ARTICLE IN PRESS
JID:YREDY AID:338 /FLA [m1+; v 1.59; Prn:6/04/2006; 12:17] P.1 (1-28)
Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Review of Economic Dynamics ••• (••••) •••–•••
Review of Economic Dynamics
www.elsevier.com/locate/red

Credibility and endogenous societal discounting

Christopher Sleet *, Şevin Yeltekin

Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213

Received 12 December 2005; revised 9 March 2006

Abstract

We consider a dynamic moral hazard economy inhabited by a planner and a population of privately informed agents. We assume that the planner and the agents share the same discount factor, but that the planner cannot commit. We show that optimal allocations in such settings solve the problems of *committed* planners who discount the future *less heavily* than agents. Thus, we provide micro-foundations for dynamic moral hazard models that assume a societal discount factor in excess of the private one. We extend the analysis to allocations that are *reconsideration-proof* in the sense of Kocherlakota [Kocherlakota, N., 1996. Reconsideration-proofness: A refinement for infinite horizon time inconsistency. Games and Economic Behavior 15, 33–54]. We show that these allocations solve the choice problem of a committed planner with a unit discount factor.

JEL classification: ???; ???

Keywords: ???; ???

1. Introduction

A large literature has explored the implications of dynamic moral hazard for inequality and insurance. One class of models considered assumes long-lived, risk-averse agents who experience privately observed shocks to their tastes, endowments or productivities. A benevolent planner provides insurance to these agents via mechanisms that give incentives for the truthful revelation of information. These models have been widely applied to analyses of optimal social insurance and taxation.

* Corresponding author.
E-mail address: csleet@andrew.cmu.edu (C. Sleet).

1094-2025/$ – see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.red.2006.03.001

Until recently, virtually all such models made two assumptions. The first was that the planner shared the same discount factor as the agents, the second that the planner could commit to implementing future allocations. Under these assumptions and some standard restrictions on shocks and preferences,[1] optimal social insurance arrangements exhibit a dramatic property. They imply severe long run inequality, with almost all agents eventually absorbed by a minimal utility, immiserating state.

Three recent papers, Farhi and Werning (FW) (2005), Phelan (2005) and Sleet and Yeltekin (SY) (2005a), have since departed from the equal discounting assumption. They suppose a planner or societal discount factor in excess of the agents' and show that this modification profoundly alters the optimal social insurance arrangement. Under mild conditions, there now exist optimal arrangements that imply ergodic processes for agent utilities and consumption. The immiseration result is overturned. Long run inequality is much less severe and these arrangements exhibit social mobility with, in the language of Phelan (2005), an absence of caste systems: the poor eventually become rich, the rich poor.

In this paper, we provide micro-foundations for the excess societal discounting supposed by FW, Phelan and SY.[2] In contrast to these contributions, we retain equal discounting, but drop the assumption of planner commitment. We then show that optimal allocations in such settings *solve the choice problems of planners who can commit, but who use perturbed discounting schemes*. These schemes discount the future less heavily than those of agents. We assume *neither* commitment *nor* differential discounting across planners and agents, but obtain allocations that solve problems in which both are present. In our model, high planner or societal discount factors are an equilibrium phenomenon, rather than an assumption.

Removal of the planner commitment assumption is of interest in its own right, independent of the connection with planner discounting. When both commitment and equal discounting are assumed, as in much of the existing literature, the immiseration result implies that the benevolent planner commits to her own eventual destitution. Thus, the optimal allocations obtained in the existing literature place extremely strong demands on the planner's (and more, generally, society's) ability to commit. It is natural to relax this ability.

Our analysis proceeds by blending the dynamic contracting concepts described above with notions from the literature on sustainable macroeconomic policies.[3] We assume a relatively simple environment in which agents receive privately observed taste shocks that affect their desire to

---

[1] See Phelan (1998).

[2] FW and Phelan provide an alternative normative rationale for this assumption, based on intergenerational considerations. See our discussion of the Rawlsian planner in Section 4.1.

[3] The literature on sustainable macroeconomic plans considers policymaking by governments that cannot commit, see for example, Chari and Kehoe (1990, 1993). Governments are modeled as being socially benevolent and as having access to a restricted set of tax mechanisms. Our modeling of a planner that cannot commit is in similar spirit, although our planner operates a mechanism constrained by incentive-compatibility considerations.

Bisin and Rampini (2005) also incorporate a lack of planner commitment into a dynamic moral hazard model. In contrast to us, they allow agents to undertake intertemporal trades that are hidden from the planner. On the other hand, Bisin and Rampini only consider finite horizon examples and explicitly rule out the reputational forces that are central to this paper. Dynamic optimal taxation in Mirrleesian environment is considered by Berliant and Ledyard (2005) and, in an early classic, by Roberts (1984). Although, enriching the environment in other ways relative to us, both of these papers also omit reputational considerations. Acemoğlu et al. (2005) do introduce reputation into a model with non-benevolent planner who cannot commit. See Remark 3 for some discussion of their paper.

consume.[4] These agents play an infinitely repeated game with a planner. In the first stage of each period, the planner selects a mechanism. This consists of a lottery, a message space that agents can use to communicate with the planner and an allocation function that awards consumption to agents based on their history of current and past messages and lottery outcomes. Agents then send their messages and the mechanism executes. Crucially, the planner cannot commit to future mechanisms. We define a credible equilibrium to be a pair of strategies for the planner and the agents that are optimal after each past history. We obtain a revelation principle for this environment and then give necessary and sufficient conditions for an allocation to be credible, i.e. an equilibrium outcome. Amongst these conditions are a collection of "sustainability" constraints. These require that the planner's continuation payoffs remain above the autarkic value that would obtain if no insurance was provided to the agents.

There are typically many credible allocations. We first focus on optimal ones. By manipulating the Lagrangian from the choice problem that defines them, we show that such allocations solve the problems of planners who can commit, but who use an endogenous sequence of discount factors in excess of the agents'. The values of these endogenous discount factors are determined by the Lagrange multipliers on the sustainability constraints. The greater the severity of the credibility problem, as measured by these multipliers, the higher the planner's effective discount factor and the more the planner behaves as if she is patient. Intuitively, by requiring that the planner's continuation payoffs remain above the autarkic value, the sustainability constraints raise the shadow value of future planner payoffs. This translates into greater effective planner patience.

We then turn to allocations that are reconsideration-proof in the sense of Kocherlakota (1996). Kocherlakota (1996). A credible equilibrium is reconsideration-proof if it is optimal amongst equilibria that deliver the same payoff to the planner after all histories. The credible equilibrium concept ensures that the planner (and the agents) have no incentive to take a unilateral deviation from their strategies. The reconsideration-proof concept strengthens this by additionally requiring that the planner has no incentive to organize a joint planner-agent deviation.

The final section of the paper establishes a close connection between reconsideration-proofness and a class of problems in which the planner can commit and has a unit discount factor. In these problems, agents are reinterpreted as altruistic dynasties. Member generations of a dynasty discount the utility of future generations, but the planner does not. In this sense, she is Rawlsian. We show that reconsideration-proof allocations solve such planning problems. Hence, we obtain micro-foundations for the Rawlsian planner that are derived from credibility considerations. Phelan (2005) assumes a Rawlsian planner directly and characterizes a solution to her problem. Our results complement his by establishing that this solution is reconsideration-proof and by providing additional motivation for a unit societal discount factor.[5]

2. The policy game

In this section we describe a policy game.

---

[4] This simple setting allows us to benchmark our results against those of the well known model of Atkeson and Lucas (1992) that assumes planner commitment. However, our results generalize from this to many other settings with dynamic private information.

[5] As Phelan (2005) acknowledges, a unit societal discount factor can seem like a rather special limiting case: "Farhi and Werning (2005) and Sleet and Yeltekin (2005a) generalize this result. That is they show a much weaker assumption than zero societal discounting is sufficient to ensure finite limiting inequality."

2.1. *Physical environment: shocks and preferences*

The environment is inhabited by a unit mass of infinitely-lived agents and a planner. Agents have preferences over stochastic consumption sequences $\{c_t\}_{t=1}^{\infty}$ of the form:

$$(1-\beta)E\left[\sum_{t=1}^{\infty}\beta^{t-1}\theta_t u(c_t)\right]. \tag{1}$$

Here $\beta \in (0,1)$ denotes the agent's discount factor, while $\theta_t \in \Theta$ is an idiosyncratic taste shock privately observed by the agent. Taste shocks are assumed to satisfy the following condition.

Assumption 1. (1) $\Theta = \{\hat{\theta}_k\}_{k=1}^{K}$, $\hat{\theta}_{k+1} > \hat{\theta}_k > 0$; (2) the $\{\theta_t\}$ are independently and identically distributed over time and across agents with distribution π.[6,7]

Let $\theta^t = \{\theta_1, \dots, \theta_t\} \in \Theta^t$ denote a t-period history of shocks and let π^t denote the corresponding probability distribution. The utility function u is assumed to satisfy:

Assumption 2. $u : \mathbb{R}_+ \to D \subseteq \mathbb{R} \cup \{-\infty\}$ is continuous, strictly increasing, strictly concave and bounded above. Let $\bar{u} = \sup D < \infty$.

Notice that Assumption 2 requires that u is bounded above, but not below. Denote the inverse of u by $C : D \to \mathbb{R}_+$.

We assume that the planner is utilitarian. She seeks to allocate a fixed quantity of resources R so as to maximize an equally weighted aggregate of agent lifetime utilities. Since agent taste shocks are not public information, the planner must induce agents to reveal information about them. She does this by rewarding agents who reveal a low current desire for consumption with higher future utility and penalizing those who reveal a high current desire with reduced future utility. Thus, future allocations are used to elicit information in the present. We assume, however, that the planner cannot commit to such allocations.

2.2. *The stage game*

In period t, each agent is publicly identified by a history of past lottery outcomes $\gamma^{t-1} \in \mathbb{R}^{t-1}$ and past messages $m^{t-1} \in \mathcal{M}^{t-1}$ that the agent has previously sent to the planner, where $\mathcal{M}^{t-1}$ denotes a history of past message spaces. The ensuing stage game consists of three subperiods. In the first, a stage mechanism S_t is selected by the planner. S_t has three components: a lottery function $\rho_t : \mathbb{R}^{t-1} \times \mathcal{M}^{t-1} \to \mathcal{P}(\mathbb{R})$, a message space $\mathcal{M}_t$ and an allocation function $\varphi_t : \mathbb{R}^t \times \mathcal{M}^t \to D$, where $\mathcal{P}(\mathbb{R})$ denotes the space of probability measures on $\mathbb{R}$. The first of these components, ρ_t, gives a history-specific lottery for each agent. We denote an agent's tth period lottery outcome by γ_t. The second, $\mathcal{M}_t$, provides agents with a set of messages that they can use to communicate with the planner. We denote an agent's tth period message by m_t. To

[6] We also interpret $\pi(\theta)$ as the fraction of agents receiving the shock θ. In doing so we rely on the argument of Judd (1985).

[7] The i.i.d. assumption is widely made in models with commitment, e.g. Atkeson and Lucas (1992), FW (2005) and Phelan (2005). Our use of it enables us to compare our results directly to those obtained from such models. We discuss some of the issues involved in extending the current model to the persistent shock case later in the paper.

avoid mathematical complications, we will assume that the planner must choose message spaces with a finite number of elements. The third, φ_t, describes a current *utility* allocation for agents contingent upon their history of lottery outcomes and messages, i.e. $\varphi_t(\gamma^t, m^t)$ gives the utility from consumption obtained by an agent in period t after the lottery-message history (γ^t, m^t). The corresponding consumption allocation can be recovered using C, the inverse of u. We assume that φ_t and, for each Borel set $B \subseteq \mathbb{R}$, $\rho_t(\cdot)(B)$ are Borel measurable. In the second sub-period, agents learn the outcome of their lottery, receive their private taste shocks and choose a probability distribution over the message space. They draw a message from this distribution and transmit it to the planner. In the third and final sub-period, the allocation functions execute, delivering utility to agents contingent upon their history of lottery outcomes and messages according to φ_t. As noted, the planner cannot commit in advance to a particular sequence of future mechanisms. However, having selected a mechanism she must execute it; she cannot deviate to some alternative current allocation after hearing the agents' messages. Thus, we assume planner commitment within, but not across periods.

