Coalition-proof incentive contracts*

Alexei Savvateev†

January 2006

Abstract

The paper presents sufficient conditions for group strategy-proofness in mechanism design and multi-agency problems. These conditions are *Spence-Mirrlees single crossing property* on the agents' preferences profiles and *order semi-invariance* and *payoff complementarity* of proposed incentive schemes. Applications include serial cost sharing of non-convex public goods, as well as scenarios of a principal with multiple agents paradigm under perfect monitoring and costly enforcement. The examples provided are those of tax evasion and pollution control problems.

Keywords: coalition-proofness, strong equilibrium, incentive contracts, multistep strategies, serial cost sharing, Spence-Mirrlees conditions.

JEL Classification Numbers: C62, C71, D78, H26.

---

*The author expresses his deepest gratitudes to three persons for their continuing interest, ideas, discussions and fruitful critique. The first is his supervisor Victor Polterovich, and other two are Isabelle Brocas and Shlomo Weber, Also, he is happy to thank Akhmed Akhmedov, Alexander Friedman, Yurii Gavrilets, Serguei Kokovin, Gregory Kosenok, Valerii Marakulin, Vladimir Matveenko, Herve Moulin, Alexander Shen, Konstantin Sonin, Alexander Tonis, Alexander Vasin, Evgenii Zhelobod'ko, and the participants of seminars at CEMI (Moscow), CORE (Louvain-La- Neuve, Belgium) and IM SB RAS (Novosibirsk), for their comments and suggestions.

†Central Economics and Mathematics Institute, Moscow, and Catholic University of Louvain, Louvain-La-Neuve, Belgium. Financial support through grants R98-0631 from the Economic Education and Research Consortium, # NSh-1939.2003.6 School Support, Russian Foundation for Basic Research No. 04-02-17227, and Russian Science Support Foundation is gratefully acknowledged. The paper was partly written while the author was a visiting fellow in Sobolev Institute for Mathemitics, Novosibirsk.

1 Introduction

Consider the problem where several agents simultaneously determine their demand for some good. After demands are satisfied, the agents pay for this good according to a general "price mechanism" where every agent is charged with a payment which depends on all agents' demands, including his/her own. There could be certain exogenous restrictions on available price mechanisms that may vary for different applications. However, there is a common feature for all applications, that, after a price mechanism is announced, agents play a simultaneous-move game by choosing their demands. The question is the nature of the solution concept under which a price mechanism is implemented and in this paper, we study mechanisms which implement the most demanding solution concept of a *strong equilibrium* (see Aumann 1959).

Applications of this setting are divided between two basic frameworks. One is the mechanism design, where the only goal is to implement stable incentive mechanisms, which, in our case, are immune against coalitional deviations by agents. The formal interpretation here is that the outcome of implementation would be a strong equilibrium. The second is the Principal-agent framework with many agents, where contracts that could be offered to agents are compared with each other, according to the Principal's objectives. Nevertheless, even in this case the relevant issue is the requirements for stability of the outcome implemented by the contract. Hence, in certain circumstances, i.e., when the agents coordinate their actions, the Principal may offer contracts that yield the strong equilibrium.

In this paper we present sufficient conditions for the strong equilibrium implementation in the general problem formulated above. These conditions include, first, Spence-Mirrlees single crossing property of agents' preferences profile; this means that agents are ordered with respect to their rate of substitution between the good and money. In fact, we need only a weaker version of Spence-Mirrlees conditions, introduced in (Milgrom, Shannon 1994) and called *single-crossing condition* there. Secondly, a price mechanism should satisfy *payoff complementarity* and *order semi-invariance*. The former means that the payment of any agent is (non-strictly) decreasing in the other agents' demands. The latter requires that the payment of an agent does not depend on demands that exceed her own (see Moulin et al. for the characterization of the serial cost sharing method).

We show that, if a price mechanism satisfies these two conditions and the Spence-Mirrlees property holds for agents' preferences profile, then the corresponding game admits a strong equilibrium. Moreover, it turns out that there is an iterative procedure of finding a strong equilibrium, which is a maximal point in the set of all Nash equilibria that is preferable to any other Nash equilibrium *for all agents*.

In the literature on mechanism design, the issue of coalitional stability has been discussed for cost and revenue sharing problems. Namely, (Moulin, Shenker 1992) and (Moulin 1994) proved that the so-called *serial formula* for sharing common costs results in a strong equilibrium as the outcome of the simultaneous choices of demands. But their assumptions are slightly different: the cost function and preferences of agents were convex, but otherwise arbitrary without any Spence-Mirrlees conditions. In our set-up the serial formula for public good still produces strong equilibrium in non-convex environments, provided a profile of agents' preferences satisfies the Spence-Mirrlees conditions. In particular, this covers the case of a *discrete public good* with a finite number of feasible levels.

If we turn to the Principal-agent framework, the matters are quite different. As stated by (Mookherjee 1984) who studied incentive contracts with many agents, *"The incentive schemes optimal for Nash equilibrium implementation < .... > may suffer from vulnerability to collusion among agents. < .... > No satisfactory theory seems available < .... > and future research in this area will be valuable".* After 20 years I am still unaware of results on characterization of coalition-proof incentive contracts. In order to explain this lacuna, let us resort to the classical Principal-agent paradigm. Usually, there is an imperfect correlation between unobservable action and observable, hence, contractible output (Mookherjee 1984). It has been demonstrated that if the actions were observable and there were no exogenous restrictions on the mechanism, then the problem of strong equilibrium implementation were *inessential*. Namely, in this case the first-best outcome is implemented via *individual peace-rate contracts*, which are tautologically coalitionally stable as their implementation involves no interaction among the agents, at all.

