possible for the lowest type, so $\left(c_1^R, k_1^R, y_1^R\right) = \left(c_1^P, k_1^P, y_1^P\right)$.

It is easy to see that the third case is satisfied as long as $\lambda_m^{TC} \geq \lambda_{m+1}^{TC}$, in essence, $u'\left(c_m^P\right) > \mu$. Note the highest productivity, θ_M, belongs to the third case.

Finally, due to the co-monotonicity of c and y (Hellwig [(2007)]) for the persistent allocation, it must be the case that only high types belong to the third case, and lower types belong to the second case. This establishes the existence of a cutoff $\bar{\theta}$. ■

Proof of Theorem 6:

Part (i) of the theorem is immediate from the first order conditions presented in the proof for Theorem 5.

For part (ii), if $\theta_m \geq \bar{\theta}$, then by Theorem 5 the TI agents are strictly worse off than TC agents of the same productivity. Hence, from the first order conditions, $\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} = 0$ and the result follows. For TC agents with θ_M productivity, by Hellwig [(2007)], it must be that $U_c\left(c_M^P, k_M^P, \frac{y_M^P}{\theta_M}\right) + U_y\left(c_M^P, k_M^P, \frac{y_M^P}{\theta_M}\right) = 0$. The result follows.

For part (iii), if $\theta_m < \bar{\theta}$, then by Theorem 5 the the TC agents are either separated from the TI agents or bunched with the TI agents of the same productivity. If the TI agents are separated from the TC agents, then from the first order conditions it must have $\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} \geq 0$. It is thus immediate that $U_c\left(c_m^R, k_m^R, \frac{y_m^R}{\theta_m}\right) + U_y\left(c_m^R, k_m^R, \frac{y_m^R}{\theta_m}\right) \geq 0$. If the TC agents are bunched with TI agents, then by Hellwig [(2007)], the distortion away from efficiency must be downward: $U_c\left(c_m^P, k_m^P, \frac{y_m^P}{\theta_m}\right) + U_y\left(c_m^P, k_m^P, \frac{y_m^P}{\theta_m}\right) > 0$, which is true also for any TC agent with $\theta_m < \theta_M$. ■

Proof of Theorem 7:

By the envelope theorem,

$$\begin{aligned}\frac{\partial W^T}{\partial \phi} = &\sum_{\theta_m} \pi_m \left[U\left(c_m^R, k_m^R, y_m^R; \theta_m\right) - U\left(c_m^P, k_m^P, y_m^P; \theta_m\right)\right] \\ &+ \mu \sum_{\theta_m} \pi_m \left[\left(y_m^R - c_m^R - k_m^R\right) - \left(y_m^P - c_m^P - k_m^P\right)\right],\end{aligned}$$

where μ is the Lagrange multiplier on the feasibility constraint. By Theorem 6, for all $\theta_m \geq \bar{\theta}$, $c_m^R < c_m^P$ and $k_m^R < k_m^P$. Since $u(\cdot)$ and $w(\cdot)$ are strictly concave, it follows that $\theta_m \geq \bar{\theta}$,

$$u\left(c_m^P\right) - u\left(c_m^R\right) < u'\left(c_m^R\right)\left(c_m^P - c_m^R\right) = \mu\left(c_m^P - c_m^R\right),$$

$$w\left(k_m^P\right) - w\left(k_m^R\right) < w'\left(k_m^R\right)\left(k_m^P - k_m^R\right) = \mu\left(k_m^P - k_m^R\right),$$

where the last equality follows from the first order necessary conditions on c_m^R and k_m^R. Following a similar argument, since $y_m^P < y_m^R$ and $h(\cdot)$ is strictly convex, the following inequality holds

$$h\left(\frac{y_m^R}{\theta_m}\right) - h\left(\frac{y_m^P}{\theta_m}\right) < \frac{1}{\theta_m} h'\left(\frac{y_m^R}{\theta_m}\right)\left(y_m^R - y_m^P\right) = \mu\left(y_m^R - y_m^P\right).$$

Substituting the inequalities above, the following relationship holds

$$\frac{\partial W^F}{\partial \phi} > \sum_{\theta_m < \bar{\theta}} \pi_m \Bigg\{ \left[u\left(c_m^R\right) - u\left(c_m^P\right)\right] - \mu\left(c_m^R - c_m^P\right) - \left[h\left(\frac{y_m^R}{\theta_m}\right) - h\left(\frac{y_m^P}{\theta_m}\right)\right] + \mu\left(y_m^R - y_m^P\right) + \left[w\left(k_m^R\right) - w\left(k_m^P\right)\right] - \mu\left(k_m^R - k_m^P\right) \Bigg\}.$$

For $\theta_m < \bar{\theta}$, by Theorem 5, either $\left(c_m^R, k_m^R, y_m^R\right) = \left(c_m^P, k_m^P, y_m^P\right)$ or $\left(c_m^R, k_m^R, y_m^R\right) \gg \left(c_m^P, k_m^P, y_m^P\right)$. It is sufficient to focus on the latter case.

