Does it Take a Tyrant to Implement a Good Reform?*

Anna Rubinchik-Pessach, Ruqu Wang†

February 16, 2005

Abstract

In our model a reform is a switch from one norm of behavior (equilibrium) to another and agents have to endure private costs of transition in case of a reform. A (local) authority, which coordinates the transition, can enforce transfers across the agents and is capable of imposing punishments upon them. A transfer/tax is limited, however, by an agent's equilibrium payoff, and a punishment can not exceed an upper bound monitored by a "third party" (international community). Implementing a good (Pareto improving) reform can be hindered by asymmetric information about the costs of transition, which are privately known to the agents and can not be observed by the authority. In this case even a benevolent authority may need to credibly threaten agents with a punishment to induce both the desired behavior and the truthtelling about the costs, as otherwise some good reforms will not be implementable, even with Bayesian mechanisms. Allowing for harsher punishments in that framework reduces to 'softening' the individual rationality constraint, thus widening the range of implementable reforms. The flip-side of increasing the admissible punishment is making 'bad' reforms feasible. With the international community setting a uniform standard of (negative) human rights (or maximal level of punishment) across countries, some will be unable to implement good reforms, while others will be prone to undesirable transitions. We, thus, formulate a trade-off between a successful implementation of good reforms from the utilitarian perspective and well-being of selected individuals in the society.

*We are grateful to Eckhard Janeba for extensive and insightful discussions. We also wish to thank Murat Iyigun, Sergio O. Parreiras as well as the participants of 'Political Economy Group' at the University of Colorado at Boulder and those at the Second Game Theory World Congress 2004. Wang acknowledges financial support from the Social Sciences and Humanity Research Council of Canada. This paper has been previously circulated under the title "Implementability of Reforms and Human Rights."

†Anna Rubinchik-Pessach, Anna.Rubinchik@colorado.edu, University of Colorado at Boulder, Department of Economics, UCB 256, Boulder, CO, 80309; Ruqu Wang, Ruqu.Wang@colorado.edu, University of Colorado at Boulder and Queen's University.

Key words: reform, mechanism design, incentive compatibility

JEL Classification numbers: D78, E61.

1 Motivation

Why are some potentially good reforms never implemented? What can explain fruitfulness of recent economic reforms in China, overwhelming success of the newly industrialized countries in the 1970's-1980's along with a turbulent and murky path followed by India and Russia in the past decade, for example?[1] Our goal is not to provide an ultimate answer, but, rather, to illuminate some possible connections between political and economic changes. We envision the role of a reformer – be it a single "dictator" or a democratic government – as a one-time intervention, with the sole purpose of changing the "norm of behavior" in a country.[2] For example, a norm could describe production-consumption choices under a given market structure, degree of openness to the international trade and a monetary regime. Even if two different norms can be ranked Pareto, a (decentralized) switch to a dominant one might not occur due to the reluctance of some individuals to cover the transition costs, which range from the effort of re-structuring one's investment portfolio to the 'loss of cultural identity' (as in Janeba (2003)). There is no doubt that a tyrant with an unlimited power can implement any change, however whimsical. We aim at describing the least severe threat necessary to "convince" all the individuals to change their actions in accordance with the new norm. This threat has to be credible, in other words, the potential punishment can not constitute a violation of the laws protecting human dignity. This means that in the presence of some exogenous constraint on punishments certain reforms will be impossible to implement. Existence of such constraints can be justified, in particular, by the concern of the international community that in the absence of human rights monitoring a "malevolent" authority will get a lee-way to implement undesirable reforms. This same constraint, however, might prevent some good reforms from going through.

In his overview of economic reforms (1960-1980) around the world, Rodrik (1996) finds it puzzling that often times sound economic reforms are not popular, moreover,

[1] For a general overview see Rodrik (1996), Velasco and Tommasi (1996) among others; Russian economic reforms are discussed in more detail in Lang and Weber (2000).

[2] In the spirit of Binmore (1998), we view a social norm to be self-sustainable in a sense that, once in place, it prescribes each individual to act in his own best interest.

he mentions that "*...the implementation of good economic policy is often viewed as requiring 'strong' and autonomous' (not to say authoritarian) leadership.*"[3] Similar observations are offered by Harberger (1993) in his overview of the Latin American reforms. He stresses personal charisma and outstanding leadership of "key group of individuals," often times acting – as in case of Roberto Campos, who is now given credit for the 'Brazilian Miracle,' – "*...in spite of adverse circumstances and at high personal cost.*" Not surprisingly, much of the recent literature have been devoted to the political determinants of reforms, thus, shifting traditional focus from normative suggestions to understanding the barriers to implementing the prescriptions. Fernandez and Rodrik (1991) show that individual specific uncertainty, i.e., inability of an individual to assess with certainty whether she will be a winner or a loser, can hinder good reforms, if no transfers are allowed. Jain and Mukand (2003) allow for the transfers that are sustainable in a citizen-candidate political equilibrium. It is the uncertainty with respect to the identity of the politician in power, and thus, to whether the transfers will eventually be realized, that drives the resistance to a reform. While these models can provide an introspection as to why actual polities that rely heavily on majority voting are unable to deliver the desired reform, we focus on feasibility of a transition under *any* social choice procedure that respects the laws protecting individual well-being. When the costs of transition are privately known to the individuals, some good reforms become infeasible for reasons that are similar to those that give rise to inefficiency in Myerson and Satterthwaite (1983), Mailath and Postlewaite (1990) and Rob (1989). Thus, it is the asymmetric information that drives the result rather than the individual-specific uncertainty stressed in the previous work on the subject. Our analysis relies on the mechanism design literature that we cite below as the model is developed.

Another related strand of literature is devoted to potential failures to coordinate actions by a large group of individuals. Morris and Shin (1998) develop a model of speculative currency attacks, in which the value of investor's holding crucially depends on the actions of the other investors. They all get an imperfect signal from the government about "fundamentals," indicating the desire of the government to support local currency. The noise in the signal destroys common knowledge that the currency is in the stable region, even when it, indeed, is. This creates a chance that the investors abandon local currency even, if the fundamentals are good, i.e., it creates a possibility of the switch to a Pareto dominated equilibrium. This can be remedied by a costly action of the "policy maker," as shown in Angeletos, Hellwig, and Pavan

[3] p.10.

(2003). Inability of individuals to synchronize their actions can also lead to the failure of a (de-centralized) switch to the efficient equilibrium, as in Morris (1995). While strategic manipulation of individual beliefs can be interesting to explore, we leave it for future investigation, resorting, instead, to a common knowledge environment. This choice is dictated, in part, by our desire to relieve the pressure on necessary punishments by adhering to the least demanding solution concept (Bayesian Nash), that allows for efficient outcomes to be implemented in some environments.

The rest of the paper is organized as follows.

After setting up the model in section 2, we proceed with the full information model, in which individual costs of transition are known to the reformer (local authority). In section 3 we show that the authority does not need to use punishments to implement good (Pareto improving) reforms, moreover a 'malevolent' authority may be incapable of forcing undesirable reforms (i.e., a switch to a Pareto dominated equilibrium) without resorting to a punishment. Under asymmetric information, introduced in section 4, the authority may need to credibly threat individuals with punishments. The punishment might be higher for more divided countries and in case of bad reforms, as illustrated in section 4.2 for the discrete distribution case. We generalize the main results for the case of continuous costs distribution thereafter. Extensions and conclusions follow. The proofs are in the appendix.

2 The Setup

A country consists of N individuals (agents). Their everyday interactions are reduced to a simultaneous move coordination game G with two actions $\{A, B\}$ and real-valued payoffs which depend on i's choice of action and are symmetric with respect to the actions of others,

$$u_i(s) = u(s_i, s_{-i}), \tag{1}$$

where s_{-i} is the action profile chosen by all the players but i, $s \in \{A, B\}^N$ is the strategy profile. If one thinks about a strategy as representing a sequence of actions over time, the payoff can be viewed then as a (discounted) sum of future payoffs that individual i receives, if he follows the chosen strategy, s_i, for example, driving on the right side of the road, or accepting local currency. We would like to introduce the simplest possible framework and analyze reforms as coordinated switches between the two pure strategy – $s^A = (A, A, .., A)$, $s^B = (B, B, .., B)$ – Nash equilibria.[4] These equilibria are 'Pareto' ranked as follows:

$$u\left(s^A\right) = a > b = u\left(s^B\right) \geq 0, \tag{6}$$

and assume both dominate the mixed strategy payoff.

Definition 1 *A reform is a switch from one equilibrium (norm) to another.*

Agent i has a cost, $c_i \in [\underline{c}, \bar{c}] \subset \mathbf{R}_+$, associated with switching her action. In this model a switch from s^B to s^A is a Pareto improving (a good) reform, provided the average cost is below the gain, $a - b$. Otherwise a switch is a bad, or an undesirable one.

An authority, however, may have distinct interests from the rest of the society. It has the ability of coordinating a switch, or announcing the reform, besides, it has an access to two tools: (1) transfers to the agents, $(t_i)_i \in \mathbf{R}^N$; (2) punishments, $(m_i)_i \in \mathbf{R}_+^N$. There are no outside sources of financing the reform so that

$$\Sigma_i t_i \leq 0 \tag{BB}$$

Both the transfers and the punishment, we assume, are anonymous, they can only depend on the observed actions and on the cost of transition (if observed).

[4]In addition, there could be "knife edge" assymetric equilibria of the form: proportion p (pN is an integer) of the agents are choosing A and the rest are choosing B :

$$\begin{aligned} u\left(\underset{pN}{A, .., A}, \underset{(1-p)N}{B, ..B}\right) &= w > u\left(B, \underset{pN-1}{A, .., A}, \underset{(1-p)N}{B, ..B}\right); && (2) \\ w &> u\left(\underset{pN+1}{A, .., A}, \underset{(1-p)N-1}{B, ..B}\right). && (3) \end{aligned}$$

In addition (could be omitted) for any $q \neq p$, such that qN is an integer, either

$$\begin{aligned} u\left(\underset{qN}{A, .., A}, \underset{(1-q)N}{B, ..B}\right) &< u\left(B, \underset{qN-1}{A, .., A}, \underset{(1-q)N}{B, ..B}\right) \text{ or} && (4) \\ u\left(\underset{qN}{A, .., A}, \underset{(1-q)N}{B, ..B}\right) &< u\left(\underset{qN+1}{A, .., A}, \underset{(1-q)N-1}{B, ..B}\right). && (5) \end{aligned}$$

There are assumed away for simplicity.

More precisely, the transfers and the punishment vary only with the action, s_i^1, taken by individual i, actions taken by the rest of the players, s_{-i}^1 (after the reformers announcement has been made) and the cost of transition, c_i

$$\begin{aligned} t_i &= t(s_i^1, s_{-i}^1, c_i, c_{-i}, I(c)); & (7) \\ m_i &= m(s_i^1, s_{-i}^1, c_i, c_{-i}, I(c)). & (8) \end{aligned}$$

In particular, costs of transition might influence the decision with respect to the reform, indicated by $I(c)$, which is unity in case the reform is announced and zero otherwise, $c \in [\underline{c}, \bar{c}]^N$.

While the authority announces its recommendation "switch to strategy A" or "continue with B," it also has to make sure that the agents are sure to follow. This implies that the prescribed action s_i^* should satisfy[5]

$$s^* \in \arg \max_{s_i^1 \in \{A,B\}} u\left(s_i^1, s_{-i}^1\right) - c_i \iota\left(s_i^0, s_i^1\right) + t_i - m_i, \forall i \tag{IC}$$

over the available (new) actions $s_i^1 \in \{A, B\}$, with $\iota(s_i^0, s_i^1)$ is the switching index, it is unity, if i switched the action, so that pre- and post- reform actions are different, $s_i^0 \neq s_i^1$; and zero otherwise.

There is no doubt that with the threat of a punishment harsh enough, any request of the authority will be "convincing enough," in other words, if the punishment (m_i) for disobeying the prescription is sufficiently large, any prescription will be followed. One of our goals is to understand just how much punishment is needed to motivate the agents to follow the suggestions of the authority.