2.3. Histories and strategies

This subsection introduces notation and definitions for the infinitely repeated game played by the planner and the agents. Define an *aggregate history* S^t, $t \geqslant 1$, to be a sequence $\{S_s\}_{s=1}^t$ of current and past stage mechanisms chosen by the planner. Let $\mathcal{S}^t$ denote the set of t-period aggregate histories and let $\mathcal{S}_t$ denote the set of t-period stage mechanisms. A planner strategy $\sigma = \{\sigma_t\}_{t=1}^\infty$ is a collection of functions that map from aggregate histories to a current mechanism, $\sigma_t : \mathcal{S}^{t-1} \to \mathcal{S}_t$, $t \geqslant 2$, and $\sigma_1 \in \mathcal{S}_1$.[8] Any given σ induces a sequence of aggregate histories recursively according to: $S^t = (S^{t-1}, \sigma_t(S^{t-1}))$.[9] Let $\sigma|S^r$ denote the continuation of σ after aggregate history S^r and, for $t \geqslant r$, let $S^t(\sigma|S^r)$ denote the period t aggregate history and $S_t(\sigma|S^r)$ the period t stage mechanism induced by $\sigma|S^r$ along its outcome path. Let $\varphi_t(\sigma|S^r)$ and $\mathcal{M}_t(\sigma|S^r)$ denote the corresponding tth period allocation function and message space.

The period t *individual public history* of an agent $(S^t, \gamma^t, m^{t-1}) \in \mathcal{S}^t \times \mathbb{R}^t \times \mathcal{M}^{t-1}$ augments S^t with the agent's history of lottery outcomes and messages. The behavior of an agent is described by a *message strategy*, $\lambda = \{\lambda_t\}_{t=1}^\infty$, where $\lambda_t : \mathcal{S}^t \times \mathbb{R}^t \times \mathcal{M}^{t-1} \times \Theta \to \mathcal{P}(\mathcal{M}_t)$ and $\mathcal{P}(\mathcal{M}_t)$ denotes the set of probability distributions over messages at date t. Thus, λ_t gives the tth period message distribution of an agent contingent on its individual public history and current shock. For each date t, message set $B \subseteq \mathcal{M}_t$ and aggregate history S^t, $\lambda_t(S^t, \cdot)(B)$ is assumed to be Borel measurable. Any message strategy λ and aggregate history S^t induces a probability distribution over t-period histories of lottery outcomes and messages which we denote μ^t, or $\mu^t(S^t, \lambda)$ if we wish to emphasize the dependence on S^t and λ.[10]

---

[8] As in Chari and Kehoe (1990), we consider an economy inhabited by a large strategic player, the planner, and a population of atomistic agents. This structure motivates our formulation of strategies; the planner's strategy is conditioned on past histories of planner actions only. The planner's strategy does not allow its treatment of an agent to depend on the past actions of other agents.

[9] Throughout, we use the following notational convention for lists containing an element superscripted by 0: $\{\ldots, x, y^0, z, \ldots\} = \{\ldots, x, z, \ldots\}$. Thus, $S^1 = (S^0, \sigma_1(S^0)) = \sigma_1$.

[10] Given λ and the sequence $S^T = \{\rho_t, \mathcal{M}_t, \varphi_t\}_{t=1}^T$, the corresponding sequence $\{\mu^t\}_{t=1}^T$ can be recovered recursively. For each Borel set $B \in \mathcal{B}(\mathbb{R} \times \mathcal{M})$, set $\mu^1(B) = \int 1_{\{(\gamma^1, m^1) \in B\}} \lambda_1(\mathrm{d}m|S^1, \gamma^1, \theta)\pi(\theta)\rho_1(\mathrm{d}\gamma^1)$. For $t > 1$ and each Borel set $B \in \mathcal{B}(\mathbb{R}^t \times \mathcal{M}^t)$, set $\mu^t(B) = \int 1_{\{(\gamma^t, m^t) \in B\}} \lambda_t(\mathrm{d}m|S^t, \gamma^t, m^{t-1}, \theta)\pi(\theta)\rho_t(\mathrm{d}\gamma|\gamma^{t-1}, m^{t-1})\mu^{t-1} \times (\mathrm{d}\gamma^{t-1}, \mathrm{d}m^{t-1})$.

A pair of strategies (σ, λ) induce a continuation payoff for the utilitarian planner after each aggregate history S^{t-1} which we denote by $W_t(\sigma, \lambda | S^{t-1})$. Similarly, they induce a continuation payoff for the agent after each history (S^t, γ^t, m^{t-1}) which we denote $U_t(\sigma, \lambda | S^t, \gamma^t, m^{t-1})$.[11]

We define a planner strategy σ to be *resource-feasible* given message strategy λ if after each aggregate history the allocations induced by (σ, λ) consume an amount of resources less than R. More formally, we have:

Definition 1. Given a message strategy λ, a mechanism $S = (\rho, \mathcal{M}, \varphi)$, is *resource-feasible* at $S^{t-1} = \{\rho_r, \mathcal{M}_r, \varphi_r\}_{r=1}^{t-1}$ if:

$$\int_{\mathbb{R}^t \times \mathcal{M}^t} C\big(\varphi(\gamma^t, m^t)\big)\mu^t\big(S^{t-1}, S, \lambda\big)\big(\mathrm{d}\gamma^t, \mathrm{d}m^t\big) \leqslant R.$$

Similarly, given λ, a sequence of mechanisms $\{S_t\}_{t=1}^T$ is *resource-feasible* if each S_t is resource-feasible at S^{t-1}. Given λ, let $\mathcal{S}^t(\lambda)$ denote the set of t-period resource-feasible aggregate histories. Finally, given λ, a planner strategy σ is *resource-feasible* if after each $S^{t-1} \in \mathcal{S}^{t-1}(\lambda)$, $\sigma_t(S^{t-1})$ is resource-feasible at S^{t-1}.

We will also impose the following bound as a primitive constraint on planner strategies.

Definition 2. A planner strategy σ is bounded if after all aggregate histories S^r its continuation $\sigma | S^r$ induces a sequence of allocation functions satisfying,

$$\forall \gamma^\infty, m^\infty, \quad (1-\beta) \lim_{t \geqslant r} \beta^{t-r} \sum_{s=1}^{\infty} \beta^{s-1} \theta_{t+s} \varphi_{t+s}\big(\sigma | S^r\big)\big(\gamma^{t+s}, m^{t+s}\big) = 0.$$

Let $\Sigma(\lambda)$ denote the set of bounded, resource-feasible planner strategies given λ.

Remarks on mechanisms

Remark 1 (*Generality of Mechanisms*). Our assumption that the planner cannot commit means that we cannot automatically invoke the revelation principle. This necessitates our initial use of a fairly general class of mechanisms. In particular, we do not a priori restrict the planner to the use of message spaces equal to the type space Θ. Later, we show that all equilibrium payoffs are attained by equilibria in which agents are truthful provided the planner has not defected from her strategy in the current period.

Remark 2 (*On the role of lotteries*). The optimal social insurance literature with commitment has often assumed that the planner confronts an initial cross-sectional distribution of agent utility promises that it must honor. Optimal allocations are then conditioned upon these promises as well as the (reported) shock histories of agents. Our introduction of a lottery in period 1 allows us to obtain equilibrium allocations with these features. More generally, it allows us to consider credible equilibria in which agents with the same message histories receive different allocations. Given that we allow the planner to use lotteries in period 1, it is then natural to allow her to

[11] The upper bound on u and the measurability restrictions on strategies ensure that this is well defined, though possibly non-finite.

use them in subsequent periods as well. Our initial formulation does this. Later, we specialize equilibrium strategies so that they do not condition on lotteries after the initial period.

3. Credible equilibria

A *credible equilibrium* is defined as follows.

Definition 3. (σ, λ) is a *credible equilibrium* if $\sigma \in \Sigma(\lambda)$ and:

(1) *(Agent optimality)* $\forall t, S^t, \gamma^t, m^{t-1}, \hat{\lambda}$,
$$U_t\big(\sigma, \lambda \mid S^t, \gamma^t, m^{t-1}\big) \geqslant U_t\big(\sigma, \hat{\lambda} \mid S^t, \gamma^t, m^{t-1}\big);$$
(2) *(Planner optimality)* $\forall t, S^{t-1} \in \mathcal{S}^{t-1}(\lambda), \hat{\sigma} \in \Sigma(\lambda)$,
$$W_t\big(\sigma, \lambda \mid S^{t-1}\big) \geqslant W_t\big(\hat{\sigma}, \lambda \mid S^{t-1}\big).$$

The first of these conditions requires that the continuation of the agent's message strategy is optimal after all public individual histories given that the planner plays according to σ in the future. The second condition demands that after all resource-feasible aggregate histories the planner is better off adhering to the strategy σ than deviating to some alternative resource-feasible strategy $\hat{\sigma}$.

Worst credible equilibria We next define autarkic planner and message strategies and show that they constitute a worst credible equilibrium. We say that a planner strategy is autarkic if its allocation functions assign each agent an amount of utility $u(R)$ in every period. Thus, an autarkic planner strategy provides no insurance against taste shocks.

Definition 4. Let σ be a planner strategy. σ is *autarkic* if for all t, $S^{t-1} = \{\rho_s, \mathcal{M}_s, \varphi_s\}_{s=1}^{t-1}$, $\gamma^t \in \mathbb{R}^t$, $m^t \in \mathcal{M}^{t-1} \times \mathcal{M}_t(\sigma|S^{t-1})$,
$$\varphi_t\big(\sigma \mid S^{t-1}\big)\big(\gamma^t, m^t\big) = u(R). \tag{2}$$

Autarkic planner strategies imply that agents' individual public histories are ignored when utility is awarded. If the planner plays in this way, no message that an agent sends can influence future allocations and it will be optimal for agents to choose message strategies that maximize their current payoff in each period. This observation motivates the following definition of an autarkic message strategy.

Definition 5. Let λ be a message strategy. λ is *autarkic* if for all $S^t = \{S^{t-1}, (\rho_t, \mathcal{M}_t, \varphi_t)\}$, γ^t, m^{t-1}, θ_t,
$$\lambda_t\big(S^t, \gamma^t, m^{t-1}, \theta_t\big) \in \arg\max_{m \in \mathcal{M}_t} \theta_t \varphi_t\big(\gamma^t, m^{t-1}, m\big).$$

Let $V^{aut} = E[\theta]u(R)$. We say that a credible equilibrium is worst if there is no other credible equilibrium that gives a lower payoff.

Proposition 1. *Let $(\sigma^{aut}, \lambda^{aut})$ be an autarkic strategy pair. $(\sigma^{aut}, \lambda^{aut})$ is a worst credible equilibrium.*

Proof. See Appendix A. □

JID:YREDY AID:338 /FLA [m1+; v 1.59; Prn:6/04/2006; 12:17] P.8 (1-28)

The intuition for the preceding result is straightforward. If agents play according to λ^{aut}, then the planner cannot use intertemporal incentives to induce those with low taste shocks to reveal themselves and accept low current consumption. She can do no better than implementing the allocation $\{u(R), u(R), \ldots\}$. On the other hand, under σ^{aut}, agents' current messages do not affect their future allocations. Recognizing this, agents will send messages that maximize their current consumption. They will play according to λ^{aut}. Thus, $(\sigma^{aut}, \lambda^{aut})$ is a credible equilibrium with payoff V^{aut}. This equilibrium is worst because the planner can always ignore agent messages, allocate the resource amount R to each agent and obtain the payoff V^{aut}.

3.1. Revelation principle

In environments with planner commitment, the revelation principle asserts that all equilibrium payoffs can be obtained with direct mechanisms that induce agents to truthfully report their shocks. Below, we obtain a revelation principle for our environment without planner commitment. It states that all equilibrium payoffs are attained by equilibria in which agents are truthful provided the planner has not defected from her strategy in the current period. Two definitions precede the result.

Definition 6. A stage mechanism $S_t = (\rho_t, \mathcal{M}_t, \varphi_t)$ is direct if $\mathcal{M}_t = \Theta$. A planner strategy σ is direct if for all t, S^{t-1}, $\sigma_t(S^{t-1})$ is direct.

Definition 7. Let S_t be a direct mechanism. λ is truthful after individual public history $(S^{t-1}, S_t, \gamma^t, m^{t-1})$ if, for all θ,

$$\lambda_t\big(S^{t-1}, S_t, \gamma^t, m^{t-1}, \theta\big) = \theta. \tag{3}$$

Proposition 2 *(Revelation Principle). Let (σ, λ) be a credible equilibrium. Then there exists a credible equilibrium $(\hat{\sigma}, \hat{\lambda})$ such that* (1) *$\hat{\sigma}$ is direct,* (2) *$\hat{\lambda}$ is truthful after all public histories $(S^{t-1}, \sigma_t(S^{t-1}), \gamma^t, m^{t-1})$ and* (3) *the planner attains the same payoff in the equilibrium $(\hat{\sigma}, \hat{\lambda})$ as in the equilibrium (σ, λ), i.e. $W_1(\hat{\sigma}, \hat{\lambda}) = W_1(\sigma, \lambda)$.*

Proof. See Appendix A. □

Our proof of Proposition 2 is in two steps. First, we construct a sequence of direct mechanisms that delivers the same expected payoff to truthful agents as they obtained in the original equilibrium. We verify that it is optimal for agents to be truthful when confronted with these mechanisms. This step follows the standard proof of the revelation principle. Second, we show that reversion to an autarkic equilibrium (with direct mechanisms and truthful reporting) is sufficient to ensure that the planner will not defect from the sequence of mechanisms obtained in the first step.