When $\left(c_m^R, k_m^R, y_m^R\right) \gg \left(c_m^P, k_m^P, y_m^P\right)$, either $\lambda_{m|\hat{m}}^{TI} = 0$ for all $\theta_{\hat{m}}$ or there exists $\theta_{\hat{m}}$ such that $\lambda_{m|\hat{m}}^{TI} > 0$. For the former case, the same procedure delineated above can be applied. For the latter, suppose there exists $\theta_m \in \Theta$ such that

$$\begin{aligned} H\left(c_m^R, k_m^R, y_m^R\right) \equiv & U\left(c_m^R, k_m^R, y_m^R; \theta_m\right) + \mu\left(y_m^R - c_m^R - k_m^R\right) \\ & - U\left(c_m^P, k_m^P, y_m^P; \theta_m\right) - \mu\left(y_m^P - c_m^P - k_m^P\right) < 0. \end{aligned}$$

Let $\Gamma = u(c) + w(k)$. From the proof of Theorem 5, if $\left(c_m^P, k_m^P, y_m^P\right) = \left(c_m^R, k_m^R, y_m^R\right)$, then there exists perturbations $d\Gamma$ and dy such that the gain in welfare from the higher net output outweighs the loss in utility for type θ_m TI agent. Also, from the first order conditions,

$$\mu = \frac{1}{\sum_{\theta_m} \frac{\pi_m}{u'\left(c_m^R\right)}}.$$

Thus, by the continuity of $H\left(c_m^R, k_m^R, y_m^R\right)$, there exists $\left(\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R\right)$ such that $H\left(\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R\right) = 0$ and $\left(c_m^R, k_m^R, y_m^R\right) \gg \left(\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R\right) \gg \left(c_m^P, k_m^P, y_m^P\right)$. Let $\bar{\mu}$ denote the multiplier on the feasibility constraint associated with $\left(\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R\right)$. Since $\left(c_m^R, k_m^R, y_m^R\right) \neq \left(\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R\right)$, then consider $\tilde{c}_m^R = \bar{c}_m^R + \alpha\epsilon$ and $\tilde{y}_m^R = \bar{y}_m^R + \epsilon$ for $\epsilon > 0$ and $\alpha = \left[u'\left(c_m^P\right)\right]^{-1} \frac{1}{\theta_{m+1}} h'\left(\frac{y_m^P}{\theta_{m+1}}\right)$. This perturbation improves welfare, so it must be the case that $H\left(\tilde{c}_m^R, \bar{k}_m^R, \tilde{y}_m^R\right) > 0$. Thus, the neighborhood around any (c, k, y) such that $H(c, k, y) = 0$ has a perturbation $d\Gamma$ and dy that improves welfare. Therefore, it can never be the case that

$H\left(c_m^R, k_m^R, y_m^R\right) < 0$.

Hence, $\frac{\partial W^F}{\partial \phi} > 0$. Furthermore, by the maximum theorem, $W^T(\phi)$ is a continuous function over $[0,1]$. Thus, welfare increases continuously from the constrained efficient level to the efficient level as ϕ increases. ■

Proof of Theorem 8:

The first part of the proof will establish the implementability of the optimal allocation from the threat mechanism. Choose the persistent and real allocations to be the same as the optimal persistent and real allocations from the threat mechanism. The deterrent allocations can be the same, or different as long as it is capable of deterring TI agents from misreporting to be TC agents. For the imaginary allocations, set $y_m^I = y_m^P$ for all $\theta_m = \Theta$, and

$$u(c_m^I) + w(k_m^I) = u(c_m^P) + w(k_m^P).$$

In particular, choose $c_1^I = c_1^P = c_1^R$ and $k_1^I = k_1^P = k_1^R$. Notice that with this construction, all of the incentive compatibility constraints are satisfied, since the persistent allocations from the threat mechanism are incentive compatible. Next, choose $\left(c_m^I, k_m^I\right)_{\theta_m \in \Theta}$ such that the fooling and executability constraints are satisfied: $\forall \theta_m \in \Theta$

