

# An Efficient Solution to the Informed Principal Problem

Sergei Severinov \*

First version: May 2004, This Version: December 2006

## Abstract

In this paper I study mechanism design by an informed principal. I show that generically this problem has an ex-post efficient solution. In the equilibrium mechanism, the informed principal appropriates all expected social surplus, with each type of her getting all expected social surplus conditional on that type. This outcome is supported as a perfect sequential equilibrium of the informed principal game when the joint probability distribution from which the agents' types are drawn satisfies two conditions: the well-known condition of Cremer & McLean and Identifiability condition introduced by Kosenok and Severinov (2002). Conversely, these conditions are necessary for an ex-post efficient outcome to be attainable in an equilibrium of the informed principal game. Under these conditions only our equilibrium outcome constitutes a neutral optimum, i.e. cannot be eliminated by any reasonable concept of blocking (Myerson 1983). Identifiability and Cremer-McLean conditions are generic when there are at least three agents, and none of them has more types than the number of type profiles of the other agents.

JEL Nos: C72, D82. Keywords: informed principal, mechanism design, Bayesian mechanisms, perfect sequential equilibrium.

---

\*Fuqua School of Business, Duke University, Durham, NC, 27708; email: [sseverin@duke.edu](mailto:sseverin@duke.edu). I am grateful to Raymond Deneckere, Grigory Kosenok, Roger Myerson and Timofei Mylovanov for comments. Special thanks are due to Richard McLean, Associate Editor, for carefully checking all arguments and for many useful suggestions to improve the paper. I also thank seminar participants at UCLA, Duke, UT-Austin, Essex, UBC, University of Wisconsin-Madison, LBS, 2004 Game Theory Congress, 2004 North American Summer Meetings of the Econometric Society, and 2006 Economic Design Conference for comments. A standard disclaimer applies

# 1 Introduction.

In this paper, I study mechanism design by an informed principal. Thus, I consider an environment in which the mechanism for aggregating private information and choosing a social decision is designed by a party who, like other participants, possesses relevant private information. Examples include the design of a selling procedure by a buyer or a seller who has private information about her valuation or cost, different types of partnership problems (task allocation, dissolution), mechanisms for providing public goods, etc. In these situations all participants typically possess private information. For several reasons, hiring an outsider to serve as a mechanism designer could be infeasible, and/or too costly and involve the dissipation of a significant share of surplus to the latter. In particular, an outsider may not be able to understand all the details of the environment and may not even recognize what constitutes an efficient decision rule. It is reasonable to presume that this would be the case in financial markets, and in professional partnerships (for example, partnerships that involve doctors, lawyers, or faculty members). Furthermore, in collusion context it is natural to presume that the collusion mechanism has to be designed by one of the informed colluding parties. In contrast, in standard mechanism design, the designer does not possess any private information about her own or other participants' types or preferences.

Since different types of the informed mechanism designer may offer different mechanisms, the choice of a mechanism itself becomes a signal to the other participants (who I will refer to as *agents*) regarding the designer's type. This generates an inference process affecting the agents' incentives and their willingness to participate in the mechanism. Consequently, the informed principal problem is significantly harder to analyze than a standard mechanism environment.

The main result of this paper establishes that the informed principal problem generically possesses an ex-post efficient solution. Genericity refers to the requirement that the probability distribution of the agents' type profiles (including the type of the informed principal herself) satisfies two generic conditions: *Identifiability* condition introduced by Kosenok and Severinov (2002), and the well-known condition of Crémer and McLean (1988). Precisely, I show that under these conditions, the informed principal game has a sequential equilibrium with an ex-post efficient outcome. In this equilibrium the informed principal obtains all expected social surplus, with each type of her getting all social surplus *conditional* on that type. This equilibrium is perfect sequential in the sense of Grossman and Perry (1986). This refinement eliminates non-credible beliefs off the equilibrium path. Furthermore, I show that Identifiability and Crémer-McLean conditions are necessary for ex-post efficiency to be attained with informed principal. If either of these conditions fails then, under some profiles of the utility functions, the informed principal game has no ex-post efficient Bayesian equilibrium.

An intuitive explanation of the Identifiability condition using the notion of the probability distribution of agents' reported type profile in a direct mechanism is provided immediately after Definition 1. Briefly, it says that for any probability distribution  $q(\cdot)$ ,  $q(\cdot) \neq p(\cdot)$ , of the agents' reported type profile, there is an agent-type such that the conditional probability distribution of the other agents' reported type profile corresponding to  $q$  could not have been induced by this agent unilaterally deviating from truth-telling and reporting this type untruthfully. This agent-type may be thought of as a non-deviator under  $q$ . Identifiability condition, as well as Crémer-McLean condition, are generic when there are at least three agents and none of them has more types than the number of type profiles of all other agents.

Having established the existence of an ex-post efficient mechanism in which each type of the informed principal gets all social surplus conditional on her type, I then investigate the issue of

uniqueness of this equilibrium outcome. Using the concept of perfect sequential equilibrium, one can eliminate the outcomes in which the expected payoff of every type of the informed principal is less than the expected social surplus conditional on that type, with some types getting strictly less than the corresponding expected social surplus. However, this refinement does not allow us to rule out the possibility of equilibrium outcomes in which the expected payoff of some type of the informed principal is strictly greater than the expected social surplus conditional on that type.

So, to obtain uniqueness, one has to turn to stronger solution concepts. In particular, I show that our equilibrium outcome is an essentially unique *neutral optimum*. Neutral optimum of Myerson (1983) constitutes a strong refinement of non-cooperative and cooperative solution concepts, and will be discussed in more detail below.

To summarize, the contribution of this paper is two-fold. First, it shows that generically the informed principal problem has an ex-post efficient solution. Second, it establishes that the informed principal is able to extract all expected social surplus. To the best of my knowledge, these results have not been exhibited in the literature before.

Necessary and sufficient conditions for full surplus extraction by an uninformed mechanism designer have been derived by Crémer and McLean (1985) and (1988), and by McAfee and Reny (1992). Thus, this paper extends the surplus extraction results to the informed principal environment.

Mechanism design by an informed principal has been studied by several authors. In the pioneering work of Myerson (1983), and Maskin and Tirole (1990) and (1992), the authors have proposed several solution concepts for the problem and have established the corresponding existence results.

Maskin and Tirole (1990) and (1992) (MT in the sequel) focus on the environments with one principal and one agent and either an independent type distribution (in private values case) or no private information on the agent's side (in common values case where the agent's utility depends on the principal's type). MT suggest two efficiency criteria: weak interim efficiency (WIE) under common values and strong unconstrained Pareto optimum (SUPO) under private values. Roughly, these concepts require the allocation to be the best for the principal (maximize the weighted sum of the payoffs of her different types) subject to the agent's incentive compatibility and individual rationality constraints. Both concepts are weaker than ex-post efficiency.

MT characterize the equilibrium outcomes with quasilinear utility functions. In this case, the fact that the principal's information is private does not affect the outcome. Specifically, there is a unique perfect Bayesian equilibrium in which every type of the principal implements a standard second-best mechanism optimal for that type. The equilibrium mechanism is incentive compatible even if the agent knew the principal's type (in Myerson's terminology, it constitutes a strong solution), but it is typically not ex-post efficient. Yilankaya (1999) establishes a similar result in bilateral trade environment where only a single unit of the good is traded. He shows that in equilibrium all types of the informed seller use a fixed-price mechanism. Again, this outcome is not ex-post efficient but is ex-ante optimal, i.e. it maximizes the seller's ex-ante expected utility subject to the buyer's incentive constraints.

Recently, Mylovanov (2005a) has extended the one-agent characterization results of MT and Yilankaya (1999) to a quasilinear environment with multiple agents. Furthermore, Mylovanov (2005b) shows that with independently distributed types and private values, the equilibrium mechanism is not even ex-ante optimal generically in the space of the utility functions, i.e. it does not maximize the ex-ante expected utility of the informed principal subject to the agents'

incentive constraints.

Thus, the results of this paper differ significantly from those in the literature. This is due to the fact that I relax the assumption of independently distributed types and allow them to be stochastically dependent. I also require that the number of players is at least three (i.e. there are at least two agents besides the informed principal) as Identifiability would not be generic otherwise, while Maskin and Tirole (1990), (1992) and Yilankaya (1999) consider two-player (one principal and one agent) situations.

The rest of the paper is organized as follows. In section 2 I develop the model. In section 3 the main result is established. Section 4 is devoted to refinements. Section 5 concludes. All the proofs are relegated to an Appendix.

## 2 Model and Preliminaries

The economy consists of  $n \geq 3$  privately informed agents, who need to take a social decision affecting everyone's utility. The economic environment, the mechanism for implementing social decisions, and the procedure for choosing the mechanism are described below.

Agent  $i$ 's privately known information, or type, belongs to the type space  $\Theta_i \equiv \{\theta_i^1, \dots, \theta_i^{m_i}\}$  of cardinality  $m_i < \infty$ . A generic element of  $\Theta_i$  will be denoted by  $\theta_i$  or  $\theta'_i$ . A state of the world is characterized by a type profile  $\theta = (\theta_1, \dots, \theta_n)$ . The set of type profiles is given by  $\Theta \equiv \prod_{i=1, \dots, n} \Theta_i$ , with cardinality  $L \equiv \prod_{i=1, \dots, n} m_i$ . When focussing on agent  $i$ , we will use the notation  $(\theta_{-i}, \theta_i)$  for the profile of agent-types, where  $\theta_{-i}$  stands for the profile of types of agents other than  $i$ . Let  $\Theta_{-i} = \prod_{l \neq i} \Theta_l$ ,  $L_{-i} = \prod_{l \neq i} m_l$ ,  $\Theta_{-i-j} = \prod_{l \notin \{i, j\}} \Theta_l$ , and  $L_{-i-j} = \prod_{l \notin \{i, j\}} m_l$ . A generic element of  $\Theta_{-i-j}$  is denoted by  $\theta_{-i-j}$ .

The (true) probability distribution of the agents' type profile  $\theta$  is common knowledge and is denoted by  $p(\theta)$ , with  $p_i(\theta_i)$  and  $p_{i,j}(\theta_i, \theta_j)$  denoting the corresponding marginal probability distribution of agent  $i$ 's type and the marginal probability distribution of types of agents  $i$  and  $j$ , respectively. Also,  $p_{-i}(\theta_{-i}|\theta_i)$  ( $p_j(\theta_j|\theta_i)$ ) denotes the probability distribution of type profiles of agents other than  $i$  (agent  $j$ 's type) conditional on the type of agent  $i$ . A similar system of notation is used for other probability distributions over  $\Theta$  that we operate with below. The set of all probability distributions over  $\Theta$  is denoted by  $\mathcal{P}(\Theta)$ .

We assume that  $p_{i,j}(\theta_i, \theta_j) > 0$  for any  $\theta_i \in \Theta_i, \theta_j \in \Theta_j$  of any two agents  $i$  and  $j$ . This condition is clearly generic.