Another way of looking at it is to assume that during the transition "human rights" constraints should abided, as those are strictly enforced by an "international community,"[6]

$$m_i \leq \bar{m}, \tag{IRH}$$

where $\bar{m} \in \mathbf{R}_+$ denotes the upper bound on credible punishment. Thus, we will be seeking to define the smallest such bound $\bar{m}$ that will allow for good reforms. This

[5]Here and in what follows we restrict attention to "weak implementation," our objective being to formulate the smallest necessary punishment, in particular, to determine whether any punishment is needed at all. Requiring "full implementation," for example, might require more pressure on the punishment stemming from a more demanding solution concept, although the latter exercise can be interesting to perform on positive grounds.

[6]See the related discussion in the introduction.

could be of interest to a benevolent international community, viewed as a "meta-mechanism designer" whose objective is to prevent bad reforms and not to inhibit good reforms with limited tools, those being just the bound on punishments, $\bar{m}$. Indeed, it might be impossible for an outsider to judge whether the "reformer" is benevolent or not and to dictate precisely how to use the transfers and whether to undertake the reform, i.e., intervening in the internal affairs of a country.

Clearly, if there are no additional constraints, and if taxes (transfers) can be expropriated by the reformer or simply burnt, the (IRH) constraint is irrelevant (not binding). However, "financial" punishments are not unlimited. We, therefore, impose another, "positive," assumption – the individual resource constraint – imposing a lower bound on the amount of transfers that can be collected.

$$t_i \geq -u\left(\tilde{s}, s_{-i}^1\right), \tag{RC}$$

It amounts to saying that no more than an individual's "income" can be extracted from each.

To sum up, a reformer formulates a mechanism, (I, t, m), where I indicates whether the reform is declared based on the profile of individual costs,

$$I : [\underline{c}, \bar{c}]^N \rightarrow \{0, 1\}, \tag{9}$$

and t is the transfer profile, and m is a punishment profile, both based on the costs reports, the decision with respect to the reform and the actions takes by the individuals:

$$\begin{aligned} t &= (t_i)_{i=1}^N; && (10) \\ m &= (m_i)_{i=1}^N, && (11) \end{aligned}$$

where t_i and m_i are functions defined in $(7, 8)$.

3 The Benchmark

We will start with the full information case, in which the authority observes individual costs of transition.

We will assume that a benevolent reformer calls for the reform only if the sum of the individual costs of transition is below the total surplus from the switch, or, if

$\mu \leq a - b$, with average cost of transition being μ :

$$I_1(c) = \begin{cases} 1, & \text{if } \mu \leq a - b \\ 0, & \text{otherwise} \end{cases} . \tag{12}$$

A malevolent reformer wants the reform no matter what the costs are, $I_2(c) = 1$.

Definition 2 *A rule I is implementable (in Nash strategies) with allowable punishment $\bar{m}$, if there exist a transfer profile t and a punishment profile m, satisfying BB, IC, IRH, RC.*

In this case an appropriate choice of the level of punishment can allow for a good reform to be implemented and, sometimes avoid bad reforms.

Proposition 3 *The objective of a benevolent planner, I_1, is implementable with allowable punishment of at least* $\max\{0, \bar{m}_1\}$, $\bar{m}_1 \equiv -a + \mu$. *The same punishment is sufficient to implement the reform always, rule I_2.*

A way to implement the reform is to redistribute the surplus so that every individual will pay the average cost of transition. In case the reform is good, $\mu \leq a - b$, there should be enough surplus to implement the reform with no punishment, $\bar{m}_1 \leq 0$. Also, if $\mu > a$, no punishment (imposing $\bar{m} = 0$) will prevent undesirable reforms.

The result hinges on several key assumptions: (1) full observability (and verifiability) of individual costs, (2) no limits on transfers, and (3) common knowledge of the timing and the decision of the reformer. If they are satisfied, there should be no reason for a country not to implement a good reform, moreover, there should be no reason to expect a tyrant, threatening agents with punishments to head the desirable transition. Interestingly, the mere presence of such threats signals that reform might be too costly and, thus, not worthwhile. As we will show in what follows this conclusion might be erroneous if the first assumption is relaxed, so that the costs are privately observable and are impossible to verify.[7] We, thus, retain the ability of a reformer to transfer utility and show that she might need to resort to a punishment, even if she has benevolent intentions.

[7] Implications of assumptions (2) and (3) have been studied in the literature, see introduction for a brief overview.

4 Asymmetric Information with respect to the Costs of Transition

Now assume costs of transition are privately known to the citizens. They share a common belief that the costs are drawn independently from distribution $F : [\underline{c}, \bar{c}] \rightarrow [0, 1]$, $\underline{c} \geq 0$. The reformer announces a mechanism, (I, t, m), where I indicates whether the reform is declared based on the *reports* of individual costs,

$$I : [\underline{c}, \bar{c}]^N \rightarrow \{0, 1\}, \tag{13}$$

and both the punishment and the tax depend on the announced valuations,

$$\begin{aligned} t_i &= t(s_i^1, s_{-i}^1, \theta_i, \theta_{-i}, I(\theta)); \qquad (14) \\ m_i &= m(s_i^1, s_{-i}^1, \theta_i, \theta_{-i}, I(\theta)). \qquad (15) \end{aligned}$$

Agents privately observe the realization of the costs of transition, and report them simultaneously to the reformer. Based on the reports, the reformer might either call for a reform or not. Endowed with the common knowledge of the reformer's decision, the agents choose one of two actions $\{A, B\}$. They get transfers and are subject to punishment according to the mechanism thereafter.

A rule is implementable if every agent is choosing his best response given his cost and his beliefs about the costs of the others, the costs are truly revealed and everybody chose the action as instructed by the authority, i.e., according to $I(\theta)$. Note that due to linearity of the agents' preferences, we can separate the incentive to choose the requested action (at the last stage) from the incentive to report the costs truthfully, therefore we can apply the revelation principle to the latter. Therefore, after the announcements are made, and thus, the true costs revealed, the reformer will have to make sure that the agents are motivated to act as she requires, i.e., the prescribed action s_i^* should maximize

$$u\left(s_i^1, s_{-i}^1\right) - c_i \iota\left(s_i^0, s_i^1\right) + t_i - m_i, \forall i$$

as under full observability. Given true revelation of costs, this condition is equivalent to (IC). In addition, the agents should be motivated to tell the truth, so that Bayesian Incentive compatibility is satisfied,

$$\begin{aligned} c_i &\in \arg\max_{\theta_i} E_{c_{-i}} U\left(I\left(\theta_i, c_{-i}\right)\right) - c_i I\left(\theta_i, c_{-i}\right) + t_i - m_i; \qquad (16) \\ U\left(I\left(\theta\right)\right) &\equiv aI\left(\theta\right) + b\left(1 - I\left(\theta\right)\right). \qquad (17) \end{aligned}$$

Here θ_i is citizen $i's$ announcement about his cost of transition and $E_{c_{-i}}[\cdot]$ denotes expectation formed by citizen i over the costs of his fellow citizens conditional on his cost being c_i.

Definition 4 *Rule I is implementable in Bayesian strategies with allowable punishment $\bar{m}$, if there exist transfers t and punishments m, satisfying BB, IC, 16, IRH, RC.*

The following lemma helps to simplify the analysis by reducing the number of constraints.

Lemma 5 *The constraints IC, IRH, RC imply*

$$a - c_i + t\left(s^A, \theta, 1\right) \geq -\bar{m} \text{ for all } i, \tag{18}$$

in case the reform is announced and

$$b + t\left(s^B, \theta, 0\right) \geq 0 \text{ for all } i \tag{19}$$

otherwise.

In particular, the latter constraints imply,

$$E_{c_{-i}}\left[U\left(I\left(\theta_i, c_{-i}\right)\right) - c_i I\left(\theta_i, c_{-i}\right) + \tau\left(\theta_i, c_{-i}\right)\right] \geq -\bar{m} E_{c_{-i}}\, I\left(\theta_i, c_{-i}\right) \text{ for all } i, \tag{IIR}$$

where

$$\tau\left(\theta_i, c_{-i}\right) \equiv t\left(s^A, \theta_i, c_{-i}, 1\right) I\left(\theta_i, c_{-i}\right) + t\left(s^B, \theta_i, c_{-i}, 1\right)\left(1 - I\left(\theta_i, c_{-i}\right)\right). \tag{20}$$

Luckily, this is nothing but an interim individual rationality constraint from standard mechanism design literature, if $\bar{m} = 0$. Allowing for $\bar{m} > 0$, thus, "softens" this constraint, undeniably "helping" the reformer.

As we demonstrate below, the minimal punishment might be above zero even for implementing a benevolent rule I_1 and it crucially depends on the shape of distribution F. However, a malevolent ruler has to be the most tyrannical, as she needs to resort to a punishment above the one pertinent to a benevolent rule. First, we calculate the latter "upper" bound, thereby proceeding with the minimal threat to be granted to a benevolent reformer in order to be always successful.

Proposition 6 *The malevolent rule I_2 is implementable with punishment of at least* $\max\{0, \bar{m}_2\}$,

$$\bar{m}_2 = \bar{c} - a, \tag{21}$$

where $\bar{c}$ is the upper support of the cost distribution.

It is worth noting that $\bar{m}_2$ is not necessarily strictly positive, so that even in the asymmetric information case a malevolent ruler might not need to resort to strictly positive punishments. For example, if the improvement, $(a - b)$, is quite small relative to the costs, but the level of the new benefit a is sufficiently high, $\bar{c} < a$, no punishment will be necessary. A mechanism supporting such a reform is very simple. Impose no transfers if an agent complies with the request to switch his action. In case an agent obeys the authority, the new payoff is then $a - c_i \geq a - \bar{c}$; in case he pursues B, all his income is transferred away and, potentially, the harshest punishment is applied. But if $a - \bar{c} > 0$, there is no need to resort to punishment, because even in its absence, the non-compliance payoff is zero.

In most of what follows we focus on the complementary case, $\bar{c} > a$.

4.1 Implementing a Good Reform

Recall that our objective is to describe the smallest punishment consistent with implementing the "first best," i.e., rule I_1 with the smallest $\bar{m}$, defined in (IIR) and consistent with truthtelling at the same time. In order to determine this bound, we will first analyze the case of a discrete distribution sections and then proceed to the continuous case in Section (4.3).

4.1.1 The Two Types Case

Suppose that each agent's switching cost is either $\underline{c}$ (with probability ρ) or $\bar{c}$ (with the complimentary probability) and is distributed independently and identically, so that the the costs are driven from distribution D :

$$D(x) = \begin{cases} 0, \text{if } x < \underline{c} \\ \rho, \text{ if } \underline{c} \leq x < \bar{c} \\ 1, \text{otherwise} \end{cases} . \tag{22}$$

If $a - \underline{c} \leq b$, then switching from s^B to s^A is never beneficial. If $a - \bar{c} \geq b$, then switching from s^B to s^A is always beneficial. Each of these two cases is straightforward.

We will focus on the most interesting case where $a - \underline{c} > b$ and $a - \bar{c} < b$. Then there exits an integer $0 < n^* < N$, such that the switch from s^B to s^A is beneficial if and only if the number of agents having $\underline{c}$ is at least n^*. In other words,

$$(n^* - 1)\underline{c} + (N - n^* + 1)\bar{c} > N(a - b) \geq n^*\underline{c} + (N - n^*)\bar{c}.$$

In most of the analysis, we ignore the integer problem for n^* and set

$$N(a - b) = n^*\underline{c} + (N - n^*)\bar{c}. \tag{23}$$

Thus, n^* can be viewed as the smallest number of low cost individuals needed to make the regime switch welfare improving.

We can now calculate the lowest punishment necessary to implement the benevolent rule.