The argument underlying Proposition 2 relies on the assumption of i.i.d. shocks. When agents adopt the autarkic message strategy, the planner is unable to induce those with a low current taste shock to reveal themselves and accept low current consumption. In the i.i.d. case, no insurance against taste shocks is possible and the planner is penalized. When shocks follow a persistent process, truthful revelation at t conveys information about an agent's likely state in future periods. Consequently, the planner can continue to provide some insurance against taste shocks and the autarky punishment is mitigated. In this case, some equilibrium payoffs may only be attainable if non-separating mechanisms are used along the equilibrium path. By revealing less information to the planner these mechanisms sharpen the autarkic penalty and deter planner defection. A reasonable conjecture is that as the persistence of shocks falls relative to the discount factor β, the incentive for the planner to defect from a given strategy after a truthful revelation of agent shock types becomes smaller. Thus, equilibria with truthful revelation along their outcome paths come closer to attaining all equilibrium payoffs. We leave analysis of this issue for future work.

Remark 3. In a recent paper, Acemoğlu et al. (2005) also consider optimal credible mechanisms. Although their environment is quite different from ours, they derive a similar revelation principle. The argument underlying their result follows ours. It establishes that a given equilibrium payoff can be attained with direct mechanisms and truthful reporting along the equilibrium path and reversion to the worst equilibrium following a planner defection. Although they allow for persistent shocks, their worst equilibrium, like ours, has a planner payoff that does not depend on past information that the agent has conveyed. Hence, non-separating mechanisms do not relax the incentive constraints on the planner and do not need to be used on the equilibrium path to attain a given equilibrium payoff.[12]

Since $W_1(\sigma, \lambda)$ gives the ex ante expected payoff to an agent, as well as to the planner, it follows from Proposition 2 that, from a welfare point of view, there is no loss of generality in restricting attention to equilibria with direct mechanisms and truthful revelation along their equilibrium paths. This restriction greatly simplifies the analysis and we make it throughout this section and the next. Thus, we limit attention to credible equilibria (σ, λ) satisfying:

Property 1. σ is direct along its outcome path and λ is truthful along this path.

It is convenient to focus upon credible equilibria (σ, λ) that satisfy a second property:

Property 2. For all t, there is a ψ_t such that for each γ_1 , γ^{t-1} and θ^t, $\varphi_t(\sigma)(\gamma_1, \gamma^{t-1}, \theta^t) = \psi_t(\gamma_1, \theta^t)$.

This property implies that equilibrium allocation functions condition allocations upon the period 1 lottery outcome and the agent's (reported) shock history only. Subsequent lottery outcomes do not influence the utility awarded to an agent and can be ignored. Again, from a welfare point of view, Property 2 is without loss of generality; we can replace any credible equilibrium that fails to satisfy it with one that does by integrating out lotteries after period 1.

The restriction to credible equilibria satisfying Properties 1 and 2 motivates the following definitions and notation for allocations. We now formally define an *allocation* to be a pair (Ψ_1, α) where $\Psi_1 \in \mathcal{P}(\mathbb{R})$ and $\alpha = \{\psi_t\}_{t=1}^{\infty}$ is a sequence of Borel measurable functions with for all t, $\psi_t : \mathbb{R} \times \Theta^t \to D$. We denote the continuation allocation obtained by an agent after the lottery-(reported) shock history (γ, θ^t) by $\alpha(\gamma, \theta^t)$ and the corresponding agent payoff by $U(\alpha(\gamma, \theta^t))$. We refer to continuation allocations $\alpha(\gamma)$ for specific γ-indexed agents and, more generally, to

[12] In their model, private information concerns agent productivity. At any time, agents may become disabled; disabled agents cannot work. The worst equilibrium entails all agents shutting the economy down by claiming to be disabled. This delivers a zero payoff to a non-benevolent government trying to extract rents, regardless of any information this government might have about the agents' pasts.

sequences of functions $\{\psi_t'\}_{t=1}^{\infty}$ with $\psi_t' : \Theta^t \to D$ as *individual allocations*. In a small abuse of notation, we let $W_t(\Psi_1, \alpha)$ denote the period t continuation payoff to the planner from the allocation (Ψ_1, α). Finally, we define a *credible allocation* to be one induced by a credible equilibrium satisfying Properties 1 and 2.

3.2. Credible allocations

In this subsection we obtain a set of conditions that are necessary and sufficient for an allocation to be credible. We state them first, then prove the desired necessity and sufficiency.

An allocation $(\Psi_1, \alpha) = (\Psi_1, \{\psi_t\}_{t=1}^{\infty})$ is *resource-feasible* if the planner consumes less than R at all dates,

$$\forall t, \quad \int \sum_{\Theta^t} C\big(\psi_t(\gamma, \theta^t)\big)\pi^t(\theta^t)\Psi_1(\mathrm{d}\gamma) \leqslant R, \tag{4}$$

bounded if,

$$\forall \gamma, \theta^{\infty}, \quad (1-\beta) \lim_{t\to\infty} \beta^t \sum_{s=1}^{\infty} \beta^{s-1}\theta_{t+s}\psi_{t+s}\big(\gamma, \theta^{t+s}\big) = 0 \tag{5}$$

and *incentive-compatible* if,

$$\begin{aligned}\forall \gamma, t, \theta^{t-1}, k, j, \quad & (1-\beta)\hat{\theta}_k\psi_t\big(\gamma, \theta^{t-1}, \hat{\theta}_k\big) + \beta U\big(\alpha\big(\gamma, \theta^{t-1}, \hat{\theta}_k\big)\big) \\ & \geqslant (1-\beta)\hat{\theta}_k\psi_t\big(\gamma, \theta^{t-1}, \hat{\theta}_j\big) + \beta U\big(\alpha\big(\gamma, \theta^{t-1}, \hat{\theta}_j\big)\big).\end{aligned} \tag{6}$$

The constraints (6) require that after each history of shocks, an agent is better off truthfully reporting its state, than lying and being truthful thereafter.

Finally, we introduce a set of *sustainability constraints* that require that the planner's continuation payoff remains above the autarkic value. We say that an allocation is *sustainable* if

$$\forall t, \quad W_t(\Psi_1, \alpha) \geqslant V^{aut}. \tag{7}$$

The next proposition establishes that conditions (4) through (7) are necessary and sufficient for an allocation to be credible.

Proposition 3. *An allocation* $(\Psi_1, \alpha) = (\Psi_1, \{\psi_t\}_{t=1}^{\infty})$ *is credible, if and only if it satisfies the resource-feasibility* (4)*, boundedness* (5)*, incentive-compatibility* (6) *and sustainability* (7) *constraints.*

Proof. See Appendix A. □

4. Optimal allocations with and without commitment

4.1. Optimal allocations with commitment and arbitrary planner discounting

The literature has largely assumed planner commitment and equal planner-agent discounting. Up to this point, we have dropped the first of these assumptions, but retained the second. We now reverse this and, instead, consider a planner who *can* commit, but who uses a different discounting scheme from the agents. At first, this seems to lead to planning problems that are unrelated to our earlier credibility game. In fact, as we show, optimal credible allocations solve such problems.

Suppose that the planner uses the discounting scheme $\{B_1^t\}_{t=1}^{\infty}$, where $B_1^t \in (0, 1]$ and $B_1^1 = 1$, to weight future payoffs, while agents continue to use the discount factor β. As before, we assume that taste shocks are privately observed by agents and that the planner has a finite quantity of resources to allocate. Thus, the planner is still restricted by the resource (4), boundedness (5) and incentive-compatibility (6) constraints. Now, however, we assume commitment, so that the planner no longer needs to respect the sustainability constraint. Her choice problem is defined as follows.

Definition 8. The *commitment problem with arbitrary planner discounting* is given by:

$$\sup_{\{\Psi_1, \{\psi_t\}_{t=1}^{\infty}\}} \liminf_{T\to\infty} \frac{1}{\sum_{t=1}^{T} B_1^t} \int_{\mathbb{R}} \sum_{t=1}^{T} B_1^t \sum_{\Theta^t} \theta_t \psi_t(\gamma, \theta^t) \pi^t(\theta^t) \Psi_1(\mathrm{d}\gamma)$$

$$\text{s.t.} \quad (4), (5) \text{ and } (6). \tag{$\mathcal{CP}$}$$

We make two preliminary remarks on this general discounting problem. First, the use of the lim inf operator ensures that the objective in $(\mathcal{CP})$ is well defined even if $\lim_{T\to\infty} \sum_{t=1}^{T} B_1^t = \infty$. For example, it is well defined when the planner has a unit discount factor and each $B_1^t = 1$. Second, if $\lim_{T\to\infty} \sum_{t=1}^{T} B_1^t < \infty$, then the upper bound on agent utility guarantees that the limiting planner payoff exists and that the objective in $(\mathcal{CP})$ can be re-expressed as

$$\int_{\mathbb{R}} \sum_{t=1}^{\infty} B_1^t \sum_{\Theta^t} \theta_t \psi_t(\gamma, \theta^t) \pi^t(\theta^t) \Psi_1(\mathrm{d}\gamma). \tag{8}$$

In this case, the strict concavity of u ensures that the planner prefers to assign agents the same individual utility allocation regardless of their γ index. $(\mathcal{CP})$ can then be reduced to a problem in which the planner selects a common individual utility allocation, $\{\psi_t^*\}_{t=1}^{\infty}$, with $\psi_t^*: \Theta^t \to D$, and the distribution over γ indices is degenerate. (In contrast, when $\sum_{t=1}^{\infty} B_1^t = \infty$, the payoff to the planner over any finite number of periods is 0 and it is often more convenient to consider solutions that are stationary and that assign different γ-indexed agents different utility allocations, see, for example, Phelan, 2005.)

We now describe three specific cases and relate them to contributions elsewhere in the literature.

Case 1: $B_1^t = \beta^{t-1}$. When $B_1^t = \beta^{t-1}$ for all $t, \beta \in (0, 1)$, the planner and the agents' discount factors coincide and $(\mathcal{CP})$ essentially reduces to a primal version of the problem considered by Atkeson and Lucas (1992).[13] In this case, for a large class of utility functions u, the solution to $(\mathcal{CP})$ satisfies a well known agent immiseration property. Denoting the individual utility allocation that solves $(\mathcal{CP})$ by $\{\psi_t^*\}_{t=1}^{\infty}$, this property can be stated as: for π^{∞}-a.e. θ^{∞}, $\lim_{t\to\infty} U(\{\psi_{t+s}^*(\theta^t, \cdot)\}_{s=1}^{\infty})$ exists and equals $E[\theta] \inf D$. When u is

[13] In Atkeson and Lucas (1992), the planner chooses a sequence of allocation functions $\{\psi_t\}_{t=1}^{\infty}$. Ψ_1 is interpreted as a distribution over utility promises and is treated as a parameter, not a choice variable. It follows that $(\mathcal{CP})$ with $B_1^t = \beta^{t-1}$ is the primal version of an Atkeson–Lucas problem in which Ψ_1 places all of its mass at the payoff obtained by agents in $(\mathcal{CP})$, $B_1^t = \beta^{t-1}$.

bounded above, this agent immiseration property automatically carries over to the planner and $\lim_{t\to\infty}\sum_{\Theta^t} U(\{\psi^*_{t+s}(\theta^t,\cdot)\}_{s=1}^{\infty})\pi^t(\theta^t) = E[\theta]\inf D$. Since $V^{aut} = E[\theta]u(R) > E[\theta]u(0) = E[\theta]\inf D$, it follows that for a large class of utility functions, the optimal allocation with commitment and equal discounting is not credible. This holds independently of the value of the common discount factor β; there is no Folk theorem available here.

Case 2: $B_1^t = \delta^{t-1} \in (\beta^{t-1}, 1)$. In this case, the planner has a discount factor δ in excess of the agents'. Problems of this general form have been considered by FW and SY. Both modify $(\mathcal{CP})$ by adding a promise-keeping constraint. This constraint requires that the planner must deliver a pre-specified utility amount $U(\gamma) \in (E[\theta]\inf D, E[\theta]\bar{u})$ to each γ-indexed agent.[14] The planner's problem is then:

$$\sup_{\{\psi_t\}_{t=1}^{\infty}} \int_{\mathbb{R}} \sum_{t=1}^{\infty} \delta^{t-1} \sum_{\Theta^t} \theta_t \psi_t\big(\gamma,\theta^t\big)\pi^t\big(\theta^t\big)\Psi_1(\mathrm{d}\gamma)$$

s.t. (4), (5), (6), and

$$U(\gamma) = (1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}\sum_{\Theta^t}\theta_t\psi_t\big(\gamma,\theta^t\big)\pi^t\big(\theta^t\big). \tag{9}$$

Now, Ψ_1 is taken as a primitive rather than a choice variable, as is the function $U(\gamma)$. FW (2005) and SY (2005a) show that problems of this form can have solutions that are ergodic. This ergodicity implies that there is social mobility even in the long run. Thus, the pathological immiseration result found when planner and agent discount factors coincide is overturned.