The set of public decisions is denoted by  $X$ , with  $x$  denoting a generic element of  $X$ . Agent  $i$ 's utility function is quasilinear in the decision  $x$  and transfer  $t_i$  that she receives, and is given by  $u_i(x, \theta) + t_i$ . Without loss of generality, an agent's reservation utility (i.e. her utility from the outside option) is normalized to zero.<sup>1</sup> A (social) decision rule  $x(\cdot)$  is a function mapping the type space  $\Theta$  into  $X$ , the set of public decisions<sup>2</sup>. Let  $t_i(\cdot) : \Theta \mapsto \mathbf{R}$  be a transfer function to agent  $i$ , and  $t(\cdot) = (t_1(\cdot), \dots, t_n(\cdot))$  denote a collection of transfer functions to all agents. An allocation profile is a combination of a decision rule  $x(\cdot)$  with a collection of transfer functions  $t(\cdot)$ . A decision rule  $x(\cdot)$  is *ex-post efficient* if  $x(\theta) \in \arg \max_{x \in X} \sum_{i=1}^n u_i(x, \theta)$  for all  $\theta \in \Theta$ , i.e.  $x(\theta)$  maximizes *ex-post social surplus*  $\sum_{i=1}^n u_i(x, \theta)$ . Since one can always cause the agents to take their outside options, we can without loss of generality assume that

<sup>1</sup>Suppose that agent  $i$ 's utility from her outside option is equal to  $w_i(\theta)$ . Such environment is equivalent to the environment where  $i$ 's utility function is given by  $u_i(x, \theta) - w_i(\theta) + t_i$  and her outside option is 0. Note that the sets of ex-post efficient decision rules and the notions of social surplus are the same in both environments.

<sup>2</sup>Randomization in public decisions is implicitly allowed, since  $X$  can be regarded as a set of probability distributions over some set of "pure" outcomes.

$\max_{x \in X} \sum_{i=1}^n u_i(x, \theta) \geq 0$  for all  $\theta \in \Theta$ .

The mechanism for aggregating the agents' private information and choosing a public decision is designed by one of the agents after she has learned her private information. This agent -acting as an informed principal- has the authority to propose and implement a mechanism. Without loss of generality, let agent 1 play the role of the informed principal.<sup>3</sup> A mechanism  $M$  offered by agent 1 consists of a set of strategy spaces  $S_1, \dots, S_n$  for all agents, including the mechanism designer - agent 1, and an outcome function  $g : \prod_{i=1, \dots, n} S_i \mapsto X \times \mathbf{R}^n$  mapping the set of agents' strategy profiles into the set of social decisions and transfers.<sup>4</sup>

The mechanism is implemented via the following informed principal game  $\Gamma$ :

- Stage 1. All agents learn their types.
- Stage 2. Agent 1 proposes mechanism  $M$ .
- Stage 3. Agents 2 to  $n$  simultaneously decide whether to participate in the mechanism, or to reject it.
- Stage 4. If all agents have agreed to participate, the mechanism  $M$  is implemented. The outcome is then determined by the agents' strategy choices and the outcome function  $g(\cdot)$  of  $M$ . If at least one of the agents rejects the mechanism, it is not implemented and all agents get their reservation payoffs.

Most contributions in the literature (e.g. Myerson (1983), Maskin and Tirole (1990) and (1992)) use the same extensive form game  $\Gamma$  in their analysis of the informed principal problem. The informed principal's strategy in this game involves choosing a mechanism  $M$  in stage 1, and a strategy  $s_1 \in S_1$  in  $M$  at stage 4. A strategy of agent  $i \in \{2, \dots, n\}$  consists of a participation decision at stage 2 and her choice of strategy  $s_i \in S_i$  in mechanism  $M$  at stage 4.

Let  $Z$  denote the set of feasible mechanisms. We require every mechanism  $M \in Z$  to possess a sequential equilibrium for any profile of agents' beliefs about each other. This can be ensured by simply assuming that all mechanisms in  $Z$  are finite, i.e. have a finite set of outcomes. Obviously, the latter assumption does not restrict the set of implementable decision rules, since the type space is finite.

The default assumption in this paper is that the set of feasible mechanisms  $Z$  is finite. However, all the results and proofs also hold in the case of an infinite  $Z$ . The only difference is that, with infinite  $Z$ , we have to use perfect Bayesian solution concept in the sufficiency part of Theorem 1, instead of sequential equilibrium. Perfect Bayesian equilibrium does not restrict the beliefs off the equilibrium path, while sequential equilibrium requires such beliefs to be consistent in the sense of Kreps and Wilson (1982). But the consistency of the off-equilibrium beliefs in sequential equilibrium is defined for finite games only.

Since there are no outsiders, any mechanism  $M \equiv (S_1, \dots, S_n, g(\cdot))$  in  $Z$  has to balance the budget. Precisely, let  $t_i^g(s_1, \dots, s_n)$  denote the transfer to agent  $i$  when the strategy profile

---

<sup>3</sup>The problem of choosing the mechanism designer is outside the scope of this paper. However, it is worth noting that our model applies if the mechanism designer is chosen at an ex-ante stage, i.e. before the agents have learned their information.

<sup>4</sup>The restriction to deterministic outcome functions is without loss of generality. First, as mentioned above, randomization over social alternatives is implicitly allowed. Second, the restriction to deterministic transfers is without loss of generality because the agents are risk-neutral and only care about the expected value of their transfers. Therefore, for every mechanism with stochastic transfer rule, there is a payoff-equivalent mechanism with a deterministic transfer rule.

$(s_1, \dots, s_n)$  is chosen by the agents. Then  $M$  has to satisfy:

$$\sum_{i=1}^n t_i^g(s_1, \dots, s_n) = 0 \quad \text{for all } (s_1, \dots, s_n) \in S_1 \times \dots \times S_n. \quad (1)$$

The assumption that the mechanism is not implemented whenever at least one agent refuses to participate in it, reflects that participation in the mechanism is voluntary. Any agent can withdraw from it and obtain her reservation utility if she wishes to do so. This assumption does not affect our results, and it is straightforward to replace it with an alternative one which still allows any agent to withdraw but does not require a dissolution when only some agents drop out.<sup>5</sup> Indeed, our equilibrium mechanism involves voluntary participation by all agent-types and transfers maximal possible social surplus to agent 1. Therefore, under any alternative assumption that does not require a unanimous acceptance for the mechanism to be implemented, agent 1 will still offer the same equilibrium mechanism and all agent types will accept this mechanism on the equilibrium path.

Let  $b_{-i}^3(\theta_{-i}|\theta_i, M)$  denote agent  $i$ 's beliefs about the type profile of the other agents at the beginning of stage 3, i.e.  $b_{-i}^3(\theta_{-i}|\theta_i, M)$  is the probability that agent-type  $\theta_i$  assigns to the type profile  $\theta_{-i}$  after agent 1 offers mechanism  $M$ . Similarly,  $b_{-i}^4(\theta_{-i}|\theta_i, M)$  denotes agent  $i$ 's beliefs about the type profile of the other agents at the beginning of stage 4.

In general, a solution to the informed principal problem is a collection of mechanisms  $\{M(\theta_1)\}_{\theta_1 \in \Theta_1}$  such that  $M(\theta_1) \in Z$  is offered by type  $\theta_1$ . However, by the *Inscrutability Principle* of Myerson (1983), we can assume that *on the equilibrium path* of  $\Gamma$  all types of agent 1 offer the same mechanism  $M$ . Then the other agents would not infer any additional information from the choice of the mechanism  $M$  at stage 2, and agent-type  $\theta_i$ 's beliefs  $b_{-i}^3(\theta_{-i}|\theta_i, M)$  on the equilibrium path would be equal to the posterior  $p_{-i}(\theta_{-i}|\theta_i)$ . The Inscrutability Principle holds because for any profile of mechanisms  $\{M(\theta_1)\}_{\theta_1 \in \Theta_1}$  offered by different types of agent 1 at stage 2, there is an outcome-equivalent mechanism  $\hat{M}$  which is offered by all types of agent 1 and which involves agent 1 revealing her type only at the implementation stage 4. Such mechanism is called *inscrutable*. For detailed proof, see Myerson (1983).

Further, by the Revelation principle, for any equilibrium of any mechanism  $M \in Z$  there exists an outcome-equivalent equilibrium of a *direct* incentive compatible mechanism.<sup>6</sup> This is easily shown by applying the proof of the Revelation Principle at stages 3 and 4 of  $\Gamma$ .

Thus, the Inscrutability and the Revelation Principles imply that, without loss of generality, we can assume that *on the equilibrium path* of  $\Gamma$ , all types of agent 1 offer a direct mechanism  $(x(\cdot), t(\cdot))$  which is incentive compatible under the beliefs  $p_{-i}(\theta_{-i}|\theta_i)$ . Since this mechanism is inscrutable, the equilibrium beliefs  $b_{-i}^3(\cdot|\theta_i, (x(\cdot), t(\cdot)))$  of any agent-type  $\theta_i$  at stage 3 of  $\Gamma$  are equal to  $p_{-i}(\theta_{-i}|\theta_i)$ .

This restriction on the equilibrium path is useful for describing an equilibrium mechanism in  $\Gamma$ . However, to prove that a certain mechanism is offered in equilibrium, we have to consider all possible deviations by agent 1, including deviations to non-direct and non-inscrutable mechanisms, i.e. mechanisms which are offered by *some, but not all* types of agent 1. Such

<sup>5</sup>For example, for every subset  $A$  of the set of agents  $\{1, \dots, m\}$ , the mechanism designer could specify a submechanism  $(S_1^A, \dots, S_n^A, g_A)$  which would be implemented when only the agents from the set  $A$  accept in stage 3. In this mechanism,  $S_i^A$  is a strategy space for agent  $i \in A$  while the outcome function  $g_A : \prod_{i \in A} S_i^A \mapsto X^A \times \mathbf{R}^{\#A}$  maps the participating agents' strategy profiles into  $X^A \subset X$ , the set of social decisions which are feasible with this set of participants.

<sup>6</sup>A direct mechanism can be represented as an outcome function  $(x(\cdot), t(\cdot))$  from the space of type profiles  $\Theta$  into the allocation space  $(X, \mathbf{R}^n)$ . Incentive compatibility in this case requires truth-telling to be an optimal strategy for each agent-type  $\theta_i$  given her beliefs  $b_{-i}^4(\theta_{-i}|\theta_i, M)$ .

a deviation would cause the other agents to update their prior beliefs in a non-trivial way in stages 3 and 4 of  $\Gamma$ .