Proposition 7 *Assume the costs are distributed D independently, with $a - \underline{c} > b$ and $a - \bar{c} < b$. Then I_1 is implementable with with allowable punishment of at least* $\max\{0, \bar{m}_1\}$,

$$\bar{m}_1 = \frac{-a\Pr\{\mathcal{A}\} - b\Pr\{\mathcal{B}\} + \bar{c}\Pr\{n_L \geq n^*\} + \underline{c}\Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\}}{\Pr\{n_L \geq n^*\}}, \tag{24}$$

where $\Pr\{\mathcal{A}\}$ *is ex-ante probability that the reform is worthwhile,* $\Pr\{\mathcal{B}\}$ *is the complementary probability, n_L is the number of low cost agents excluding agent 1.*

Note that if $\bar{c}$ is high and a or b are sufficiently small, $\bar{m}_1$ is positive. This is because it is expensive to make the high cost agents to switch, and the tax that is available for transfers is not enough to cover the expense. In this case, some punishments have to be imposed to make the switch implementable, so that more "costly" reforms might require higher punishments. Conversely, higher gross payoffs from the transition, $a - b$, decrease the necessary threat. As will be shown for the continuous case, the bound $\bar{m}_1$ is closely related to the "informational rents" (that a low cost agent can extract), often blamed for the "inefficiency" in the standard public good provision problem,[8] i.e., the impossibility to implement (for some environments)

[8] See Ledyard and Palfrey (2003), Mailath and Postlewaite (1990), Myerson and Satterthwaite (1983).

the benevolent objective in the presence of the hard individual rationality constraint, which corresponds to setting $\bar{m}_1 = 0$ in this framework.

Interestingly, the boundaries on punishments can be ordered,[9]

$$\bar{m}_1 \leq \bar{m}_2. \tag{27}$$

If the allowable punishment is in the range $[\bar{m}_1, \bar{m}_2]$, the benevolent rule I_1 is implementable (for all realizations of individual costs), while the (constant) malevolent rule is sometimes not. If the "third party" (international community) is interested in the well-being of the citizens of the country, they should not set the allowable boundary above $\bar{m}_1$. However, this boundary hinges on the exact knowledge of the distribution F, from which the costs are driven. If a uniform bound is to be set for a group of countries with different underlying primitives (reduced to the cost generating distribution), this boundary might be set too high for some countries and too low for the others.

Moreover, as we show below, 'big' reforms and more dispersed distributions of costs (describing a "more divided country") require more severe punishment $\bar{m}_1$ in order to be implemented, thus, potentially calling for $\bar{m}_1 > 0$, thereby setting the "right" bound on the allowable punishment a non-trivial task.

4.2 Comparative Statics

One would think that a reform that requires harsher punishments (or a threat thereof) will be harder to implement. This assertion can be justified, if one envisions an international authority setting a uniform standard of permissible punishments (effective limit on m). In this case, too high a required bound for a given reform ($\bar{m}_1$) will make it unimplementable. Thus, we will say that a reform is "easier" to put through, if it calls for a smaller punishment. This is another reason to perform the comparative statics on the upper bound of credible punishments. These comparisons are reproduced for the continuous case in section (4.3).

[9]Indeed, it amount to showing that $a \Pr\{\mathcal{A}\} + b \Pr\{\mathcal{B}\} - \underline{c} \Pr\{c_1 = \underline{c}\} \Pr\{n_L = n^* - 1\} > a \Pr\{n_L \geq n^*\}$, which follows from the inequality $a > \underline{c}$ and

$$\begin{aligned} \Pr\{\mathcal{A}\} - \Pr\{n_L \geq n^*\} &= \Pr\{c_1 = \underline{c}\} \Pr\{n_L \geq n^* - 1\} - \Pr\{c_1 = \underline{c}\} \Pr\{n_L \geq n^*\} = \quad (25) \\ &= \Pr\{c_1 = \underline{c}\} \Pr\{n_L = n^* - 1\}. \quad (26) \end{aligned}$$

4.2.1 Reforms in Divided Countries are Harder to Implement

We can compare two 'countries' that differ by the shape of their costs distribution. One is more 'divided' than the other, if the possible realizations of costs are further apart. This, for example, corresponds to the variation in attitudes towards the reform: if some people favor the transition (view its costs as rather small), while others perceive it as undesirable, or very costly. The bigger is this gap – we show – the harder it is to implement the reform. It happens as higher difference in costs increases the "informational rents," which, in turn, call for a higher minimal punishment. As an illustration one could rely on economic success of (relatively) homogenous Far Eastern countries (Taiwan, Singapore) in the mid-1980's and challenges of economic reforms in the vastly diverse India.

As in the previous case, we want to keep the social decision with respect to reform, i.e., the smallest number of low cost announcements to execute the reform, n^*, constant. In order to do so, we can only consider cases in which low cost and high cost realizations are equally likely and the gain from reform is exactly between the costs, thus making the "majority rule" an optimal decision.

Lemma 8 *Assume the costs are distributed D independently, with $a - \underline{c} > b$ and $a - \bar{c} < b$. Assume, in addition, that $\rho = 1/2$ and $a - b = N\left(\frac{\bar{c}+\underline{c}}{2}\right)$. Then a mean preserving spread of the costs, i.e., if an individual cost either $\bar{c}+\delta$ or $\underline{c}-\delta$ with equal probabilities for any $\delta > 0$, leads to an increase in the required punishment, $\bar{m}_1$, to implement the corresponding benevolent rule.*

4.2.2 Smaller Reforms are Easier to Implement

In this section we show that smaller reforms are easier to implement as opposed to big leaps. Relatively successful reforms in China and a painful transition in Russia can be seen as an illustration of this relation.

To make such a comparison we have to introduce "intermediate steps," or to extend the initial coordination game to generate additional equilibria. Let the initial action set in game G now include action X, and we assume, that every agent choosing action X constitutes a new (pure strategy) Nash equilibrium, s^X, in that game with the corresponding payoff $x \in (b, a)$ to each. Therefore, switching from s^B to s^X captures a proportion of the benefit of the big switch (s^B to s^A). Let α denote this proportion. That is, $x - b = \alpha(a - b)$. We think about a transition to X as a "scaled down" reform, so it is natural to think that the costs for this transition are also proportionally smaller. Let $\underline{c}(x)$ and $\bar{c}(x)$ denote that switching cost respectively for a low cost agent and a high cost agent to s^X. Then

$$c_j(x) = \alpha c_j\,(a) = \alpha c_j,\ j \in \{L, H\}\,. \tag{28}$$

In this set up the smallest number of the low cost agents needed for a reform to be worthwhile stays constant from switch to switch. Indeed, let $n^*(x)$ denote the minimum number of low-cost agents required for the switch to s^X to be beneficial. Then

$$N(x-b) = n^*(x)\underline{c}(x) + (N - n^*(x))\bar{c}(x).$$

Since $x - b = \alpha(a-b)$, $\underline{c}(x) = \alpha\underline{c}$, and $\bar{c}(x) = \alpha\bar{c}$, we can conclude that $n^*(x) = n^*$.

Define the benevolent rule for small reforms, I_1^α, accordingly, with x replacing a and the new average cost being $\alpha\mu$.

Proposition 9 *Let $b > 0$. Assume the costs are distributed D independently, with $a - \underline{c} > b$ and $a - \bar{c} < b$ and that an agent's switching cost is proportional to the gain from a switch.*

Then I_1^α is implementable with with allowable punishment of at least $\max\{0, \bar{m}_1^\alpha\}$,

$$\bar{m}_1^\alpha = \alpha\bar{m}_1 - (1-\alpha)\,b, \tag{29}$$

where $\bar{m}_1$ is the punishment needed for a big (original) reform. Therefore for any $a, b, \underline{c}, \bar{c}$ there exists $\alpha^* > 0$, rule I_1^α is implementable with no punishment $(\bar{m}_1^\alpha \leq 0)$.

One might not find it surprising that a smaller reform requires less coercion. As the size of the reform decreases, gross payoff from the switch, $a-b$, as well as the switching costs all decrease at the same proportion, thereby decreasing the gain from misrepresenting one's costs (informational gains). What is striking, however, is that for a scaled down version of any reform the implementation of the benevolent rule requires no punishment whatsoever. Interestingly, a malevolent authority can not avoid threats altogether if the big leap requires a positive punishment, i.e., if $\bar{m}_2 = \bar{c} - a > 0$. The best she can do is to decrease the punishment for a small reform to $\alpha\,(\bar{c} - a)$, which is, clearly, positive for any $\alpha > 0$.

Along the same lines, if one compares a reform with a big gain and a high cost with another reform, in which both the gain and the costs are smaller (not necessarily proportional to the first one), then the punishment needed to implement a bigger reform is higher provided the threshold n^* of the low cost agents is identical in both cases. See lemma 14 in the appendix.

4.3 Continuous costs case

Let the private costs of transition c_i be i.i.d. with compact support. Note that in the discussion of the bad reform we did not rely on the discreteness of the distribution, so it is left to derive $\bar{m}_1$, the the bound on the minimal punishment, for a benevolent reformer.[10]

First, we follow the standard mechanism design argument (Mirrlees (1971), as presented in Fudenberg and Tirole (1996)) to develop implications of incentive compatibility and then incorporate all other constraints to calculate the bound on the minimal punishment, $\bar{m}_1$.

Proposition 10 *Let c_i be i.i.d. F on $[\underline{c}, \bar{c}]$ with corresponding marginal distribution f. Then I_1 is implementable with with allowable punishment of at least* $\max\{0, \bar{m}_1\}$,

$$\bar{m}_1 = -\frac{1}{I(\bar{c})}\left\{\int_{\underline{c}}^{\bar{c}}\left[a - \left(\frac{F(s)}{f(s)} + s\right)\right]\bar{I}(s) f(s)\, ds + (1 - Q(\Delta))\, b\right\}, \tag{30}$$

where Q is the cumulative distribution of the sum of the costs, and $\Delta = N(a-b)$, i.e., the sum of the gains from transition; $\bar{I}(\theta_i) = Q(\Delta | c_i = \theta_i)$ probability of reform conditional on individual i's announcement of his cost, θ_i.

It is possible to demonstrate that this is, indeed, a counterpart of $\bar{m}_1$ derived for the two types case.[11]

---

[10] We use the same notation for the upper bound of the punishment in the continuous case. This is justified by the next footnote.

[11] Note that $Q(\Delta) = \Pr\{\mathcal{A}\}$. Let $\bar{c} = \bar{c}$, and $\underline{c} = \underline{c}$. Furthermore, for the two type case

$$\begin{aligned} & \int_{\underline{c}}^{\bar{c}} s f(s)\, \bar{I}_F(s)\, ds = \\ = \;& \bar{c}\Pr\{\mathcal{A}|c_1 = \bar{c}\}\Pr\{c_1 = \bar{c}\} + \underline{c}\Pr\{\mathcal{A}|c_1 = \underline{c}\}\Pr\{c_1 = \underline{c}\} \\ = \;& \bar{c}\Pr\{n_L \geq n^*\}\Pr\{c_1 = \bar{c}\} + \underline{c}\Pr\{n_L \geq n^* - 1\}\Pr\{c_1 = \underline{c}\}. \end{aligned}$$

For the two point distribution, $D(s) = \Pr\{c_1 = \underline{c}\}$ for $s \in (\underline{c}, \bar{c})$. Furthermore, $\bar{I}_D(s) = \Pr\{n_L \geq n^*\}$

Note that this bound, $\bar{m}_1$, is the negative of two terms. The first is the expected 'virtual' payoff in case of reform, and the second one is its counter-part in case no reform is undertaken. The first term is familiar from the mechanism design literature. Assume $b = 0$, then $\bar{m}_1 > 0$ only when the objective is *not* implementable in the standard framework, i.e., if the standard individual rationality constraint is incompatible with incentive compatibility and budget balance constraints.[12] Softening restrictions on the punishment, is identical (in this case) to relaxing the individual rationality constraint, thus, it extends the range of implementable reforms. Recall that without the individual rationality constraint, benevolent rule is implementable using d'Aspremont and Gérard-Varet (1979) mechanism.

The next proposition generalizes some of the comparative statics results for this case.

If costs distributions can be ordered according to the first order stochastically dominance criterion, then the dominating distribution corresponds to a more 'expensive' reform, in particular, with higher average cost of transition. In particular, it asserts that 'bad' reforms require harsh punishments. The second part of the proposition compares punishments under two distributions that are ordered by "more peaked" order. The following definition is adopted from Shaked and Shanthikumar (1994), p.77.