Case 3: The Rawlsian planner. Phelan (2005) considers an economy inhabited by altruistic dynasties. Each generation of a dynasty uses the discount factor β to value the utility of future generations. Thus, the tth generation of a dynasty with history (γ,θ^t) attaches value

$$(1-\beta)\sum_{s=0}^{\infty}\beta^s\sum_{\Theta^s}\psi_{t+s}\big(\gamma,\theta^{t+s}\big)\pi^s\big(\theta^s\big)$$

to the allocation $(\Psi_1,\{\psi_t\}_{t=1}^{\infty})$. The planner is Rawlsian in the sense that her objective places equal weight on generations:

$$\liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}}\sum_{s=0}^{\infty}\beta^s\sum_{\Theta^{t+s}}\theta_{t+s}\psi_{t+s}\big(\gamma,\theta^{t+s}\big)\pi^{t+s}\big(\theta^{t+s}\big)\Psi_1(\mathrm{d}\gamma). \tag{10}$$

While not immediately of the form assumed in $(\mathcal{CP})$, this objective coincides with that of a planner who applies a unit discount factor to the utility streams obtained by infinitely-lived agents:

$$\liminf_{T\to\infty}\frac{1}{T}\int_{\mathbb{R}}\sum_{t=1}^{T}\sum_{\Theta^t}\theta_t\psi_t\big(\gamma,\theta^t\big)\pi^t\big(\theta^t\big)\Psi_1(\mathrm{d}\gamma) \tag{11}$$

[14] In SY (2005a), the promise-keeping constraints are implicit. They assume that the planner uses a sequence of personalized discount factors to evaluate the future payoffs of agents. Their planner objective is equivalent to a Lagrangian from a problem in which the planner uses a single discount factor, different from the agents', to evaluate future payoffs, but is constrained by promise-keeping constraints.

when restricted to the set of allocations satisfying the bound[15]:

$$\sup_t \left\{ \sum_{r=0}^{\infty} \beta^{t+r-1} \sum_{\Theta^{t+r}} \theta_{t+r} \left|\psi_{t+r}\left(\gamma, \theta^{t+r}\right)\right| \pi^{t+r}\left(\theta^{t+r}\right) \Psi_1(\mathrm{d}\gamma) \right\} < \infty. \tag{12}$$

Moreover, the solution obtained by Phelan (2005) to the Rawlsian planner's problem (i.e. the problem with objective (10) subject to the resource, boundedness and incentive-compatibility constraints) solves the planner's problem in which (11) replaces (10). Importantly, this solution, like those considered by FW and SY, does not entail immiseration and is ergodic, with the implications for social mobility and long run inequality previously described.

4.2. *Optimal allocations without commitment*

We now return to the credibility game. This game typically admits many credible allocations, although, as mentioned in the discussion of Case 1, the optimal allocation with commitment is not amongst them. In this section, we focus on credible allocations that maximize the planner's payoff.

We begin by considering the following *no commitment problem* in which the planner's (utilitarian) objective is maximized subject to the resource (4), boundedness (5), incentive-compatibility (6), and sustainability (7) constraints.

Definition 9. The planner's *optimal no commitment problem* is given by:

$$\sup_{\{\psi_t\}_{t=1}^{\infty}} (1-\beta) \sum_{t=1}^{\infty} \beta^{t-1} \sum_{\Theta^t} \psi_t\left(\theta^t\right) \pi^t\left(\theta^t\right)$$
$$\text{s.t.} \quad (4), (5), (6), \text{and } (7). \tag{$\mathcal{NCP}$}$$

As in the commitment case ($\mathcal{CP}$), a solution to this problem assigns the same individual allocation to agents; to simplify the exposition, we drop the γ indices from the notation. In light of Proposition 3, ($\mathcal{NCP}$)'s solution clearly implies an optimal credible allocation.

4.3. *Credibility and societal patience*

In this section we show that the no commitment problem ($\mathcal{NCP}$) can be reformulated as one in which the planner *can* commit, *but* has an endogenously perturbed discounting scheme $\{B_1^t\}_{t=1}^{\infty}$. Thus, ($\mathcal{NCP}$) is equivalent to a problem in the class ($\mathcal{CP}$). The perturbed scheme discounts the future less heavily than that of the agents, $B_1^t \geqslant \beta^{t-1}$, and so, in the credible equilibria underlying ($\mathcal{NCP}$), the planner behaves as if she can commit and is more patient than the agents. In this specific sense, credibility is a force for societal patience.

To establish the equivalence between ($\mathcal{NCP}$) and problems of the form ($\mathcal{CP}$), we first construct a Lagrangian that incorporates the sustainability constraints. The Lagrangian is given by:

$$(1-\beta)\mathcal{L}^{NC}\left(\alpha, \phi^{\infty}\right) = (1-\beta) \sum_{t=1}^{\infty} \beta^{t-1} \sum_{\Theta^t} \psi_t\left(\theta^t\right) \pi^t\left(\theta^t\right)$$

[15] The proof of this claim is a simple application of Abel's (summation by parts) lemma and the monotone convergence theorem.

$$+\sum_{t=2}^{\infty}\beta^{t-1}\prod_{s=2}^{t-1}(1+\phi_s)\phi_t \times\left[(1-\beta)\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\psi_{t+r}\left(\theta^{t+r}\right)\pi^{t+r}\left(\theta^{t+r}\right)-V^{aut}\right] \tag{13}$$

where $\phi^\infty=\{\phi_t\}_{t=2}^{\infty}\in\mathbb{R}_+^\infty$ denotes a sequence of multipliers on the sustainability constraints.[16] Since we assume that the agent's utility function is bounded above, each term $(1-\beta)\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\psi_{t+r}(\theta^{t+r})\,\pi^{t+r}(\theta^{t+r})-V^{aut}$ is also bounded above and the Lagrangian is well defined (though possibly infinitely-valued). We now provide conditions such that this Lagrangian can be rearranged to give an objective of the form (8).

Lemma 1. *Assume that* $\sum_{t=1}^{\infty}\beta^{t-1}\prod_{s=2}^{t}(1+\phi_s)<\infty$ *and* $\sup_t\{\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\theta_r\times|\psi_{t+r}(\theta^{t+r})|\pi^{t+r}(\theta^{t+r})\}<\infty$*, then*

$$\mathcal{L}^{NC}(\alpha,\phi^\infty)=\sum_{t=1}^{\infty}B_1^t\sum_{\Theta^t}\theta_t\psi_t\left(\theta^t\right)\pi^t\left(\theta^t\right)+K\left(\phi^\infty\right) \tag{14}$$

where $B_1^t=\beta^{t-1}\prod_{s=2}^{t}(1+\phi_s)$, $t>1$, $B_1^1=1$ *and* $K(\phi^\infty)=[-\sum_{t=1}^{\infty}B_1^t+1]\frac{V^{aut}}{1-\beta}$.

Proof. Recall Abel's Lemma (see, for example, Rudin, 1976, Theorem 3.41, p. 70):

$$\sum_{t=1}^{T}a_tb_t=\sum_{t=1}^{T-1}\left(\sum_{s=1}^{t}a_s\right)(b_t-b_{t+1})+\left(\sum_{s=1}^{T}a_s\right)b_T. \tag{15}$$

Let $b_t=\beta^{t-1}\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\theta_{t+r}\psi_{t+r}(\theta^{t+r})\pi^{t+r}(\theta^{t+r})$, $a_1=1$ and for $t\geqslant 2$, $a_t=\prod_{s=2}^{t-1}(1+\phi_s)\phi_t$. Substituting these expressions into (15) and using the definition of B_1^t yields:

$$\begin{aligned}&\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{1+r}}\theta_{1+r}\psi_{1+r}\left(\theta^{1+r}\right)\pi^{1+r}\left(\theta^{1+r}\right)\\&\quad+\sum_{t=2}^{T}\left(\prod_{s=2}^{t-1}(1+\phi_s)\phi_t\right)\beta^{t-1}\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\theta_{t+r}\psi_{t+r}\left(\theta^{t+r}\right)\pi^{t+r}\left(\theta^{t+r}\right)\\&=\sum_{t=1}^{T-1}B_1^t\left(\sum_{\Theta^t}\theta_t\psi_t\left(\theta^t\right)\pi^t\left(\theta^t\right)\right)+B_1^T\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{T+r}}\theta_{T+r}\psi_{T+r}\left(\theta^{T+r}\right)\pi^{T+r}\left(\theta^{T+r}\right).\end{aligned} \tag{16}$$

By assumption $\sum_{t=1}^{\infty}B_1^t<\infty$ and each $B_1^t\geqslant 0$, so $\lim_{T\to\infty}B_1^T=0$. Also, $\sup_T\{\sum_{r=0}^{\infty}\beta^r\times\sum_{\Theta^{T+r}}\theta_{T+r}|\psi_{T+r}(\theta^{T+r})|\pi^{T+r}(\theta^{T+r})\}<\infty$. Hence, $\lim_{T\to\infty}B_1^T\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{T+r}}\theta_{T+r}\times\psi_{T+r}(\theta^{T+r})\pi^{T+r}(\theta^{T+r})=0$. Taking limits on both sides of (16) then gives:

$$\sum_{r=1}^{\infty}\beta^{r-1}\sum_{\Theta^r}\theta_r\psi_r\left(\theta^r\right)\pi^r\left(\theta^r\right)$$

[16] To keep the notation concise, we set the term $\prod_{s=2}^{1}(1+\phi_s)$ equal to 1. We do not impose the sustainability constraint at date 1 explicitly. If a solution to $(\mathcal{NCP})$ exists, which we assume, it is redundant.

$$+\sum_{t=2}^{\infty}\left(\prod_{s=2}^{t-1}(1+\phi_s)\phi_t\right)\beta^{t-1}\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\theta_{t+r}\psi_{t+r}\big(\theta^{t+r}\big)\pi^{t+r}\big(\theta^{t+r}\big)$$
$$=\sum_{t=1}^{\infty}B_1^t\left(\sum_{\Theta^t}\theta_t\psi_t\big(\theta^t\big)\pi^t\big(\theta^t\big)\right). \tag{17}$$

Substituting this equality and the definition of $K(\phi^\infty)$ into (13) gives the desired expression for the Lagrangian. □

Remark 4. Lemma 1 imposes bounds on both multiplier sequences and utility allocations. Absent these, the rearrangement of the Lagrangian in the lemma need not be valid. The bound on utility allocations in the lemma is distinct from our earlier pointwise bound on individual allocations (5). The latter was used to ensure the sufficiency of the temporary incentive constraints (6) for full incentive-compatibility.

With Lemma 1 in hand, we give our main results. The first establishes that any (bounded) solution to ($\mathcal{NCP}$) solves a commitment problem ($\mathcal{CP}$) with a perturbed discounting scheme. The proof is in two steps. The first verifies that the solution solves a Lagrangian, the second invokes Lemma 1 and rearranges this Lagrangian to obtain a commitment problem.

Proposition 4. *If $\alpha^{NC}=\{\psi_t^{NC}\}_{t=1}^{\infty}$ solves the no commitment problem ($\mathcal{NCP}$) and satisfies*

$$\sup_t\left\{\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\theta_{t+r}\left|\psi_{t+r}^{NC}\big(\theta^{t+r}\big)\right|\pi^{t+r}\big(\theta^{t+r}\big)\right\}<\infty, \tag{18}$$

then it solves

$$\sup_{\{\psi_t\}_{t=1}^{\infty}}\sum_{t=1}^{\infty}B_1^t\sum_{\Theta^t}\theta_t\psi_t\big(\theta^t\big)\pi^t\big(\theta^t\big)$$
$$\text{s.t.}\quad \text{the resource (4), boundedness (5), and incentive-compatibility constraints (6),} \tag{19}$$

for some sequence $\{B_1^t\}_{t=1}^{\infty}$, with $B_1^t\in[0,1]$ and $B_1^1=1$.

