INFORMED-PRINCIPAL PROBLEM WITH MORAL HAZARD, RISK NEUTRALITY, AND NO LIMITED LIABILITY

CHRISTOPH WAGNER, TYMOFIY MYLOVANOV, AND THOMAS TRÖGER

ABSTRACT. We consider a principal-agent moral-hazard problem with risk-neutral parties and no limited liability in which the principal has private information. The principal's private information creates signaling considerations that may distort the implemented outcome. These distortions can explain, e.g., efficiency wages (Beaudry 1994) and muted incentives (Inderst 2001). We show that in a large class of environments these distortions vanish if the principal is allowed to offer sufficiently rich contracts.

1. INTRODUCTION

In the standard principal-agent model, the principal has no private information and the optimal contract can be found by solving a constrained optimization problem. If the principal has private information, the model becomes a signaling game and the contract's value is determined endogenously in equilibrium. The signaling incentives introduce novel distortions, which may explain, e.g., efficiency wages (Beaudry 1994) and muted incentives (Inderst 2001).

We consider a principal-agent model with risk-neutral parties and no limited liability in which the principal has private information (her type) about the technology that translates the agent's effort into observables (Beaudry 1994, Inderst 2001). This environment is useful for understanding distortions that can be generated by the privacy of the principal's information since the benchmark environment in which the principal has no private information features no distortions: the optimal contract is to "sell the firm" to the agent (see e.g. Grossman and Hart (1983), Proposition 3(2)).

Date: February 1, 2017.

Christoph Wagner: Department of Economics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany, email: cwagner@uni-bonn.de; Tymofiy Mylovanov: Department of Economics, University of Pittsburgh, 230 South Bouquet St, Pittsburgh, PA 15260, USA, email: mylovanov@gmail.com; Thomas Tröger, Department of Economics, University of Mannheim, L7 3-5, 68131 Mannheim, Germany, email: troeger@uni-mannheim.de. We would like to thank the editor and two excellent referees for the comments that substantively improved the paper. This work has benefited from detailed comments by Jörg Budde, Paul Heidhues, Daniel Kräbmer, Dezső Szalay, and Venuga Yokeeswaran, Rafael Aigner, Carsten Dahremoller, Deniz Dizdar, Markus Fels, Fabian Herweg, Sebastian Kranz, Matthias Lang, Patrick Rey, Christian Seel, Axel Stahmer, and Philipp Strack. We gratefully acknowledge financial support from the National Science Foundation, grant 1024683, and from the German Science Foundation (DFG) through SFB/TR 15 "Governance and the Efficiency of Economic Systems."

Under asymmetric information, the value of the firm is uncertain to the agent and the environment becomes that of the informed principal (Myerson 1983, Maskin and Tirole 1992). The model, however, differs from the literature because there exists an ex-post verifiable variable, e.g., profit, that is correlated both with the agent's effort and the principal's private information.

In Section 3, we consider environments in which the first-best effort is constant across all principal types, and we provide two logically independent sufficient conditions under which the first best can be implemented. These conditions are (1) the first-best effort is the most costly action for the agent and (2) a linear independence condition on the distribution of the ex-post verifiable variable, which is satisfied generically if the support of the variable is sufficiently large. In Section 4, we allow the first-best effort to vary with the principal's type and show that the first best can be implemented if a stronger linear independence condition holds, which is still generically satisfied if the support of the ex-post verifiable variable is sufficiently large.

This result stands in contrast with the results in Beaudry (1994) and Inderst (2001), in which these independence conditions are violated. The conditions hold if the dimensionality of the support of the ex-post verifiable variable is sufficiently large relative to the dimensionality of the type and effort spaces. Hence, the dimensionality is the key to the distortions that can be caused by the signaling considerations and, in sufficiently rich contractual environments, the privacy of the principal's information does not impose any costs on the principal. This observation can manifest itself in other moral hazard environments and, thus, it might prove useful when developing applications.