We start by describing the incentive compatibility and individual rationality constraints which an inscrutable direct mechanism  $(x(\cdot), t(\cdot))$  has to satisfy. First, an agent-type  $\theta_i$  will accept the mechanism  $(x(\cdot), t(\cdot))$  only if her expected payoff from this mechanism is no less than her reservation payoff. Since the agents do not update their beliefs when an inscrutable mechanism is offered, this implies that  $(x(\cdot), t(\cdot))$  must satisfy the following *Interim Individual Rationality* constraints  $IR_i(\theta_i)$  for all  $i \in \{1, \dots, n\}$  and  $\theta_i \in \Theta_i$ :

$$U_i(\theta_i | (x(\cdot), t(\cdot))) \equiv \sum_{\theta_{-i} \in \Theta_{-i}} (u_i(x(\theta_{-i}, \theta_i), (\theta_{-i}, \theta_i)) + t_i(\theta_{-i}, \theta_i)) p_{-i}(\theta_{-i} | \theta_i) \geq 0. \quad (2)$$

Further,  $(x(\cdot), t(\cdot))$  is incentive compatible if it satisfies the following  $IC_i(\theta_i, \theta'_i)$  constraints for all  $i \in \{1, \dots, n\}$  and  $\theta_i, \theta'_i \in \Theta_i$ :

$$\sum_{\theta_{-i} \in \Theta_{-i}} (u_i(x(\theta_{-i}, \theta_i), (\theta_{-i}, \theta_i)) + t_i(\theta_{-i}, \theta_i) - u_i(x(\theta_{-i}, \theta'_i), (\theta_{-i}, \theta_i)) - t_i(\theta_{-i}, \theta'_i)) p_{-i}(\theta_{-i} | \theta_i) \geq 0. \quad (3)$$

Note that  $IR$  and  $IC$  constraints must also hold for agent 1, the mechanism designer, as otherwise some types of agent 1 would either prefer not to offer any mechanism, rather than to offer  $(x(\cdot), t(\cdot))$ , or would not report truthfully in stage 4. Finally, given that  $(x(\cdot), t(\cdot))$  is a direct mechanism, the budget balance constraint (1) can be rewritten as follows:

$$\sum_{i=1}^n t_i(\theta) = 0 \quad \text{for all } \theta \in \Theta. \quad (4)$$

We will say that a direct mechanism  $(x(\cdot), t(\cdot))$  is *admissible* if it is budget balanced, incentive compatible and individually rational, i.e. satisfies (2)-(4) for all  $\theta \in \Theta$ ,  $i \in \{1, \dots, n\}$ , and  $\theta_i \in \Theta_i$ .<sup>7</sup> The main issue in the context of the informed principal problem is which admissible mechanism will be chosen by the informed principal. Since different types of the informed principal may have different, sometimes opposite, preferences among the social alternatives, the equilibrium mechanism has to carefully balance the interests of all her types. Of particular interest is whether ex-post efficiency is attainable in the informed principal framework. I will address this question in the next Section.

### 3 Main Result

This section demonstrates that, generically, the informed principal problem has an ex-post efficient solution, with all expected social surplus from the mechanism obtained by agent 1, the informed principal. More specifically, each type  $\theta \in \Theta_1$  of agent 1 gets all expected social surplus from an ex-post efficient decision rule conditional on her type. The latter is denoted by  $V_1(\theta_1)$  and is formally given by the following expression:

$$V_1(\theta_1) \equiv \sum_{\theta_{-1} \in \Theta_{-1}} \left[ \max_{x \in X} \sum_{i \in \{1, \dots, n\}} u_i(x, (\theta_{-1}, \theta_1)) \right] p_{-1}(\theta_{-1} | \theta_1). \quad (5)$$

<sup>7</sup>In the sequel, incentive compatibility and individual rationality constraints of a mechanism are understood to be with respect to posterior beliefs  $p_{-i}(\cdot | \theta_i)$ , unless indicated otherwise.

I will show that this outcome is supported as a sequential equilibrium of  $\Gamma$  surviving the perfection refinement of Grossman and Perry (1986) and as a unique *neutral optimum*. These results hold under two generic conditions on the prior  $p(\cdot)$ , the Identifiability condition introduced by Kosenok and Severinov (2002) and the well-known condition of Crémer and McLean (1985) and (1988).

**Definition 1 *Identifiability*.** *The probability distribution  $p(\cdot)$  of the agents' type profile is identifiable if for any probability distribution  $q(\cdot) \in \mathcal{P}(\Theta)$ ,  $q(\cdot) \neq p(\cdot)$ , there is an agent  $i$  and her type  $\theta'_i$ , with  $q_i(\theta'_i) > 0$ , such that for any collection of nonnegative coefficients  $\{c_{\theta_i, \theta'_i}\}$ ,  $\theta_i, \theta'_i \in \Theta_i$ , we have:*

$$q_{-i}(\cdot|\theta'_i) \neq \sum_{\theta_i \in \Theta_i} c_{\theta_i, \theta'_i} p_{-i}(\cdot|\theta_i). \quad (6)$$

To describe the Identifiability condition intuitively, consider the agents' type reporting strategies in a direct mechanism. Agent  $i$ 's reporting strategy is a collection of  $m_i$  probability distributions over  $i$ 's type space  $\Theta_i$  -one for each type of  $i$ . From the ex-ante point of view, i.e. given that the agents' type profile is distributed according to the prior  $p(\cdot)$ , a collection of all agents' reporting strategies induces a probability distribution over reported type profiles. So, consider the set of all probability distributions over  $\Theta$ . The Identifiability condition requires that for each probability distribution  $q(\cdot)$  over  $\Theta$ ,  $q(\cdot) \neq p(\cdot)$ , there exists an agent  $i$  and her type  $\theta'_i$  such that  $i$  cannot induce the reported type profile of the other agents to be distributed according to  $q_{-i}(\cdot|\theta'_i)$  by following any reported strategy, when all other agents report truthfully. That is, agent  $i$  does not have a reporting strategy such that when  $i$  reports type  $\theta'_i$  according to this strategy and the other agents report their types truthfully, the conditional probability distribution of the other agents' type profiles is equal to  $q_{-i}(\cdot|\theta'_i)$ . Thus, agent-type  $\theta'_i$  could be thought of as a non-deviator type under  $q(\cdot)$ . Identifiability of  $p(\cdot)$  requires that for any  $q(\cdot) \neq p(\cdot)$ , the set of non-deviator agent-types is non-empty. Additional details can be found in Kosenok and Severinov (2002).

Next, consider the condition of Crémer and McLean (1985) and (1988) under which the uninformed mechanism designer can extract all social surplus in a Bayesian mechanism:

**Definition 2** *Say that Crémer-McLean condition holds for agent  $i$  if for any type  $\theta'_i \in \Theta_i$ ,  $p_{-i}(\cdot|\theta'_i)$  cannot be expressed as a convex combination of  $p_{-i}(\cdot|\theta_i)$ ,  $\theta_i \neq \theta'_i$ , i.e. for any collection of nonnegative coefficients  $c_{\theta_i, \theta'_i}$ , where  $\theta_i, \theta'_i \in \Theta_i$ , we have:*

$$p_{-i}(\cdot|\theta'_i) \neq \sum_{\theta_i \in \Theta_i \setminus \theta'_i} c_{\theta_i, \theta'_i} p_{-i}(\cdot|\theta_i).$$

Kosenok and Severinov (2002) have shown that Identifiability condition holds generically when there are at least three agents and at least two of them have weakly less types than the number of type profiles of all other agents (when there are three agents, it is also required that at least one of them has at least three types). It is well-known that Crémer-McLean condition for agent  $i$  holds generically when  $m_i \leq \prod_{j \neq i} m_j$ . So, a generic  $p(\cdot)$  is identifiable and satisfies Crémer-McLean condition for all  $i$  if  $\prod_{j \neq i} m_j \geq m_i$  also for all  $i$ .

Kosenok and Severinov (2002) have established the following result for the case when the mechanism is designed by an uninformed principal:

**Theorem** (Kosenok and Severinov). *An ex-post efficient, interim individually rational, ex-post budget balanced Bayesian mechanism exists under any profile of the utility functions (quasi-linear in transfers) if and only if  $p(\cdot)$  is identifiable and Crémer-McLean condition holds for all agents  $i = 1, \dots, n$ . Furthermore, when these conditions hold, it is possible to attain any distribution of expected social surplus between agent-types.*

An implication of this Theorem is that, under Identifiability and Crémer-McLean conditions, there exists an ex-post efficient, incentive compatible, budget-balanced mechanism  $M^* = (x^*(\cdot), t^*(\cdot))$  in which the expected payoff of any type  $\theta_1 \in \Theta_1$  of agent 1 is equal to  $V_1(\theta_1)$ , the expected social surplus conditional on agent 1's type  $\theta_1$  (see expression (5)), while every type of any other agent gets her reservation utility of zero.<sup>8</sup>

Mechanism  $M^*$  is a natural candidate for a solution to the informed principal game  $\Gamma$ . My first result shows that this mechanism can be supported as part of a sequential equilibrium of  $\Gamma$ .

Let us explain why this is so intuitively. To show that the mechanism  $M^*$  is offered in equilibrium, I have to consider all possible deviations in the choice of a mechanism at stage 2 and ascertain that no type of agent 1 can benefit from any such deviation. So suppose that at stage 2 agent 1 deviates from  $M^*$  and offers some mechanism  $M \in \mathcal{Z}$ . Then, there is a system of agents' beliefs  $\{\hat{b}_{-i}^3(\theta_{-i}|\theta_i, M)\}_{\theta_i \in \Theta_i, i \in \{2, \dots, n\}}$  and an equilibrium strategy profile  $\hat{\sigma}$  in mechanism  $M$  under these beliefs, such that the beliefs  $\{\hat{b}_{-i}^3(\theta_{-i}|\theta_i, M)\}_{\theta_i \in \Theta_i, i \in \{2, \dots, n\}}$  put a positive weight only on those types of agent 1 who obtain the largest payoff increment when equilibrium  $\hat{\sigma}$  is played in  $M$  compared to their putative equilibrium payoffs  $V_1(\cdot)$  that they obtain in  $M^*$ . A standard fixed point argument can be used to show this. But then any type  $\theta_1$  of agent 1 who is assigned a positive probability by the beliefs  $\{\hat{b}_{-i}^3(\theta_{-i}|\theta_i, M)\}_{\theta_i \in \Theta_i, i \in \{1, \dots, n\}}$  cannot get a strictly higher payoff in  $M$  than the payoff  $V_1(\theta_1)$  which she gets in the mechanism  $M^*$  and which is equal to the expected maximal social surplus conditional on agent 1's type being  $\theta_1$ . For, if the opposite was true, then all types of agent 1 who deviate to  $M$  would get more than the total social surplus generated in  $M$ , and therefore some agent-type  $\theta'_j$ ,  $j \in \{2, \dots, n\}$ , would get a strictly negative expected payoff in  $M$ . However, this would violate the participation constraint of  $\theta'_j$ , as she would be better off refusing to participate in  $M$ . Hence, after a deviation to  $M$  no type  $\theta_1$  of agent 1 gets a payoff strictly exceeding  $V_1(\theta_1)$ , and so no type has an incentive to deviate from  $M^*$ . Formally, we have:

**Theorem 1 Sufficiency.** *Suppose that  $p(\cdot)$  is identifiable and Crémer-McLean condition holds for all  $i \in \{1, \dots, n\}$ . Then the informed principal game  $\Gamma$  possesses a sequential equilibrium,  $E^*$ , in which all types of agent 1 offer ex-post efficient mechanism  $M^*$  and the expected payoff of each type  $\theta_1 \in \Theta_1$  of agent 1 is equal to  $V_1(\theta_1)$ , the expected maximal social surplus conditional on  $\theta_1$ .*

**Necessity.** *If either  $p(\cdot)$  is not identifiable or Crémer-McLean condition fails for some  $i \in \{1, \dots, n\}$ , then for some profiles of the utility functions the informed principal game does not have an ex-post efficient Bayesian equilibrium.<sup>9</sup>*

<sup>8</sup>For detailed proof and the construction of the mechanism, see Kosenok and Severinov (2002).