$$u\left(c_m^I\right) - h\left(\frac{y_m^I}{\theta_m}\right) + \hat{\beta} w\left(k_m^I\right) \geq u\left(c_m^R\right) - h\left(\frac{y_m^R}{\theta_m}\right) + \hat{\beta} w\left(k_m^R\right),$$

$$u\left(c_m^R\right) - h\left(\frac{y_m^R}{\theta_m}\right) + \beta w\left(k_m^R\right) \geq u\left(c_m^I\right) - h\left(\frac{y_m^I}{\theta_m}\right) + \beta w\left(k_m^I\right).$$

By the construction of the imaginary allocations, the fooling and executability constraints can be rewritten as

$$w\left(k_m^I\right) \leq \frac{U\left(c_m^P, k_m^P, y_m^P; \theta_m\right) - U\left(c_m^R, k_m^R, y_m^R; \theta_m\right)}{1 - \hat{\beta}} + w\left(k_m^R\right),$$

$$w\left(k_m^I\right) \geq \frac{U\left(c_m^P, k_m^P, y_m^P; \theta_m\right) - U\left(c_m^R, k_m^R, y_m^R; \theta_m\right)}{1 - \beta} + w\left(k_m^R\right).$$

By the incentive compatibility of the persistent and real allocations and by Theorem 5, $U\left(c_m^P, k_m^P, y_m^P; \theta_m\right) > U\left(c_m^R, k_m^R, y_m^R; \theta_m\right)$ for all $\theta_m \in \Theta \setminus \theta_1$. Therefore, it is possible to find $\left(c_m^I, k_m^I\right)$ for all productivity types to satisfy the fooling constraints when $1 > \hat{\beta} > \beta$.

The second part of the proof will argue the optimality of the optimal threat allocation in a fooling mechanism. Note that a direct truth-telling fooling mechanism has extra potentially

binding constraints that a direct truth-telling threat mechanism does not have: $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$U\left(c_m^P, k_m^P, y_m^P; \theta_m\right) \geq U\left(c_{\hat{m}}^I, k_{\hat{m}}^I, y_{\hat{m}}^I; \theta_m\right).$$

For a threat mechanism, a TC agent would never choose the threat allocations. However, in a fooling mechanism, the government needs to deter the TC agent from pretending to be a TI agent and choose the imaginary allocations. Hence, the government's problem in a fooling mechanism has more constraints than a threat mechanism. Since the optimal allocation from a direct threat mechanism is implementable in a direct fooling mechanism, then the optimal allocations in both mechanisms must be equivalent. ■

Proof of Theorem 9:

Similar to the proof of Theorem 5, the proof will focus on a relaxed problem. In essence, only the local downward incentive compatibility constraints are considered along with a monotonicity constraints on y^P. The incentive compatibility constraints are (31), (32) and (33). Notice the fooling constraint is included in (33), when $\theta_{\hat{m}} = \theta_m$. The executability constraint is

$$u\left(c_m^R\right) - h\left(\frac{y_m^R}{\theta_m}\right) + \beta w\left(k_m^R\right) \geq u\left(c_m^P\right) - h\left(\frac{y_m^P}{\theta_m}\right) + \beta w\left(k_m^P\right).$$

Notice the fooling and executability constraints imply that

$$k_m^P \geq k_m^R, \tag{34}$$

and

$$u\left(c_m^R\right) - h\left(\frac{y_m^R}{\theta_m}\right) \geq u\left(c_m^P\right) - h\left(\frac{y_m^P}{\theta_m}\right). \tag{35}$$

Let ω_m be the Lagrange multiplier associated with the executability constraint for TC agents of type θ_m. The rest of the Lagrange multipliers are the same as in the proof for Theorem 5.