Proof. Since α^{NC} satisfies the bound (18), there is no loss of generality in augmenting the no commitment problem with the constraint:

$$\sup_t\left\{\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\theta_{t+r}\left|\psi_{t+r}\big(\theta^{t+r}\big)\right|\pi^{t+r}\big(\theta^{t+r}\big)\right\}<\infty. \tag{20}$$

Denote the set of allocations satisfying the resource (4), boundedness (5), incentive-compatibility (6) constraints and the new bound (20) by $\mathcal{U}$. Define $G:\mathcal{U}\to\mathbb{R}^\infty$ by:

$$G\big(\{\psi_t\}_{t=1}^{\infty}\big)=\left\{(1-\beta)\sum_{r=0}^{\infty}\beta^r\sum_{\Theta^{t+r}}\theta_{t+r}\psi_{t+r}\big(\theta^{t+r}\big)\pi^{t+r}\big(\theta^{t+r}\big)-V^{aut}\right\}_{t=2}^{\infty}. \tag{21}$$

Given the definition of $\mathcal{U}$, it is immediate that $G:\mathcal{U}\to\ell_\infty$, where $\ell_\infty=\{\{x_t\}_{t=1}^{\infty}\colon \sup_t|x_t|<\infty\}$. We now show that $\{\{\psi_t\}_{t=1}^{\infty}\in\mathcal{U}\colon G(\{\psi_t\}_{t=1}^{\infty})\geqslant 0\}$ has an interior point. To do this, we perturb the autarkic allocation as follows. In each odd period, an additional amount of resources δ_{odd}

is given to any agent who reports the highest shock value and an amount equal to $\delta_{\text{odd}} \frac{\pi(\hat{\theta}_N)}{1-\pi(\hat{\theta}_N)}$ is deducted from all other agents. In each even period, an amount δ_{even} is deducted from any agent who reported the highest shock in the previous period (regardless of their current report) and an additional amount of resources equal to $\delta_{\text{even}} \frac{\pi(\hat{\theta}_N)}{1-\pi(\hat{\theta}_N)}$ is given to all other agents. We choose $\frac{\delta_{\text{even}}}{\delta_{\text{odd}}} \in (\frac{\hat{\theta}_{N-1}}{\beta E[\theta]}, \frac{\hat{\theta}_N}{\beta E[\theta]})$. If δ_{odd} is sufficiently small, this perturbed allocation satisfies the incentive-compatibility constraints. It is clearly resource-feasible and satisfies both of the bounds on allocations. For δ_{odd} small it raises the planner's payoff (above the autarkic one) by approximately $\Delta V = \frac{1}{1+\beta}(\sum_{k=1}^{N-1} (\hat{\theta}_N - \hat{\theta}_k) \frac{\pi(\hat{\theta}_k)}{1-\pi(\hat{\theta}_N)}) u'(R) \delta_{\text{odd}} \pi(\hat{\theta}_N)$ in odd periods and $\beta \Delta V$ in even ones.

The planner's objective and the function G are both (weakly) concave and, by the previous argument, $\{\{\psi_t\}_{t=1}^{\infty} \in \mathcal{U}\colon G(\{\psi_t\}_{t=1}^{\infty}) \geqslant 0\}$ has an interior point. Since α^{NC} satisfies the bound (18), the planner's payoff is finite. It then follows from Luenberger (1969, Theorem 1, p. 217) that there exists a sequence $\phi = \{\phi_t\}_{t=2}^{\infty}$ such that α^{NC} attains the supremum in:

$$\sup_{\{\psi_t\}_{t=1}^{\infty}} (1-\beta) \sum_{t=1}^{\infty} \beta^{t-1} \sum_{\Theta^t} \psi_t(\theta^t) \pi^t(\theta^t)$$
$$+ \sum_{t=2}^{\infty} \beta^{t-1} \prod_{s=2}^{t-1} (1+\phi_s) \phi_t \left[(1-\beta) \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \psi_{t+r}(\theta^{t+r}) \pi^{t+r}(\theta^{t+r}) - V^{aut} \right], \tag{22}$$

$\sum_{t=2}^{\infty} \beta^{t-1} \phi_t \prod_{s=2}^{t-1} (1+\phi_s) < \infty$ and for all t, $\phi_t \geqslant 0$. The desired result then follows from Lemma 1. □

We now prove a converse result. We show that if α^{NC} is a bounded allocation that (1) solves a commitment problem ($\mathcal{CP}$) with discount factors constructed from a bounded multiplier sequence and (2), along with this sequence, satisfies a complementary slackness condition, then α^{NC} solves the no commitment problem ($\mathcal{NCP}$).

Proposition 5. *Let* $\alpha^{NC} = \{\psi_t^{NC}\}_{t=1}^{\infty}$ *with* $\psi_t^{NC} : \Theta^t \to D$ *and let* $\phi^{\infty} = \{\phi_t\}_{t=2}^{\infty} \in \mathbb{R}_+^{\infty}$. *Define* $B_1^t = \beta^{t-1} \phi_t \prod_{s=2}^{t-1}(1+\phi_s)$, $t \geqslant 2$ *and* $B_1^1 = 1$. *Suppose* α^{NC} *and* ϕ^{∞} *satisfy the following conditions*:

(1) *(Boundedness of multipliers)*:

$$\sum_{t=2}^{\infty} \beta^{t-1} \phi_t \prod_{s=2}^{t-1} (1+\phi_s) < \infty;$$

(2) *(Boundedness of allocations)*:

$$\sup_t \left\{ \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \left| \psi_{t+r}^{NC}(\theta^{t+r}) \right| \pi^{t+r}(\theta^{t+r}) \right\} < \infty;$$

(3) *(Sustainability)*: *for* $t \geqslant 2$,

$$\phi_t \left(\sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \psi_{t+r}^{NC}(\theta^{t+r}) \pi^{t+r}(\theta^{t+r}) - \frac{V^{aut}}{1-\beta} \right) = 0 \quad \textit{and}$$

$$\sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \psi_{t+r}^{NC}\big(\theta^{t+r}\big)\pi^{t+r}\big(\theta^{t+r}\big) - \frac{V^{aut}}{1-\beta} \geqslant 0;$$

(4) *(Planner optimality)*: α^{NC} *solves*

$$\sup_{\{\psi_t\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} B_1^t \sum_{\Theta^t} \theta_t \psi_t\big(\theta^t\big)\pi^t\big(\theta^t\big)$$

s.t. resource (4), *boundedness* (5), *and incentive-compatibility constraints* (6). (23)

Then α^{NC} *solves the no commitment problem* $(\mathcal{NCP})$.

Proof. Suppose the proposition is false, then there exists a $\{\psi_t'\}_{t=1}^{\infty}$ that is feasible for the no commitment problem $(\mathcal{NCP})$ and that satisfies: $\sum_{t=1}^{\infty} \beta^{t-1} \sum_{\Theta^t} \theta_t \psi_t'(\theta^t)\pi^t(\theta^t) > \sum_{t=1}^{\infty} \beta^{t-1} \sum_{\Theta^t} \theta_t \psi_t^{NC}(\theta^t)\pi^t(\theta^t)$. Since $\{\psi_t'\}_{t=0}^{\infty}$ is feasible for $(\mathcal{NCP})$, it satisfies the sustainability constraints. Hence, for all $t > 1$, ϕ_t $(\sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \psi_{t+r}'(\theta^{t+r})\pi^{t+r}(\theta^{t+r}) - \frac{V^{aut}}{1-\beta}) \geqslant 0 = \phi_t(\sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \psi_{t+r}^{NC}(\theta^{t+r})\pi^{t+r}(\theta^{t+r}) - \frac{V^{aut}}{1-\beta})$. We then obtain:

$$\begin{aligned}
\infty &> \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{1+r}} \theta_{1+r} \psi_{1+r}'\big(\theta^{1+r}\big)\pi^{1+r}\big(\theta^{1+r}\big) \\
&\quad + \sum_{t=2}^{\infty} \beta^{t-1} \prod_{s=2}^{t-1} (1+\phi_s)\phi_t \left(\sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \psi_{t+r}'\big(\theta^{t+r}\big)\pi^{t+r}\big(\theta^{t+r}\big) - \frac{V^{aut}}{1-\beta} \right) \\
&> \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{1+r}} \theta_{1+r} \psi_{1+r}^{NC}\big(\theta^{1+r}\big)\pi^{1+r}\big(\theta^{1+r}\big) \\
&= \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{1+r}} \theta_{1+r} \psi_{1+r}^{NC}\big(\theta^{1+r}\big)\pi^{1+r}\big(\theta^{1+r}\big) \\
&\quad + \sum_{t=2}^{\infty} \beta^{t-1} \prod_{s=2}^{t-1} (1+\phi_s)\phi_t \left(\sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \psi_{t+r}^{NC}\big(\theta^{t+r}\big)\pi^{t+r}\big(\theta^{t+r}\big) - \frac{V^{aut}}{1-\beta} \right) \\
&= \sum_{t=1}^{\infty} B_1^t \sum_{\Theta^t} \theta_t \psi_t^{NC}(\theta^t)\pi^t\big(\theta^t\big).
\end{aligned}$$

The first inequality above stems from the boundedness above of the agent's utility function and the boundedness of multipliers (property (1)), the second from the definition of $\{\psi_t'\}$ and the earlier discussion. The first equality follows from the sustainability condition (property (3)), the second from the boundedness of the multipliers ϕ^{∞} and the allocation $\{\psi_t^{NC}\}_{t=1}^{\infty}$ (properties (1) and (2)) and the earlier Lemma 1. Next, we have:

$$\begin{aligned}
&\sum_{t=1}^{\infty} B_1^t \sum_{\Theta^t} \theta_t \psi_t^{NC}\big(\theta^t\big)\pi^t\big(\theta^t\big) \\
&\quad \geqslant \sum_{t=1}^{\infty} B_1^t \sum_{\Theta^t} \theta_t \psi_t'\big(\theta^t\big)\pi^t\big(\theta^t\big)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{1+r}} \theta_{1+r} \psi'_{1+r}\big(\theta^{1+r}\big) \pi^{1+r}\big(\theta^{1+r}\big) \\
&\quad + \sum_{t=2}^{\infty} \left(\prod_{s=2}^{t-1} (1+\phi_s)\phi_t \right) \beta^{t-1} \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \psi'_{t+r}\big(\theta^{t+r}\big) \pi^{t+r}\big(\theta^{t+r}\big) \\
&\quad - \lim_{T\to\infty} B_1^T \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{T+r}} \theta_{T+r} \psi'_{T+r}\big(\theta^{T+r}\big) \pi^{T+r}\big(\theta^{T+r}\big) \\
&\geqslant \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{1+r}} \theta_{1+r} \psi'_{1+r}\big(\theta^{1+r}\big) \pi^{1+r}\big(\theta^{1+r}\big) \\
&\quad + \sum_{t=2}^{\infty} \left(\prod_{s=2}^{t-1} (1+\phi_s)\phi_t \right) \beta^{t-1} \sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{t+r}} \theta_{t+r} \psi'_{t+r}\big(\theta^{t+r}\big) \pi^{t+r}\big(\theta^{t+r}\big).
\end{aligned}
$$

Here, the first inequality stems from the planner optimality condition (property (4)). The first equality is obtained by applying Abel's Lemma (as in Lemma 1), taking limits, using the finiteness of $\sum_{r=0}^{\infty} \beta^r \sum_{\Theta^{1+r}} \theta_{1+r} \psi'_{1+r}(\theta^{1+r}) \pi^{1+r}(\theta^{1+r}) + \sum_{t=2}^{\infty} (\prod_{s=2}^{t-1} (1+\phi_s)\phi_t) \beta^{t-1} \sum_{r=0}^{\infty} \beta^r \times \sum_{\Theta^{t+r}} \theta_{t+r} \psi'_{t+r}(\theta^{t+r}) \pi^{t+r}(\theta^{t+r})$ (which follows from our earlier sequence of inequalities) and the boundedness above of u. The final inequality also uses the boundedness above of u and the boundedness of multipliers. Combining the two sequences of inequalities yields a contradiction. □

Discussion The sustainability constraints from problem ($\mathcal{NCP}$) no longer appear in (19) or (23), they have been absorbed into a modified objective. This objective corresponds to that of a planner who uses a perturbed discounting scheme $\{B_1^t\}$ constructed from the (optimal) Lagrange multipliers ϕ^∞ to value future allocations. By construction, for all t, $B_1^t \geqslant \beta^t$, and so the behavior of a planner facing the sustainability constraints (7) resembles that of a *more patient* planner who faces no such constraints and uses the perturbed discounting scheme. Additionally, the planner's high discount factors in the reformulation are directly tied to her inability to commit via the optimal multipliers ϕ^∞. If the planner's sustainability constraint binds in period $t+1$, then $\phi_{t+1} > 0$ and her perturbed discount factor between periods t and $t+1$ will be $\beta(1+\phi_{t+1}) > \beta$.