The result that the principal can implement the same outcome regardless of whether her information is private or publicly known to the agent has been observed in other environments. It holds in independent private value environments with risk-neutral players (Myerson 1985, Maskin and Tirole 1990, Tan 1996, Yilankaya 1999, Balestrieri 2008, Skreta 2009, Mylovanov and Tröger forthcoming) if payoff functions satisfy a monotonicity condition (Mylovanov and Tröger forthcoming).

The contracts that attain the first best are strongly neologism-proof (Maskin and Tirole 1992, Mylovanov and Tröger 2012): they extract the entire surplus conditional on each type of the principal and cannot be dominated by any other contract. Severinov (2008) offers a full-surplus-extraction informed-principal result in environments with adverse selection, no moral hazard, and correlated types. Fleckinger (2007) presents a full-surplus-extraction result for an informed principal in an environment with adverse selection, no moral hazard, and countervailing incentives.

In addition to Beaudry (1994) and Inderst (2001), the informed principal problem with moral hazard is studied in Jost (1996), Bond and Gresik (1997), Mezzetti and Tsoulouhas (2000), Chade and Silvers (2002), Benabou and Tirole (2003), Kaya (2010), Karle and Staat (2013), and Fong and Lee (2013). These papers feature risk-averse players, allow for dynamic interaction, impose limited-liability constraints, or introduce moral hazard on the part of the principal.

2. MODEL

There is a principal ($i = 0$) and an agent ($i = 1$) who operate a firm. The principal has a type $t \in T = \{1, \dots, \tau\}$. The type distribution is denoted $\lambda^0 = (\lambda_1^0, \dots, \lambda_\tau^0)$, where the probability $\lambda_t^0 > 0$ for all t . The agent chooses an effort $a \in A = \{1, \dots, m\}$ which determines the distribution $\mathbf{f}(t, a) = (f_1(t, a), \dots, f_n(t, a))$ of an ex-post verifiable variable $s \in \{1, \dots, n\}$. Alternatively, the agent may choose not to work for the firm, which we denote as $a = 0$. In addition, the parties can exchange a verifiable monetary payment or wage $w \in \mathbb{R}$ from the principal to the agent.

Parties are risk-neutral and have private values, with the payoff functions

$$\begin{aligned} u_0(s, w) &= \pi_s - w, \\ u_1(a, w) &= w - c(a), \end{aligned}$$

where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$ is the vector of the expected profits of the firm conditional on s , $c(a)$ is the cost of effort $a \in A$. If the agent decides not to work for the firm ($a = 0$), both players obtain 0 payoff (formally, suppose that there is an additional realization $s = 0$ with $\pi_0 = 0$ that occurs for sure if $a = 0$, and $c(0) = 0$).

2.1. First-best efforts. Let $a^*(t) \in A$ be the (assumed unique and deterministic) first-best effort that maximizes the difference between expected firm profits,¹ $\mathbf{f}(t, a) \cdot \boldsymbol{\pi}$, and effort cost, $c(a)$:

$$a^*(t) = \arg \max_{a \in A} \mathbf{f}(t, a) \cdot \boldsymbol{\pi} - c(a).$$

Let

$$U_0^*(t) = \mathbf{f}(t, a^*(t)) \cdot \boldsymbol{\pi} - c(a^*(t))$$

denote the first-best payoff of type t , and let $\mathbf{U}_0^* = (U_0^*(t_1), \dots, U_0^*(t_\tau))$. We assume $U_0^* \geq 0$, that is, each type generates a weakly positive surplus.

In the benchmark environment without uncertainty about the principal's type, each type can obtain $U_0^*(t)$ by “selling the firm” to the agent. The contract makes the agent the residual claimant of the surplus by asking the agent to pay the principal the expected value of the firm, $U_0^*(t)$, and transfers to the agent the property rights over the profits of the firm, $\boldsymbol{\pi}$.

If the principal type is uncertain, the selling-the-firm contract is not feasible because of adverse selection. The market can unravel, and the parties might fail to implement the efficient outcome (Inderst 2001).