The first order conditions for the real allocations are

$$\left(\phi\pi_m - \sum_{\theta_{\hat{m}}} \lambda^{TI}_{m|\hat{m}} + \omega_m\right) u'\left(c^R_m\right) = \phi\pi_m\mu,$$
$$\left(\phi\pi_m - \sum_{\theta_{\hat{m}}} \lambda^{TI}_{m|\hat{m}} + \beta\omega_m\right) w'\left(k^R_m\right) = \phi\pi_m\mu$$
$$\left(\phi\pi_m + \omega_m\right)\frac{1}{\theta_m}h'\left(\frac{y^R_m}{\theta_m}\right) - \sum_{\theta_{\hat{m}}} \lambda^{TI}_{m|\hat{m}}\frac{1}{\theta_{\hat{m}}}h'\left(\frac{y^R_m}{\theta_{\hat{m}}}\right) = \phi\pi_m\mu.$$

The first order conditions for the persistent allocations are

$$\left[(1-\phi)\,\pi_m + \lambda_m - \lambda^{TC}_{m+1} - \omega_m\right] u'\left(c^P_m\right) = (1-\phi)\,\pi_m\mu,$$
$$\left[(1-\phi)\,\pi_m + \lambda_m - \lambda^{TC}_{m+1} - \beta\omega_m\right] w'\left(k^P_m\right) = (1-\phi)\,\pi_m\mu,$$
$$\left[(1-\phi)\,\pi_m + \lambda_m - \omega_m\right]\frac{1}{\theta_m}h'\left(\frac{y^P_m}{\theta_m}\right) - \lambda^{TC}_{m+1}\frac{1}{\theta_{m+1}}h'\left(\frac{y^P_m}{\theta_{m+1}}\right) = (1-\phi)\,\pi_m\mu + \gamma_m - \gamma_{m+1}.$$

The first order condition on x_m gives the following relationship for Lagrange multipliers

$$\lambda_m = \lambda^{TC}_m + \sum_{\theta_{\hat{m}}} \lambda^{TI}_{\hat{m}|m}.$$

By the Kuhn-Tucker Theorem, the solution is characterized by the first order conditions above and the complementary slackness conditions. Notice that if $\omega_m > 0$, then there will be intertemporal distortions for type θ_m. Next, the proof will proceed to show how $\omega_m > 0$ for all $\theta_m \in \Theta \setminus \theta_1$, there by demonstrating part (i).

Suppose there exists $\theta_m \in \Theta \setminus \theta_1$ such that $\omega_m = 0$. By the first order conditions and (34), it must be the case that $c^P_m \geq c^R_m$ and with strict inequality only if $k^P_m > k^R_m$. There are two cases: if $k^P_m > k^R_m$ or $k^P_m = k^R_m$.

If $k^P_m > k^R_m$, then $c^P_m > c^R_m$ and

$$1 + \frac{\lambda_m - \lambda^{TC}_{m+1}}{(1-\phi)\,\pi_m} > 1 - \frac{\sum_{\theta_{\hat{m}}} \lambda^{TI}_{m|\hat{m}}}{\phi\pi_m}.$$

Thus by (35), it must be the case that $y^P_m > y^R_m$. If $\sum_{\theta_{\hat{m}}} \lambda^{TI}_{m|\hat{m}} = 0$, then

$$\left(1 + \frac{\lambda_m - \lambda^{TC}_{m+1}}{(1-\phi)\,\pi_m}\right)\frac{1}{\theta_m}h'\left(\frac{y^P_m}{\theta_m}\right) + \frac{\gamma_{m+1} - \gamma_m}{(1-\phi)\,\pi_m} < \left(1 - \frac{\sum_{\theta_{\hat{m}}} \lambda^{TI}_{m|\hat{m}}}{\phi\pi_m}\right)\frac{1}{\theta_m}h'\left(\frac{y^R_m}{\theta_m}\right), \qquad (36)$$

which implies $\gamma_m > 0$ if $y^P_m > y^R_m$ is true. Thus, it follows that $y^P_m = y^P_{m-1}$ with $c^P_m = c^P_{m-1}$

and $k_m^P = k_{m-1}^P$. Hence, with $\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} = 0$, then either a θ_{m-1} agent has an incentive to misreport to be a θ_m agent, or vice versa, which is a contradiction, so $\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} > 0$. If this is the case, then it must be that $\lambda_{m|m}^{TI} > 0$ with $\lambda_{m|\hat{m}}^{TI} = 0$ for all $\theta_{\hat{m}} \in \Theta \setminus \theta_m$. This is because if $\lambda_{m|\hat{m}}^{TI} > 0$ for $\theta_{\hat{m}} > \theta_m$, then $y_m^R \geq y_m^P$ and if $\lambda_{m|\hat{m}}^{TI} > 0$ for $\theta_{\hat{m}} < \theta_m$, then there is a welfare enhancing perturbation by increasing the output and consumption. However, if $\lambda_{m|m}^{TI} > 0$, then (36) must be true, which is a contradiction. Hence, if $\omega_m = 0$ for some $\theta_m \in \Theta \setminus \theta_1$, then it can not be the case that $k_m^P > k_m^R$.