<sup>9</sup>In the necessity part of Theorem, I use a weaker solution concept of Bayesian equilibrium, in contrast to sequential equilibrium used in the sufficiency part. This is done to ascertain that the non-existence of an efficient outcome in the informed principal game is caused by the fundamental implementability problem, rather than by specific requirements of sequential or perfect Bayesian equilibrium concepts.

Theorem 1 is stated under our default assumption that the space of feasible mechanisms  $Z$  is finite. As mentioned above, it also holds when  $Z$  is infinite, but in the latter case we need to use perfect Bayesian instead of sequential equilibrium solution concept in the sufficiency part. The proof applies verbatim, with the exception of establishing the consistency of the off-equilibrium beliefs, which is required by the definition of sequential equilibrium, but becomes redundant under perfect Bayesian concept.

## 4 Refinements

It is well-known that the sequential equilibrium concept allows a broad leeway in the specification of beliefs off the equilibrium path, which could give rise to a multiplicity of equilibria. To address this issue, in this section I will consider two refinements of sequential equilibrium. First, I will demonstrate that our equilibrium  $E^*$  is perfect sequential (Grossman and Perry 1986).<sup>10</sup> In fact, all equilibria with the same outcome as in  $E^*$ , where agent 1 offers mechanism  $M^*$  with probability 1 and all agents truthfully report their types, survive this refinement. Second, I will show that this equilibrium outcome is a unique “neutral optimum.” The latter solution concept was introduced by Myerson (1983) and will be described in detail below.

The concept of perfect sequential equilibrium augments the notion of sequential equilibrium by imposing an additional restriction on the beliefs (‘credibility’) and strategies (‘perfection’) off the equilibrium path. This restriction requires subjecting an equilibrium  $E'$  of  $\Gamma$  to the following ‘test.’ For any mechanism  $M$  off equilibrium in  $E'$ , we look for a set of types  $K \subseteq \Theta_1$  of agent 1 and a strategy profile  $\hat{\sigma}^M$  in  $M$  s.t.: (i) the strategy profile  $\hat{\sigma}^M$  constitutes an equilibrium in  $M$  given that all agents other than 1 believe that agent 1’s type belongs to  $K$ <sup>11</sup>, (ii) every type of agent 1 from the set  $K$  gets a higher payoff, and at least one type from  $K$  gets a strictly higher payoff, in mechanism  $M$  when the strategy profile  $\hat{\sigma}^M$  is played than in the equilibrium  $E'$ . An equilibrium  $E'$  is perfect sequential when no such mechanism  $M$  and set  $K$  exist. Applying this refinement, we obtain:

**Theorem 2** *The equilibrium  $E^*$  of Theorem 1 and any sequential equilibrium of  $\Gamma$  with the same outcome as  $E^*$  i.e. any equilibrium in which agent 1 offers  $M^*$  and all agents follow truthtelling strategies with probability 1, is perfect sequential in the sense of Grossman and Perry (1986).*

A couple of remarks are in order at this point. Applying perfect sequential equilibrium refinement, we can also eliminate any equilibrium  $\tilde{E}$  of  $\Gamma$  in which the expected payoff of every type  $\theta_1 \in \Theta_1$  of agent 1 is less than  $V_1(\theta_1)$ , the expected maximal surplus conditional on her type, with at least one type,  $\bar{\theta}_1$ , earning strictly less than  $V_1(\bar{\theta}_1)$ . Specifically, such an equilibrium is undermined by all types of agent 1 offering mechanism  $M^*$  followed by acceptance of  $M^*$  and truthful type announcements by all agents. This continuation play is supported by ‘credible’ beliefs at stage 3 that all types of agent 1 have offered  $M^*$ .<sup>12</sup>

<sup>10</sup>Grossman and Perry (1986) define perfect sequential equilibrium for a 2-player signalling game in which only one party, the sender, has private information. However, it is straightforward to extend their concept to an  $n$ -player setting and apply it to our informed principal game  $\Gamma$ . Farrell (1993) proposes a similar refinement, neologism-proofness, for cheap talk games.

<sup>11</sup>The belief formation procedure is described precisely in the proof of Theorem 2 in the Appendix.

<sup>12</sup>If the mechanism  $M^*$  is offered on the equilibrium path of  $\tilde{E}$  with a positive probability, then to eliminate  $\tilde{E}$  we need to expand our set of feasible mechanisms by allowing cheap talk messages to accompany an offer of a mechanism by agent 1. Then  $\tilde{E}$  is undermined by a new mechanism  $M^{**}$  which is identical to  $M^*$  and,

However, this does not imply the uniqueness of a perfect sequential equilibrium outcome of  $\Gamma$ . In particular, we cannot rule out equilibria in which the expected payoff of at least one type  $\bar{\theta}_1 \in \Theta_1$  of agent 1 strictly exceeds  $V_1(\bar{\theta}_1)$ , the expected social surplus conditional on  $\bar{\theta}_1$ . This is so because agent-type  $\bar{\theta}_1$  would strictly lose from a deviation to  $M^*$ , if prior beliefs were maintained and truth-telling strategies were played in  $M^*$ . She would also strictly lose from a deviation to any other mechanism  $M'$ , if such deviation was associated with prior beliefs and an equilibrium  $e^{M'}$  in which every type  $\theta_1$  of agent 1 earns  $V_1(\theta_1)$ . Therefore, the credibility restriction would require agents' beliefs to be different from the prior after a deviation to the mechanism  $M^*$  or the mechanism  $M'$ . But with credible beliefs different from the prior, truth-telling may no longer be an equilibrium in  $M^*$  and  $e^{M'}$  may no longer be an equilibrium in  $M'$ , while new equilibria in these mechanisms may not be attractive to any types.

Nevertheless, uniqueness of our equilibrium outcome obtains once we turn to a stronger solution concept - that of *neutral optimum* of Myerson (1983). This solution concept is based on the notion of blocking. The idea is to identify a set of incentive compatible mechanisms in which the payoffs obtained by the principal (agent 1) are sufficiently high that no subset of her types would prefer a coordinated deviation to some other mechanism.

To define neutral optimum, let us abstract from the specifics of the game form of  $\Gamma$  for a moment, and consider some incentive-compatible direct mechanism  $\tilde{M} \equiv (\tilde{x}(\cdot), \tilde{t}(\cdot))$ . As pointed out above, any outcome implemented through any mechanism can be attained as an outcome of a direct incentive compatible inscrutable mechanism. Therefore we can restrict consideration to such mechanisms. Let  $\{U_i(\theta_i|\tilde{M})\}_{\theta_i \in \Theta_i}$  denote the vector of expected payoffs of agent  $i \in \{1, \dots, n\}$  in  $\tilde{M}$ , i.e.  $U_i(\theta_i|\tilde{M}) = \sum_{\theta_{-i} \in \Theta_{-i}} (u_i(\tilde{x}(\theta_{-i}, \theta_i), \theta_{-i}, \theta_i) + \tilde{t}_i(\theta_{-i}, \theta_i))p(\theta_{-i}|\theta_i)$ .

Further, let  $B(\Gamma)$  denote the set of blocked expected payoff vectors of agent 1 in  $\Gamma$  according to some notion of blocking  $B$ . Following Myerson (1983), we will allow for any notion of blocking  $B(\Gamma)$  that satisfies the following four axioms.

**Axiom 1 (Domination)** For any vectors  $w(\cdot)$  and  $z(\cdot)$  in  $\mathbb{R}^{m_1}$ , if  $w(\cdot) \in B(\Gamma)$ , and  $z(\theta_1) \leq w(\theta_1)$  for every  $\theta_1 \in \Theta_1$ , then  $z(\cdot) \in B(\Gamma)$ .

In words, if the vector  $w(\cdot)$  of expected payoffs of agent 1 is blocked according to  $B(\Gamma)$ , and the vector  $z(\cdot)$  is dominated by  $w(\cdot)$ , then  $z(\cdot)$  must also be blocked according to  $B(\Gamma)$ .

**Axiom 2 (Openness)**  $B(\Gamma)$  is open in the set of feasible expected payoff vectors  $FS = \{z(\cdot) \in \mathbf{R}_+^{m_1} : \sum_{\theta_1 \in \Theta_1} z(\theta_1)p_1(\theta_1) \leq \sum_{\theta \in \Theta} p(\theta) \max_{x \in X} \sum_{i \in \{1, \dots, n\}} u_i(x, \theta)\}$ .<sup>13</sup>

**Axiom 3 (Extension)** Let  $\bar{\Gamma}$  be an informed principal game which differs from  $\Gamma$  only because its feasible action set  $\bar{X}$  includes the feasible action set  $X$  of  $\Gamma$ , i.e.  $X \subset \bar{X}$ . Then  $B(\Gamma) \subset B(\bar{\Gamma})$ .

**Axiom 4 (Strong Solutions)** Suppose that mechanism  $M = (x(\cdot), t(\cdot))$  is incentive compatible and individually rational given the type of agent 1, i.e. it satisfies IR and IC constraints

additionally, has a cheap-talk message  $m^n$  associated with it. This cheap talk message  $m^n$  from agent 1, a 'neologism' in the terminology of Farrell (1993), should mean the following: "I offer mechanism  $M^{**}$  with probability 1, no matter what my type is." When all agents believe this message,  $M^{**}$  possesses a truth-telling equilibrium in which a type  $\theta_1$  of agent 1 earns  $V_1(\theta_1)$ , and so offering  $M^*$  is optimal for any  $\theta_1 \in \Theta_1$ .

<sup>13</sup>Here we make a slight departure from Myerson's definition which requires  $B(\Gamma)$  to be open in  $\mathbb{R}^{m_1}$ . Myerson's proof of existence of a neutral optimum and characterization results apply verbatim with our notion of openness. It appears more natural to require  $B(\Gamma)$  to be open relative to the set of feasible payoff vectors  $FS$ .

(2)-(3) of agent 1 for all  $\theta_1 \in \Theta_1$ , as well as the following individual rationality and incentive constraints for all  $i \in \{2, \dots, n\}$ ,  $\theta_i, \theta'_i \in \Theta_i$  and  $\theta_1 \in \Theta_1$ :

$$\sum_{\theta_{-1-i} \in \Theta_{-1-i}} (u_i(x(\theta_1, \theta_{-1-i}, \theta_i), (\theta_1, \theta_{-1-i}, \theta_i)) + t_i(\theta_1, \theta_{-1-i}, \theta_i)) p_{-1-i}(\theta_{-1-i} | \theta_1, \theta_i) \geq \max \left\{ 0, \sum_{\theta_{-1-i} \in \Theta_{-1-i}} (u_i(x(\theta_1, \theta_{-1-i}, \theta'_i), (\theta_1, \theta_{-1-i}, \theta_i)) - t_i(\theta_1, \theta_{-1-i}, \theta'_i)) p_{-1-i}(\theta_{-1-i} | \theta_1, \theta_i) \right\} \quad (7)$$

Suppose also that there does not exist another admissible mechanism  $M'$  satisfying  $U_1(\theta_1 | M) \leq U_1(\theta_1 | M')$  for all  $\theta_1 \in \Theta_1$  with strict inequality for at least one  $\theta_1$ . Then mechanism  $M$  is called a strong solution. A strong solution cannot be blocked i.e.,  $U_1(\theta_1 | M) \notin B(\Gamma)$ .