At the same time, if actions are unobservable and the output depends on a realization of a random variable as well as on the action undertaken, a quest for coalition-proof incentive contracts seems to be fairly difficult. Even the Nash equilibrium implementation is challenging as demonstrated by Groves 1973; Holmstrom 1982; Mookherjee 1984; Malcomsom 1986; Ma 1988. The case where actions are observable, is obvious due to the observation made in (Mookherjee 1984).

However, this observation crucially depends on the absence of any restrictions placed on incentive contracts as with these restrictions, say a sort of a budget constraint, the first-best outcome were no longer feasible. Among numerous environments in which actions are observable but budget is limited, we focus on two applications: tax evasion auditing and pollution control. We show that under a budget constraint, a question of strong equilibrium implementation becomes an essential one — piece-rate contracts become inferior even in the class of coalition-proof incentive schemes. At the same time, Nash equilibrium implementation is still trivial even under budget constraint, as in this case the first-best outcome is achieved by the first rank-order contract when only agent(s) with the highest demand should pay.

We generalize tax evasion and pollution frameworks into so called *n-inspection problem* (I owe this term to V. Polterovich): agents choose levels of *cheating*, the Principal observes the agents' choices, but he cannot charge infinite prices for cheating; the sum of all payment is bounded from above. Then, we study a certain class of implementation schemes, where by considering cheating as a (generalized) good, we offer the Principal to fix a certain cost function for this good, as in (Fridman, Moulin 1999): this cost function may depend on the individual demands in an arbitrary way. Then, agents divide "costs" according to the serial formula. If the good of cheating is a *discrete public good*, the corresponding incentive contract is called a *multistep incentive scheme*. Main theorem is then applied to guarantee strong equilibrium implementation by any multistep scheme.

Note that throughout this paper we do not show that the class we examine contains the optimal contract regardless of the Principal objectives. We also do not compare alternative multistep incentive schemes with each other and optimality is resolved separately for each of possible applications of the general n-inspection problem. All these issues are left for the future research.

The rest of the paper is organized as follows. Section 2 provides the statement of the main theorem, whose proof is relegated to the Appendix. Section 3 applies the main theorem to cost sharing problems. Section 4 introduces the so-called *n-inspection problem*, a basic framework for the applications of the main result to the Principal — many agents paradigm. Also, the examples of the tax evasion auditing and pollution control are considered. A numerical example is presented which demonstrates essentiality of the problem in our context. Section 5 establishes important links between cost sharing and n-inspection problems, which result in a certain class of coalition-proof incentive contracts, called *multistep strategies*. Conclusion is a quest for further research in the field. Appendix contains the proof of the main result.

2 Main theoretical result

Consider the following "equal-treatment" mechanism design problem. There are n agents indexed by $i = 1, \ldots, n$ who simultaneously choose levels of some "general good", $0 \leq z_i \leq 1$ (or from a given compact, maybe finite subset $Z \subset \mathbf{R}$; in the latter case, let $1 \in Z$ be a maximal element, without loss of generality). Then, each agent i pays for his demand, according to a price correspondence,

$$p\left(z_i; \{z_j\}_{j \neq i}\right), \tag{1}$$

which non-decreases in his own demand, and also depends on the unordered collection of other agents' demands. In different applications, there could be alternative restrictions on the feasible price mechanism, $p\left(\cdot, \cdot\right)$; we do not consider this issue here, aiming at common insights into all possible applications.

Agents have preference profiles $\succeq_i$ over a set of pairs $\{(z,p)\}_{z\in Z, p\in\mathbf{R}}$; preferences increase in z and decrease in p. For instance, we could have assumed that each agent has a utility representation, $U_i(z,p)$, though all the assertions made below are of an ordinal nature.

The aim is to design coalition-proof price mechanisms, in the sense that the strategic outcome of the game the agents play by choosing their demands for good, is a em strong equilibrium. Let me give a definition in purely ordinal terms (see Aumann 1959 for a standard treatment):

Definition 1: Consider a game in normal form with the set N of players, strategy sets $\{Z_i\}_{i\in N}$ and preference profiles $\{\preceq_i\}_{i\in N}$ defined over the Cartesian product $\times_{i\in N} Z_i$.

A profile $z^* = \{z_i^*\}_{i\in N} \in \times_{i\in N} Z_i$ of strategies is called a *strong equilibrium* if there exists no coalition $S \subset N$ of players, together with a collection of alternative strategies $\{\tilde{z}_i\}_{i\in S}$ for members of S, such that for every member $i \in S$ we have

$$\{\{\tilde{z}_i\}_{i\in S}; \{z_i^*\}_{i\notin S}\} \succ_i z^*. \tag{2}$$

This is much stronger requirement than that of being Nash equilibrium, for in the Definition 2, not only all the individuals are allowed to deviate from z^*, but also arbitrary groups of individuals. The notion of a strong equilibrium gives a formal representation of coalition-proofness. The need in coalition-proofness arises each time when the agents coordinate their actions and there is a threat of collusion.