Now, examine the case when $k_m^P = k_m^R$, then $c_m^P = c_m^R$. By the fooling and executability constraints, this means that $y_m^P = y_m^R$. However, this is not optimal. Consider the perturbation: $\tilde{c}_m^R = c_m^R + \alpha\epsilon$, $\tilde{y}_m^R = y_m^R + \epsilon$ and $\tilde{k}_m^R = k_m^R$, with $\alpha = \left[u'\left(c_m^P\right)\right]^{-1} \frac{1}{\theta_{m+1}} h'\left(\frac{y_m^P}{\theta_{m+1}}\right)$. As was demonstrated in the proof for Theorem 5, the gain in welfare from the higher net output more than offsets the loss in utility for the type θ_m TI agent. Hence, for all $\theta_m \in \Theta_m \setminus \theta_1$, it must be that $\omega_m > 0$. This implies the TI agents under-save, and the TC agents over-save.

For part (ii), notice part (i) implies $\omega_M > 0$. Suppose $\lambda_{M|M}^{TI} > 0$, then both the fooling and executability constraints for type θ_M are binding by the complementary slackness condition. The binding constraints imply $k_m^P = k_m^R$ with $c_M^P = c_M^R$ and $y_M^P = y_M^R$. From the first order condition on c, it implies that

$$\frac{\lambda_M - \omega_M}{1-\phi} = \frac{\omega_M - \lambda_{M|M}^{TI}}{\phi},$$

which implies $k_m^P > k_m^R$ from the first order conditions on k. Thus, $\lambda_{M|M}^{TI} = 0$. If $\lambda_{M|M}^{TI} = 0$, notice that it can not be the case that $k_M^P = k_M^R$, since by the first order conditions, this would imply $c_M^R > c_M^P$ and $y_M^R < y_M^P$ which violates the fooling (incentive compatibility) constraint. Hence, $k_M^P > k_M^R$. The case with $c_m^P \geq c_m^R$ and $y_m^P \geq y_m^R$ can be ruled out since this violates the first order conditions. It follows that $c_M^P < c_M^R$ and $y_M^P > y_M^R$. This proves part (ii).

For part (iii), first notice that if $\omega_1 = 0$, then by the same argument in the proof of Theorem 5, $\left(c_1^R, k_1^R, y_1^R\right) = \left(c_1^P, k_1^P, y_1^P\right)$.

Suppose $\omega_1 > 0$, then by (34) and the first order condition on savings, it must be that

$$\frac{-\lambda_2^{TC} - \beta\omega_1}{1-\phi} \geq \frac{\beta\omega_1 - \sum_{\theta_{\hat{m}}} \lambda_{1|\hat{m}}^{TI}}{\phi}. \tag{37}$$

Hence, there exists $\theta_{\hat{m}}$ such that $\lambda_{1|\hat{m}}^{TI} > 0$.

If $c_1^P \geq c_1^R$, then $y_1^P \geq y_1^R$. by (35). Thus, it must be the case that $\lambda_{1|1}^{TI} > 0$. By complementary slackness condition, both the executability and fooling constraints for type θ_1 must bind. It follows that $k_1^R = k_1^P$, which implies that (37) holds with equality. However,

from the first order conditions on consumption, $c_1^P \geq c_1^R$ is not possible.

Now check $c_1^R > c_1^P$, which implies

$$\frac{-\lambda_2^{TC} - \omega_1}{1-\phi} < \frac{\omega_1 - \sum_{\theta_{\hat{m}}} \lambda_{1|\hat{m}}^{TI}}{\phi}. \tag{38}$$