A blocking concept based on Axioms 1-4 identifies some set of blocked payoff vectors  $B(\Gamma)$  of agent 1. However, these Axioms do not pin down  $B(\Gamma)$  uniquely. Rather, there may be several different sets of blocked payoff vectors. So, let  $H$  be the set of all blocking concepts satisfying Axioms 1-4, and let  $B^*(\Gamma) = \cup_{B(\Gamma) \in H} B(\Gamma)$ , i.e.  $B^*(\Gamma)$  denotes the union of all sets of blocked payoff vectors. Thus, a payoff vector is not in  $B^*(\Gamma)$  if it cannot be blocked by any concept of blocking satisfying Axioms 1-4.

**Definition 3** (Myerson 1983) A mechanism  $\tilde{M}$  is a neutral optimum of the informed principal game  $\Gamma$  if it is admissible (i.e. satisfies (2)-(4))<sup>14</sup> and the vector  $\{U(\theta_1 | \tilde{M})\}_{\theta_1 \in \Theta_1}$  of expected payoffs of agent 1 does not belong to  $B^*(\Gamma)$ .

Axioms 1-4 are fairly natural and intuitive, and so the concept of a neutral optimum is not unnecessarily restrictive. Myerson (1983) has shown that a neutral optimum exists for a class of Bayesian games which includes our informed principal game  $\Gamma$  (see Theorem 6 in his paper). He has also provided a characterization of neutral optima, and established that a neutral optimum outcome can be supported as a sequential equilibrium. At the same time, the relationship between the set of neutral optima and the set of perfect sequential equilibrium outcomes is unclear. Therefore the next Theorem -which establishes generic uniqueness of the neutral optimum outcome in the game  $\Gamma$ - complements Theorem 2 without subsuming it.

**Theorem 3** Suppose that  $p(\cdot)$  is identifiable and Crémer-McLean condition holds. Then a mechanism  $(x(\theta), t(\theta))$  is a neutral optimum of the informed principal games  $\Gamma$  if and only if it is admissible, ex-post efficient and the expected payoff of any type  $\theta_1 \in \Theta_1$  of agent 1, the informed principal, is equal to  $V_1(\theta_1)$ , the expected maximal social surplus conditional on her type.

Thus, generically, only mechanism  $M^*$  or its equivalent, i.e. another ex-post efficient mechanism which provides the same expected payoff to each agent-type as in the mechanism  $M^*$ ,

<sup>14</sup>Myerson's definition of neutral optimum does not explicitly require individual rationality which is a part of our definition of admissibility. However, Myerson's notion of incentive compatibility includes individual rationality. See p. 1772 of Myerson (1983). Also note that our definition of an admissible mechanism in  $\Gamma$  includes ex- post budget balance (4). This does not introduce an additional restriction on the notion of neutral optimum. Rather, this is a restriction on the set of feasible allocations in our informed principal game  $\Gamma$ . Specifically, the set of feasible allocations of  $\Gamma$  is given by  $\{(x, t_1, \dots, t_n) | x \in X, (t_1, \dots, t_n) \in \mathbb{R}^n, \sum_{i=1, \dots, n} t_i = 0\}$ .

constitute neutral optima.<sup>15</sup> It follows that all types of all agents 2 to  $n$  obtain their reservation payoffs of zero in a neutral optimum.

To conclude, it is worth noting that in Bayesian mechanism design uniqueness of an equilibrium outcome is typically hard to obtain. So, the uniqueness of an equilibrium outcome in our case, even under a stronger solution concept, provides a desirable robustness check confirming that we have focussed on the right equilibrium and the right mechanism  $M^*$ .

## 5 Conclusions.

This paper demonstrates that generically an informed principal will implement an ex-post efficient mechanism and extract all expected social surplus. For this outcome to obtain as a perfect sequential equilibrium and a unique neutral optimum of the mechanism-choice game, it is necessary and sufficient that the probability distribution of the agents' type profile satisfies Identifiability and Crémer-McLean conditions. These conditions are generic when there are at least three players (i.e. at least two other agents besides the informed principal), and none of them has more types than the number of type profiles of the others.

These results can be applied to study a number of economic settings where informed principal problem arises naturally, such as auctions, resale, dissolution of partnerships and allocations of tasks in them. I intend to explore these applications in future research.

## 6 Appendix

### Proof of Theorem 1.

**Sufficiency Part.** A sequential equilibrium is an assessment, i.e. a collection of sequentially rational strategies and consistent beliefs of all agents (see Kreps and Wilson (1982)). Agent 1's strategy in the informed principal game  $\Gamma$  is an offer of a mechanism in stage 2 and a type report in stage 4. A strategy of agent  $i \in \{2, \dots, n\}$  is an acceptance/rejection decision in stage 3 and a type report in stage 4. The agents form beliefs at stage 3 after a mechanism is offered, and in stage 4 if every agent accepts the offered mechanism.

Our candidate equilibrium strategies prescribe the following play on the equilibrium path. All types of agent 1 offer the mechanism  $M^*$  in stage 2. In stage 3, all types of agents 2 to  $n$  agree to participate in  $M^*$ . In stage 4, if the mechanism  $M^*$  has been accepted by all agents, then all agents including agent 1 report their types truthfully. Equilibrium beliefs  $b_{-i}^3(\theta_{-i}|\theta_i, M^*)$  in stage 3 after  $M^*$  is offered and  $b_{-i}^4(\theta_{-i}|\theta_i, M^*)$  in stage 4 after all types have accepted  $M^*$  and before the type announcements are made are given by  $p_{-i}(\cdot|\theta_i)$ , for any type  $\theta_i \in \Theta_i$  of agent  $i \in \{1, \dots, n\}$ . These beliefs are consistent with all types of agent 1 offering  $M^*$  and all types of agents 2 to  $n$  accepting  $M^*$ .

Under these equilibrium beliefs, it is sequentially rational for every type of any agent  $i \in \{2, \dots, n\}$  to accept  $M^*$  in stage 3 and to report her type truthfully in stage 4. This is so because  $M^*$  is individually rational and incentive compatible with respect to the prior  $p(\cdot)$ .

Hence, to complete the sufficiency part of the proof, we only need to show that there exists a profile of sequentially rational strategies and consistent beliefs off the equilibrium path in the game  $\Gamma$  such that no type of agent 1 can profitably deviate at stage 2 by offering some mechanism  $M$  different from  $M^*$ . This will be established in four steps below.

---

<sup>15</sup>A degree of freedom in specifying the transfers in each state of the world arises since a neutral optimum outcome determines only the *expected* payoff of each agent-type, not state-by-state transfers in the mechanism, and there may be more than one ex-post efficient decision rule.

*Step 1. Definition of the auxiliary game  $\Gamma(M)$ .*

Fix an arbitrary mechanism  $M \equiv (S_1^M, \dots, S_n^M, x^M(\cdot), t^M(\cdot)) \in Z$ ,  $M \neq M^*$ , and define the auxiliary game  $\Gamma(M)$  as follows. Let  $\Gamma(M)$  be identical to the game  $\Gamma$ , except for stage 2. Particularly, in stage 2 of  $\Gamma(M)$  agent 1 can choose only between two actions. She can either exit the game and get a payoff equal to  $V_1(\theta_1)$  if her type is  $\theta_1$ , or alternatively she can offer the mechanism  $M$ . If agent 1 offers  $M$ , then stages 3 and 4 of the game  $\Gamma(M)$  are the same as in the original game  $\Gamma$  after  $M$  is offered.

*Step 2. Characterizing a sequential equilibrium of the auxiliary game  $\Gamma(M)$ .*

Since the mechanism  $M$  is finite, the auxiliary game  $\Gamma(M)$  possesses a sequential equilibrium (see Kreps and Wilson (1982)).<sup>16</sup> So let us fix an arbitrary sequential equilibrium of  $\Gamma(M)$  and denote it by  $\nu^M \equiv \{\rho_i(M|\theta_i), \sigma_i(\cdot|\theta_i, M), b_{-i}^3(\cdot|\theta_i, M), b_{-i}^4(\cdot|\theta_i, M)\}_{i=1, \dots, n, \theta_i \in \Theta_i}$ , where  $\rho_1(M|\theta_1) \in [0, 1]$  is the probability with which type  $\theta_1$  of agent 1 offers mechanism  $M$  at stage 2, and  $\rho_i(M|\theta_i) \in [0, 1]$ , for  $i = 2, \dots, n$ , is the probability with which type  $\theta_i \in \Theta_i$  of agent  $i$  accepts mechanism  $M$  at stage 3. Further,  $\sigma_i(\cdot|\theta_i, M)$  is a probability measure over  $S_i^M$  representing the strategy of agent-type  $\theta_i$ ,  $i \in \{1, \dots, n\}$ , in mechanism  $M$ . Finally,  $b_{-i}^3(\cdot|\theta_i, M)$  and  $b_{-i}^4(\cdot|\theta_i, M)$  are probability measures over  $\Theta_{-i}$  that represent, respectively, the beliefs of agent-type  $\theta_i \in \Theta_i$ ,  $i \in \{1, \dots, n\}$  in stage 3 after agent 1 has offered mechanism  $M$  and the beliefs of agent-type  $\theta_i \in \Theta_i$  in stage 4 after agent 1 has offered  $M$  and all agents have accepted it.