Definition 11 *Let X, Y be random variables with distributions symmetric about μ and ν correspondingly. Let distribution of $|X - \mu|$ be F and that of $|Y - \nu|$ be H. Then X is more peaked about μ than Y about ν iff H first order stochastically dominates F.*

This order indicates which distribution is more spread, thus, corresponding to a more 'heterogeneous' society. Therefore, this is a counter-part of proposition (8).

for $s \in (\underline{c}, \bar{c}]$. Therefore,

$$\int_{\underline{c}}^{\bar{c}} F(s)\, \bar{I}_F(s) ds = \Pr\{c_1 = \underline{c}\} \Pr\{n_L \geq n^*\}(\bar{c} - \underline{c}).$$

[12] In terms of Myerson and Satterthwaite (1983), all the individuals are "sellers" that obtain the same "price" (a) and have private information about their costs.

Proposition 12 *1. Assume H first order stochastically dominates (FOSD) F, then the allowable punishment necessary to implement I_1 under H is higher than that under F;*

2. The same conclusion is also true if

(a) f, h are symmetric and unimodal around $d = \frac{1}{2}(\bar{c} + \underline{c}) = (a - b)$;

(b) F is more peaked than H.

To formulate a generalization of proposition (9), note that a 'small' step reform that generates a fraction $\alpha \in (0,1)$ of the original gain, $(a-b)$, and requires a fraction α of the original costs, $c_i^\alpha \equiv \alpha c_i$ for all agents i will require a punishment

$$\bar{m}_1^\alpha \equiv \int_{\alpha\underline{c}}^{\alpha\bar{c}} \left[F\left(\frac{1}{\alpha}s_\alpha\right) + s_\alpha \frac{1}{\alpha} f\left(\frac{1}{\alpha}s_\alpha\right)\right] \bar{I}\left(\frac{1}{\alpha}s_\alpha\right) ds_\alpha - b - \alpha(a-b)\,Q(\Delta) \quad (31)$$

$$= \alpha\bar{m}_1 - (1-\alpha)\,b. \quad (32)$$

Clearly, with α small enough and $b > 0$, the small step reform will require no punishments, $m_1^\alpha \leq 0$, then identical argument to that in proposition (9) establishes the rest of the result.

5 Extensions

5.1 Outcome Uncertainty

Here we demonstrate that it is easy to re-formulate this model to capture some cases of common uncertainty with respect to the outcome of a reform for the two types case.

Suppose that there are two different levels of payoffs (a_H and a_L) in the outcome of s^A, both of which are higher than b. Each agent's payoff in s^A is independently and identically distributed. Every agent has the same switching cost c. Let $\tilde{n}_H$ be the number of agents other than agent 1 that have the high payoffs. Let $\tilde{n}^*$ be the cut-off number of agents with high payoffs such that the switch is beneficial. Thus,

$$\tilde{n}^* a_H + (N - \tilde{n}^*) a_L = Nc.$$

Then we have the following result.

Proposition 13 *Assume that there are only two levels of benefits. When $a_H - c > b$ and $a_L - c < b$, the minimum level of punishment to always implement the first best outcome is given by* $\max\{0, \bar{m}_1\}$,

$$\bar{m}_1 = -a_H \Pr\{a_1 = a_H\} \Pr\{\tilde{n}_H = \tilde{n}^* - 1\} - a_L \Pr\{\tilde{n}_H \geq \tilde{n}^*\} - b \Pr\{\mathcal{B}\} + c \Pr\{\mathcal{A}\}. \quad (33)$$

Consider another situation where the payoff is the same for every agent in s^A but different agent receives different information about it. Suppose that all agents have the same switching cost. Let a be the common payoff that every of them will receive in s^A, and agent i receives a signal a_i. It is easy to see that it is a dominant strategy for them to reveal their information in the absence of any transfers. Therefore, the truthtelling condition is automatically satisfied. Suppose that conditional on the revealed information $(a_1, a_2, ..., a_N)$, it is beneficial to switch to s^A. Then the incentive compatibility constraint for switching becomes

$$E(a|a_1, a_2, ..., a_N) - c \geq -\bar{m}.$$

Therefore, we require that

$$\bar{m} \geq c - E(a|a_1, a_2, ..., a_N)$$

for all combinations of $(a_1, a_2, ..., a_N)$ such that switching is beneficial in expectation term. Since switching is beneficial only if

$$E(a|a_1, a_2, ..., a_N) - c \geq b,$$

the above determined $\bar{m}$ is negative. That is, no positive punishment is required.

Therefore, provided the reforming authority can credibly announce all the individual signals (example being "free press"), a worthwhile reform that imposes identical costs across agents should be implementable with no punishments.

5.2 Foreign Aid

As an additional extension, one could also easily introduce a source of an outside financing or aid, T, so as to modify the budget constraint BB to read $\Sigma_i t_i \leq T$. It is easy to check that in this case, the necessary punishments, $\bar{m}_1$ and $\bar{m}_2$, (if positive) each decrease by $T/(NI(\bar{c}))$, thus, making it easier to implement both a benevolent and a malevolent rule.[13] Provided the interim well-being of the highest cost individual is exactly equal to $-\bar{m}_i$, this outside transfer T, may improve the (expected) utility of the least fortunate.[14] Clearly, this improvement is not guaranteed by the mere presence of transfer, as the reformer has to be encouraged to use it appropriately. If the latter is assured, our results can be re-interpreted as suggesting that preserving the level of "human rights" might be costly (i.e., require external financial assistance) especially at the outset of a reform involving substantial individual adjustments, besides, more aid might be required for more heterogeneous societies.[15] The model then introduces a way to formulate a trade-off between the standards of human rights and foreign aid. Besides, it also provides a rationale to condition international aid on human rights protection in the recipient country.

6 Discussion

We chose a standard Bayes-Nash implementation framework to understand why resorting to punishment in the presence of "limited liability" constraint (restrictions on the individual tax) can "improve efficiency." We have deliberately avoided formulating "second best" solutions that depend on the severity of available punishments, as that necessitates formulating the objective of the designer (reformer, in this case) which is not a-priori clear. Our goal could be reduced to characterizing the lowest bound on punishments that make the set of implementable ex-post efficient mechanisms feasible, i.e., (ex-ante) budget balanced and (ex-post) individually rational (in terms of lemma 5). The "possibility" results (d'Aspremont and Gérard-Varet (1979)) with no individual rationality constraints, and the "impossibility" results (Myerson and Satterthwaite (1983), Mailath and Postlewaite (1990) and Rob (1989)) imposing those constraints suggest, in particular, that there should be just the right way to 'soften' the individual rationality restrictions, which amounts to increasing the lower bound on punishments in our environment.

We used Bayesian implementation, as it is easier to implement the objective of a designer, when the choice of the agents' strategies is not restricted (to dominant strategies, say), thus, imposing less pressure on the threats that the designer needs.[16]

[13]Follows from the appropriately modified proofs of proposition 10 and IIR correspondingly.

[14]With high enough transfers (in this model) an individual will be induced to switch to any action, so that *any* reform, however whimsical, can be implemented.

[15]under the assumptions of proposition 12.

[16]The assumption of common knowledge of distribution of types (costs) shared by all the agents

Even in this framework a designer has to credibly threaten with a punishment in order to implement good reforms under some circumstances.

As usual, many of the assumptions were made for simplicity. Introducing additional aspects that vary across individuals will hardly simplify the problem of finding the appropriate lower bound on punishment. However, in some environments multi-dimensionality of the relevant individual characteristics might relieve the pressure on this boundary. Bundling, or linking independent decisions (public goods) can improve efficiency, see Jackson and Sonnenschein (2003), Fang and Norman (2003). However in the context of this model, provided the reform is interpreted as a single (global) public good the above results are not applicable.

So far we have assumed that a reform entails a coordinated response of all the agents in a society. No doubt, it might be a close description of some real-life transitions, for example, altering the alphabet, or exchanging the acceptable currency, a switch from driving on the left to driving on the right hand side and vice versa. However, some other reforms, say, privatization, rely on just a subset of individuals to substantially alter their actions for the reform to be "successful." It could be interesting to extend the framework by allowing some of the agents to retain their old action, for example, if their costs are high enough, i.e., to incorporate partial reforms.

In the view of the contribution by Ledyard and Palfrey (2003), we conjecture that our results can be re-formulated for independent but not identical distributions. Introducing correlation in individual costs of transition, however, can substantially change the results. It is well known that in this case it is possible to (approximately) achieve ex-post efficiency, as in Crémer and McLean (1985) and McAfee and Reny (1992). Although the implementation abides interim individual rationality constraint, it might violate the ex-post one (requiring infinitely high taxes), which is crucial to the current framework, imposing the "limited liability" constraint (RC) and restricting available punishments (IRH). Similar investigation for correlated environments, therefore, is left for the future research.

may, indeed, be problematic (see Myles (1995)), but it "helps" the designer, thus, decreases the necessary punishment. Provided lower bound on punishment is our objective, this assumption is reasonable.

7 Conclusions

We found that reforms can be hindered by asymmetric information about costs of transition if some individuals have to be compensated for the transition, if there is no outside funding for a reform, and if individual resources' available for taxing are limited. In this case a reformer might need to be able to credibly threaten an agent with a punishment to assure compliance. The minimal level of credible punishment should be higher for 'big' transitions and in more divided societies, as we have illustrated.

This might explain the puzzling link between economic success of reforms and the 'authoritarian' rulers in power mentioned in the introduction. However, even in our simple set up, the threat of punishment (associated with these rulers) does not necessarily have to be severe in order for *any* desirable reform to be implemented. Moreover, unfortunately, as was mentioned above, allowing for more punishments makes it easier to implement a rule that we called malevolent.

It is then natural to expect that the international community will come up with some mechanisms to protect individuals against bad reforms in their countries. With direct foreign intervention (determining which reforms to undertake) being often impossible and undesirable, the outsiders can settle on enforcing human rights protection instead. As our results suggest, human rights, indeed, may be a sensible indicator to monitor. If the level of maximal individual punishment is set at a 'correct' level, bad reforms will be impossible to implement, while good ones can still go through. This punishment level hinges on the knowledge of local costs of transition and their distribution, which may differ by country. If this level is to be set internationally, i.e., it is to be the same across all countries, it will prevent good reforms in some countries and enable bad ones in the others.

This work also suggests that there is a trade-off between a successful implementation of good reforms from the utilitarian perspective and well-being of selected individuals in the society. This trade-off is not driven by ad-hoc restrictions on transfers, but, rather, by asymmetric information with respect to private costs of transition. We are far from being able to contribute to the 'moral calculus' resolving the trade-off, but the credit is ours for its formulation.

A Proofs for the Benchmark

Proof of proposition 3. First, assume the benevolent reformer observes $\mu \leq a - b$. Then, to make the good reform (from norm b to norm a) implementable, the following conditions should hold: $BB\ IRH, RC$ and the incentive constraint becomes (restricting attention to pure strategies)

$$\begin{aligned} U\left(s^A, c_i, c_{-i}, 1\right) &\geq U\left(B, s_{-i}^A, c_i, c_{-i}, 0\right) \text{ for all } i, \text{ where} \\ U\left(s, c_i, c_{-i}, y\right) &= u(s) - c_i y + t\left(s, c_i, c_{-i}, y\right) - m\left(s, c_i, c_{-i}, y\right), \end{aligned} \tag{34}$$

where s_{-i}^A is subprofile of actions corresponding to the case in which all players but i choose action A, so that

$$\left(B, s_{-i}^A\right) = (B, A, .., A). \tag{35}$$

Then the reformer will set

$$m\left(s^A, c, 1\right) = 0; \tag{36}$$

$$m\left(B, s_{-i}^A, c, 0\right) = \bar{m}; \tag{37}$$

$$t\left(B, s_{-i}^A, c, 0\right) = -u\left(B, s_{-i}^A\right) \tag{38}$$

Therefore the incentive constraint becomes

$$a - c_i + t\left(s^A, c_i, c_{-i}, 1\right) \geq u\left(B, s_{-i}^A\right) - u\left(B, s_{-i}^A\right) - \bar{m} \text{ for all } i, \tag{39}$$

which implies

$$a - c_i + t\left(s^A, c_i, c_{-i}, 1\right) \geq -\bar{m} \text{ for all } i, \tag{40}$$

Sum up over i and divide by N, and get

$$a - \mu + \bar{t} \geq -\bar{m}, \tag{41}$$

where $\bar{t}$ is the average tax and μ is the average switching cost. But in the view of (BB), $\bar{t} = 0$, so if all the incentive constraints hold if

$$\bar{m} \geq -a + \mu. \tag{42}$$

conversely, if (42) holds, the reformer can pick taxes in such a way as to equalize the after tax switching cost across agents and thus, implement the reform. Note that this condition is independent on whether the reform is a good one ($\mu < a - b$), or not.