Throughout this paper, we will assume that the profile of agents' preference relations satisfies the *Spence-Mirrlees single crossing condition.* Roughly speaking, this is a requirement that agents with higher index have higher rate of substitution between the good and money (i.e. stronger preference towards a good). Formally, let us introduce lexicographic order on the set of pairs $\{(z,p)\}$:

$$(z,p) \succ_L (z',p') \;\Leftrightarrow\; z > z', \text{ or } z = z' \text{ and } p > p'. \tag{3}$$

Definition 2: We say that the profile of agents' preference relations satisfies the *Spence-Mirrlees* *single crossing condition* if

$$\forall j > i \, \forall (z,p) \succ_L (z',p')$$

$$(z,p) \succ_i (z',p') \Rightarrow (z,p) \succ_j (z',p'); \tag{4}$$

$$(z,p) \succeq_i (z',p') \Rightarrow (z,p) \succeq_j (z',p').$$

(Milgrom, Shannon 1994) used this restriction to characterize monotone comparative statics of a certain variable f with respect to parameter t; here, we placed the same restriction on the parameterized family of preference profiles, $\{\preceq_i\}$, with respect to the agents' index, i. Like in (Milgrom, Shannon 1994), it is easy to demonstrate that if the agents' preferences are represented by continuously differentiable utility functions, $U_i(z,p)$, then to satisfy Spence-Mirrlees single crossing condition, the parametrized family of utility functions, $\{U_i\}$ should satisfy the following property (which is in fact the original form used by Spence and Mirrlees in signaling and taxation models):

$$\left| \frac{dU_i/dz}{dU_i/dp} \right| \tag{5}$$

is non-decreasing in i for all pairs (z,p).

Our approach is purely ordinal; however, it is sometimes easier to think in terms of utility representations, in which case the last requirement is much more intuitive; it roughly says that indifference curves of different agents cross at most once. In what follows, we maintain the assumption that Spence-Mirrlees conditions are satisfied on the agents' part.

Now, we are ready to state the main theorem. Below, I give the two conditions which are sufficient for price mechanism to be implementable in strong equilibrium.

Theorem 1. Consider a mechanism design problem formulated above, where the profile of agents' preferences satisfies the Spence-Mirrlees single crossing condition. If the price mechanism $p\left(z_i; \{z_j\}_{j\neq i}\right)$ possesses the next two properties

- *order semi-independence:*

$$p\left(z_i; \{z_j\}_{j\neq i}\right) = p\left(z_i; \{\max\{z_i, z_j\}\}_{j\neq i}\right). \tag{6}$$

- *payoff complementarity*:

$$p\left(z_i; \{z_j\}_{j\neq i}\right) \quad \text{is non-increasing in} \quad z_j,\, j \neq i, \tag{7}$$

then there exists a strong equilibrium $z^* = (z_1^*, \ldots, z_n^*)$ in a game the agents play by choosing their demands, z_i; moreover, this equilibrium is

(a) *monotonic in* i, i.e. $z_1^* \leq \cdots \leq z_n^*$;

(b) *a maximum point* of the set of all Nash equilibria:

$$\forall i \quad z_i^* \geq \tilde{z}_i,$$

if $\tilde{z} = (\tilde{z}_1, \ldots, \tilde{z}_n)$ — another Nash equilibrium;

(c) (non-strictly) better than any other Nash equilibrium *for all* agents:

$$\forall i \quad (z_i^*, p_i^*) \succeq_i (\tilde{z}_i, \tilde{p}_i)$$

(d) can be approached by no more than n steps by a simple iterative procedure (see the Appendix).

Here, $p_i^* = p\left(z_i^*; \{z_j^*\}_{j\neq i}\right)$ and $\tilde{p}_i = p\left(\tilde{z}_i; \{\tilde{z}_j\}_{j\neq i}\right)$.

Proof is delegated to the Appendix. I just stress here that the Spence-Mirrlees conditions are quite popular in the mechanism design, and the fact that they also are responsive for strong equilibrium gives them an additional support. As for the two other conditions placed on the price mechanism, the first ("payoff complementarity") is immanent to the Bertrand oligopoly and is a counterpart of the famous strategic complementarity of (Milgrom and Roberts 1990), and the order semi-independence is adopted from serial cost sharing (Moulin 1994; Moulin and Shenker 1992). The combination of these by no means newly introduced requirements gives the desired coalitionary stable implementation mechanisms.

Now, armed with this theoretical result, we turn to its applications. We start with applications to cost sharing problems, which not only are interesting by itself, but also will have direct relation to observations made later within the Principal and multiple agents framework.

3 Application to serial cost sharing: an overview

First, let me remind some basic facts about serial cost sharing. In (Moulin and Shenker 1992), the authors offered a way to share costs of joint production of a normal good q characterized by a production function $c(q)$. They offered the next *serial formula*: if we place demands for goods of n agents in increasing order: $q_1 \leq \cdots \leq q_n$, then the agents share the overall cost

$$c(q_1 + \cdots + q_n) \tag{8}$$

of production in the following way (m_j denotes a contribution of agent j towards the cost of production):

$$\begin{gathered} m_j = \frac{c(nq_1)}{n} + \frac{c\left(q_1 + (n-1)q_2\right) - c(nq_1)}{n-1} + \cdots + \\ \frac{c\left(q_1 + (n-1)q_2 + \cdots + (n-j+1)q_j\right) - c\left(q_1 + (n-1)q_2 + \cdots + (n-j)q_{j-1}\right)}{n-j+1}. \end{gathered} \tag{9}$$

Authors proved that if agents' preferences $\succeq_i$ on the set $\mathbf{R}^2_+$ of pairs (q, m), as well as the cost function $c(\cdot)$ are *convex*, then this mechanism is coalition strategy-proof, both in the sense of truth-telling, and in that its strategic outcome is a strong equilibrium. Serial formula is designed in such a way that the payment charged on a given agent is independent of demands for this good which exceed his own demand, which is just an order semi-independence property introduced above.