2.2. The informed-principal game. We consider a mechanism-selection game in which any *generalized mechanism* as defined in Myerson (1983) may be proposed by the principal. Following Myerson's terminology, the “public actions” that may be specified by the mechanism are the payment profiles (that is, the mappings from the ex-post verifiable variable s into payments, formally the vectors in $\mathbb{R}^n$), and the agent's “private action” is $a \in \{0\} \cup A$. A generalized mechanism specifies a

¹Let $\cdot$ denote the inner product on $\mathbb{R}^n$.

message space for the principal together with a space of possible recommendations for the agent, and maps messages into probability distributions over public actions and recommendations. The timing is as follows. After privately observing her type t , the principal offers a generalized mechanism M . Then each player chooses her strategy in M (consisting of a type-dependent message for the principal, and a private action for the agent that depends on the recommendation received). If the agent has decided to work for the firm ($a \neq 0$), the ex-post verifiable variable is realized, and the payment according to the payment profile specified by M is implemented; otherwise both players obtain 0 payoff.

A *perfect-Bayesian equilibrium* for the mechanism-selection game specifies (i) for each type of the principal, an optimal (possibly randomized) mechanism proposal, (ii) for each mechanism, a belief about the principal's type that is computed via Bayes rule if the mechanism is proposed by at least one type, and (iii) for each mechanism, a strategy profile that is a sequential equilibrium in the continuation game that follows the proposal of the mechanism.

2.3. Direct mechanisms. Equilibrium outcomes can be described by direct mechanisms. A direct mechanism or allocation ρ is a mapping from the principal's announcements of her type into a payment profile and an effort recommendation to the agent.²

$$\begin{aligned}\rho(\cdot) &= (\alpha(\cdot), w(\cdot)), \\ \alpha(\cdot) : T &\rightarrow \mathcal{A}, \quad \mathbf{w}(\cdot) : T \rightarrow \mathbb{R}^n,\end{aligned}$$

where $\mathcal{A}$ is the set of probability distributions over $\{0\} \cup A$. With some abuse of notation, we use $\alpha(a|t)$ to denote the probability of recommending a after report t , and we use $w_s(t)$ to denote the payment conditional on signal s and report t .

Given a direct mechanism ρ , the expected payoff of type t of the principal if she announces type $\hat{t}$ is denoted

$$U_0^\rho(\hat{t}, t) = \sum_{a \in A} \mathbf{f}(t, a) \cdot (\boldsymbol{\pi} - \mathbf{w}(\hat{t})) \alpha(a|\hat{t}).$$

Similarly, given a direct mechanism ρ and a belief λ about the principal, the expected payoff of the agent who, for each $a \in \{0\} \cup A$, chooses effort $\epsilon(a)$ if he is recommended to choose a , is denoted

$$U_1^{\rho, \lambda}(\epsilon) = \sum_{a \in A} \sum_{t \in T} (\mathbf{f}(t, \epsilon(a)) \cdot \mathbf{w}(t) - c(\epsilon(a))) \alpha(a|t) \lambda_t.$$

We will use the shortcut $U_0^\rho(t) := U_0^\rho(t, t)$.

²Due to additive separability of payoffs, we can restrict attention to mechanisms in which, conditional on each type, the payment profile and the effort recommendation are stochastically independent. Due to risk-neutrality, we do not need to allow randomization over payment profiles.

An allocation ρ is called λ -feasible if both players' incentive constraints—truthful reporting by the principal (1) and obedience by the agent (2)—are satisfied:

$$\forall t, \hat{t} \in T : \quad U_0^\rho(t) \geq U_0^\rho(\hat{t}, t), \quad (1)$$

$$\forall \epsilon : \{0\} \cup A \rightarrow \{0\} \cup A : \quad U_1^{\rho, \lambda}(a) \geq U_1^{\rho, \lambda}(\epsilon), \quad (2)$$

where with some abuse of notation a stands for obeying recommendation about effort choice.

By the revelation principle, any equilibrium outcome can be represented by a λ^0 -feasible allocation.