For a large class of utility functions, optimal allocations with commitment and equal discounting are immiserating; they eventually violate the sustainability constraints. Consequently, in these cases, there must be some date at which the sustainability constraint binds and the planner's perturbed discount factor exceeds the agents'. Thereafter, the sustainability constraints bind infinitely often, and the planner's perturbed discount factor is repeatedly greater than that of the agents'.[17]

[17] It is difficult to derive further analytical characterization of the optimal multipliers and the planner's perturbed discount factors at this level of generality. Numerical calculations suggest that both converge to constants, with the limiting multiplier being positive and the limiting effective planner discount factor exceeding that of the agents'. The corresponding optimal allocation converges to one characterized by a stationary distribution over agent utilities and consumptions. Thus, the computed optimal allocation and multiplier sequence satisfies the boundedness conditions in Proposition 5. See SY (2005b) for more numerical details.

4.4. General Pareto optimal allocations without commitment

The optimal no commitment problem ($\mathcal{NCP}$) identified the best credible allocation from both the planner's utilitarian perspective and from the ex ante perspective of an agent who had not yet been assigned her γ-index. More generally, one can define a family of Pareto optimal no commitment problems. In these, a credible allocation is chosen to maximize a Pareto-weighted sum of agent utilities. Such allocations are the outcomes of credible equilibria in which the initial lottery over γ-indices serves as a lottery over Pareto weights. These problems are formally defined as follows.

Definition 10. Let Ψ_1 have support in $\mathbb{R}_+$. The Ψ_1-*no commitment problem* is given by:

$$\sup_{\{\psi_t\}_{t=1}^{\infty}} (1-\beta)\int \gamma \sum_{t=1}^{\infty} \beta^{t-1} \sum_{\Theta^t} \theta_t \psi_t(\gamma, \theta^t)\pi^t(\theta^t)\Psi_1(\mathrm{d}\gamma)$$
$$\text{s.t.}\quad (4),\ (5),\ (6),\ \text{and}\ (7). \qquad (\mathcal{NCP}(\Psi_1))$$

As in the previous section, a Lagrangian that incorporates the sustainability constraints can be constructed for this problem and rearranged to give a new objective with a perturbed discounting scheme. Now, however, a further rearrangement coupled with the addition of a constant term allows us to reinterpret the Pareto weights, γ, as Lagrange multipliers on a family of promise-keeping constraints. These constraints require that the planner deliver a lifetime utility amount $U^*(\gamma)$ to each agent. More precisely, suppose that (($\mathcal{NCP}\Psi_1$)) has a solution and let $U^*(\gamma)$ denote the payoff obtained by the γth agent from this solution. Adding the constant term $\int_\gamma (\gamma - 1)\frac{U^*(\gamma)}{1-\beta}\Psi_1(\mathrm{d}\gamma)$ to the rearranged Lagrangian allows us to reformulate (($\mathcal{NCP}\Psi_1$)) as[18]:

$$\sup_{\{\psi_t\}_{t=1}^{\infty}} \int_\gamma \sum_{t=1}^{\infty} B_1^t \sum_{\Theta^t} \theta_t \psi_t(\gamma, \theta^t)\pi^t(\theta^t)\Psi_1(\mathrm{d}\gamma)$$
$$+ \int_\gamma (\gamma - 1)\left\{\sum_{t=1}^{\infty} \beta^{t-1} \sum_{\Theta^t} \psi_t(\gamma, \theta^t)\pi^t(\theta^t) - \frac{U^*(\gamma)}{1-\beta}\right\}\Psi_1(\mathrm{d}\gamma)$$
$$\text{s.t.}\quad (4),\ (5),\ \text{and}\ (6). \qquad (24)$$

Here, as before, $B_1^t = \prod_{s=2}^{t-1}(1+\phi_s)\phi_t$, $t \geqslant 2$, $B_1^1 = 1$, where $\{\phi_t\}$ are the optimizing Lagrange multipliers from the sustainability constraints. Thus, the Pareto-weighted no commitment problem can be reformulated as one from the class of problems considered by FW (2005) and SY (2005a), i.e. as one with a committed planner who uses discount factors that exceed those of the agents and must implement a particular distribution of utility promises. It follows that credibility considerations provide microfoundations for the sorts of problems considered by FW and SY and, in particular, for the high planner discount factors assumed in both of their papers.

As noted previously, FW and SY show that when the planner's discount factor takes a fixed value in excess of the agents', problems of the form (24) can have ergodic allocations as solutions. Such allocations have the economic properties previously described in Section 4.1, Case 2.

[18] The reformulation is subject to the qualification that the optimal allocation and the optimal multipliers must be appropriately bounded, as in our previous discussion of the optimal non-commitment problem.

Numerical calculations (e.g. SY, 2005a, 2005b) indicate that there exists an ergodic Pareto optimal credible allocation with a constant planner discount factor endogenously generated by the optimal sustainability multipliers. Moreover, they also indicate that the optimal non-commitment allocation converges to this ergodic one.

5. Reconsideration proofness

Payoff stationary equilibria We now turn to allocations induced by equilibria that satisfy the reconsideration-proofness refinement of Kocherlakota (1996). As a precursor, we introduce the notion of a *payoff stationary* credible equilibrium (PSCE):

Definition 11. (σ, λ) is a PSCE if it is a credible equilibrium and if there exists some constant W such that after all resource-feasible histories S^{t-1}, the continuation payoff to the planner equals W, i.e. for all t, $S^{t-1} \in \mathcal{S}^{t-1}(\lambda)$,

$$W_t(\sigma, \lambda | S^{t-1}) = W.$$

In Proposition 2, we presented a version of the revelation principle for credible equilibria. This proposition asserted that any equilibrium payoff for the planner could be attained by a credible equilibrium with direct mechanisms and, provided the planner had not deviated in the current period, truthful revelation. These credible equilibria relied on strategies that reverted to autarky following a planner defection. PSCE do not generally incorporate such a reversion and, consequently, we need to modify our earlier argument to make it applicable to them. We do so in Proposition 6.

Proposition 6. *Let (σ, λ) be a PSCE with payoff W. Then there exists a PSCE, $(\hat{\sigma}, \hat{\lambda})$, also with payoff W such that* (1) *$\hat{\sigma}$ is direct after all histories S^t,* (2) *$\hat{\lambda}$ is truthful after all individual histories $(S^{t-1}, \hat{\sigma}_t(S^{t-1}), \gamma^t, m^{t-1}, \theta)$.*

Proof. See Appendix A. □

It follows that from a welfare point of view, there is, once again, no loss of generality in invoking Property 1 and limiting attention to PSCE with direct mechanisms and truthful revelation along their equilibrium path. Additionally, we invoke Property 2 and limit attention to PSCE that use a lottery in the initial period only. We now treat this lottery as being over utility promises to agents. To emphasize this interpretation, we replace γ with w in the notation. We further restrict attention to equilibria in which these initial promises are kept. Since we do not constrain the planner's lottery choice, this restriction is also without loss of generality.[19]

We will refer to allocations as *payoff stationary credible allocations* (PSCAs) if they are the outcomes of PSCE (with promise-keeping). The following lemma gives necessary and sufficient conditions for a PSCA.

[19] It is in the spirit of a normalization, since many combinations of lotteries and allocation functions induce the same conditional distributions over consumptions and utilities.

Lemma 2. *An allocation* (Ψ_1, α) *is a PSCA if and only if it satisfies resource-feasibility* (4)*, boundedness* (5)*, incentive-compatibility* (6)*, promise-keeping,*

$$\forall w \in D, \qquad w = U\big(\alpha(w)\big), \tag{25}$$

and there exists some $W \geqslant V^{aut}$ *such that for all* t*,* $W_t(\Psi_1, \alpha) = W$*.*

The key change here is the final condition that refines the sustainability constraint (7) by requiring *constant* planner payoffs in excess of autarky. That the conditions in Lemma 2 are necessary for a PSCA is obvious. Sufficiency is established by an argument in the proof of Proposition 6. This argument shows that allocations satisfying the conditions of Lemma 2 can be supported by a PSCE in which the planner uses a strategy that "restarts" the allocation following any defection and agents use message strategies that are truthful unless there has been a planner defection in the current period, in which case agents give the message that maximizes their current payoff.

Reconsideration-proofness In the context of single player decision problems with time inconsistent preferences, Kocherlakota (1996) defines the concept of reconsideration-proofness. He identifies an agent strategy as symmetric subgame perfect if it is subgame perfect and delivers the same payoff after all histories. He labels a strategy *reconsideration-proof* if it is the best symmetric subgame perfect strategy. The restriction to symmetric subgame perfect equilibria may be viewed as an internal consistency restriction in the spirit of Farrell and Maskin (1989). Any subgame perfect equilibrium implies a family of continuation equilibria and continuation equilibrium payoffs. If the player is free to reconsider the equilibrium after any history and choose one of these, she would choose a continuation equilibrium that delivers the highest payoff. To be robust to this sort of reconsideration, any equilibrium must give the same continuation payoff after all histories. The restriction to *best* symmetric subgame perfect equilibria can be viewed as an external consistency condition, in the sense that once restricted to symmetric equilibria, the player would choose the best. Kocherlakota gives examples that illustrate the reasonableness of this concept.

Kocherlakota argues that reconsideration-proofness is a useful refinement for macroeconomic policy games since there is a mapping between the credible equilibria of such games and the subgame perfect equilibria of games played by single players with time inconsistent preferences. Our credible equilibrium concept rules out joint planner-agent strategy revisions. If the planner can initiate such joint revisions, then the definition of equilibrium must be modified to ensure that she has no incentive to do so. The logic underpinning reconsideration-proofness in single player settings can then be applied. To ensure that the planner does not reconsider and organize a joint planner-agent revision, credible equilibria should be best payoff stationary.

The definition below formally re-expresses reconsideration-proofness in terms of our earlier notion of a PSCE.

Definition 12. A credible equilibrium (σ, λ) is *reconsideration-proof* if it is payoff stationary and attains the greatest planner payoff amongst PSCE. We define an allocation as *reconsideration-proof* if it is the outcome of a reconsideration-proof equilibrium.

Note that while the set of credible equilibrium payoffs may be large, the set of reconsideration-proof equilibrium payoffs is a singleton. Notice also that while a reconsideration-proof allocation must deliver the same payoff to the planner at each date, there is no presumption that such an allocation is stationary in any other sense.

5.1. Reconsideration-proof allocations and Rawlsian planners

We now link reconsideration-proofness to planner problems with commitment and high societal discount factors. Roughly speaking, we show that any reconsideration-proof allocation is optimal for a Rawlsian planner who can commit. Conversely, any payoff stationary solution to the Rawlsian planner's problem is reconsideration-proof. The link between reconsideration-proofness and the Rawlsian planner is via a stationary allocation problem analyzed by Phelan (2005). If a solution to this problem exists, it is both reconsideration-proof and optimal for the Rawlsian planner. Moreover, a simple linear programming problem gives a numerical approximation to this solution.

We describe three problems successively. The first is the problem of the Rawlsian planner who can commit:

Rawlsian Problem

$$W^{\text{rawls}} = \sup_{\{\Psi_1,\{\psi_t\}_{t=1}^{\infty}\}} \left[\liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \int_D \sum_{s=0}^{\infty} \beta^s \sum_{\Theta^{t+s}} \theta_{t+s}\psi_{t+s}\big(w,\theta^{t+s}\big)\pi^{t+s}\big(\theta^{t+s}\big)\Psi_1(\mathrm{d}w) \right] \tag{26}$$

subject to the resource (4), boundedness (5), incentive-compatibility (6) and promise-keeping (25) constraints.