Later on, (Moulin 1994) modified this serial formula to the case of an excludable public good q, for which, if demands are $q_1, \ldots, q_n$, then the overall cost needed to satisfy these demands is just $c\left(\max\limits_{i=1}^{n} q_i\right)$. Moulin suggested the following generalization of (9) to the case of an excludable public good (here, demands are again ordered in a way such that $q_1 \leq \cdots \leq q_n$, and hence, to satisfy demands of the first i agents, one needs to cover cost of $c(q_i)$):

$$m_i = \frac{c(q_1)}{n} + \frac{c(q_2) - c(q_1)}{n-1} + \cdots + \frac{c(q_i) - c(q_{i-1})}{n-j+1}. \tag{10}$$

This is done, again, in order to maintain the *order semi-independence* principle, and in fact this method for sharing costs of a public good was originally suggested in (Littlechild and Owen 1973) in the so-called "airport game": when several companies share the same landing road, a *demand* of a company is a landing distance of its company's crafts. The idea has originated even earlier, since in fact the way of sharing costs explored in (10) is just the Shapley value of the corresponding cooperative game, see (Shapley 1953).

(Littlechild and Owen 1973) were not interested in the strategic properties of the proposed mechanism; they just demonstrated that the game they considered is supermodular, hence, Shapley value is at the core, which justifies its use. (Moulin 1994) gives an additional justification for the serial formula (which coincides with the Shapley value of the cooperative game which follows a choice of demands, $q_1, \ldots, q_n$), on the grounds of group strategy-proofness. Namely, Moulin showed that again, if agents' preferences $\succeq_i$ on the set $\mathbf{R}_+^2$ of pairs (q, m) are convex, while production function $c(\cdot)$ is concave, then the serial mechanism's strategic outcome is necessarily a strong equilibrium. This time, both assumptions placed on the price mechanism in Theorem 1 are satisfied; however, instead of assuming Spence-Mirrlees conditions on the agents' part, (Moulin 1994) kept the convexity of all the ingredients into the model, just as in (Moulin and Shenker 1992).

While convexity of preference relations seems to be rather innocent assumption, we cannot firmly say the same about the cost function. In particular, the assumption of convexity eliminates a possibility of a public good to be *discrete*, i.e. available only in finite number of levels, which is often the case in reality.

Using Theorem 1, it can be easily demonstrated that group strategy-proofness still holds for non-convex environments, *provided* a profile of agents' preferences satisfies Spence-Mirrlees condition. More generally, for an arbitrary public good (with convex or non-convex costs), formula (10) is being implemented under strong equilibrium, if the Spence-Mirleez conditions hold on the consumption part. Summing this up, we state the following result.

Theorem 2. Consider a cost sharing problem with an arbitrary non-decreasing cost function $c(q)$, and n agents indexed by $i = 1, \ldots, n$ whose preference relations are defined over the set of pairs $\{(q, m)\}_{(q,m)\in\mathbf{R}_+{}^2}$. If the profile of agents' preferences satisfies the Spence-Mirrlees single crossing condition, then the serial formula (10) implements strong equilibrium in the game the agents play by shoosing levels of a public good they demand.

Proof is a direct consequence of Theorem 1 (compare to Moulin 1994).

Now, let us turn to the Principal — many agents paradigm, and apply the main theorem in order to design coalition-proof incentive contracts. In what follows, we adopt the formula (10) to introduce a class of coalition-proof incentive schemes named *multistep strategies*.

4 n-inspection problem

In this section, we consider the model specification for the study of coalition-proof incentive contracts. We will simplify the general Principal — many agents framework to that in which actions are observable (hence, contractible), in order to concentrate on strategic properties of incentive contracts. Also, we will *not* dwell upon finding the optimal contract, among collusion-proof ones. However, I will show below that individual, *piece-rate* contracts which reduce to the combination of n independent Principal — agent problems could be *inferior* in the whole class of coalition-proof contracts.

In the introduction, we have discussed the relevance of strong equilibrium implementation for multi-agency problems; the observation due to (Mookherjee 1984) is that, in absence of any restrictions on the incentive contract, first-best is being implemented via piece-rate contracts, hence, there is no case for introducing more sophisticated schemes. However, as will be argued below, there are some situations in which observability comes hand-in-hand with restrictions on contracts being used by the Principal. These very situations are those in which Theorem 1 could help by suggesting a class of coalition-proof incentive contracts. Here are our two main examples.

Example 1. Tax evasion and auditing (Vasin 2005):

$$u_i(z, p) = b_i(z) - pz, \tag{11}$$

where $b_i(z) := b(i, z)$ is increasing, and increase in i with its derivative. Here, z is a percentage of hidden transactions (as in Vasin 2005).