Therefore the boundary $\bar{m}_1 \equiv \mu - a$ is also the smallest punishment to introduce any reform (bad ones included).

If $\mu > a - b$, a benevolent reformer, implementing I_1, has to make sure the agents are discouraged from switching, thus implying

$$b + t\left(s^B, c_i, c_{-i}, 0\right) \geq -\bar{m} - c_i \tag{43}$$

so that

$$\bar{m} \geq -b - \mu. \tag{44}$$

Note that as $\mu > a - b$, this boundary is below $-a + \mu$, in other words

$$-a + \mu > -b - \mu. \tag{45}$$

Therefore,

$$\bar{m} \geq \bar{m}_1 = -a + \mu, \tag{46}$$

■

B Proofs for the Discrete Case

Proof of lemma 5. ¿From IC, for any profile of announcements θ that call for the reform, in order to convince the agents to play the desired action, the authority should set

$$\begin{aligned} m\left(s^A, \theta, 1\right) &= 0; & (47)\\ m\left(B, s^A_{-i}, \theta, 0\right) &= \bar{m}; & (48)\\ t\left(B, s^A_{-i}, \theta, 0\right) &= -u\left(B, s^A_{-i}\right). & (49) \end{aligned}$$

It implies, that

$$a - c_i + t\left(s^A, \theta, 1\right) \geq -\bar{m} \text{ for all } i. \tag{50}$$

In case the reform is not to be implemented based on the announced valuations the same argument implies that the corresponding incentive constraint should be of the form

$$b + t\left(s^B, \theta, 0\right) \geq -\bar{m} \text{ for all } i. \tag{51}$$

Note however, that this latter constraint always holds as long as $\bar{m} \geq 0$, and another constraint (RC) is satisfied, i.e.,

$$b + t\left(s^B, \theta, 0\right) \geq 0. \tag{52}$$

The conclusion then follows. ■

Proof of proposition 6. By lemma 5,

$$a - c_i + t(s^A, \theta_i, \theta_{-i}, 1) \geq -\bar{m}. \tag{53}$$

It implies that

$$t(s^A, \theta_i, \theta_{-i}, 1) \geq c_i - a - \bar{m}, \tag{54}$$

which has to be satisfied for any announcements and any cost c_i.

In addition, truthtelling constraint should be satisfied, in other words, compensation should be formulated in such a way that nobody has a motivation to lie about the costs of his transition. As the decision rule is constant, i.e., independent of the profile of the announced costs, so should the transfer, as otherwise every agent would announce the cost corresponding to the highest transfer. This implies that $t(s^B, \theta_i, \theta_{-i}, 1)$ should not vary with $\theta_i \in \mathbf{R}$,

$$t(s^B, \theta_i, \theta_{-i}, 1) = \hat{t} \tag{55}$$

Combining with (54), it implies that

$$\hat{t} \geq c_i - a - \bar{m} \tag{56}$$

for all $c_i \in [\bar{c}, \underline{c}]$. To minimize the transfer while still satisfying the incentive compatibility constraint for all $c_i \in [\bar{c}, \underline{c}]$, we set

$$\hat{t} = \bar{c} - a - \bar{m}.$$

To balance the budget, the sum of the transfers has to be non-negative. That is,

$$\bar{m} \geq \bar{c} - a.$$

Therefore, the minimum of $\bar{m}$ is

$$\bar{m}_2 = \bar{c} - a.$$

Note that $\hat{t} = 0$ for all $c_i \in [\bar{c}, \underline{c}]$ is individually feasible, satisfying (RC). It also satisfies the rest of the constraints for $\bar{m} \geq \bar{m}_2$. ■

Proof of proposition 13. First, denote

$$\Pr\{\mathcal{A}\} = \Pr\{c_1 = \underline{c}\}\Pr\{n_L \geq n^* - 1\} + \Pr\{c_1 = \bar{c}\}\Pr\{n_L \geq n^*\} \tag{57}$$

as the ex-ante probability that s^A should be enforced and

$$\Pr\{\mathcal{B}\} = \Pr\{c_1 = \underline{c}\}\Pr\{n_L < n^* - 1\} + \Pr\{c_1 = \bar{c}\}\Pr\{n_L < n^*\}$$

as the probability that s^B should be enforced ex-ante. Note that $\Pr\{\mathcal{A}\}$ can also be expressed as

$$\Pr\{\mathcal{A}\} = \Pr\{n_L \geq n^*\} + \Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\}.$$

This is because

$$\Pr\{n_L \geq n^* - 1\} = \Pr\{n_L \geq n^*\} + \Pr\{n_L = n^* - 1\}.$$

Suppose that agent 1 has switching cost c_1. Let $E_A\tau(c_1)$ denote the expected transfer agent 1 receives conditional on that s^A should be implemented. Similarly, let $E_B\tau(c_1)$ denote the expected transfer agent 1 receives conditional on that s^B should be implemented. Thus,

$$E_A\tau(\underline{c}) = E\{\tau(\underline{c}, c_2, ..., c_N)|n_L \geq n^* - 1\},$$

$$E_A\tau(\bar{c}) = E\{\tau(\bar{c}, c_2, ..., c_N)|n_L \geq n^*\},$$

$$E_B\tau(\underline{c}) = E\{\tau(\underline{c}, c_2, ..., c_N)|n_L < n^* - 1\},$$

and

$$E_B\tau(\bar{c}) = E\{\tau(\bar{c}, c_2, ..., c_N)|n_L < n^*\}.$$

There are two stages in this game. First, the government announces transfers to the agents contingent on their report of the costs. Second, the government announces whether or not a reform will take place and punishments for anyone who does not follow the instructions. Due to the revelation principle, we concentrate on direct mechanisms. We characterize the conditions for an efficient equilibrium to exist, and then select the minimum punishment for such an equilibrium to exist.

First, we consider the incentive compatibility constraints for agent 1 to follow the instruction of whether or not to switch from B to A. Consider an efficient equilibrium.

If he receives $\bar{c}$ and $n_L \geq n^*$, he is required to switch to A and gets $a - \bar{c} + E_A\tau(\bar{c})$, where $E_A\tau(\bar{c})$ is the transfer in this case. (As we can easily see below, having a constant transfer of $E_A\tau(\bar{c})$ helps to satisfy the incentive compatibility constraint.) If he refuses to switch, all of his income will be taxed away plus he is punished to the most extend. Therefore, he receives $-\bar{m}$ in this case. That is, for $c_1 = \bar{c}$ and $n_L \geq n^*$,

$$a - \bar{c} + E_A\tau(\bar{c}) \geq -\bar{m}. \tag{58}$$

For $c_1 = \bar{c}$ and $n_L < n^*$, no switch is required, and

$$b + E_B\tau(\bar{c}) \geq -\bar{m}. \tag{59}$$

Similarly, for $c_1 = \underline{c}$ and $n_L \geq n^* - 1$, switching to A is required, and

$$a - \underline{c} + E_A\tau(\underline{c}) \geq -\bar{m}. \tag{60}$$

For $c_1 = \underline{c}$ and $n_L < n^* - 1$, no switch is required, and

$$b + E_B\tau(\underline{c}) \geq -\bar{m}. \tag{61}$$

Now consider the information revelation in the first stage. Suppose that agent 1's switching cost is $\underline{c}$. Then the incentive compatibility constraint for him to report $\underline{c}$ is given by

$$\begin{aligned} & [a - \underline{c} + E_A\tau(\underline{c})] \Pr\{n_L \geq n^* - 1\} + [b + E_B\tau(\underline{c})] \Pr\{n_L < n^* - 1\} \\ \geq \ & [a - \underline{c} + E_A\tau(\bar{c})] \Pr\{n_L \geq n^*\} + [b + E_B\tau(\bar{c})] \Pr\{n_L < n^*\} \end{aligned} \tag{62}$$

Suppose that agent 1's switching cost is $\bar{c}$. Then the incentive compatibility constraint for him to report $\bar{c}$ is given by

$$\begin{aligned} & [a - \bar{c} + E_A\tau(\bar{c})] \Pr\{n_L \geq n^*\} + [b + E_B\tau(\bar{c})] \Pr\{n_L < n^*\} \\ \geq \ & [a - \bar{c} + E_A\tau(\underline{c})] \Pr\{n_L \geq n^* - 1\} + [b + E_B\tau(\underline{c})] \Pr\{n_L < n^* - 1\} \end{aligned} \tag{63}$$

The assumption that the government cannot tax more than one's income implies that

$$E_A\tau(\bar{c}) \geq -a, \qquad E_A\tau(\underline{c}) \geq -a, \qquad E_B\tau(\bar{c}) \geq -b, \qquad \text{and } \ E_B\tau(\underline{c}) \geq -b.$$

These inequalities imply that (59) and (61) are automatically satisfied.

Let $\bar{\tau}(c_1)$ denote the expected transfer agent 1 receives when his reported switching cost is c_1. That is,

$$\bar{\tau}(\underline{c}) = E_A\tau(\underline{c}) \Pr\{n_L \geq n^* - 1\} + E_B\tau(\underline{c}) \Pr\{n_L < n^* - 1\}$$

and

$$\bar{\tau}(\bar{c}) = E_A\tau(\bar{c})\Pr\{n_L \geq n^*\} + E_B\tau(\bar{c})\Pr\{n_L < n^*\}.$$

The two incentive compatibility constraints for truthful reporting (62) and (63) can then be simplified as

$$\bar{\tau}(\bar{c}) \leq \bar{\tau}(\underline{c}) + (a - \underline{c} - b)\Pr\{n_L = n^* - 1\} \equiv \beta \tag{64}$$

and

$$\bar{\tau}(\bar{c}) \geq \bar{\tau}(\underline{c}) + (a - \bar{c} - b)\Pr\{n_L = n^* - 1\} \tag{65}$$

We argue that $E_B\tau(\bar{c}) = -b$ and that $E_B\tau(\underline{c}) = -b$. This is because $E_B\tau(\bar{c})$ and $E_B\tau(\underline{c})$ cannot be lower than $-b$ from the tax constraint. If we raise them while lowering $E_B\tau(\bar{c})$ and $E_B\tau(\underline{c})$ to keep the expected transfers $\bar{\tau}(\bar{c})$ and $\bar{\tau}(\underline{c})$ constant, it would not affect (64) and (65), but make (58) and (59) more difficult to hold.

We want to characterize the minimum $\bar{m}$ such that the budget is balanced ex ante, that is, $E(\bar{\tau}(c)) \leq 0$. In order to do so, we fix $\bar{m}$ and characterize the minimum expected transfer that still implement the efficient equilibrium outcome. This transfer is a decreasing function of $\bar{m}$. We then set the expected transfer to zero to obtain the minimum feasible $\bar{m}$.