Observe that (2) requires that, conditional on each recommendation a , the agent cannot gain by choosing not to work for the firm; in particular,

$$U_1^{\rho, \lambda}(a) \geq 0. \quad (3)$$

2.4. Strongly neologism-proof mechanism. Given allocations ρ and ρ' and a belief λ , we say that ρ is λ -dominated by ρ' if ρ' is λ -feasible and

$$\begin{aligned} \forall t \in \text{supp}(\lambda) : \quad & U_0^{\rho'}(t) \geq U_0^\rho(t), \\ \exists t \in \text{supp}(\lambda) : \quad & U_0^{\rho'}(t) > U_0^\rho(t). \end{aligned}$$

Definition 1. An allocation ρ is strongly neologism-proof if (i) ρ is λ^0 -feasible and (ii) ρ is not λ -dominated for any belief λ with support in T .

Strong neologism-proofness is introduced in (Mylovanov and Tröger 2012)³ and is a generalization of the concept of “strongly unconstrained Pareto optimal” (SUPO) allocations of Maskin and Tirole (1990).

2.5. Perfect Bayesian equilibrium. Mylovanov and Tröger (2012), Proposition 1, show that any strongly neologism-proof allocation is a perfect-Bayesian equilibrium allocation. Their result is obtained for environments with private values in which the only private-action choice is the agent's participation decision. Extending the proof to the common-value environments with agent's effort choice that are considered here is straightforward. We present the sketch of the proof and omit the details.

Proposition 1 (Mylovanov and Tröger (2012)). *Any strongly neologism-proof allocation is a perfect-Bayesian equilibrium outcome of the mechanism-selection game.*

Sketch of the proof. The proof is by contradiction. Take a strongly neologism-proof allocation ρ . Suppose that there exists a deviating mechanism M such that, for every belief λ and every λ -feasible allocation implemented by M , at least one type strictly prefers M over ρ . The principal's type-dependent payoff can be summarized as a vector $\mathbf{U}_0 : T \rightarrow \mathbb{R}$. For any belief λ , M determines the set of implementable payoff vectors $M(\lambda)$. On the other hand, for any payoff vector $\mathbf{U}_0$ we can determine the set of beliefs that are Bayesian-consistent with the observation that the principal chooses her payoff from $\mathbf{U}_0$ over her strongly neologism-proof payoff $\mathbf{U}_0^\rho$. By Glicksberg and

³The definition in (2012) includes provisions about types who obtain the highest feasible payoff. Here, there are no such types because payments can be arbitrarily high.

Fan's (or Kakutani's) fixed point theorem, there exist a belief λ^* and a payoff vector $\mathbf{U}_0^*$ that fit together in the sense that both implementability and Bayesian consistency are satisfied. But this contradicts the strong neologism-proofness of ρ because some type strictly prefers U_0^* over $\mathbf{U}_0^\rho$, and hence the allocation that implements $\mathbf{U}_0^*$ λ^* -dominates ρ . $\square$

Corollary 1. *If the first-best payoffs $\mathbf{U}_0^*$ are obtained in a λ^0 -feasible allocation ρ , then ρ is a perfect-Bayesian equilibrium outcome of the mechanism-selection game.*

To see this, observe that, for any belief λ and any λ -feasible allocation ρ' ,

$$\sum_{t \in T} U_0^{\rho'}(t) \lambda_t + U_1^{\rho', \lambda}(a) \leq \sum_{t \in T} U_0^*(t) \lambda_t$$

by definition of the first best. Using (3) with ρ replaced by ρ' , one sees that ρ is not λ -dominated by ρ' , showing that ρ is strongly neologism-proof.

3. FIRST BEST: TYPE-INDEPENDENT EFFORT

In this section, we consider the environment in which the first-best effort is independent of the principal's type. An example of such an environment is a model with two levels of effort, shirking ($a = 1$) and working hard ($a = 2$), in which working hard is the efficient effort level for each type, that is,

$$A = \{1, 2\} \text{ with } \mathbf{f}(t, 2) \cdot \boldsymbol{\pi} - c(2) > \mathbf{f}(t, 1) \cdot \boldsymbol{\pi} - c(1) \quad (t \in T).$$

The results below establish sufficient conditions for the existence of a perfect Bayesian equilibrium in which each type obtains her first-best payoff or, equivalently, existence of a λ^0 -feasible allocation in which the agent undertakes the first-best effort a^* and is paid the expected wage of $c(a^*)$.