Example 2. Pollution control (thanks to H.Moulin). Here, I assume that there is a

$$u_i(z,p) = b_i(z) - p \cdot \varphi(z), \tag{12}$$

where $\varphi(\cdot)$ is a penalty when the pollution index is z. Mathematically, this is a generalization of the previous specification.

(formally, a generalization of the previous example)

Tax evasion; there are n agents, with benefit functions

$$\frac{z_1}{n}; \frac{z_2}{n-1}; \dots \frac{z_{n-1}}{2}; z_n \tag{13}$$

To enforce $z_i \equiv 0$ under peace-rate contract, one needs

$$P = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + 1; \tag{14}$$

To enforce $z_i \equiv 0$ under strong equilibrium, one needs $P = 1 + \varepsilon$! Incentive contract is *to check with equal probability all* $z_i > 0$, and in the next section it will be shown that this contract implements no-cheating as a strong equilibrium.

Both examples illustrate the Principal — many agents framework, with *effort* being replaced by *cheating level*, while *award* replaced by *punishment*. This language will be maintained throughout the rest of the paper.[1] Namely, we introduce the following *n-inspection problem* as the basic framework of our analysis.

There is a principal who hires n agents (indexed by $i = 1, \dots, n$) for some task. Agents could cheat to some degree $z \in [0,1]$,[2], and the fact of cheating (and its degree, as well) will be revealed by the principal. However, only limited number of inspections could be organized,[3] due to lack of

[1]There are essential reasons for using the "crime" —punishment paradigm, instead of effort — award one. Those reasons come not only from similarity with cost sharing agenda which is crucial in what follows, but also in strategic aspects of the game the agents play in the former versus the latter paradigm of the Principal — multiple agents model. But this issue stands outside the scope of the current paper.

[2]Nothing changes if we replace $[0,1]$ by an arbitrary compact subset $Z \in \mathbf{R}$.

[3]Less than the number of agents, otherwise the problem becomes trivial.

time, and/or to the complexity of the punishment procedure. One could treat punishment as a subtraction of the corresponding share of a salary. The question is to reduce cheating by designing (and committing to) a coalitionary stable *punishment scheme*, or alternatively an *incentive contract* which attains a probability distribution of being inspected to every possible profile of cheating parameters.

Formally, this contract is a mapping from the set $[0,1]^n$ of all the possible cheating profiles to the set Δ_n of all the generalized probability distributions

$$\Delta_n = \{(\lambda_1, \dots, \lambda_n)\} \mid \sum_{i=1}^{n} \lambda_i \le P \tag{15}$$

over the n points (i.e. agents):

$$\lambda : \quad [0,1]^n \to \Delta_n. \tag{16}$$

The set of all such mappings reflects the principal's strategic opportunities.

However, according to the *equal-treatment requirement*, we confine ourselves to incentive contracts which are unanimous, in the sense that there exists a common *punishment rule* $p(\cdot,\cdot)$, such that, $\forall i = 1, \dots, n$ we have

$$p_i = p\left(z_i; \{z_j\}_{j \ne i}\right), \tag{17}$$

where $\lambda = (p_1, \dots, p_n)$. The Principal's constraint on incentive contracts used now could be formulated as follows: *For any profile* $z = (z_1, \dots, z_n)$ we must observe that

$$\sum_{i=1}^{n} p\left(z_i; \{z_j\}_{j \ne i}\right) \le P. \tag{18}$$

Agents are assumed to be risk-neutral with respect to penalties they pay, so we can assume that there is only one parameter of punishment, namely, *expected loss* from being penalized. Another parameter of the agent's payoff is his cheating level. We assume that agents' preference relations over the set of pairs $\{(z,p)\}$ are as described in the conditions of Theorem 1, particularly we assume that the preferences' profile satisfies Spence-Mirrlees property, as defined in section 2. For simplicity only, let there be utility representations $U_i(z,p)$ of the agents' preference relations.

Being informed of the contract used by the Principal, agents simultaneously choose their parameters of cheating $z_i \in [0,1]$. Then, payoffs are realized, and the Principal faces the profile $z = (z_1, \ldots, z_n) \in [0,1]^n$ (negatively) reflecting the efficiency of the task fulfilled by the agents. The latter ones, in turn, get their payoffs which are equal to

$$u_i(z) = u_i(z_1, \ldots, z_n) = U_i(z_i, p_i) = U_i\left(z_i, p\left(z_i; \{z_j\}_{j \neq i}\right)\right). \tag{19}$$

Notice that, under our assumption of perfect observability, even with the restriction (18), the Principal can implement first-best outcome (i.e. no cheating equilibrium $z_i \equiv 0$) as *Nash* equilibrium. For that matters, he has to commit himself to a simple *first-rank contract* (see Malcomson 1986):

$$p_i = p\left(z_i; \{z_j\}_{j \neq i}\right) = 0, \tag{20}$$

whenever $z_i \neq \max\{z_j | j = 1, \ldots, n\}$. (i.e. by inspecting *only* those whose cheating levels are maximal, with equal probabilities).[4] *But this contract is prone to coalitional deviations!*

In the next section, we use one technique, in order to introduce a sufficiently broad class of coalition-proof incentive contracts.