If $y_1^R < y_1^P$, then it has to be the case that $\lambda_{1|1}^{TI} > 0$. Hence, it must be the case that $k_1^R = k_1^P$ by the complementary slackness conditions. This violates the executability constraint. Finally, it remains to check $\left(c_1^R, y_1^R\right) \gg \left(c_1^P, y_1^P\right)$. Since $y_1^R > y_1^P$, from the first order conditions, it must be that

$$\frac{\omega_1}{\phi\pi_1}\frac{1}{\theta_1}h'\left(\frac{y_1^R}{\theta_1}\right) + \frac{\omega_1}{(1-\phi)\,\pi_1}\frac{1}{\theta_1}h'\left(\frac{y_1^P}{\theta_1}\right) < \frac{\lambda_{1|\hat{m}}^{TI}}{\phi\pi_1}\frac{1}{\theta_{\hat{m}}}h'\left(\frac{y_1^R}{\theta_{\hat{m}}}\right) - \frac{\lambda_2^{TC}}{(1-\phi)\,\pi_1}\frac{1}{\theta_2}h'\left(\frac{y_1^P}{\theta_2}\right),$$

which contradicts (38). Thus, at the optimum, $\omega_1 = 0$. This implies perfect consumption smoothing for type θ_1. ■

Proof of Theorem 10:

By Theorem 8, it is without loss of generality to focus on a fooling mechanism for $W(\hat{\beta} < 1)$. Let $\mathcal{A}(\hat{\beta} < 1)$ and $\mathcal{A}(\hat{\beta} = 1)$ denote the sets of real and persistent allocations that are implementable under a fooling mechanism with partially naïve and fully naïve agents respectively.

First, I will show that $\mathcal{A}(\hat{\beta} = 1) \subseteq \mathcal{A}(\hat{\beta} < 1)$. Note that for any $\mathfrak{a} \in \mathcal{A}(\hat{\beta} = 1)$, the persistent allocation is equivalent to the imaginary allocation and the incentive compatibility constraints for the persistent allocation is trivially satisfied. Also, the incentive compatibility constraints of the real allocations hold. Hence, $\mathfrak{a} \in \mathcal{A}(\hat{\beta} < 1)$, which implies $W(\hat{\beta} < 1) \geq W(\hat{\beta} = 1)$.

Finally, I will show that the optimal allocation $a^* \in \mathcal{A}(\hat{\beta} < 1)$ is not implementable when agents are fully naïve, in essence, $a^* \notin \mathcal{A}(\hat{\beta} = 1)$. To implement a^* with fully naïve agents, the government has to set the imaginary allocations equal to the persistent allocations. By Theorem 5 and Theorem 8, a^* has the following property for θ_M : $c_M^P > c_M^R$, $k_M^P > k_M^R$ and $y_M^P < y_M^R$. Hence, the following fooling constraint is violated

$$u\left(c_M^R\right) - h\left(\frac{y_M^R}{\theta_M}\right) + \beta w\left(k_M^R\right) \geq u\left(c_M^P\right) - h\left(\frac{y_M^P}{\theta_M}\right) + \beta w\left(k_M^P\right).$$

Therefore, the real allocation in a^* is not implementable when agents are fully naïve. This implies that $W(\hat{\beta} < 1) > W(\hat{\beta} = 1)$. ■

Proof of Proposition 3:

This result is shown in the text of the paper. ■

Proof of Proposition 4:

Consider the following repayment plan for a projection $y^e \in \left[y_m^*, y_{m+1}^*\right)$,

$$R\left(b, y^e, y\right) = \begin{cases} \eta\left(y^e\right) - \rho^* b & \text{if } y_m^* \leq y < \bar{y}\left(y^e\right) \\ \hat{\eta}\left(y^e, y\right) - \hat{\rho}\left(y^e\right) b & \text{if } y \geq \bar{y}\left(y^e\right) \\ b & \text{if } y < y_m^* \end{cases}$$

and $\bar{y}\left(y^e\right) = \infty$ if $y^e \geq y_M^*$. The repayment plan consists of the principal, η, and a variable payment contingent on savings, ρb. The construction of the repayment plan ensures the agent would not produce less than y_m^* if the initial projection was $y^e \in \left[y_m^*, y_{m+1}^*\right)$. The government sets $\rho^* = \frac{1}{\beta} - 1$, so consumption is smoothed on the equilibrium path. To see this, notice that the sequential budget constraints (13) and (14) can be consolidated, and the budget constraint on the equilibrium path is

$$c + \frac{k}{1+\rho^*} = y + L\left(y^e\right) - T(y) - {}^{\eta(y^e)}/_{1+\rho^*}.$$