Given $E_B\tau(\bar{c})$ and $E_B\tau(\underline{c})$, from (58) and (60), we have

$$\begin{aligned}
\bar{\tau}(\bar{c}) &= E_A\tau(\bar{c})\Pr\{n_L \geq n^*\} + E_B\tau(\bar{c})\Pr\{n_L < n^*\} \\
&\geq \gamma \equiv (\bar{c} - a - \bar{m})\Pr\{n_L \geq n^*\} - b\Pr\{n_L < n^*\},
\end{aligned}$$

and

$$\begin{aligned}
\bar{\tau}(\underline{c}) &= E_A\tau(\underline{c})\Pr\{n_L \geq n^* - 1\} + E_B\tau(\underline{c})\Pr\{n_L < n^* - 1\} \\
&\geq \delta \equiv (\underline{c} - a - \bar{m})\Pr\{n_L \geq n^* - 1\} - b\Pr\{n_L < n^* - 1\}.
\end{aligned}$$

At $\bar{\tau}(\underline{c}) = \delta$, noting (64),

$$\begin{aligned}
\beta &= \bar{\tau}(\underline{c}) + (a - \underline{c} - b)\Pr\{n_L = n^* - 1\} \\
&= (\underline{c} - a - \bar{m})\Pr\{n_L \geq n^* - 1\} - b\Pr\{n_L < n^* - 1\} \\
&\qquad + (a - \underline{c} - b)\Pr\{n_L = n^* - 1\} \\
&< (\bar{c} - a - \bar{m})\Pr\{n_L \geq n^* - 1\} - b\Pr\{n_L < n^* - 1\} \\
&\qquad + (a - \bar{c} - b)\Pr\{n_L = n^* - 1\} \\
&< (\bar{c} - a - \bar{m})\Pr\{n_L \geq n^*\} - b\Pr\{n_L < n^*\} \\
&= \gamma
\end{aligned}$$

because $\Pr\{n_L \geq n^* - 1\} - \Pr\{n_L = n^* - 1\} = \Pr\{n_L \geq n^*\}$.

Therefore, to minimize the expected transfer, we raise $\bar{\tau}(\underline{c})$ such that $\beta = \gamma$, and set $\bar{\tau}(\bar{c}) = \gamma$. The first condition becomes

$$\begin{aligned} & \bar{\tau}(\underline{c}) + (a - \underline{c} - b) \Pr\{n_L = n^* - 1\} \\ = \; & (\bar{c} - a - \bar{m}) \Pr\{n_L \geq n^*\} - b \Pr\{n_L < n^*\}, \end{aligned}$$

which gives us

$$\begin{aligned} \bar{\tau}(\underline{c}) \; = \; & -(a - \underline{c} - b) \Pr\{n_L = n^* - 1\} \\ & +(\bar{c} - a - \bar{m}) \Pr\{n_L \geq n^*\} + (-b - \bar{m}) \Pr\{n_L < n^*\} \\ = \; & -\bar{m} \Pr\{n_L \geq n^*\} - a \Pr\{n_L \geq n^* - 1\} - b \Pr\{n_L < n^* - 1\} \\ & +\bar{c} \Pr\{n_L \geq n^*\} + \underline{c} \Pr\{n_L = n^* - 1\} \end{aligned} \tag{66}$$

Therefore, the minimum expected transfer is

$$\begin{aligned} E(\bar{\tau}(\cdot)) \; = \; & \Pr\{c_1 = \underline{c}\}\bar{\tau}(\underline{c}) + \Pr\{c_1 = \bar{c}\}\bar{\tau}(\bar{c}) \\ = \; & \Pr\{c_1 = \underline{c}\}[-\bar{m} \Pr\{n_L \geq n^*\} - a \Pr\{n_L \geq n^* - 1\} \\ & -b \Pr\{n_L < n^* - 1\} + \bar{c} \Pr\{n_L \geq n^*\} + \underline{c} \Pr\{n_L = n^* - 1\}] \\ & + \Pr\{c_1 = \bar{c}\}[(\bar{c} - a - \bar{m}) \Pr\{n_L \geq n^*\} - b \Pr\{n_L < n^*\}] \\ = \; & -\bar{m} \Pr\{n_L \geq n^*\} - a \Pr\{\mathcal{A}\} - b \Pr\{\mathcal{B}\} + \bar{c} \; \Pr\{n_L \geq n^*\} \\ & +\underline{c} \Pr\{c_1 = \underline{c}\} \Pr\{n_L = n^* - 1\} \end{aligned} \tag{67}$$

The ex ante budget balance $E(\bar{\tau}(c)) \leq 0$ implies

$$\bar{m} \geq -a \frac{\Pr\{\mathcal{A}\}}{\Pr\{n_L \geq n^*\}} - b \frac{\Pr\{\mathcal{B}\}}{\Pr\{n_L \geq n^*\}} + \bar{c} + \underline{c} \frac{\Pr\{c_1 = \underline{c}\} \Pr\{n_L = n^* - 1\}}{\Pr\{n_L \geq n^*\}}.$$

By taking the minimum of $\bar{m}$, we obtain

$$\bar{m}_1 = -a \frac{\Pr\{\mathcal{A}\}}{\Pr\{n_L \geq n^*\}} - b \frac{\Pr\{\mathcal{B}\}}{\Pr\{n_L \geq n^*\}} + \bar{c} + \underline{c} \frac{\Pr\{c_1 = \underline{c}\} \Pr\{n_L = n^* - 1\}}{\Pr\{n_L \geq n^*\}}.$$

We still need to verify that the tax constraints are satisfied when s^A is implemented; that is, no one is taxed more than his income. First note that $\bar{m}_1 < \bar{c}$, since $a > \underline{c}$. From $\bar{\tau}(\bar{c}) = \gamma$, we have

$$\begin{aligned} & E_A \tau(\bar{c}) \Pr\{n_L \geq n^*\} + E_B \tau(\bar{c}) \Pr\{n_L < n^*\} \\ = \; & (\bar{c} - a - \bar{m}) \Pr\{n_L \geq n^*\} - b \Pr\{n_L < n^*\}. \end{aligned}$$

That is,

$$E_A\tau(\bar{c}) = \bar{c} - a - \bar{m} > -a.$$

Therefore, the tax constraint is satisfied for the high cost agents.

From (66) and the definition of $\bar{\tau}(\underline{c})$, we have

$$\begin{aligned} & E_A\tau(\underline{c})\Pr\{n_L \geq n^* - 1\} + E_B\tau(\underline{c})\Pr\{n_L < n^* - 1\} \\ = & -\bar{m}\Pr\{n_L \geq n^*\} - a\Pr\{n_L \geq n^* - 1\} - b\Pr\{n_L < n^* - 1\} \\ & \quad +\bar{c}\Pr\{n_L \geq n^*\} + \underline{c}\Pr\{n_L = n^* - 1\}. \end{aligned}$$

Note that $E_B\tau(\underline{c}) = -b$ and substitute $\bar{m}_1$ for $\bar{m}$. We can easily show that $E_A\tau(\underline{c}) > -a$. Therefore, the tax constraint for the low cost agent is satisfied as well. So the characterization we obtained indeed satisfies all of the conditions. ■

Proof of proposition 9. Rewrite (24) in Proposition 7, we have

$$\begin{aligned} \Pr\{n_L \geq n^*\}\bar{m}_1 &= -b - (a-b)\Pr\{\mathcal{A}\} + \bar{c}\Pr\{n_L \geq n^*\} \\ & \quad +\underline{c}\Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\}. \end{aligned} \tag{68}$$

Applying Proposition 7 to the small reform (s^B to s^X), we have

$$\begin{aligned} \Pr\{n_L \geq n^*\}\bar{m}_1^{\alpha} &= -b - (x-b)\Pr\{X\} + \bar{c}(x)\Pr\{n_L \geq n^*\} \\ & \quad +\underline{c}(x)\Pr\{c_1 = \underline{c}(x)\}\Pr\{n_L = n^* - 1\} \\ &= -b - \alpha(a-b)\Pr\{\mathcal{A}\} + \alpha\bar{c}\Pr\{n_L \geq n^*\} \\ & \quad +\alpha\underline{c}\Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\} = \qquad (69) \\ &= \alpha\bar{m}_1 - (1-\alpha)\,b. \qquad (70) \end{aligned}$$

noting that $\Pr\{X\} = \Pr\{\mathcal{A}\}$.

Let α^* be the α such that $\bar{m}_1^{\alpha^*} = 0$ in (69). As $b > 0$, $\alpha^* > 0$. ■

Lemma 14 *Assume the costs are distributed D independently, with $a - \underline{c} > b$ and $a - \bar{c} < b$. Let $\bar{c} = \frac{N(a-b) - n^*\underline{c}}{N - n^*}$, so that n^* is preserved if either of the other parameters change. Then*

$$\begin{aligned} \frac{\partial \bar{m}_1}{\partial a} &> 0, \quad \frac{\partial \bar{m}_1}{\partial b} < 0, \\ \frac{\partial \bar{m}_1}{\partial \underline{c}} &< 0. \end{aligned}$$

Proof.

$$\begin{aligned}
\Pr\{n_L \geq n^*\}\bar{m}_1 &= -a\Pr\{\mathcal{A}\} - b\Pr\{\mathcal{B}\} + \frac{N(a-b) - n^*\underline{c}}{N-n^*}\Pr\{n_L \geq n^*\} \\
&\quad + \underline{c}\Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\}.
\end{aligned}$$

$$\begin{aligned}
\Pr\{n_L \geq n^*\}\frac{\partial \bar{m}_1}{\partial a} &= -\Pr\{\mathcal{A}\} + \frac{N}{N-n^*}\Pr\{n_L \geq n^*\} \\
&= -[\Pr\{n_L \geq n^*\} + \Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\}] \\
&\qquad + \left(1 + \frac{n^*}{N-n^*}\right)\Pr\{n_L \geq n^*\} \\
&= -\rho \begin{pmatrix} N-1 \\ n^*-1 \end{pmatrix} \rho^{n^*-1}(1-\rho)^{N-n^*} \\
&\qquad + \frac{n^*}{N-n^*}\left[\begin{pmatrix} N-1 \\ n^* \end{pmatrix}\rho^{n^*}(1-\rho)^{N-n^*-1} + \cdots\right] \\
&> -\begin{pmatrix} N-1 \\ n^*-1 \end{pmatrix}\rho^{n^*}(1-\rho)^{N-n^*} \\
&\qquad + \frac{n^*}{N-n^*}\frac{N-n^*}{n^*}\begin{pmatrix} N-1 \\ n^*-1 \end{pmatrix}\rho^{n^*}(1-\rho)^{N-n^*-1} \\
&> 0.
\end{aligned}$$

Similarly, we can prove that

$$\Pr\{n_L \geq n^*\}\frac{\partial \bar{m}_1}{\partial \underline{c}} = -\frac{n^*}{N-n^*}\Pr\{n_L \geq n^*\} + \Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\} < 0.$$

It is clear that

$$\Pr\{n_L \geq n^*\}\frac{\partial m_1}{\partial b} = -\Pr\{\mathcal{B}\} - \frac{N}{N-n^*}\Pr\{n_L \geq n^*\} < 0.$$

■

Proof of proposition 8. By the assumptions $n^* = \frac{1}{2}N$ under any such spread. By definition of $\bar{m}_1$, (24), it is enough to show that $\Pr\{n_L \geq n^*\} > \Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\}$. Indeed,

$$\begin{aligned}
\Pr\{n_L &\geq n^*\} > \frac{(N-n^*)}{n^*}\begin{pmatrix} N-1 \\ n^*-1 \end{pmatrix}\rho^{n^*}(1-\rho)^{N-n^*-1} = \left[\text{if } n^* = \frac{1}{2}N\right] && (71) \\
&= \begin{pmatrix} N-1 \\ n^*-1 \end{pmatrix}\rho^{n^*}(1-\rho)^{N-n^*-1} > \begin{pmatrix} N-1 \\ n^*-1 \end{pmatrix}\rho^{n^*}(1-\rho)^{N-n^*} = && (72) \\
&= \Pr\{c_1 = \underline{c}\}\Pr\{n_L = n^* - 1\}. && (73)
\end{aligned}$$

■

Proof of proposition 13. If we replace $a-\underline{c}$, $a-\bar{c}$, c_1, n_L, n^*, $\tau(\underline{c})$ and $\tau(\underline{c})$ in the proof of Proposition 5 by a_H-c, a_L-c, a_1, $\tilde{n}_H$, $\tilde{n}^*$, $\tau(a_H)$ and $\tau(a_H)$ respectively, then the proof goes through perfectly. From equation (66), we have

$$\begin{aligned}\bar{\tau}(a_H) &= -(a_H-c-b)\Pr\{\tilde{n}_H=\tilde{n}^*-1\} \\ &\quad +(c-a_L-\bar{m})\Pr\{\tilde{n}_H\geq\tilde{n}^*\}-b\Pr\{n_H<\tilde{n}^*\}\end{aligned}$$