Let  $W_i(\theta_1, \dots, \theta_n | M, \nu^M)$  denote the expected payoff of agent  $i$  conditional on  $M$  being offered by agent 1 in stage 2 when the agents play strategies prescribed by  $\nu^M$  and the type profile is given by  $(\theta_1, \dots, \theta_n)$ . We have:

$$W_i(\theta_1, \dots, \theta_n | M, \nu^M) = \left( \sum_{(s_1, \dots, s_n) \in S^M} (u_i(x^M(s_1, \dots, s_n), (\theta_1, \dots, \theta_n)) + t_i^M(s_1, \dots, s_n)) \prod_{i=1, \dots, n} \sigma_i(s_i | \theta_i, M) \right) \prod_{i=2, \dots, n} \rho_i(M | \theta_i) \quad (8)$$

Next, let  $U_i(\theta_i | M, \nu^M)$  denote the expected payoff of agent-type  $\theta_i \in \Theta_i$ ,  $i \in \{1, \dots, n\}$  conditional on  $M$  being offered by agent 1 in stage 2 and the agents following the strategies and holding beliefs prescribed by  $\nu^M$ . Then,

$$U_i(\theta_i | M, \nu^M) = \sum_{\theta_{-i} \in \Theta_{-i}} W_i(\theta_{-i}, \dots, \theta_i | M, \nu^M) b_{-i}^3(\theta_{-i} | \theta_i, M) \quad (9)$$

Note that sequential rationality of agent 1's mechanism-choice strategy  $\rho_1(M|\theta_1)$  in the auxiliary game  $\Gamma(M)$  implies the following:

$$\rho_1(M|\theta_1) = \begin{cases} 1 & \text{if } U_1(\theta_1 | M, \nu^M) > V_1(\theta_1) \\ \text{any } x \in [0, 1] & \text{if } U_1(\theta_1 | M, \nu^M) = V_1(\theta_1) \\ 0 & \text{if } U_1(\theta_1 | M, \nu^M) < V_1(\theta_1). \end{cases} \quad (10)$$

*Let us establish the following important property of the equilibrium  $\nu^M$ :*

$$U_1(\theta_1 | M, \nu^M) \leq V_1(\theta_1) \quad \text{for all } \theta_1 \in \Theta_1. \quad (11)$$

<sup>16</sup>If we allow mechanisms with infinite strategy space, then sequential equilibrium is not defined for such games and so perfect Bayesian equilibrium is an appropriate solution concept. Hence we need to show that  $\Gamma(M)$  possesses a perfect Bayesian equilibrium. The existence of such equilibrium follows directly from our assumption regarding the class of feasible mechanisms  $Z$ .

The proof of (11) is by contradiction, so let us suppose that there exists  $\hat{\theta}_1 \in \Theta_1$  such that  $U_1(\hat{\theta}_1|M, \nu^M) > V_1(\hat{\theta}_1)$ . From (10), we have  $\rho_1(M|\hat{\theta}_1) = 1$ .

Note that agent-type  $\theta_i$ 's beliefs at stage 3 are given by  $b_{-i}^3(\theta_{-i}|\theta_i, M) = \frac{\rho_1(M|\theta_1)p(\theta_{-i}, \theta_i)}{\sum_{\theta'_1 \in \Theta_1} \rho_1(M|\theta'_1)p_{1,i}(\theta'_1, \theta_i)}$  for agent  $i \in \{2, \dots, n\}$  and  $b_{-1}^3(\theta_{-1}|\theta_1, M) = p_{-1}(\theta_{-1}|\theta_1)$  for agent 1.

Next, let  $TS(\nu^M, M)$  denote the expected social surplus generated in the equilibrium  $\nu^M$  of  $\Gamma(M)$  conditional on mechanism  $M$  being offered. To express  $TS(\nu^M, M)$  in terms of the agents' expected utilities, we need some additional notation. So, let  $Pr[\theta_1, \dots, \theta_n|M, \nu^M]$  denote the probability of the type profile  $(\theta_1, \dots, \theta_n)$  conditional on agent 1 offering mechanism  $M$  in equilibrium  $\nu^M$ . Let  $Pr[\theta_i|M, \nu^M]$  denote the probability that agent  $i$ 's type is  $\theta_i$ , for  $i \in \{1, \dots, n\}$ , when mechanism  $M$  is offered in equilibrium  $\nu^M$ . Note that  $Pr[\theta_1, \dots, \theta_n|M, \nu^M] = \frac{\rho_1(M|\theta_1)p(\theta_1, \dots, \theta_n)}{\sum_{\theta'_1 \in \Theta_1} \rho_1(M|\theta'_1)p_1(\theta'_1)}$  for all  $(\theta_1, \dots, \theta_n) \in \Theta$ ,  $Pr[\theta_1|M, \nu^M] = \frac{\rho_1(M|\theta_1)p_1(\theta_1)}{\sum_{\theta'_1 \in \Theta_1} \rho_1(M|\theta'_1)p_1(\theta'_1)}$  for  $\theta_1 \in \Theta_1$ , and  $Pr[\theta_i|M, \nu^M] = \frac{\sum_{\theta'_1 \in \Theta_1} \rho_1(M|\theta'_1)p_{1,i}(\theta'_1, \theta_i)}{\sum_{\theta'_1 \in \Theta_1} \rho_1(M|\theta'_1)p_1(\theta'_1)}$  for  $\theta_i \in \Theta_i$ ,  $i \in \{2, \dots, n\}$ . Then we have:

$$\begin{aligned}
TS(M) &\equiv \sum_{(\theta_1, \dots, \theta_n) \in \Theta} Pr[\theta_1, \dots, \theta_n|M, \nu^M] \sum_{i \in \{1, \dots, n\}} W_i(\theta_1, \dots, \theta_n|M, \nu^M) \\
&= \sum_{i \in \{1, \dots, n\}} \sum_{\theta_i \in \Theta_i} Pr[\theta_i|M, \nu^M] \sum_{\theta_{-i} \in \Theta_{-i}} b_{-i}^3(\theta_{-i}|\theta_i, M) W_i(\theta_i, \theta_{-i}|M, \nu^M) \\
&= \sum_{i \in \{1, \dots, n\}} \sum_{\theta_i \in \Theta_i} Pr[\theta_i|M, \nu^M] U_i(\theta_i|M, \nu^M) \\
&> \sum_{\theta_1 \in \Theta_1} Pr[\theta_1|M, \nu^M] V_1(\theta_1) + \sum_{i \in \{2, \dots, n\}} \sum_{\theta_i \in \Theta_i} Pr[\theta_i|M, \nu^M] U_i(\theta_i|M, \nu^M) \tag{12}
\end{aligned}$$

The equivalence sign reflects the definition. The first equality holds because  $Pr[\theta_1, \dots, \theta_n|M, \nu^M] = b_{-i}^3(\theta_{-i}|\theta_i, M)Pr[\theta_i|M, \nu^M]$  for all  $(\theta_1, \dots, \theta_n) \in \Theta$ . The second equality holds by definition (see equation (9)). The inequality in the last line holds because  $Pr[\theta_1|M, \nu^M] > 0$  only if  $U_1(\theta_1|M, \nu^M) \geq V_1(\theta_1)$  (this follows from (10) and from the definition of  $Pr[\theta_1|M, \nu^M]$ ) and  $Pr[\hat{\theta}_1|M, \nu^M] > 0$  and  $U_1(\hat{\theta}_1|M, \nu^M) > V_1(\hat{\theta}_1)$  for some  $\hat{\theta}_1 \in \Theta_1$ .

Next, since  $\sum_{i \in \{1, \dots, n\}} t^M(s_1, \dots, s_n) = 0$  for all  $(s_1, \dots, s_n) \in S^M$ , we can sum the expressions in (8) over all  $i \in \{1, \dots, n\}$  to obtain:

$$\begin{aligned}
&\sum_{i \in \{1, \dots, n\}} W_i(\theta_1, \dots, \theta_n|M, \nu^M) = \\
&\left( \sum_{(s_1, \dots, s_n) \in S^M} \left( \sum_{i \in \{1, \dots, n\}} u_i(x^M(s_1, \dots, s_n), (\theta_1, \dots, \theta_n)) \right) \prod_{i=1, \dots, n} \sigma_i(s_i|\theta_i, M) \right) \prod_{i=2, \dots, n} \rho_i(M|\theta_i) \\
&\leq \sum_{i \in \{1, \dots, n\}} u_i(x^*(\theta_1, \dots, \theta_n), (\theta_1, \dots, \theta_n)) \tag{13}
\end{aligned}$$

The last inequality holds because: (i) the decision rule  $x^*(\cdot)$  is socially efficient and so  $\sum_{i=1, \dots, n} u_i(x^M(s_1, \dots, s_n), (\theta_1, \dots, \theta_n)) \leq \sum_{i=1, \dots, n} u_i(x^*(\theta_1, \dots, \theta_n), (\theta_1, \dots, \theta_n))$  for all  $(\theta_1, \dots, \theta_n) \in \Theta$ ; (ii)  $\sigma_i(\cdot|\theta_i, M)$  is a probability distributions over  $S_i$  and so  $\sum_{(s_1, \dots, s_n) \in S^M} \prod_{i \in \{1, \dots, n\}} \sigma_i(s_i|\theta_i, M) = 1$ ; (iii)  $\prod_{i=2, \dots, n} \rho_i(M|\theta_i) \leq 1$  for all  $(\theta_1, \dots, \theta_n) \in \Theta$ .

Then, combining the definition of  $TS(M)$  in the first line of (12) with (13), we obtain:

$$\begin{aligned}
TS(M) &\leq \sum_{(\theta_1, \dots, \theta_n) \in \Theta} Pr[\theta_1, \dots, \theta_n | M, \nu^M] \sum_{i \in \{1, \dots, n\}} u_i(x^*(\theta_1, \dots, \theta_n), (\theta_1, \dots, \theta_n)) \\
&= \sum_{\theta_1 \in \Theta_1} Pr[\theta_1 | M, \nu^M] \sum_{\theta_{-1} \in \Theta_{-1}} b_{-1}^3(\theta_{-1} | \theta_1, M) \sum_{i \in \{1, \dots, n\}} u_i(x^*(\theta_1, \theta_{-1}), (\theta_1, \theta_{-1})) \\
&= \sum_{\theta_1 \in \Theta_1} Pr[\theta_1 | M, \nu^M] V_1(\theta_1)
\end{aligned} \tag{14}$$

Comparing (12) with (14), we conclude that the inequalities in (12) and (14) can hold simultaneously only if there exists some  $\theta_i \in \Theta_i$  for  $i \in \{2, \dots, n\}$  such that  $U_i(\theta_i | M, \nu^M) < 0$ . The latter can hold only if  $\rho_i(M | \theta_i) > 0$  i.e., agent-type  $\theta_i$  accepts mechanism  $M$  with a strictly positive probability. However, this acceptance decision is not sequentially rational since agent-type  $\theta_i$  can obtain zero payoff by rejecting  $M$  with probability 1, i.e. by following a strategy  $\rho_i(M | \theta_i) = 0$ . This contradiction implies that inequality (11) must hold for all  $\theta_1 \in \Theta_1$ .

*Step 3. The auxiliary game  $\Gamma(M)$  possesses a sequential equilibrium,  $\bar{\nu}^M$ , in which with probability 1 every type  $\theta_1 \in \Theta_1$  of agent 1 selects the outside option with payoff  $V_1(\theta_1)$ .*

Suppose that  $\nu^M$  is a sequential equilibrium of  $\Gamma(M)$  (which exists by Step 1 and has properties characterized in Step 2). If  $\nu^M$  prescribes that  $\rho_1(M | \theta_1) = 0$  for all  $\theta_1 \in \Theta_1$ , then define  $\bar{\nu}^M = \nu^M$ .

If there exists  $\tilde{\theta}_1 \in \Theta_1$  s.t.  $\rho_1(M | \tilde{\theta}_1) > 0$ , then to complete this step we will show that  $\Gamma(M)$  possesses another sequential equilibrium,

$\bar{\nu}^M \equiv \{(\bar{\rho}_1(M | \theta_1), \rho_2(M | \theta_2), \dots, \rho_n(M | \theta_n)), \sigma_i(\cdot | \theta_i, M), b_{-i}^3(\cdot | \theta_i, M), b_{-i}^4(\cdot | \theta_i, M))\}_{i=1, \dots, n, \theta_i \in \Theta_i}$ , that differs from  $\nu^M$  only in agent 1's acceptance strategy  $\bar{\rho}_1(M | \cdot)$  which satisfies  $\bar{\rho}_1(M | \theta_1) = 0$  for all  $\theta_1 \in \Theta_1$ .