In order to illustrate our construction, let $\underline{t} \in \arg \min_T U_0^*(t)$ be a type with minimum surplus. We define an allocation in which type $t = \underline{t}$ of the principal sells the firm to the agent by setting $w_s(\underline{t}) = \pi_s + c(a^*) - \mathbf{f}(\underline{t}, a^*) \cdot \boldsymbol{\pi}$, and all other types t keep the firm and pay a fixed wage $w_s(t) = c(a^*)$.

This allocation is incentive compatible for type $\underline{t}$ because she obtains her first-best payoff regardless of her report. It is also incentive compatible for all other types $t \neq \underline{t}$ because all reports $\hat{t} \neq \underline{t}$ yield the first-best payoff of $U_0^*(t)$, while a deviation to $\hat{t} = \underline{t}$ sells the firm at the lower price of $U_0^*(\underline{t})$.

This direct mechanism is non-transparent for the agent in the sense that the agent cannot infer the type of the principal and, hence, the wage profile when choosing his effort based on the mechanism's recommendation to take action a^* . The agent's incentive constraints (2) are satisfied with a slack conditional on type $\underline{t}$ and are violated otherwise. It follows that if $\underline{t}$ is sufficiently likely, then the mechanism is incentive compatible for the agent; otherwise, the mechanism's construction is more involved and includes "overselling" the firm conditional on type $\underline{t}$ such that the incentive constraints for the agent conditional on this type are satisfied with a sufficiently

large slack. The proof of Proposition 2 extends this construction by allowing general contracts that generate slack in the agent's incentive constraints in addition to a sell-the-firm contract.

Proposition 2. *Suppose that the first-best effort is constant in type, $a^*(t) = a^*$, and that there exists a type $\underline{t}$ and a vector $\mathbf{k} = (k_1, \dots, k_n)$ such that*

$$\mathbf{f}(\underline{t}, a^*) \cdot \mathbf{k} = 0, \quad \mathbf{f}(t, a^*) \cdot \mathbf{k} \geq 0 \text{ for all } t \neq \underline{t}, \text{ and } \mathbf{f}(\underline{t}, a) \cdot \mathbf{k} < 0 \text{ for all } a \in A \setminus \{a^*\}.$$

Then, there exists a perfect-Bayesian equilibrium in which each type obtains her first-best payoff.

Proof. Consider the allocation ρ that always recommends action a^* and pays

$$w_s(t) = \begin{cases} c(a^*) + \gamma k_s & \text{if } t = \underline{t}, \\ c(a^*) & \text{otherwise,} \end{cases}$$

where $\gamma > 0$ is some positive number to be determined.

The allocation satisfies the principal's incentive compatibility constraints (1). Indeed, type $\underline{t}$ obtains the same payoff regardless of her report,

$$U_0^\rho(\underline{t}, \hat{t}) = \mathbf{f}(\underline{t}, a^*) \cdot \boldsymbol{\pi} - c(a^*) = U_0^*(\underline{t}) = U_0^\rho(\underline{t}).$$

For types $t \neq \underline{t}$, (1) writes

$$\mathbf{f}(t, a^*) \cdot \boldsymbol{\pi} - c(a^*) \geq \mathbf{f}(t, a^*) \cdot (\boldsymbol{\pi} - \mathbf{w}(\underline{t}))$$

or, equivalently,

$$\gamma \mathbf{f}(t, a^*) \cdot \mathbf{k} \geq 0,$$

which holds by definition of $\mathbf{k}$.