Consequently, with $\rho^* = \frac{1}{\beta} - 1$, the doer would choose the appropriate savings amount on the equilibrium path. In essence, ρ^* is a discount on the repayment contingent on savings. Also, the government can set $L\left(y^e\right) = \beta\eta\left(y^e\right)$, which means the agents do not need to pay any interest on the commitment loan if they choose the appropriate savings b and income $y < \bar{y}\left(y^e\right)$. The income tax is

$$T\left(y\right) = \begin{cases} y - c^* - \beta k^* & \text{if } y \in [y_1^*, \infty) \\ y & \text{if } y \in [0, y_1^*) \end{cases}$$

so agents would never produce less than y_1^*. The agents would choose the efficient allocation, $\left(c^*, k^*, y_m^*\right)$, on the equilibrium path. To see why this is the case, first notice the doer solves the following problem on the equilibrium path:

$$\max V(c, k, y; \theta) \text{ s.t. } c + {}^{k}/_{1+\rho^*} = y + L\left(y^e\right) - T(y) - {}^{\eta(y^e)}/_{1+\rho^*}.$$

With the constructed policies, an agent would never produce more than y_m^* on the equilibrium path, and the on-equilibrium budget constraint is $\left(c - c^*\right) + \beta\left(k - k^*\right) = 0$. Thus, for the

first order conditions and the budget constraint to hold, the unique solution to the doer's problem is the efficient allocation.

Next, to make sure a type θ_m agent would choose a commitment loan for projection $y^e \in \left[y_m^*, y_{m+1}^*\right)$, construction of the off-equilibrium path policy, $\left\{\left(\hat{\eta}\left(y^e, y\right), \hat{\rho}\left(y^e\right), \bar{y}\left(y^e\right)\right)\right\}_{m=1}^{M}$, along with the fixed principal payment $\eta\left(y^e\right)$ will be demonstrated.

To prevent deviation, the off-equilibrium path policy needs to deter agents from choosing $L\left(y^e\right)$ that is not intended for their productivity. First, for any $y_1^e, y_2^e \in \left[y_m^*, y_{m+1}^*\right)$, set $\eta\left(y_1^e\right) = \eta\left(y_2^e\right)$, $\hat{\rho}\left(y_1^e\right) = \hat{\rho}\left(y_2^e\right)$, $\bar{y}\left(y_1^e\right) = \bar{y}\left(y_2^e\right)$ and $\hat{\eta}\left(y_1^e, y\right) = \hat{\eta}\left(y_2^e, y\right)$. Next, set $\hat{\eta}\left(y^e, y\right) = \left(y - \bar{y}\left(y^e\right)\right)\left(1 + \hat{\rho}\left(y^e\right)\right)$ so the agents would never choose to produce more than $\bar{y}\left(y^e\right)$. In essence, if the agent has output $y = \bar{y}\left(y^e\right)$, then the agent does not need to pay the principal η, but will instead repay an amount proportional to his savings b.

Substituting in the policies, a type θ_{m+1} agent contemplating a commitment loan for $y^e \in \left[y_m^*, y_{m+1}^*\right)$ predicts his doer would solve the following program if $y \in \left[y_m^*, \bar{y}\left(y^e\right)\right)$

$$\max_{c,y,k} u\left(c\right) - h\left(\frac{y}{\theta_{m+1}}\right) + \hat{\beta} w\left(k\right) \text{ s.t. } c + \beta k = \left(c^* + \beta k^*\right).$$

It is obvious the optimal output is y_m^*, and let $\left(\tilde{c}, \tilde{k}\right)$ denote the solution for consumption and savings. Notice the solution to the problem does not depend on the agent's actual productivity θ_{m+1} nor θ_m. If $y = \bar{y}\left(y^e\right)$, the type θ_{m+1} agent would solve the following problem

$$\max_{c,k} u\left(c\right) + \hat{\beta} w\left(k\right) \text{ s.t. } c + \frac{k}{1 + \hat{\rho}\left(y^e\right)} = c^* + \beta k^* + \beta \eta\left(y^e\right),$$

and let $\left(\hat{c}_m, \hat{k}_m\right)$ denote the solution, which also does not depend on θ_{m+1}.