To establish that  $\bar{\nu}^M$  is a sequential equilibrium, first, note that  $\bar{\rho}_1(M | \theta_1) = 0$  is sequentially rational for any  $\theta_1 \in \Theta_1$ . This is so because the expected payoff  $U_i(\theta_i | M, \bar{\nu}^M)$ , which type  $\theta_1$  of agent 1 obtains after offering  $M$  in  $\bar{\nu}^M$ , is the same as in  $\nu^M$ , i.e.  $U_i(\theta_i | M, \bar{\nu}^M) = U_i(\theta_i | M, \nu^M)$ , and so property (11) of Step 2 implies that  $U_i(\theta_i | M, \bar{\nu}^M) \leq V_1(\theta_1)$ .

Further, since  $\sigma_i(\cdot | \theta_i, M)$  is the equilibrium strategy in  $\nu^M$ , and  $b_{-i}^3(\cdot | \theta_i, M)$  and  $b_{-i}^4(\cdot | \theta_i, M)$  are equilibrium beliefs in  $\nu^M$ ,  $\sigma_i(\cdot | \theta_i, M)$  must be sequentially rational given beliefs  $b_{-i}^3(\cdot | \theta_i, M)$ , while beliefs  $b_{-i}^4(\cdot | \theta_i, M)$  must be consistent given  $\sigma_i(\cdot | \theta_i, M)$  and  $b_{-i}^3(\cdot | \theta_i, M)$ .

It remains to show that beliefs  $b_{-i}^3(\cdot | \theta_i, M)$ ,  $i \in \{2, \dots, n\}$ , are consistent with equilibrium acceptance strategies  $\bar{\rho}_1(M | \cdot)$  satisfying  $\bar{\rho}_1(M | \theta_1) = 0$ .<sup>17</sup> To see this, note that the consistency of beliefs  $b_{-i}^3(\cdot | \theta_i, M)$ ,  $i \in \{2, \dots, n\}$ , with the participation strategy  $\rho_1(M | \cdot)$  in  $\nu^M$  implies that  $b_{-i}^3(\theta_1, \theta_{-i-1} | \theta_i, M) = \frac{\rho_1(M | \theta_1) p(\theta_1, \theta_{-i-1}, \theta_i)}{\sum_{\theta'_1 \in \Theta_1} \rho_1(M | \theta'_1) p_{1,i}(\theta'_1, \theta_i)}$ . This expression is well-defined because  $\rho_1(M | \hat{\theta}_1) > 0$  for some  $\hat{\theta}_1 \in \Theta_1$ .

Then consider a sequence of strictly positive strategies  $\rho_1(M | \theta_1, t) = \frac{\rho_1(M | \theta_1)}{tD} + \frac{1}{t^2D}$  where  $D > 2$  is a positive constant, and  $t = 1, 2, \dots, \infty$ . Then we have:  $\bar{\rho}_1(M | \theta_1) = 0 = \lim_{t \rightarrow \infty} \rho_1(M | \theta_1, t)$  for all  $\theta_1 \in \Theta$ , and for all  $i \neq 1$  and  $\theta_i \in \Theta_i$  we have:

$$b_{-i}^3(\theta_1, \theta_{-i-1} | \theta_i, M) = \frac{\rho_1(M | \theta_1) p(\theta_1, \theta_{-i-1}, \theta_i)}{\sum_{\theta'_1 \in \Theta_1} \rho_1(M | \theta'_1) p_{1,i}(\theta'_1, \theta_i)} = \lim_{t \rightarrow \infty} \frac{\rho_1(M | \theta_1, t) p(\theta_1, \theta_{-i-1}, \theta_i)}{\sum_{\theta'_1 \in \Theta_1} \rho_1(M | \theta'_1, t) p_{1,i}(\theta'_1, \theta_i)} \tag{15}$$

<sup>17</sup>The following proof of consistency as well as that in the last paragraph of the sufficiency proof below are redundant when  $Z$  is infinite. In this case, we apply the concept of perfect Bayesian equilibrium which does not restrict the beliefs off equilibrium path.

So, the beliefs  $b_{-i}^3(\cdot|\theta_i, M)$  are consistent with strategies  $\bar{\rho}_1(M|\cdot)$ .

*Step 4. Constructing a sequential equilibrium of  $\Gamma$  in which every type of agent 1 offers mechanism  $M^*$  with probability 1.*

Recall that Steps 1-3 were established for an arbitrary  $M \in Z$ ,  $M \neq M^*$ . Therefore, we can choose a collection of sequential equilibria  $\{\bar{\nu}^M\}_{M \in Z}$  for the auxiliary games  $\Gamma(M)$ ,  $M \in Z$ ,  $M \neq M^*$ , with the participation strategies  $\bar{\rho}_1(M|\theta_1) = 0$  for all  $M \neq M^*$  and all  $\theta_1 \in \Theta_1$ . We will use this collection to construct a sequential equilibrium  $E^*$  of  $\Gamma$ .

The equilibrium  $E^*$  is as follows. All types of agent 1 offer the mechanism  $M^* = (x^*(\theta), t^*(\theta))$  with probability 1. Also with probability 1, all agent-types accept  $M^*$  and report their types truthfully at stage 4. The agents' beliefs in stage 3 after mechanism  $M^*$  has been offered and in stage 4 after  $M^*$  has been accepted are given by  $p_{-i}(\cdot|\theta_i)$  for all  $i \in \{1, \dots, n\}$ . If agent 1 offers mechanism  $M \in Z$ ,  $M \neq M^*$ , then agent-type  $\theta_i$ ,  $i \in \{1, \dots, n\}$ , plays the strategy  $\sigma_i(\cdot|\theta_i, M)$  used in  $\bar{\nu}^M$ , and holds beliefs  $b_{-i}^3(\cdot|\theta_i, M)$  and  $b_{-i}^4(\cdot|\theta_i, M)$  in stages 3 and 4. That is, her beliefs are the same as in equilibrium  $\bar{\nu}^M$  of the auxiliary game  $\Gamma(M)$  after  $M$  is offered.

Let us show that  $E^*$  is a sequential equilibrium. First, it is optimal for agent-type  $\theta_1$  to offer  $M^*$  because her expected payoff in  $M^*$  is equal to  $V_1(\theta_1)$ , while her payoff in some mechanism  $M \in Z$ ,  $M \neq M^*$ , is equal to  $U_i(\theta_i|M, \bar{\nu}^M)$ , and by Step 3,  $U_i(\theta_i|M, \bar{\nu}^M) \leq V_1(\theta_1)$ .

Further, it is sequentially rational for any agent-type  $\theta_i$  to accept  $M^*$  and follow a truthtelling strategy in it, because  $M^*$  is incentive compatible and individually rational under the beliefs  $p_{-i}(\cdot|\theta_i)$ . For any  $M \in Z$ , the sequential rationality of the strategy  $\sigma_i(\cdot|\theta_i, M)$  of agent-type  $\theta_i$  follows from Step 3.

The consistency of beliefs  $b_{-i}^3(\cdot|M^*, \theta_i) = p_{-i}(\cdot|\theta_i)$  and  $b_{-i}^4(\cdot|\theta_i, M)$  for all  $M \in Z$  is immediate. The consistency of beliefs  $b_{-i}^3(\cdot|\theta_i, M)$ ,  $M \neq M^*$ , can be established as in Step 3. That is, equation (15) holds if we set  $\rho_1(M|\theta_1, t) = \frac{\rho_1(M|\theta_1)}{tD} + \frac{1}{t^2D}$  and  $\rho_1(M|\theta_1, t) = 1 - \sum_{M \in Z, M \neq M^*} \rho_1(M|\theta_1, t)$  with  $D = 2\#Z$ , where  $\#Z < \infty$  is the cardinality of the space  $Z$ .

**Necessity.** In the standard mechanism design environment with an uninformed principal, an ex-post efficient, individually rational Bayesian mechanism fails to exist under some profiles of the utility functions when either Identifiability or Crémer-McLean condition fails (see Theorem 1 in Kosenok and Severinov (2002)). Therefore the informed principal game  $\Gamma$  does not have an ex-post efficient Bayesian equilibrium in such cases either. For suppose otherwise. Then in the standard environment, an uninformed principal can implement an efficient, individually rational, budget-balanced mechanism by committing to always delegate the design of the mechanism to agent 1. This would contradict Theorem 1 in Kosenok and Severinov (2002). *Q.E.D.*

### Proof of Theorem 2.

Let us show that equilibrium  $E^*$  is perfect sequential. The same proof works for any other sequential equilibrium with the same outcome as  $E^*$ , i.e. in which agent 1 offers mechanism  $M^*$  with probability 1 and all agents follow truthtelling strategies in  $M^*$ .<sup>18</sup> Recall that in the equilibrium  $E^*$ , each type  $\theta_1$  of agent 1 earns  $V_1(\theta_1)$ , the expected maximal social surplus conditional on  $\theta_1$ .

To prove our claim, we need to establish that off equilibrium beliefs in  $E^*$  are credible in the sense of Grossman and Perry (1986). Specifically, we need to show that there is no mechanism  $\tilde{M} \in Z$ ,  $\tilde{M} \neq M^*$ , for which there exists a profile of agents' strategies  $\tilde{\sigma}$  and two disjoint sets of agent 1's types  $K^s \subseteq \Theta_1$ ,  $K^s \neq \emptyset$ , and  $K^w \subseteq \Theta_1$ , and a system of beliefs  $\tilde{b}_{-i}^3(\theta_{-i}|\theta_i, \tilde{M})$  that have the following properties:

<sup>18</sup>Such equilibria may differ with respect to off equilibrium play and beliefs.

(a) Let  $U_1(\theta_1|\tilde{M}, \tilde{\sigma})$  denote the payoff of a type  $\theta_1 \in \Theta_1$  in the mechanism  $\tilde{M} \in Z$  when the strategy profile  $\tilde{\sigma}$  is played. Then for any type  $\theta_1 \in K^s$ , we have  $U_1(\theta_1|\tilde{M}, \tilde{\sigma}) > V_1(\theta_1)$ . For any type  $\theta'_1 \in K^w$ , we have  $U_1(\theta'_1|\tilde{M}, \tilde{\sigma}) = V_1(\theta'_1)$ . For any type  $\theta''_1 \notin (K^s \cup K^w)$ , we have  $U_1(\theta''_1|\tilde{M}, \tilde{\sigma}) < V_1(\theta''_1)$ .