The agent's payoff from following the recommendation of the direct mechanism is equal to 0 conditional on each type (because $\mathbf{f}(\underline{t}, a^*) \cdot \mathbf{k} = 0$). Hence, the principal obtains the first-best payoffs $\mathbf{U}_0^*$, and the agent's incentive constraint (2) with $\epsilon(a^*) = 0$ is satisfied. Consider now $a = \epsilon(a^*) \neq a^*$ and $\neq 0$. Then

$$U_1^{\rho, \lambda^0}(\epsilon) = \sum_{t \in T} (\mathbf{f}(t, a) \cdot \mathbf{w}(t) - c(a)) \lambda_t^0 = c(a^*) - c(a) + \gamma \lambda_{\underline{t}}^0 \mathbf{f}(\underline{t}, a) \cdot \mathbf{k}. \quad (4)$$

Since $\mathbf{f}(\underline{t}, a) \cdot \mathbf{k} < 0$ by definition of $\mathbf{k}$, setting $\gamma > \frac{c(a^*) - c(a)}{\lambda_{\underline{t}}^0 |\mathbf{f}(\underline{t}, a) \cdot \mathbf{k}|}$ for all $a \notin \{0, a^*\}$ ensures that the expression in (4) is non-positive, (2) is satisfied, and the allocation ρ is λ^0 -feasible. ■

Corollary 2. *Suppose that the first-best effort is constant in type and is the most costly effort for the agent, i.e., $a^*(t) = a^*$ and $c(a^*) \geq c(a)$ for all $a \in A$. Then there exists a perfect-Bayesian equilibrium in which each type obtains her first-best payoff.*

Proof. Let $\underline{t} \in \arg \min_{t \in T} U_0^*(t)$ be a type with minimum surplus. Define $k_s = \pi_s - \mathbf{f}(\underline{t}, a^*) \cdot \boldsymbol{\pi}$.⁴ By construction, $\mathbf{f}(\underline{t}, a^*) \cdot \mathbf{k} = 0$. Furthermore, $\mathbf{f}(t, a^*) \cdot \mathbf{k} =$

⁴Note that if we set $w_s(t) = c(a^*) + k_s$, then $\pi_s - w_s(t) = U_0^*(t)$ is constant in s and the allocation conditional on $\underline{t}$ is "sell the firm".

$(\mathbf{f}(t, a^*) - \mathbf{f}(\underline{t}, a^*)) \cdot \boldsymbol{\pi} \geq 0$ by definition of $\underline{t}$. Finally, for all $a \in A \setminus \{a^*\}$, $\mathbf{f}(\underline{t}, a) \cdot \mathbf{k} = (\mathbf{f}(\underline{t}, a) - \mathbf{f}(t, a^*)) \cdot \boldsymbol{\pi} < 0$ from the assumption that a^* is the unique first-best effort and $c(a) \leq c(a^*)$. Thus, the conditions of Proposition 2 are satisfied. ■

Corollary 3. *Suppose that the first-best effort is constant in type, $a^*(t) = a^*$, and there exists $\underline{t} \in T$ such that $\{\mathbf{f}(t, a^*)\}_{t \in T} \cup \{\mathbf{f}(\underline{t}, a)\}_{a \in A \setminus \{a^*\}}$ is a set of $\tau + m - 1$ linearly independent vectors. Then there exists a perfect-Bayesian equilibrium in which each type obtains her first-best payoff.*

Proof. Fix $x < 0$ and consider the system of linear equations in $\mathbf{k}$

$$\begin{aligned} \mathbf{f}(t, a^*) \cdot \mathbf{k} &= 0 \text{ for all } t \in T, \text{ and} \\ \mathbf{f}(\underline{t}, a) \cdot \mathbf{k} &= x, \text{ for all } a \in A \setminus \{a^*\}. \end{aligned}$$

By the linear-independence assumption, this system has a solution. By construction, the solution satisfies the conditions of Proposition 2. ■

Observe that a necessary condition for the linear-independence condition is that $n \geq \tau + m - 1$.