Let $\bar{y}\left(y^e\right) = y_m^* + \alpha_m$, with $\alpha_m > 0$. The government constructs $\hat{\rho}\left(y^e\right) < \rho^*$ for all productivity types to tempt the doer to increase c and decrease k, which is undesirable to the planner. With $\hat{\rho}\left(y^e\right) < \rho^*$, the government can construct $\eta\left(y^e\right) > 0$ to ensure the credible threat constraint holds, given $\alpha_m > 0$:

$$u\left(\hat{c}_m\right) - h\left(\frac{y_m^* + \alpha_m}{\theta_{m+1}}\right) + \hat{\beta} w\left(\hat{k}_m\right) \geq u\left(\tilde{c}\right) - h\left(\frac{y_m^*}{\theta_{m+1}}\right) + \hat{\beta} w\left(\tilde{k}\right). \tag{39}$$

For any given $\hat{\rho}\left(y^e\right) < \rho^*$ and $\alpha_m > 0$, construct $\eta\left(y^e\right)$ such that (39) binds. Let $\eta_m\left(\hat{\rho}\left(y^e\right), \alpha_m\right)$ denote the solution for $y^e \in \left[y_m^*, y_{m+1}^*\right)$. Also, with (39) binding, higher productivity types would strictly prefer producing output $y = y_m^* + \alpha_m$ if they chose the loan for projected income $y^e \in \left[y_m^*, y_{m+1}^*\right)$.

Next, the government chooses $\hat{\rho}(y^e)$ to satisfy the incentive compatibility constraint:

$$u(\tilde{c}) - h\left(\frac{y_M^*}{\theta_M}\right) + w\left(\tilde{k}\right) \geq u(\hat{c}_m) - h\left(\frac{y_m^* + \alpha_m}{\theta_M}\right) + w\left(\hat{k}_m\right),$$

where $\left(\tilde{c}, \tilde{k}, y_M^*\right)$ is the solution to the following problem for taking out a loan with an initial projection of $y^e \in [y_M^*, \infty)$,

$$\max_{c,y,k} u(c) - h\left(\frac{y}{\theta_M}\right) + \hat{\beta} w(k) \text{ s.t. } c + \frac{k}{1+\rho^*} = (c^* + \beta k^*).$$

If the θ_M agent can be deterred from picking the loan for θ_m, then all other agents would be deterred as well. Since (39) binds, the incentive compatibility constraint can be rewritten as

$$\left(1 - \hat{\beta}\right)\left[w\left(\tilde{k}\right) - w\left(\hat{k}_m\right)\right] \geq h\left(\frac{y_M^*}{\theta_M}\right) - h\left(\frac{y_m^*}{\theta_M}\right).$$

Notice that, for a given $\alpha_m > 0$, $\hat{k}_m$ is an increasing function of $\hat{\rho}(y^e)$ and is the only object dependent on $\hat{\rho}(y^e)$. The government can then decrease $\hat{\rho}(y^e)$ till the incentive compatibility constraint holds for any $\alpha_m > 0$. Let $\hat{\rho}_m(\alpha_m)$ denote the level of $\hat{\rho}(y^e)$ such that the incentive compatibility constraint holds. Hence, the government can set $\eta_m(\hat{\rho}_m(\alpha_m), \alpha_m)$ and $\hat{\rho}_m(\alpha_m)$ for (39) and the incentive compatibility constraints to hold.

Finally, to pin down α_m, the government can increase α_m till the executability constraint for θ_m type agent holds. The executability constraint is

$$u(c^*) - h\left(\frac{y_m^*}{\theta_m}\right) + \beta w(k^*) \geq u(c'_m) - h\left(\frac{y_m^* + \alpha_m}{\theta_m}\right) + \beta w(k'_m).$$

where (c'_m, k'_m) is the solution to the following problem

$$\max u(c) + \beta w(k) \text{ s.t. } c + \frac{k}{1 + \hat{\rho}_m(\alpha_m)} = c^* + \beta k^* + \beta \eta_m(\hat{\rho}_m(\alpha_m), \alpha_m).$$

By construction, the government budget is balanced. Also, since (39) binds, the actual θ_m agent would predict an output of $y < \bar{y}(y^e)$ so the credible threat constraints for θ_m hold. To see this, notice the following $h(y_m^* + \alpha_m) - h\left(\frac{y_m^*}{\theta_m}\right) > h\left(\frac{y_m^* + \alpha_m}{\theta_{m+1}}\right) - h\left(\frac{y_m^*}{\theta_{m+1}}\right)$, which implies $u(\hat{c}_m) - h\left(\frac{y_m^*}{\theta_m}\right) + \hat{\beta} w\left(\hat{k}_m\right) < u(\tilde{c}) - h\left(\frac{y_m^*}{\theta_m}\right) + \hat{\beta} w\left(\tilde{k}\right)$. The process above can be repeated for all productivity types to ensure global incentive compatibility. ■