(b) The beliefs  $\tilde{b}_{-i}^3(\theta_{-i}|\theta_i, \tilde{M})$  of any agent-type  $\theta_i \in \Theta_i$ ,  $i \in \{2, \dots, n\}$ , in stage 3 after agent 1 offers mechanism  $\tilde{M}$ , are “credible.” That is, agent-type  $\theta_i \in \Theta_i$  believes that agent-type  $\theta_1 \in K^s$  offers  $\tilde{M}$  with probability 1, agent-type  $\theta_1 \in K^w$  offers  $\tilde{M}$  with some probability  $h(\theta_1) \in [0, 1]$ , and agent-type  $\theta_1 \notin (K^s \cup K^w)$  offers  $\tilde{M}$  with probability 0. Thus, for all  $\theta_{-i-1} \in \Theta_{-i-1}$ , we have:

$$\tilde{b}_{-i}^3(\theta_1, \theta_{-i-1}|\theta_i, \tilde{M}) = \begin{cases} \frac{p(\theta_1, \theta_{-i-1}, \theta_i)}{\sum_{\theta'_1 \in K^s} p_{1,i}(\theta'_1, \theta_i) + \sum_{\theta'_1 \in K^w} h(\theta'_1) p_{1,i}(\theta'_1, \theta_i)} & \text{if } \theta_1 \in K^s \\ \frac{p(\theta_1, \theta_{-i-1}, \theta_i) h(\theta_1)}{\sum_{\theta'_1 \in K^s} p_{1,i}(\theta'_1, \theta_i) + \sum_{\theta'_1 \in K^w} h(\theta'_1) p_{1,i}(\theta'_1, \theta_i)} & \text{if } \theta_1 \in K^w \\ 0 & \text{if } \theta_1 \in \Theta_1 \setminus (K^s \cup K^w) \end{cases} \quad (16)$$

(c) Under the beliefs  $\tilde{b}_{-i}^3(\theta_{-i}|\theta_i, \tilde{M})$ , the strategy profile  $\tilde{\sigma}$  is sequentially rational in  $\tilde{M}$ .

To rule out the existence of such mechanism  $\tilde{M}$ , strategy profile  $\tilde{\sigma}$ , system of credible beliefs  $\tilde{b}_{-i}^3(\theta_{-i}|\theta_i, \tilde{M})$ , and subsets  $K^s \neq \emptyset$  and  $K^w$  of agent 1’s types that satisfy these properties, we argue by contradiction. So suppose that such exist.

Note that credible beliefs  $\{\tilde{b}_{-i}^3(\theta_{-i}|\theta_i, \tilde{M})\}_{\theta_i \in \Theta_i, 2 \in \{1, \dots, n\}}$  assign positive probabilities only to those types of agent 1 that belong to  $K^s$  and  $K^w$ . Recall that for any  $\theta_1$  in the non-empty set  $K^s$ ,  $U_1(\theta_1|\tilde{M}, \tilde{\sigma}) > V_1(\theta_1)$ , and for any  $\theta'_1 \in K^w$ ,  $U_1(\theta'_1|\tilde{M}, \tilde{\sigma}) = V_1(\theta'_1)$ .

But then we can use the same argument as in the proof of Theorem 1, and in particular, the sequences of inequalities (12) and (14), to show that there must exist some agent-type  $\theta_i \in \Theta_i$ ,  $i \in \{2, \dots, n\}$ , who earns a strictly negative expected payoff in the mechanism  $\tilde{M}$ . However, then the strategy profile  $\tilde{\sigma}$  cannot be sequentially rational, because  $\theta_i$  can obtain her reservation payoff of zero by dropping out at stage 3 after agent 1 offers  $\tilde{M}$ . *Q.E.D.*

**Proof of Theorem 3.** By Theorem 6 of Myerson (1983) a neutral optimum exists. By definition, a neutral optimum mechanism must be admissible, i.e. incentive compatible, individually rational and ex-post budget-balanced. Our mechanism  $M^* = (x^*(\theta), t^*(\theta))$  satisfies these three properties. So to prove the Theorem, we need to rule out any mechanism in which the expected payoff of some type  $\theta_1$  of agent 1 differs from  $V_1(\theta_1)$ . Precisely, since there are no admissible mechanisms in which every type  $\theta_1 \in \Theta_1$  gets an expected payoff exceeding  $V_1(\theta_1)$  with at least one type  $\tilde{\theta}_1$  getting strictly more than  $V_1(\tilde{\theta}_1)$ , we only need to show that any mechanism in which the expected payoff of some type  $\theta_1$  of agent 1 is strictly less than  $V_1(\theta_1)$  is not a neutral optimum.

So, take some other admissible mechanism  $M \in Z$ , let  $\{U_1(\theta_1|M)\}_{\theta_1 \in \Theta_1}$  denote the vector of agent 1’s expected payoffs in  $M$ , and suppose that  $U_1(\theta'_1|M) < V_1(\theta'_1)$  for some  $\theta'_1 \in \Theta_1$ . Next, define blocking concept  $\hat{B}(\theta'_1)$  as follows:

$$\hat{B}(\theta'_1) \equiv \left\{ z(\cdot) \in \mathbb{R}_+^{m_1} \mid \sum_{\theta_1 \in \Theta_1} z(\theta_1) p_1(\theta_1) \leq \sum_{i: \theta \in \Theta} u_i(x^*(\theta), \theta) p(\theta); \quad z(\theta'_1) < V_1(\theta'_1) \right\}$$

Let us show that  $\hat{B}(\theta'_1)$  satisfies Axioms 1-4. It satisfies the Domination Axiom by definition. Further, it satisfies the Openness Axiom, because  $\hat{B}(\theta'_1)$  is open in the set of feasible expected payoff vectors of agent 1  $\{y(\cdot) \in \mathbb{R}_+^{m_1} \mid 0 \leq \sum_{\theta_1 \in \Theta_1} y(\theta_1) p_1(\theta_1) \leq \sum_{i: \theta \in \Theta} u_i(x^*(\theta), \theta) p(\theta)\}$ .

To see that the Extension Axiom holds, suppose that we expand the set  $X$  by adding more elements to it. This modification may change the set of ex-post efficient decision rules. As

a result, the expected social surplus  $\sum_{i=1, \dots, n; \theta \in \Theta} u_i(x^*(\theta), \theta)p(\theta)$  can increase, but it cannot decrease. Similarly,  $V_1(\theta'_1)$ , the expected social surplus conditional on type  $\theta'_1 \in \Theta_1$  of agent 1, cannot decrease after we expand  $X$ . Therefore, by definition of  $\hat{B}(\theta'_1)$ , the outcomes which were blocked prior to an expansion of  $X$ , remain blocked after such expansion.

Finally, let us show that  $\hat{B}(\theta'_1)$  satisfies Axiom 4, i.e. it does not block a strong solution. Suppose that mechanism  $M^s$  is a strong solution, and let  $U_i(\theta_i|M^s)$  ( $U_i(\theta_i|M^s, \theta_1)$ ) denote the expected payoff of agent-type  $\theta_i$  ( $\theta_i$ 's expected payoff conditional on agent 1's type  $\theta_1$ ) in  $M^s$ .

Then it must be the case that  $U_1(\theta_1|M^s) \leq V_1(\theta_1)$  for all  $\theta_1 \in \Theta_1$ . To see this, suppose otherwise, i.e.  $U_1(\bar{\theta}_1|M^s) > V_1(\bar{\theta}_1)$  for some  $\bar{\theta}_1 \in \Theta_1$ . Then, since  $M^s$  is budget-balanced and  $V_1(\bar{\theta}_1)$  is equal to the expected maximal social surplus conditional on agent 1's type  $\bar{\theta}_1$ , there must exist some agent-type  $\theta'_i \in \Theta_i$ , with  $i \in \{2, \dots, n\}$ , whose expected payoff in  $M^s$  when agent 1's type is  $\theta_1$  (i.e.,  $U_i(\theta'_i|M^s, \bar{\theta}_1)$ ) is strictly negative. But this contradicts condition 7 in Axiom 4 which requires that the expected payoff of any type  $\theta_i$  of agent  $i \in \{2, \dots, n\}$ , conditional on any type of agent 1, should be above the reservation payoff of zero.

On the other hand, we cannot have  $U_1(\theta'_1|M^s) < V_1(\theta'_1)$  for some  $\theta'_1 \in \Theta_1$ . For otherwise,  $M^s$  would be dominated by the mechanism  $M^*$  as we would have  $U_1(\theta_1|M^s) \leq U_1(\theta_1|M^*)$  for all  $\theta_1 \in \Theta_1$ , with strict inequality for  $\theta'_1$ . So, we must have  $U_1(\theta_1|M^s) = V_1(\theta_1)$  for all  $\theta_1 \in \Theta_1$ , and therefore  $M^s$  is not blocked by  $\hat{B}(\theta'_1)$ .

We have shown that  $\hat{B}(\theta'_1)$  satisfies Axioms 1-4, i.e. it is an admissible blocking concept. So, a neutral optimum outcome cannot belong to  $\hat{B}(\theta'_1)$ . Recall that  $\theta'_1 \in \Theta_1$  was chosen arbitrarily. Hence, the expected payoff of each type  $\theta_1 \in \Theta_1$  of agent 1 in a neutral optimum mechanism must be equal to  $V_1(\theta_1)$ . This implies that a neutral optimum mechanism must be ex-post efficient, and the expected payoff of every agent-type  $\theta_i \in \Theta_i$ ,  $i \in \{2, \dots, n\}$ , must be equal to zero. Otherwise, by our standard argument relying on the sequences of inequalities (12) and (14), the individual rationality of some other agent-type will fail. Thus, the set of ex-post efficient admissible mechanisms in which agent-type  $\theta_1$  obtains payoff  $V_1(\theta_1)$  includes  $M^*$  and other ex-post efficient mechanisms with the same expected payoff of every agent-type as in  $M^*$ . Q.E.D.

## References

- Cremer, J., and R. McLean (1985), "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, 53, pp. 345-361.
- Cremer, J., and R. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, pp. 1247-1257.
- Farrell, J. "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior*, 5, pp. 514-531.
- Grossman, S., and M. Perry (1986), "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39, pp. 97-119.
- Kosenok, G., and S. Severinov (2002), "Individually Rational, Budget-Balanced Mechanisms and Allocation of Surplus," forthcoming in *Journal of Economic Theory*
- Kreps, D., and R. Wilson (1982), "Sequential Equilibria," *Econometrica*, 50, pp. 863-894.
- Maskin, E., and J. Tirole (1990), "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values," *Econometrica*, 58, pp. 379-409.

- Maskin, E., and J. Tirole (1992), "The Principal-Agent Relationship with an Informed Principal, II: Common Values," *Econometrica*, 60, pp. 1-42.
- McAfee, P., and P. Reny (1992), "Correlated Information and Mechanism Design," *Econometrica*, 60, pp. 395-421.
- Myerson, R. (1983), "Mechanism Design by an Informed Principal," *Econometrica*, 51, pp. 1767-1797.
- Mylovanov, T. (2005a), "The Informed Principal Problem with Independent Private Values and Quasilinear Preferences," mimeo., Bonn University.
- Mylovanov, T. (2005b), "Negative Value of Information in the Informed Principal Problem: the Case of Private Values," mimeo., Bonn University.
- Yilankaya, O. (1999), "A Note on the Seller's Optimal Mechanism in Bilateral Trade with Two-Sided Incomplete Information," *Journal of Economic Theory*, 87, pp. 267-271.