4. FIRST BEST: TYPE-DEPENDENT EFFORT

In this section, we allow the first-best effort to vary in type and assume it is distinct for each type. Therefore, in any mechanism that implements the first best, the agent will infer the type from the effort recommendation and, unlike in the case of a constant first-best effort considered in the previous section, the principal cannot exploit uncertainty about the wage in order to relax the agent's incentive constraint. As a result, the mechanism implementing the first-best outcome must provide incentives and extract the entire surplus from the agent conditional on each type. Naturally, this requires a somewhat stronger condition on the environment.

Proposition 3. *Suppose that for all $t \in T$ there exists a vector $\mathbf{k}(t) = (k_1(t), \dots, k_n(t))$ such that*

$$\begin{aligned} \mathbf{f}(t, a^*(t)) \cdot \mathbf{k}(t) &= 0, \\ \mathbf{f}(\hat{t}, a^*(t)) \cdot \mathbf{k}(t) &\geq 0 \text{ for all } \hat{t} \neq t, \text{ and} \\ \mathbf{f}(t, a) \cdot \mathbf{k}(t) &< 0 \text{ for all } a \in A \setminus \{a^*(t)\}. \end{aligned}$$

Then there exists a perfect-Bayesian equilibrium in which each type obtains her first-best payoff.

Proof. The proof follows the proof of Proposition 2 and is relegated to the Appendix. ■

Corollary 4. *Suppose that, for all $t \in T$, the set $\{\mathbf{f}(\hat{t}, a^*(t))\}_{\hat{t} \in T} \cup \{\mathbf{f}(t, a)\}_{a \in A \setminus \{a^*(t)\}}$ consists of $\tau + m - 1$ linearly independent vectors. Then there exists a perfect-Bayesian equilibrium in which each type obtains her first-best payoff.*

Proof. The proof is analogous to that of Corollary 3 and is skipped. ■

5. CONCLUSIONS

In this paper, we provide sufficient conditions under which the privately informed principal obtains the same payoff as when her information is commonly known. The conditions hold if the dimensionality of the support of the ex-post verifiable variable is sufficiently large relative to the dimensionality of the type and effort spaces. The observation that assumptions about the relative dimensionality of the effort, the type, and the ex-post verifiable information determine whether principal private information creates signaling distortions can prove useful in studying applications in other moral hazard environments.

APPENDIX

Proof of Proposition 3. Consider the allocation ρ that, for all $t \in T$, recommends effort $a^*(t)$ following the type-announcement t and pays

$$w_s(t) = c(a^*(t)) + \gamma(t)k_s(t),$$

where $\gamma(t) > 0$ is some positive number to be determined.

The principal's incentive-compatibility constraints (1) can be written as

$$\mathbf{f}(t, a^*(t)) \cdot (\boldsymbol{\pi} - c(a^*(t))) \geq \mathbf{f}(t, a^*(\hat{t})) \cdot (\boldsymbol{\pi} - \mathbf{w}(\hat{t})) \text{ for all } \hat{t} \neq t$$

or, equivalently,

$$U_0^*(t) - (\mathbf{f}(t, a^*(\hat{t})) \cdot \boldsymbol{\pi} - c(a^*(\hat{t}))) \geq -\gamma(t)\mathbf{f}(t, a^*(\hat{t})) \cdot \mathbf{k}(\hat{t}) \text{ for all } \hat{t} \neq t.$$

Observe that the left-hand side is non-negative because $a^*(t)$ is the first-best effort. The right-hand side is non-positive by definition of $\mathbf{k}(\hat{t})$. Thus, (1) is satisfied.

The agent's payoff from following the recommendation of the direct mechanism ρ is equal to 0 conditional on each type (because $\mathbf{f}(t, a^*(t)) \cdot \mathbf{k}(t) = 0$). Hence, the principal obtains the first-best payoffs U_0^* .