By setting $\bar{\tau}(a_L)$ at its minimum level γ, we obtain the minimum (ex-ante) expected transfer, $E\left(\bar{\tau}\left(\cdot\right)\right)$, as follows:

$$\begin{aligned}E\left(\bar{\tau}\left(\cdot\right)\right) &= \Pr\{a_1=a_H\}\bar{\tau}(a_H)+\Pr\{a_1=a_L\}\bar{\tau}(a_L) \\ &= \Pr\{a_1=a_H\}[-(a_H-c-b)\Pr\{\tilde{n}_H=\tilde{n}^*-1\} \\ &\quad +(c-a_L-\bar{m})\Pr\{\tilde{n}_H\geq\tilde{n}^*\}-b\Pr\{n_H<\tilde{n}^*\}] \\ &\quad +\Pr\{a_1=a_L\}[(c-a_L-\bar{m})\Pr\{\tilde{n}_H\geq\tilde{n}^*\}-b\Pr\{\tilde{n}_H<\tilde{n}^*\}] \\ &= -\bar{m}\Pr\{\tilde{n}_H\geq\tilde{n}^*\}-a_H\Pr\{a_1=a_H\}\Pr\{\tilde{n}_H=\tilde{n}^*-1\} \\ -a_L\Pr\{\tilde{n}_H &\geq \tilde{n}^*\}-b\Pr\{\mathcal{B}\}+c\Pr\{\mathcal{A}\}\end{aligned}$$

To balance the budget ex ante, we require $E\left(\bar{\tau}\left(c\right)\right)\leq 0$. By taking the minimum of $\bar{m}$, we obtain

$$\bar{m}_1=\frac{-a_H\Pr\{a_1=a_H\}\Pr\{\tilde{n}_H=\tilde{n}^*-1\}-a_L\Pr\{\tilde{n}_H\geq\tilde{n}^*\}-b\Pr\{\mathcal{B}\}+c\Pr\{\mathcal{A}\}}{\Pr\{\tilde{n}_H\geq\tilde{n}^*\}}$$

■

C Proofs for the Continuous Case

Proof of proposition 10. Let the net payoff of player i, who follows the equilibrium strategy in the coordination game and has a cost of transition c_i be

$$v\left(I\right)-c_iI+\tau_i, \tag{74}$$

where

I is the indicator function for reform, unity (reform) or zero (status-quo);

τ_i is $i's$ transfer (provided he chooses the action suggested by the reformer);

equilibrium payoffs in the coordination game to follow;

$v(1) = a;$

$v(0) = b.$

Consider a social choice function $\phi(\theta) = (I(\theta), \tau_1(\theta), .., \tau_N(\theta))$, where I is either unity (reform) or zero (status-quo). Let

$$\begin{aligned} \bar{\tau}(\theta_i) &\equiv E_{c_{-i}}(\tau(\theta_i, c_{-i})), & (75) \\ \bar{I}(\theta_i) &\equiv E_{c_{-i}}(I(\theta_i, c_{-i})), & (76) \\ \bar{v}(\theta_i) &\equiv E_{c_{-i}}(v(I(\theta_i, c_{-i}))) & (77) \end{aligned}$$

be expected payment of i, probability of reform, and expected "equilibrium" payoff (in coordination game next period), conditional on i reporting θ_i, and all the rest are telling the truth. Note that

$$\bar{v}(\theta_i) = \bar{I}(\theta_i)\, a + \left(1 - \bar{I}(\theta_i)\right) b. \tag{78}$$

Let

$$V(c_i) \equiv \bar{v}(c_i) - c_i \bar{I}(c_i) + \bar{\tau}(c_i). \tag{79}$$

By a standard argument, Bayesian Incentive compatibility implies that $\bar{I}(c_i)$ is non-increasing (weakly decreasing). Moreover,

$$V'(c_i) = -\bar{I}(c_i), \tag{80}$$

which implies

$$V(c_i) = V(\underline{c}) - \int_{\underline{c}}^{c_i} \bar{I}(s)\, ds. \tag{81}$$

Let us incorporate additional constraints.

Recall that by $(50, 51)$,

$$\begin{aligned} a - c_i + \tau_i(c_i, c_{-i}) &\geq -\bar{m}, \text{ if } \sum_{j=1}^{N} c_j \leq \Delta = N(a-b) & (82) \\ b + \tau_i(c_i, c_{-i}) &\geq 0, \text{ if } \sum_{j=1}^{N} c_j > \Delta = N(a-b) & (83) \end{aligned}$$

This implies that the (soft) interim individual rationality constraint should be satisfied:

$$V(c_i) = \bar{I}(c_i)(a - c_i + \tau_A(c_i)) + \left(1 - \bar{I}(c_i)\right)(b + E_B\tau(c_i)) \geq -\bar{m}\bar{I}(c_i), \tag{84}$$

where, as in the discrete case

$$E_A\tau(c_i) \equiv E_{c_{-i}}\left(\tau_i(c_i, c_{-i}) \mid \sum_{j=1}^{N} c_j \leq \Delta\right); \tag{85}$$

$$E_B\tau(c_i) \equiv E_{c_{-i}}\left(\tau_i(c_i, c_{-i}) \mid \sum_{j=1}^{N} c_j > \Delta\right). \tag{86}$$

Combining (81) with (84), we get for all i

$$V(\underline{c}) - \int_{\underline{c}}^{c_i} \bar{I}(s)\, ds \geq -\bar{m}\bar{I}(c_i), \tag{87}$$

$$\int_{\underline{c}}^{c_i} \bar{I}(s)\, ds - \bar{m}\bar{I}(c_i) \leq V(\underline{c}). \tag{88}$$

When is this constraint binding?

As the left hand side in increasing in c_i, it is enough to verify that the constraint holds for the highest possible realization of cost, $c_i = \bar{c}$:

$$\int_{\underline{c}}^{\bar{c}} \bar{I}(s)\, ds - \bar{m}\bar{I}(\bar{c}) \leq V(\underline{c}). \tag{89}$$

Inequality (89) provides a lower bound on m :

$$\left(\int_{\underline{c}}^{\bar{c}} \bar{I}(s)\, ds - V(\underline{c})\right)\frac{1}{\bar{I}(\bar{c})} = \bar{m}_1 \tag{90}$$

By definition

$$V(\underline{c}) = \bar{I}(\underline{c})(a - b - \underline{c}) + b + \bar{\tau}(\underline{c}). \tag{91}$$

It implies that

$$\bar{I}(\bar{c})\bar{m}_1 = \int_{\underline{c}}^{\bar{c}} \bar{I}(s)\, ds - V(\underline{c}) \tag{92}$$

$$= \int_{\underline{c}}^{\bar{c}} \bar{I}(s)\, ds - \bar{I}(\underline{c})(a - b - \underline{c}) - b - \bar{\tau}(\underline{c}) \tag{93}$$

$$= \bar{c}\bar{I}(\bar{c}) - \int_{\underline{c}}^{\bar{c}} \frac{\partial \bar{I}(s)}{\partial s} s\, ds - \bar{I}(\underline{c})(a - b) - b - \bar{\tau}(\underline{c}) \tag{94}$$

We can express $\bar{\tau}(\underline{c})$ using the ex-ante budget balancedness,

$$\sum_{i=1}^{N} E_{i}\left[\bar{\tau}\left(c_{i}\right)\right] \leq 0 \tag{95}$$

Recall that by (81) $\bar{\tau}(c_i)$ can be represented as follows

$$\bar{\tau}\left(c_{i}\right)=\bar{\tau}(\underline{c})+(a-b)\left(\bar{I}(\underline{c})-\bar{I}\left(c_{i}\right)\right)-\underline{c}\bar{I}(\underline{c})+c_{i}\bar{I}\left(c_{i}\right)-\int_{\underline{c}}^{c_{i}} \bar{I}(s)\, ds. \tag{96}$$

Let $x=\sum_{i=1}^{N} c_i$ be distributed Q on $[\underline{c}, \bar{c}]$, with Q derived from $\{F_i\}_{i=1}^{N}$. Then

$$\begin{aligned} \bar{I}\left(\theta_{i}\right) &= Q\left(\Delta \mid c_{i}=\theta_{i}\right), & (97)\\ E_{i}\left(\bar{I}\left(\theta_{i}\right)\right) &= Q(\Delta). & (98) \end{aligned}$$

Note that

$$\int_{\underline{c}}^{c_{j}} \bar{I}(s)\, ds=c_{i}\bar{I}\left(c_{i}\right)-\underline{c}\bar{I}(\underline{c})-\int_{\underline{c}}^{c_{i}} \frac{\partial \bar{I}(s)}{\partial s} s ds, \tag{99}$$

and

$$\begin{aligned} E_{i} \int_{\underline{c}}^{c_{i}} \frac{\partial \bar{I}(s)}{\partial s} s ds &= \int_{\underline{c}}^{\bar{c}}\left[\int_{\underline{c}}^{c_{i}} \frac{\partial \bar{I}(s)}{\partial s} s ds\right] f\left(c_{i}\right) d\left(c_{i}\right)= & (100)\\ &= \int_{\underline{c}}^{\bar{c}}[1-F(s)] \frac{\partial \bar{I}(s)}{\partial s} s ds & (101) \end{aligned}$$

Combining, the three observations above with (96), we get

$$E_{i}\bar{\tau}\left(c_{i}\right)=\bar{\tau}(\underline{c})+(a-b)\left((\bar{I}(\underline{c})-Q(\Delta)\right)+\int_{\underline{c}}^{\bar{c}}[1-F(s)] \frac{\partial \bar{I}(s)}{\partial s} s ds. \tag{102}$$

Therefore, (95) implies

$$\tau(\underline{c}) \leq-\int_{\underline{c}}^{\bar{c}}[1-F(s)] \frac{\partial \bar{I}(s)}{\partial s} s ds-(a-b)\left((\bar{I}(\underline{c})-Q(\Delta)\right) \tag{103}$$

Substituting into (92), we get

$$\begin{aligned} \bar{I}(\bar{c}) \bar{m}_{1} &= \bar{c}\bar{I}(\bar{c})-\int_{\underline{c}}^{\bar{c}} \frac{\partial \bar{I}(s)}{\partial s} s ds-\bar{I}(\underline{c})(a-b)-b+ & (104)\\ &\quad +\int_{\underline{c}}^{\bar{c}}[1-F(s)] \frac{\partial \bar{I}(s)}{\partial s} s ds+(a-b)\left((\bar{I}(\underline{c})-Q(\Delta)\right) & (105)\\ &= \bar{c}\bar{I}(\bar{c})-b-(a-b) Q(\Delta)-\int_{\underline{c}}^{\bar{c}} \frac{\partial \bar{I}(s)}{\partial s} s F(s)\, ds & (106) \end{aligned}$$

Substituting

$$\int_{\underline{c}}^{\bar{c}} \frac{\partial \bar{I}(s)}{\partial s} sF(s)\, ds = \bar{c}I(\bar{c}) - \int_{\underline{c}}^{\bar{c}} [F(s) + sf(s)]\, \bar{I}(s)\, ds, \tag{107}$$

$$\begin{aligned} I(\bar{c})\, \bar{m}_1 &= \int_{\underline{c}}^{\bar{c}} [F(s) + sf(s)]\, \bar{I}(s)\, ds - b - (a-b)\, Q(\Delta) && (108) \\ &= \int_{\underline{c}}^{\bar{c}} \left[\frac{F(s)}{f(s)} + s\right] \bar{I}(s)\, f(s)\, ds - (a-b) \int_{\underline{c}}^{\bar{c}} \bar{I}(s)\, f(s)\, ds - b && (109) \\ &= \int_{\underline{c}}^{\bar{c}} \left[\frac{F(s)}{f(s)} + s - a\right] \bar{I}(s)\, f(s)\, ds - (1 - Q(\Delta))\, b. && (110) \end{aligned}$$

It is left to check that feasibility constraints (RC) are satisfied. Recall that

$$\begin{aligned} \bar{\tau}_i(c_i) &= \bar{\tau}_i(\underline{c}) + (a-b)\left(\bar{I}(\underline{c}) - \bar{I}(c_i)\right) - \underline{c}\bar{I}(\underline{c}) + c_i\bar{I}(c_i) - \int_{\underline{c}}^{c_i} \bar{I}(s)\, ds = && (111) \\ &= \bar{\tau}_i(\underline{c}) + (a-b)\left(\bar{I}(\underline{c}) - \bar{I}(c_i)\right) - \underline{c}\bar{I}(\underline{c}) + c_i\bar{I}(c_i) && (112) \\ &\quad - \left(c_i\bar{I}(c_i) - \underline{c}\bar{I}(\underline{c}) - \int_{\underline{c}}^{c_i} \frac{\partial \bar{I}(s)}{\partial s} s ds\right) && (113) \\ &= \bar{\tau}_i(\underline{c}) + (a-b)\left(\bar{I}(\underline{c}) - \bar{I}(c_i)\right) + \int_{\underline{c}}^{c_i} \frac{\partial \bar{I}(s)}{\partial s} s ds && (114) \end{aligned}$$

which implies that the (interim) payoff schedule is quasiconcave in $c_i \in [\underline{c}, \bar{c}]$, as

$$\begin{aligned} \bar{\tau}_i'(c_i) &= -\bar{I}'(c_i)\,(a - b - c_i) = && (115) \\ &= g(\Delta - c_i)\,(a - b - c_i) \begin{array}{l} \geq 0, \text{ if } c_i < a - b \\ \leq 0, \text{otherwise.} \end{array} && (116) \end{aligned}$$

Therefore, it is sufficient to verify the constraint (RC) for $c_i = \bar{c}$ and $c_i = \underline{c}$.