Our second problem, is a related cost minimization. In this the planner minimizes a cost aggregate subject to various constraints including a lower bound on the payoff to a Rawlsian planner. Formally, the cost minimization problem at W is given by:

Cost Problem

$$J^{\text{dual}}(W) = \inf_{\{\Psi_1,\{\psi_t\}_{t=1}^{\infty}\}} \left[\lim\sup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \int_D \sum_{\Theta^t} C\big(\psi_t\big(w,\theta^t\big)\big)\pi^t\big(\theta^t\big)\Psi_1(\mathrm{d}w) \right] \tag{27}$$

subject to the boundedness (5), incentive-compatibility (6) and promise-keeping (25) constraints, and the aggregate utility constraint:

$$W \leqslant \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \int_D \sum_{s=0}^{\infty} \beta^s \sum_{\Theta^{t+s}} \theta_{t+s}\psi_{t+s}\big(w,\theta^{t+s}\big)\pi^{t+s}\big(\theta^{t+s}\big)\Psi_1(\mathrm{d}w). \tag{28}$$

Our third and final problem is the recursive stationary cost problem at W. In this the planner chooses an invariant measure over agent utility promises Ψ and a pair of measurable policy functions $\psi : D \times \Theta \to D$ and $w' : D \times \Theta \to D$. The first of the policy functions gives an agent's current utility as a function of its current promise and shock, the second gives an agent's continuation promise as a function of these variables. The policy functions are chosen to satisfy recursive versions of the boundedness, promise keeping and incentive compatibility constraints. The function w' and the probability distribution over shocks π imply a Markov process for utility promises. Ψ is required to be an invariant measure for this process. Finally, the mean utility implied by Ψ must exceed W. Formally, the stationary cost problem at W is given by:

Stationary Cost Problem

$$\inf_{\{\Psi,\psi,w'\}} \int_{D\times\Theta} C\big(\psi(w,\theta)\big)\Psi(\mathrm{d}w)\pi(\theta) \tag{29}$$

subject to the recursive promise-keeping condition:

$$\forall w \in D, \quad w = \sum_{\Theta} \big[(1-\beta)\theta\psi(w,\theta) + \beta w'(w,\theta)\big]\pi(\theta), \tag{30}$$

the recursive incentive-compatibility condition:

$$\begin{aligned} &\forall w \in D, k, j \in \{1,\dots,K\}, \\ &(1-\beta)\hat{\theta}_k\psi\big(w,\hat{\theta}_k\big) + \beta w'\big(w,\hat{\theta}_k\big) \geqslant (1-\beta)\hat{\theta}_k\psi\big(w,\hat{\theta}_j\big) + \beta w'\big(w,\hat{\theta}_j\big), \end{aligned} \tag{31}$$

the recursive boundedness condition:

$$\forall w \in D, \theta^\infty \in \Theta^\infty, \quad \lim_{t\to\infty} \beta^t w_t\big(w,\theta^t\big) = 0, \tag{32}$$

where $w_t(w,\theta^t) = w'(w_{t-1}(w,\theta^{t-1}),\theta_t)$, $t > 1$ and $w_1(w,\theta) = w'(w,\theta)$, the steady state condition:

$$\forall B \in \mathcal{B}(D), \quad \Psi(B) = \int\limits_{D\times\Theta} 1_{\{w'(w,\theta)\in B\}} \Psi(\mathrm{d}w)\pi(\theta), \tag{33}$$

and the recursive aggregate utility condition:

$$W \leqslant \int\limits_{D} w\Psi(\mathrm{d}w). \tag{34}$$

Any triple (Ψ, ψ, w') induces an allocation $(\Psi, \alpha(\psi, w'))$, where each component function of $\alpha(\psi, w')$ is given by $\psi_t(w,\theta^t) = \psi(w_{t-1}(w,\theta^{t-1}),\theta_t)$. It is routine to check that this allocation satisfies the (non-recursive) boundedness and incentive-compatibility constraints.

With these definitions in hand, we show if there exist solutions to the Rawlsian and the stationary cost problems, then any reconsideration-proof allocation $(\Psi_1^{\mathrm{rc}}, \alpha^{\mathrm{rc}})$ solves the Rawlsian problem. The key step in the proof is to show that the reconsideration-proof payoff W^{rc} is weakly greater than the Rawlsian payoff W^{rawls}. To do this, we use an argument from Phelan (2005) to construct a PSCA that attains the planner payoff W^{rawls}. Since reconsideration-proof allocations are optimal amongst PSCAs, we have $W^{\mathrm{rc}} \geqslant W^{\mathrm{rawls}}$. The reverse inequality is trivial. We are then able to use the equality $W^{\mathrm{rc}} = W^{\mathrm{rawls}}$ to deduce the desired result. The converse of this result is not true. Since the Rawlsian planner cares only about the limiting allocation, there may be solutions to her problem that are not payoff stationary. However, it is easy to show that any payoff stationary solution to her problem is reconsideration-proof.

Proposition 7. *Suppose that the Rawlsian problem and the stationary cost problem at W^{rawls} have solutions. Denote the latter solution by (Ψ, ψ, w'). Then,*

(1) *if $(\Psi_1^{\mathrm{rc}}, \alpha^{\mathrm{rc}})$ is a reconsideration-proof allocation, it solves the Rawlsian problem,*
(2) *if $(\Psi_1^{\mathrm{rawls}}, \alpha^{\mathrm{rawls}})$ is a payoff-stationary solution to the Rawlsian problem, it is reconsideration-proof,*
(3) *$(\Psi, \alpha(\psi, w'))$ is a reconsideration-proof allocation and a solution to the Rawlsian problem.*

Proof. Suppose that $(\Psi_1^{\mathrm{rc}}, \alpha^{\mathrm{rc}})$ is a reconsideration-proof allocation with payoff W^{rc}. By definition, $(\Psi_1^{\mathrm{rc}}, \alpha^{\mathrm{rc}})$ satisfies the resource (4), boundedness (5), incentive compatibility (6) and promise-keeping (25) constraints. Hence, it is feasible for the Rawlsian problem. Also, since

$(\Psi_1^{\mathrm{rc}}, \alpha^{\mathrm{rc}})$ is reconsideration-proof, and, hence payoff stationary, it delivers a payoff of W^{rc} to the Rawlsian planner. Thus, $W^{\mathrm{rawls}} \geqslant W^{\mathrm{rc}}$.

We now seek to show the reverse inequality. To do so, we use two arguments in Phelan (2005). by Lemma 1 of Phelan (2005), if (Ψ_1', α') solves the Rawlsian problem, then it also solves the cost problem at aggregate utility amount W^{rawls} and does so with cost objective R. Second, we can use the construction in Lemma 2, Phelan (2005), to obtain a stationary allocation from (Ψ_1', α') that (1) attains a per-period aggregate cost less than $R + \varepsilon$ for arbitrary $\varepsilon > 0$, (2) has invariant measure Ψ_1' and (3) delivers a payoff of at least W^{rawls} to the Rawlsian planner. It follows that if a solution (Ψ, ψ, w') exists to the stationary cost problem at W^{rawls}, then it has a cost of less than or equal to R. Hence, the allocation induced by this solution, $(\Psi, \alpha(\psi, w'))$, is resource-feasible. It is also bounded, incentive-compatible, keeps promises and attains a constant planner payoff of at least W^{rawls}. But, since reconsideration-proof allocations attain the highest payoff amongst PSCAs, $W^{\mathrm{rc}} \geqslant W^{\mathrm{rawls}}$. Combining this inequality with that obtained in the first paragraph of the proof, $W^{\mathrm{rc}} = W^{\mathrm{rawls}}$. Thus, $(\Psi_1^{\mathrm{rc}}, \alpha^{\mathrm{rc}})$ solves the Rawlsian problem.

For the second part of the proposition, simply note that any payoff stationary solution to the Rawlsian problem is a PSCA and that, as previously shown, any reconsideration-proof allocation is feasible for the Rawlsian problem. Thus, any payoff stationary solution to the Rawlsian problem is reconsideration-proof.

For the third part of the proposition, note that $(\Psi, \alpha(\psi, w'))$ is payoff stationary and, by the argument given in the first part of the proof, attains the planner payoff $W^{\mathrm{rc}} = W^{\mathrm{rawls}}$. Hence, it is reconsideration-proof. Additionally, since it satisfies the constraints of the Rawlsian planner's problem, it solves this problem as well. □

Phelan (2005) establishes that the solution to the stationary allocation problem involves an ergodic, non-degenerate invariant measure Ψ. Consequently, if a solution to this and the Rawlsian problem exist, then there is an ergodic, non-degenerate reconsideration-proof allocation. These properties imply that the reconsideration-proof allocation exhibits social mobility and an absence of immiseration.

6. Conclusion

In this paper, we provide micro-foundations for high societal discount factors in dynamic moral hazard economies. Most contributions to the dynamic optimal social insurance literature have assumed that the discount factors of the planner and the agents coincide and that the planner can commit. Several recent papers, appealing to normative considerations, have relaxed the first assumption, and have shown the sensitivity of optimal arrangements to it. In contrast, we advance no normative arguments in favor of high societal discounting. Rather, we relax the commitment assumption and derive such discounting as an equilibrium phenomenon in a credibility game.

A second contribution of this paper is its formalization of credible policy games played by a large planner or government and a population of privately informed agents. Many papers in the macroeconomics literature have examined the credibility problems implied by Ramsey models. In these a planner-government is constrained to use simple mechanisms that rely on the linear taxation of a representative agent's income and wealth. In contrast, there has been relatively little work, especially outside of the Ramsey setting, on credibility issues in heterogeneous agent economies[20] yet a message of this paper is that they can be crucially important in shaping the distribution of consumption and utility obtained by agents.

Acknowledgments

We thank the editor, Narayana Kocherlakota, for suggestions that significantly improved the exposition of the paper. We thank seminar participants at the 2004 SED Florence meetings, and at the Universities of Pittsburgh, Princeton and Wisconsin-Madison.

Appendix A. Proofs

Proof of Proposition 1. Suppose that σ^{aut} is an autarkic planner strategy and λ^{aut} an autarkic message strategy. We first verify that the pair is a credible equilibrium then that it is the worst. Under σ^{aut}, agents' current messages do not affect their future allocations. Recognizing this, agents will send messages that maximize their current consumption. Hence, λ^{aut} is optimal for each agent.

Clearly, σ^{aut} is resource feasible after all aggregate histories given λ^{aut}, i.e. $\sigma^{aut} \in \Sigma(\lambda^{aut})$. Moreover, given that agents play according to λ^{aut}, the planner is unable to use intertemporal incentives to induce any agent to reveal its current shock type and accept a low current utility. In particular, it cannot induce agents with low current taste shocks to do so. Since the planner cannot, directly or indirectly, condition current utility awards on an agent's shock, the best she can do is award all agents the same utility amount $u(R)$ after all histories. Other feasible planner strategies $\sigma \in \Sigma(\lambda^{aut})$ may condition utility awards on past histories of messages or lottery outcomes, but these are clearly (weakly) dominated by σ^{aut}. Hence, $(\sigma^{aut}, \lambda^{aut})$ is a credible equilibrium.

The planner can always select the strategy σ^{aut}. Independently of λ, it is feasible and delivers a payoff V^{aut}. Thus, the planner can always attain the payoff V^{aut} and this acts as a lower bound on its equilibrium payoffs. Since $(\sigma^{aut}, \lambda^{aut})$ attains this payoff, it is a worst credible equilibrium. □

Proof of Proposition 2. Given the credible equilibrium (σ, λ), we construct a strategy pair $(\hat{\sigma}, \hat{\lambda})$ satisfying the three properties in the proposition. We then verify that it is a credible equilibrium.

Let $\lambda^t(m^t|S^t, \gamma^t, \theta^t)$ denote the distribution over message histories at t implied by the message strategy λ conditional on the aggregate history S^t and the lottery and shock history (γ^t, θ^t). Let $S_t(\hat{\sigma}) = (\Theta, \rho_t(\hat{\sigma}), \varphi_t(\hat{\sigma}))$, where for each γ^t and θ^t, $\varphi_t(\hat{\sigma})(\gamma^t, \theta^t) = \sum_{m^t \in \mathcal{M}^t(\sigma)} \varphi_t(\sigma)(\gamma^t, m^t)\lambda^t(m^t|S^t(\sigma), \gamma^t, \theta^t)$ and for each γ^{t-1}, θ^{t-1} and Borel set B, $\rho_t(\hat{\sigma})(\gamma^{t-1}, \theta^{t-1})(B) = \sum_{m^t \in \mathcal{M}^t(\sigma)} \rho_t(\sigma)(\gamma^{t-1}, m^{t-1})(B)\lambda^{t-1}(m^{t-1}|S^{t-1}(\sigma), \gamma^{t-1}, \theta^{t-1})$.

Define for all $t > 1$, $S^{t-1}(\hat{\sigma}) = \{S_r(\hat{\sigma})\}_{r=1}^{t-1}$, $\hat{\sigma}_t(S^{t-1}(\hat{\sigma})) = S_t(\hat{\sigma})$ and $\hat{\sigma}_1 = S_1(\hat{\sigma})$. Set the message strategy after such histories according to $\hat{\lambda}_t(S^t(\hat{\sigma}), \gamma^t, \theta^t) = \theta_t$. For any aggregate history $\tilde{S}^t$ such that $\tilde{S}^t$ agrees with $S^t(\hat{\sigma})$ prior to date $r \in \{1, \ldots, t\}$, but differs from it at r, set $\hat{\sigma}_t(\tilde{S}^t)$ equal to a direct autarkic mechanism (i.e. one with an allocation function that takes the constant value $u(R)$, message space equal to the type space and arbitrary lottery function). Also, set $\hat{\lambda}_t(\tilde{S}^t, \cdot) = \lambda_t^{aut}(\tilde{S}^t, \cdot)$, where $\{\lambda_t^{aut}\}_{t=1}^{\infty}$ is an autarkic message strategy satisfying $\lambda_t^{aut}(\tilde{S}^t, \cdot, \theta) = \theta$ if $\tilde{S}^t = (\tilde{S}^{t-1}, \sigma_t(\tilde{S}^{t-1}))$. It is immediate that $\hat{\sigma}$ is resource-feasible given $\hat{\lambda}$

[20] Acemoğlu et al. (2005), Berliant and Ledyard (2005), Bisin and Rampini (2005) and Roberts (1984) are exceptions.

and that $(\hat{\sigma}, \hat{\lambda})$ induces the same planner payoffs along its outcome path as (σ, λ) (and the same allocation if λ is deterministic).