To verify the agent's incentive constraint (2), it is sufficient to show that, for each type t and $a = \epsilon(a^*(t)) \neq a^*$ and $\neq 0$,

$$\mathbf{f}(t, a^*(t)) \cdot \mathbf{w}(t) - c(a^*(t)) \geq \mathbf{w}(t) \cdot \mathbf{f}(t, a) - c(a).$$

To see this, observe that

$$\mathbf{f}(t, a) \cdot \mathbf{w}(t) - c(a) = c(a^*(t)) - c(a) + \gamma(t)\mathbf{f}(t, a) \cdot \mathbf{k}(t) \quad (5)$$

Since $\mathbf{f}(t, a) \cdot \mathbf{k}(t) < 0$ by definition of $\mathbf{k}(t)$, setting $\gamma(t) > \frac{c(a^*(t)) - c(a)}{|\mathbf{f}(t, a) \cdot \mathbf{k}(t)|}$ for all $a \notin \{0, a^*\}$ ensures that the expression in (5) is non-positive, (2) is satisfied, and allocation ρ is λ^0 -feasible. $\square$

REFERENCES

- BALESTRIERI, F. (2008): "A modified English auction for an informed buyer," mimeo.
 BEAUDRY, P. (1994): "Why an Informed Principal May Leave Rents to an Agent," *International Economic Review*, 35(4), 821–832.
 BENABOU, R., AND J. TIROLE (2003): "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70(3), 489–520.

- BOND, E. W., AND T. A. GRESIK (1997): "Competition between asymmetrically informed principals," *Economic Theory*, 10(2), 227–240.
- CHADE, H., AND R. SILVERS (2002): "Informed principal, moral hazard, and the value of a more informative technology," *Economics Letters*, 74(3), 291–300.
- FLECKINGER, P. (2007): "Informed principal and countervailing incentives," *Economics Letters*, 94(2), 240–244.
- FONG, Y., AND F. Z. X. LEE (2013): "Signaling by An Expert," *mimeo*.
- GROSSMAN, S. J., AND O. D. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51(1), 7–45.
- INDERST, R. (2001): "Incentive Schemes as a Signaling Device," *Journal of Economic Behavior & Organization*, 44(4), 455–465.
- JOST, P.-J. (1996): "On the Role of Commitment in a Principal-Agent Relationship with an Informed Principal," *Journal of Economic Theory*, 68(2), 510–530.
- KARLE, H., AND C. STAAT (2013): "Signaling Quality with Initially Reduced Royalty Rates," *mimeo*.
- KAYA, A. (2010): "When Does It Pay To Get Informed?," *International Economic Review*, 51(2), 533–551.
- MASKIN, E., AND J. TIROLE (1990): "The principal-agent relationship with an informed principal: The case of private values," *Econometrica*, 58(2), 379–409.
- (1992): "The principal-agent relationship with an informed principal, II: Common values," *Econometrica*, 60(1), 1–42.
- MEZZETTI, C., AND T. TSOULOUHAS (2000): "Gathering information before signing a contract with a privately informed principal," *International Journal of Industrial Organization*, 18(4), 667–689.
- MYERSON, R. B. (1983): "Mechanism design by an informed principal," *Econometrica*, 51(6), 1767–1798.
- (1985): "Analysis of Two Bargaining Problems with Incomplete Information," in *Game Theoretic Models of Bargaining*, ed. by A. Roth, pp. 59–69. Cambridge University Press.
- MYLOVANOV, T., AND T. TRÖGER (2012): "Informed-principal problems in environments with generalized private values," *Theoretical Economics*, 7, 465–488.
- (forthcoming): "Mechanism design by an informed principal: Private-values with transferable utility," *Review of Economic Studies*.
- SEVERINOV, S. (2008): "An efficient solution to the informed principal problem," *Journal of Economic Theory*, 141(1), 114–133.
- SKRETA, V. (2009): "On the informed seller problem: optimal information disclosure," *Review of Economic Design*, 15(1), 1–36.
- TAN, G. (1996): "Optimal Procurement Mechanisms for an Informed Buyer," *Canadian Journal of Economics*, 29(3), 699–716.
- YILANKAYA, O. (1999): "A note on the seller's optimal mechanism in bilateral trade with two-sided incomplete information," *Journal of Economic Theory*, 87(1), 125–143.