5 Virtual cost functions and multistep strategies

If one looks at the formulations in sections 3 and 4, he probably finds out that there is much in common between cost sharing problems and the Principal — many agents paradigm (as it appears in n-inspection problem). Indeed, in both cases, there are n agents, with perfectly identical preferences, in the first case over the set of pairs (public good, payment), while in the second story it is the set of pairs (cheating level, expected penalty). Further on, in both formalizations it is price mechanisms which is in the focus of our study, and the question is the same: Which price mechanisms are coalition-proof?

[4] This contract implements no-cheating as Nash equilibrium under quite general conditions; reasoning here is identical to that in the standard Bertrand oligopoly.

However, there is a difference between these two research programs. Namely, restrictions on the set of available mechanisms are different. In the first case, restrictions are obvious: $\forall z \in Z^n$ we must have the *system of inequalities*

$$\sum_{i=1}^{n} p\left(z_i; \{z_j\}_{j \neq i}\right) \leq c\left(\max_{i \in N} z_i\right), \tag{21}$$

whereas in the second case, there is *only one inequality* (18).

At the same time, in the first case, coalitionary-proof mechanisms were studied to a large extent, whereas in the second case no theory seems to exist, by now.

How can we make use of these observations? The answer is, rather, straightforward: we suggest that the Principal chooses the functional form for a *virtual cost function* (this term is suggested by S. Weber) of producing a good of *cheating*. We will restrict ourselves with the case when cheating is a "public good", because the serial formula looks most simply in this case; however, this is not the only possible choice.

Let us now recall Theorem 2 (from section 3). It says, particularly, that the serial formula (9) for sharing cost of a discrete public good implements the outcome in strong equilibrium. Replace $c(z)$ with a discrete-valued function, such that there is a finite number of levels $\bar{z}_l$, $l = 1, \ldots, k$ with "incremental costs" A_l, $l = 1, \ldots, k$:

$$c(\bar{z}_l) = \sum_{h=1}^{l} A_h. \tag{22}$$

The Principal's constraint requires that $c(1) \leq P$, or $\sum\limits_{h=1}^{k} A_h \leq P$.

Rewriting the formula (9) using (22), we get the following *multistep incentive contract* parametrized by a family of parameters $(k; \{\bar{z}_1, \ldots, \bar{z}_k\}; \{A_1, \ldots, A_k\})$.

Definition 3: Multistep incentive scheme, or *strategy*, consists of several threshold levels

$$0 \leq \bar{z}_1 < \cdots < \bar{z}_k < 1, \tag{23}$$

and probabilities

$$(A_1, \ldots, A_k) \tag{24}$$

attached to these thresholds.

Implementing this strategy means that the penalty structure looks as follows:

$$p\left(z_i; \{z_j\}_{j\neq i}\right) = \sum_{\{l:\bar{z}_l < z_i\}} \frac{A_l}{\#\{j : \bar{z}_l < z_j\}}. \tag{25}$$

That is, agents with $z > \bar{z}_l$ divide together costs A_l for increasing the cheating degree above the threshold z_l.

Next theorem is the obvious corollary of Theorem 2.

Theorem 3: For any multistep strategy (23), (24), the game the agents play in the n-inspection problem by choosing levels of cheating results in strong equilibrium implementation.

We are not going to analyse the optimal multistep strategies. However, no matter which objective functional is being used by the Principal, there is a question concerning the Pareto-efficiency of a given strategy: whether it is the least costly way to implement a given profile, z under strong equilibrium. Next theorem answers to this question.

Theorem 4: For every strong equilibrium outcome z^* of a multistep strategy (23), (24), the following n-stepped strategy $(z_1^*, \ldots, z_n^*; \bar{A}_1, \ldots, \bar{A}_n)$ implements a profile $\tilde{z}^* \preceq z^*$ (possibly, it coincides with z^*):

$$\bar{A}_i = \sum_{\bar{z}_l \in [z_i^*, z_{i+1}^*)} A_l, \tag{26}$$

Proof is skipped, as it is straightforward.

Corollary. In characterizing the optimal multistep strategy for principal with n agents, it is sufficient to consider the $(2n-1)$-dimensional space of n-stepped strategies.

Next, concluding, section presents some clarifications, extensions etc.

6 Conclusions, future prospects and challenges

In the current paper, sufficient conditions for coalition-proof mechanism design were given, which incorporate payoff complementarity and order semi-independence into the Spence-Mirrlees environment.

Then, these conditions were applied to serial cost sharing; precisely it was shown that for an arbitrary cost function of public good production, the serial formula from (Moulin 1994) results in strong equilibrium among agents. New insight is that a cost function need not be convex; for instance, it could be discrete, with a finite number of levels available for production.

Next step was to apply this result in the Principal — many agents framework with observable actions and costly monitoring. Namely, the Principal chooses the “cost function” for producing cheating, and then prescribes the agents to share costs in the serial manner. Resulting punishment schemes were called *multistep strategies*, and every such strategy implements the outcome under strong equilibrium.

Finally, it was demonstrated that, whenever the objective functional of the Principal could be, it is always sufficient to look for the optimal multistep contract within the finite-dimentional subspace of multistep incentive contracts.

Here are some challenges for future research.

1. Do multistep strategies expire all the incentive contracts with the strong equilibrium implementation condition?

2. How to compare welfare properties of alternative multistep strategies? Which multistep strategy is the best one? This is just to solve the problem explicitly for alternative objectives of the Principal, and is the continuation of the previous puzzle.