Let us start with the former. Recall that the incentive constraint $a - c_i + E_A\tau(c_i) \geq -m_1$ is satisfied as equality only for $c_i = \bar{c}$.

It implies that

$$\begin{aligned} I(\bar{c})\,(a - \bar{c} + E_A\tau(\bar{c})) &= -\left[\int_{\underline{c}}^{\bar{c}} [F(s) + sf(s)]\, \bar{I}(s)\, ds - b - (a-b)\, Q(\Delta)\right] && (117) \\ I(\bar{c})\, E_A\tau(\bar{c}) &= \bar{c} - a - M, \\ M &\equiv \int_{\underline{c}}^{\bar{c}} [F(s) + sf(s)]\, \bar{I}(s)\, ds - aQ(\Delta) - (1 - Q(\Delta))\, b. \end{aligned}$$

But $M \leq \bar{c} - a$. (This can be shown by employing the argument from proposition (12) demonstrating that $\int_{\underline{c}}^{\bar{c}} [F(s) + sf(s)]\, \bar{I}(s)\, ds$ increases with the first order stochastic dominance, and then showing (as in footnote (11) that under a limiting (Dirac) distribution with all the mass on the highest realization $\bar{c}$ the expression $\int_{\underline{c}}^{\bar{c}} [F(s) + sf(s) - a]\, \bar{I}(s)\, ds$ becomes $\bar{c} - a$, which completes the argument, as

$$M \leq \int_{\underline{c}}^{\bar{c}} [F(s) + sf(s) - a]\, \bar{I}(s)\, ds = \int_{\underline{c}}^{\bar{c}} [F(s) + sf(s)]\, \bar{I}(s)\, ds - aQ(\Delta).) \quad (118)$$

So $E_A\tau(\bar{c}) \geq 0 > -a$, as $I(\bar{c}) \geq 0$, the former being in compliance with (RC).

It is then left to verify that $E_A\tau(\underline{c}) \geq -a$. Note that it has to be the case that the net payoff of the lowest cost type is strictly positive in case of reform, i.e., $a - \underline{c} + \tau_A(\underline{c}) > 0$. Indeed, if it is not the case, then, provided this type gets the highest interim utility (which also implies in this case highest interim utility conditional on reform, as it is always feasible to set the tax in case of no reform, $E_B\tau(c_i)$, to be $-b$), a non-positive payoff $a - \underline{c} + E_A\tau(\underline{c})$ for the lowest cost type will imply that everybody else gets negative payoff from the reform, contradicting it being worthwhile in the first place (in the view of balanced budget). But if $a - \underline{c} + E_A\tau(\underline{c}) > 0$, then, clearly, (RC) constraint, $E_A\tau(\underline{c}) \geq -a$, is satisfied. ■

Proof of proposition 12. Let G be the distribution of the sum of $N-1$ valuations excluding i :

$$G(y) = \Pr\left\{\sum_{j \neq i} \theta_j \leq y\right\}. \quad (119)$$

Then $\bar{I}(s) = G(\Delta - s)$. Let G_F be the cumulative distribution of the sum of $N-1$ independent random variables (costs), where each variable is distributed F.Note that first stochastic dominance order is closed under convolution by theorem 1.A.3 in Shaked and Shanthikumar (1994) (similar is true for the peakedness order by theorem 2.C.3 in the same book.). This implies that if H FOSD F then G_H FOSD G_F.

Integrating by parts the first expression in (108), or recalling (106) it is possible to show that

$$\bar{m}_1 = \bar{c} - \frac{b + (a-b)\,Q(\Delta)}{G(\Delta - \bar{c})} + \frac{\int_{\underline{c}}^{\bar{c}} g(\Delta - s)\, sF(s)\, ds}{G(\Delta - \bar{c})} \quad (120)$$

Note that $\frac{Q(\Delta)}{G(\Delta - \bar{c})}$ is probability of the sum of the valuations to be below Δ conditional on the sum of the rest $N-1$ variables is below $\Delta - \bar{c}$. No doubt, it is unity.

Note that $G(\Delta - \bar{c})$ is non-increasing under both transformations.[17] Therefore, for both statements we can concentrate on the following part of the punishment bound,

$$m_F \equiv \int_{\underline{c}}^{\bar{c}} g_F(\Delta - s)\, sF(s)\, ds, \tag{121}$$

having to show that

$$m_H \geq m_F. \tag{122}$$

Assume H FOSD F (so that H describes a more expensive transition than F), so $F(t) \geq H(t)$, implying the first inequality below,

$$\begin{aligned} m_H &= \int_{\underline{c}}^{\bar{c}} H(t)\, g_H(\Delta - t)\, t dt = -\int_{t=\underline{c}}^{\bar{c}} H(t)\, t dG_H(\Delta - t) \geq & (123)\\ &\geq \int_{t=\underline{c}}^{\bar{c}} (-F(t)\, t)\, dG_H(\Delta - t) \geq \int_{t=\underline{c}}^{\bar{c}} (-F(t)\, t)\, dG_F(\Delta - t) = m_F, & (124) \end{aligned}$$

while the second inequality is due to closedness of this stochastic order under convolutions, which implies that G_H puts more weight on values with lower t and $-F(t)\, t$ decreases in t.

Now let us prove the second part of the proposition. Assume that F is more peaked than H, so that H is "more spread" than F.

Let us define a random variable $U = |X - d|$. Then if X is distributed with a distribution F on $[\underline{c}, \bar{c}]$ symmetric around $d = \frac{1}{2}(\underline{c} + \bar{c})$, U is distributed $Z_F(u) = 1 - 2F(d - u)$, $u = (d - x) \in [0, d - \underline{c}]$ with the corresponding marginal $z_F(u) = 2f(d - u)$.

By symmetry, letting

$$\begin{aligned} \delta &\equiv \Delta - d, & (125)\\ u &\equiv |t - d|, & (126) \end{aligned}$$

[17] It is an implication of (1) closedness of both orders under convolution and (2) the definition of the corresponding order in each case.

we have

$$
\begin{aligned}
m_F &= \int_0^k (d-u)\, F(d-u)\, g_F(\delta+u)\, du + \qquad (127)\\
&\quad + \int_0^k (d+u)\,(1-F(d-u))\, g_F(\delta-u)\, du\\
&= \chi_F + \omega_F, \qquad (128)\\
\chi_F &\equiv \int_0^k d g_F(\delta-u)\, du \qquad (129)\\
\omega_F &\equiv \int_0^k u\,(1-2F(d-u))\, g_F(\delta-u)\, du, \qquad (130)
\end{aligned}
$$

where $k \equiv d - \underline{c} = \frac{1}{4}(\underline{c} + \bar{c}) = \bar{c} - d$.

As F is more peaked than H, Z_H FOSD Z_F, and the order "more peaked" is closed under convolution for unimodal (symmetric) distributions with the same mean,[18] we have

$$\chi_F = 2\int_0^k d g_{Z_F}(u)\, du \leq 2\int_0^k d g_{Z_H}(u)\, du = \chi_H, \qquad (131)$$

where $g_{Z_F}(u)$ is the marginal distribution of $N-1$ independent random variables distributed Z_F, as before.

As for the second term,

$$\omega_F = 2\int_0^k u Z_F(u)\, g_{Z_F}(u)\, du \leq 2\int_0^k u Z_H(u)\, g_{Z_H}(u)\, du = \omega_H, \qquad (132)$$

where the inequality is due to the first part of the proposition. ■

References

Angeletos, G.-M., C. Hellwig, and A. Pavan (2003). Coordination and Policy Traps. NBER Working Paper No. w9767.

Binmore, K. (1998). *Game Theory and the Social Contract*, Volume 2: Just Playing. MIT.

[18]see Shaked and Shanthikumar (1994).

Crémer, J. and R. McLean (1985). Optimal Selling Strategies Under Uncertainty for a Discriminating Monopolist When Demands Are Interdependent. *Econometrica 53*, 345–361.

d'Aspremont, C. and L.-A. Gérard-Varet (1979). Incentives and Incomplete Information. *Journal of Public Economics 11*, 25–45.

Fang, H. and P. Norman (2003). An Efficiency Rationale for Bundling of Public Goods. mimeo.

Fernandez, R. and D. Rodrik (1991). Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty. *The American Economic Review 81*, 1146 – 1155.

Fudenberg, D. and J. Tirole (1996). *Game Theory* (5 ed.). Cambridge and London: The MIT Press.

Harberger, A. C. (1993). Secrets of Success: A Handful of Heroes. *The American Economic Review 83*, 343 – 350. Papers and Proceedings of the Hundred and Fifth Annual Meeting of the American Economic Association.

Jackson, M. O. and H. Sonnenschein (2003). Overcoming Incentive Constraints by Linking Decisions. mimeo.

Jain, S. and S. W. Mukand (2003). Redistributive promises and the adoption of economic reform. *The American Economic Review 93*, 256 – 264.

Janeba, E. (2003). International Trade and Cultural Identity. mimeo.

Lang, C. and S. Weber (2000). Ten Years of Economic Reforms in Russia: Windows in a Wall. In R. M. Lastra (Ed.), *The Reform of the International Financial Architecture*, pp. 413 – 430. London: Kluwer Law International.

Ledyard, J. O. and T. R. Palfrey (2003, November). A General Characterization of Interim Efficient Mechanisms for Independent Linear Environments. Caltech Working Paper 1186.

Mailath, G. J. and A. Postlewaite (1990). Asymmetric information bargaining problems with many agents. *Review of Economic Studies 57*, 351–367.

McAfee, R. P. and P. J. Reny (1992). Correlated Information and Mechanism Design. *Econometrica 60*, 395–421.

Mirrlees, J. (1971). An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies 38*, 175–208.

Morris, S. (1995). Co-operation and Timing. CARESS Working Paper 95-05, University of Pennsylvania.

Morris, S. and H. S. Shin (1998). Unique equilibrium in a model of self-fulfilling attacks. *The American Economic Review 88*, 587 – 597.

Myerson, R. and M. A. Satterthwaite (1983). Efficient Mechanisms for Bilateral Trading. *Journal of Economic Theory 29*, 265–281.

Myles, G. D. (1995). *Public Economics*. Cambridge, New York, Melbourne: Cambridge University Press.

Rob, R. (1989). Pollution Claim Settlements under Private Information. *Journal of Economic Theory 47*, 307–333.

Rodrik, D. (1996). Understanding Economic Policy Reform. *Journal of Economic Literature 34*, 9 – 41.

Shaked, M. and J. G. Shanthikumar (1994). *Stochastic Orders and Their Applications*. Boston, San Diego, New York: Academic Press.

Velasco, A. and M. Tommasi (1996, April). Where Are We in the Political Economy of Reform? *Journal of Policy Reform*.