We now verify that $\hat{\lambda}$ is a best response to $\hat{\sigma}$. If the planner sequentially implements $S_t(\hat{\sigma})$, then $\hat{\lambda}$ calls for truth-telling. By construction, if the planner behaves in this way, each agent receives the same expected lifetime payoff from truth-telling as from using the message strategy λ in the original equilibrium. Moreover, any alternative message strategy that is non-truthful along the equilibrium path $\{S_t(\hat{\sigma})\}_{t=1}^{\infty}$ generates a payoff that the agent could have attained by deviating from λ in the original equilibrium. Since the agent chose not to deviate in this way, truth-telling must be optimal for the agent given $\{S_t(\hat{\sigma})\}_{t=1}^{\infty}$. After any history $S^t \neq S^t(\hat{\sigma})$, $\hat{\sigma}$ and $\hat{\lambda}$ are autarkic. By Proposition 1, an autarkic message strategy is a best response to future autarkic play by the planner.

Next, we verify that $\hat{\sigma}$ is a best response to $\hat{\lambda}$. Suppose the planner adheres to $\hat{\sigma}$ up until period t, and then defects to an alternative sequence of mechanisms $\{S'_{t+r}\}_{r=0}^{\infty}$ that are resource-feasible given $\hat{\lambda}$. Since $\hat{\lambda}$ prescribes an autarkic message strategy following a planner defection, by the argument in Proposition 1, the planner's continuation payoff from t is less than or equal to V^{aut}. But the planner's continuation payoff if it does not defect, $W_t(\hat{\sigma}, \hat{\lambda} | S^{t-1}(\hat{\sigma}))$, equals the continuation payoff $W_t(\sigma, \lambda | S^{t-1}(\sigma))$ from the original equilibrium. Since the continuation of a credible equilibrium is a credible equilibrium and since V^{aut} is the worst credible equilibrium payoff, it follows that $W_t(\hat{\sigma}, \hat{\lambda} | S^{t-1}(\hat{\sigma})) \geqslant V^{aut}$. Thus, the planner has no incentive to defect. After a defection, $\hat{\sigma}$ and $\hat{\lambda}$ are autarkic. By Proposition 1, autarkic play by the planner is a best response to an autarkic message strategy. It follows that $(\hat{\sigma}, \hat{\lambda})$ is a credible equilibrium as desired. □

Proof of Proposition 3. Suppose that (σ, λ) is a credible equilibrium. Let $(\mathcal{M}_t(\sigma), \rho_t(\sigma), \varphi_t(\sigma))$ denote the sequence of mechanisms along the outcome path of σ. By Property 1, each $\mathcal{M}_t(\sigma) = \Theta$. By Property 2, for each t, there is a ψ_t such that for all $\gamma^t = (\gamma_1, \gamma^{t-1})$ and θ^t, $\varphi_t(\sigma)(\gamma_1, \gamma^{t-1}, \theta^t) = \psi_t(\gamma_1, \theta^t)$. Set $\Psi_1 = \rho_1(\sigma)$, then $(\Psi_1, \alpha) = (\Psi_1, \{\psi_t\}_{t=1}^{\infty})$ is the allocation induced by (σ, λ). From the definition of a credible equilibrium, (Ψ_1, α) clearly satisfies resource-feasibility (4) and boundedness (5). Since, we restrict attention to credible equilibria that are truthful along their equilibrium path, when confronted with the sequence of mechanisms $\{(\Theta, \rho_t(\sigma), \varphi_t(\sigma))\}_{t=1}^{\infty}$ it is optimal for the agent to truthfully report her type. In particular, at all dates and after all message histories, she is better off being truthful, than lying once and being truthful thereafter. Hence, (Ψ_1, α) satisfies (6). Finally, since (σ, λ) is credible, the planner is better off adhering to the strategy σ (and implementing the sequence of mechanisms $\{(\Theta, \rho_t(\sigma), \varphi_t(\sigma))\}_{t=1}^{\infty}$), than deviating to an autarkic strategy and receiving the payoff V^{aut}. This verifies (7).

Suppose $(\Psi_1, \alpha) = (\Psi_1, \{\psi_t\}_{t=1}^{\infty})$ is an allocation satisfying the conditions in the proposition. For each $t, \gamma_1, \gamma^{t-1}, \theta^t$, let $\varphi_t(\gamma_1, \gamma^{t-1}, \theta^t) = \psi_t(\gamma_1, \theta^t)$. Let $\{\rho_t\}$ be set so that $\rho_1 = \Psi_1$ and each subsequent ρ_t is arbitrary. Define the planner strategy σ as follows: $\sigma_t(S^{t-1}) = (\Theta, \rho_t, \varphi_t)$ for all $S^{t-1} = \{\Theta, \rho_s, \varphi_s\}_{s=1}^{t-1}$ and $\sigma_t(S^{t-1}) = (\Theta, \rho_t, \varphi_t^{aut})$, where φ_t^{aut} maps all past individual histories to $u(R)$, otherwise. Set $\lambda_t(S^t, \gamma^t, \theta^t) = \theta_t$ if $S^t = \{\Theta, \rho_s, \varphi_s\}_{s=1}^{t}$ and $\lambda_t(S^t, \gamma^t, m^{t-1}, \theta) = \lambda_t^{aut}(S^t, \gamma^t, m^{t-1}, \theta)$ otherwise, where $\{\lambda_t^{aut}\}_{t=1}^{\infty}$ is an autarkic message strategy. It is routine to check that this is a credible equilibrium under the conditions in the proposition. □

Proof of Proposition 6. Given the PSCE (σ, λ), we construct a new strategy pair $(\hat{\sigma}, \hat{\lambda})$ satisfying the properties in the proposition. We then verify that it is a PSCE.

For $t \geqslant 1$, let $S_t(\hat{\sigma})$ equal the triple $(\rho_t(\hat{\sigma}), \Theta, \varphi_t(\hat{\sigma}))$, where, for each $(\gamma^{t-1}, \theta^{t-1})$ and Borel set $B \subseteq \mathbb{R}$, $\rho_t(\hat{\sigma})(\gamma^{t-1}, \theta^{t-1})(B) = \sum_{\mathcal{M}^t(\sigma)} \rho_t(\sigma)(\gamma^{t-1}, m^{t-1})(B)\lambda^{t-1}(S^t(\sigma)) \times (m^{t-1}|\gamma^{t-1}, \theta^{t-1})$ and for each γ^t and θ^t, $\varphi_t(\hat{\sigma})(\gamma^t, \theta^t) = \sum_{\mathcal{M}^t(\sigma)} \psi_t(\sigma)(\gamma^t, m^t)\lambda_t(S^t(\sigma)) \times (m^t|\gamma^t, \theta^t)$. To define $(\hat{\sigma}, \hat{\lambda})$, we proceed period by period. Suppose that we have defined $\hat{\sigma}$ and $\hat{\lambda}$ up until period $t-1$. For each past history S^{t-1}, let $r \in \{0, \ldots, t-1\}$ denote the date of the planner's last defection from $\hat{\sigma}$. If the planner has never defected from $\hat{\sigma}$ before we set $r = 0$. Let $\hat{\sigma}_t(S^{t-1}) = (\rho'_t, \Theta, \varphi'_t)$, where for all $\gamma^t = (\gamma^r, \gamma^{t-r})$, $m^t = (m^r, \theta^{t-r})$, $\varphi'_t(\gamma^t, m^t) = \varphi_{t-r}(\hat{\sigma})(\gamma^{t-r}, \theta^{t-r})$ and for all $\gamma^{t-1} = (\gamma^r, \gamma^{t-r-1})$, $m^{t-1} = (m^r, \theta^{t-r-1})$, $\rho'_t(\gamma^{t-1}, m^{t-1}) = \rho_{t-r}(\hat{\sigma})(\gamma^{t-r-1}, \theta^{t-r-1})$. Similarly, for each past history S^t, let $r \in \{0, \ldots, t\}$ denote the date of the planner's last defection from $\hat{\sigma}$ and if the planner has never defected from $\hat{\sigma}$ before, set $r = 0$. If $r < t$, set for all $(\gamma^t, m^{t-1}) \in \mathbb{R}^t \times \mathcal{M}^{t-1}$, $\hat{\lambda}_t(S^t, \gamma^t, m^{t-1}, \theta) = \theta$. If $r = t$, then set $\hat{\lambda}_t = \lambda_t^{aut}$, where $\{\lambda_t^{aut}\}_{t=1}^{\infty}$ is an autarkic message strategy. In words, the planner's strategy prescribes a sequence of direct mechanisms $\{S_t(\hat{\sigma})\}_{t=1}^{\infty}$. The agent's message strategy is truthful, provided the planner adheres to her strategy. If the planner defects in some period t, then agents adopt their autarkic strategy for that period, i.e. they give the message that maximizes their current payoff. After a defection, the planner's strategy calls for the planner to begin playing the sequence $\{S_t(\hat{\sigma})\}_{t=1}^{\infty}$ from the beginning. If the planner reverts to her strategy, the agents revert to truth-telling. It is immediate that this strategy gives the payoff W after all histories given that the original PSCE did. Moreover, it satisfies conditions 1 and 2 in the proposition.

We now check that $(\hat{\sigma}, \hat{\lambda})$ is a credible equilibrium. Suppose the planner sequentially implements $\{S_t(\hat{\sigma})\}_{t=1}^{\infty}$. Then, by construction, the payoff obtained by an agent from truth-telling equals the agent's expected payoff in the original equilibrium (σ, λ). Moreover, any alternative message strategy that is non-truthful along the equilibrium path $\{S_t(\hat{\sigma})\}_{t=1}^{\infty}$ generates a payoff that the agent could have attained by deviating from λ in the original equilibrium. Since the agent chose not to deviate in this way, truth-telling must be optimal for the agent given $\{S_t(\hat{\sigma})\}_{t=1}^{\infty}$. By a similar argument, the agent looses nothing from truth-telling after a reversion by the planner to its strategy following a prior defection. If the planner does defect in some period t, the strategy $\hat{\sigma}$ calls for the planner to ignore past messages of the agents in period $t+1$. Hence, it is optimal for an agent to choose the message function λ_t^{aut} at t.

In each period t that the planner defects, the agents play λ_t^{aut}. The planner is then unable to deliver different current utility amounts to agents contingent on their current messages. Thus, if the planner defects in t, it is optimal for her to award each agent $u(R)$. If the planner defects for k periods beginning in period t, she can do no better than obtain the payoff: $(1-\beta^k)E[\theta]u(R) + \beta^{k+1}W \leqslant W$. Thus, the planner has no incentive to defect. □

References

Acemoğlu, D., Golosov, M., Tsyvinski, A., 2005. Markets versus governments: Political economy of mechanisms. Mimeo. MIT.

Atkeson, A., Lucas, R., 1992. On efficient distribution with private information. Review of Economic Studies 59, 427–453.

Berliant, M., Ledyard, J., 2005. Optimal dynamic non-linear income taxes with no commitment. Mimeo. University of Washington.

Bisin, A., Rampini, A., 2005. Markets as beneficial constraints on the government. Journal of Public Economics, in press.

Chari, V.V., Kehoe, P., 1990. Sustainable plans. Journal of Political Economy 98, 617–636.

Chari, V.V., Kehoe, P., 1993. Sustainable plans and mutual default. Review of Economic Studies 60, 175–195.

Farhi, E., Werning, I., 2005. Inequality, social discounting and progressive estate taxation. Working paper. MIT.

Farrell, J., Maskin, E., 1989. Renegotiation in repeated games. Games and Economic Behavior 1, 327–360.

Judd, K., 1985. The law of large numbers with a continuum of IID random variables. Journal of Economic Theory 35, 19–25.

Kocherlakota, N., 1996. Reconsideration-proofness: A refinement for infinite horizon time inconsistency. Games and Economic Behavior 15, 33–54.

Luenberger, D., 1969. Optimization by Vector Space Methods. Wiley, New York.

Phelan, C., 1998. On the long run implications of repeated moral hazard. Journal of Economic Theory 79, 174–191.

Phelan, C., 2005. Opportunity and social mobility. Review of Economic Studies, in press.

Roberts, K., 1984. On the theoretical limits to redistribution. Review of Economic Studies 51, 177–195.

Rudin, W., 1976. Principles of Mathematical Analysis, third ed. McGraw-Hill.

Sleet, C., Yeltekin, Ş., 2005a. Social credibility, social patience and long run inequality. Mimeo. Carnegie Mellon University.

Sleet, C., Yeltekin, Ş., 2005b. Notes on credible social insurance. Mimeo. Carnegie Mellon University.