3. What are the conditions sufficient for the existing of a coalition-proof incentive contract implementing a zero-cheating equilibrium? This question is particularly important, since it does not depend on the objective functional.

4. What if we incorporate penalties into the objective functional of the principal, as it is assumed in a number of tax-evasion models? Do our results change?

5. Finally, there is an open quest for extending the theory on the standard case with unobservable actions and observable output.

7 Appendix

Here, I present a proof for Theorem 1 (the main theoretical contribution of the paper). Before starting it, I need the following lemma which deserves attention on its own.

Lemma A.1. Consider a game Γ in a normal form with N players, strategy sets $Z_1, \dots, Z_n$ and utility functions $u_i(z_1, \dots, z_n)$. Let $Z = Z_1 \times \dots \times Z_n$.

Assume that a profile $z^* = (z_1^*, \dots, z_n^*) \in Z$ satisfies the following system of inequalities:

$$\begin{aligned}
u_1(z_1^*, \dots, z_n^*) &= \max_{z \in Z} u_1(z); \\
u_2(z_1^*, \dots, z_n^*) &= \max_{(z_2,\dots,z_n) \in Z_2 \times \dots \times Z_n} u_2(z_1^*, z_2, \dots, z_n); \\
&\dots \\
u_i(z_1^*, \dots, z_n^*) &= \max_{(z_i,\dots,z_n) \in Z_i \times \dots \times Z_n} u_i(z_1^*, \dots, z_{i-1}^*, z_i, \dots, z_n); \\
&\dots \\
u_n(z_1^*, \dots, z_n^*) &= \max_{z_n \in Z_n} u_n(z_1^*, \dots, z_{n-1}^*, z_n).
\end{aligned} \tag{27}$$

Then, z^* is a strong equilibrium.

Proof of Lemma 1 is straightforward. Indeed, consider a coalition S, and assume it deviates by choosing a bundle $\{\tilde{z}_j\}_{j \in S}$. Find $\tilde{j} = \min\{j | \, j \in S$. Then, $\tilde{j}$ is not better off after the deviation, by (27) — a contradiction.

Proof of Theorem 1 now proceeds in two steps. Firstly, I run a certain iteratice process which will result in monotone $z^* \in Z^n$. Second, I look at this iterative process in details and show that this is the strong equilibrium.

First step in details. Denote by $MBest(i; z_-)$ the maximal best responce of agent i on the profile $z_- \in Z^{n-1}$ of other agents' choices, and recall that 1 is by definition the maximal element of the complete lattice Z. Introduce

$$\begin{aligned} z_1^* &= MBest(1; 1, \dots, 1); \\ z_2^* &= MBest(2; z_1^*, 1, \dots, 1); \\ &\dots \\ z_i^* &= MBest(i; z_1^*, \dots, z_{i-1}^*, 1, \dots, 1); \\ &\dots \\ z_n^* &= MBest(n; z_1^*, \dots, z_{n-1}^*). \end{aligned} \tag{28}$$

That is, the profile z^* is being approached iteratively, by firstly allowing everyone to choose the maximal element $1 \in Z$, and then, step by step from the agent 1 to n, offering them the maximal best responce on the existing profile.

I state that z^* is monotonic. This is the consequence of the Spence-Mirrlees conditions (I omit the details). First stage is complete.

Proof of lemma 2. We first prove that the allocation $(z_1^*, \dots, z_n^*)$ defined by the iterative procedure (28) is a monotone profile. We proceed inductively: assume that the sub-profile $(z_1^*, \dots, z_i^*)$ is monotone, and consider z_{i+1}^*. We have

$$\begin{gathered} z_{i+1}^* = z[i+1; z_1^*, \dots, z_i^*, S, \dots, S] \geq \\ z[i; z_1^*, \dots, z_i^*, S, \dots, S] = z[i; z_1^*, \dots, z_{i-1}^*, S, \dots, S] = z_i^*, \end{gathered} \tag{29}$$

where the last equality follows easily from ordinal semi-dependence axiom, in a way similar to that used in proving theorem 1.

Now, we establish that the allocation $(z_1^*, \dots, z_n^*)$ is a Nash equilibrium. Again, we use the same argument: inductively, and using monotonicity, we have that

$$\begin{gathered} z_i^* = z[i; z_1^*, \dots, z_i^*, S, \dots, S] = z[i; z_1^*, \dots, z_{i-1}^*, z_i^*, \dots, z_i^*] = \\ z[i; z_1^*, \dots, z_{i-1}^*, z_{i+1}^*, \dots, z_n^*]. \end{gathered} \tag{30}$$

And lastly, we will prove (again, inductively) that the allocation $q^* \succeq \tilde{q}$ for every Nash equilibrium allocation $\tilde{q}$. Namely, if it is proved that $z^*_j \geq \tilde{z}_j$ for $j = 1, \dots, i-1$, then

$$\begin{aligned} z^*_i = z[i; z^*_1, \dots, z^*_{i-1}, S, \dots, S] \geq \\ z[i; \tilde{z}_1, \dots, \tilde{z}_{i-1}, \tilde{z}_{i+1}, \dots, \tilde{z}_n] \geq \tilde{z}_i, \end{aligned} \tag{31}$$

because $z[i; q_{-i}] = max\{f[i; \tilde{q}_{-i}]\}$, and $\tilde{z}_i \in f[i; \tilde{q}_{-i}]$.

Summing up, we have proved that q^*, obtained by the inductive procedure (28), is the maximal Nash equilibrium. Proof of lemma 2 is complete.

Second step in details. Now, starting from this iterative procedure, let S be the coalition, and $\{\tilde{z}_i\}_{i \in S}$ — corresponding profile of strategies which makes all its members better off. I claim that $\forall i \in S\, \tilde{z}_i \leq z^*_i$. Denote by $\tilde{z}$ the profile in which all the agents $i \in S$ replace their choices by $\tilde{z}_i$, others held unchanged.

Suppose the converse is true, and let $i \in S$ be the first agent whose choice $\tilde{z}_i > z^*_i$. Then we have the following chain of inequalities:

$$u(i; \tilde{z}_i; \tilde{z}_-) \leq u(i; \tilde{z}_i; z^*_1, \dots, z^*_{i-1}, I, \dots, I) < u(i; z^*_i; z^*_1, \dots, z^*_{i-1}, I, \dots, I) = u(i; z^*). \tag{32}$$

The first inequality is the proclamation of monotonicity property, because we have non-decreased all the other agents' strategies; second, strict inequality follows from the fact that z^*_i is the maximal best responce on the profile in parenthesis, hence, all the higher choices are strictly worse than z^*_i; the last is due to iterativity.

But this chain of inequalities brings us to the point that i becomes strictly worse, which is absurdic[5]. Hence, we must have $\forall i \in S\, \tilde{z}_i \leq z^*_i$, and, as a result, $\tilde{z} \succeq z^*$.

But then no one get better position:

$$\forall i\, u(i; \tilde{z}_i; \tilde{z}_{-i}) \leq u(i; \tilde{z}_i; z^*_{-i}) \leq u(i; z^*), \tag{33}$$

of which the first follows from monotonicity, while the second — from the definition of Nash equilibrium.

[5] Notice that I proved the stronger result: z^* is super-strong equilibrium, that is, no coalition could improve no its member's position, without worsening somebody else.

The proof is complete.

References

Aumann, R.J. Acceptable points in general cooperative n-person games. Contributions to the theory of games, Volume 4. Princeton University Press, 1959.

Bernheim B.D. Rationalizable Strategic Behavior. Econometrica, Volume 52, pp. 1007-1028 (1984).

Bernheim B.D., Whinston M., and Peleg B. Coalition-Proof Nash Equilibria.I. Concepts. Journal of Economic Theory, Volume 42, 1-12 (1987).

Bagnoli M., and Lipman B.L. Provision of Public Goods: Fully Implementing The Core Through Private Contributions. The Review of Economic Studies, Volume 56, Issue 4 (Oct., 1989), 583-601.

Chander P., Wilde L. Corruption in Tax Administration. Journal of Public Economics, 49 (1992), pp. 333-349.

Danilov V., Sotskov A. Social choice mechanisms. In: Studies in Economic Design, Springer, 2002.

Friedman E.J., and Moulin H. Three Methods tTo Share Joint Costs or Surplus. Journal of Economic Theory, 87, pp. 275-312 (1999).

Hart, O.D., and Holmstrom B. (1987). The theory of contracts. In *Advances in Economic Theory*, Fifth World Congress, edired by T.Bewley. New York: Cambridge University Press.

Holmstrom, B. Moral Hazard in Teams. Bell Journal of Economics, 324-340 (1982).

Ichiishi, T. The cooperative nature of the firm. Cambridge University Press, 1993.

Ma, C. Unique Implementation of Incentive Contract with Many Agents. The Review of Economic Studies, Volume 55, Issue 4 (Oct., 1988), 555-571.

Malcomson, J-M. Work Incentives, Hierarchy, and Internal Labor Markets. The Journal of Political Economy, Volume 92, Issue 3 (Jun., 1984), 486-507.

Malcomson, J-M. Rank-Order Contracts for the Principal with Many Agents. The Review of

Economic Studies, Volume 53, Issue 5 (Oct., 1986), 807-817.

Milgrom P., and Roberts, J. Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities. Econometrica, Volume 58, No. 6 (November, 1990), 1255-1277.

Milgrom P., and Shannon C. Monotone Comparative Statics. Econometrica, Volume 62, No. 1 (January, 1994), 157-180.

Mookherjee, D. Optimal Incentive Schemes with Many Agents. The Review of Economic Studies, Volume 51, Issue 3 (Jul., 1984), 433-446.

Moulin, H. Serial Cost Sharing of Excludable Public Goods. The Review of Economic Studies, Volume 61, Issue 2 (Apr., 1994), 305-325.

Moulen H., and Shenker S. Serial Cost Sharing. Econometrica, Volume 60, No. 5 (September, 1992), 1009-1037.

Savvateev, A. Optimalnye strategii podavleniya corruptsii. Ekonomika i matematicheskie metody, forthcoming in vol. 39, pp. 62-75 (February, 2003).

Tarski, A. A Lattice-Theoretical Fixpoint Theorem and Its Applications. Pacific Journal of Mathematics, 5, 285-309 (1955).

Topkis, Donald M. Minimizing a Submodular Function on a Lattice. Operations Research, 26, 305-321 (1978).

Torkis, Donald M. Supermodularity and Complementarity. Princeton University Press, 1998.

Vasin A., Panova E. Tax Collection and Corruption in Fiscal Bodies. Working Paper No. 99/10, EERC WP Series, 2000.