GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 437

Mechanism Design by an Informed Principal: The Quasi-Linear Private-Values Case

Tymofiy Mylovanov *
Thomas Tröger **

* University of Pennsylvania
** University of Mannheim

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

MECHANISM DESIGN BY AN INFORMED PRINCIPAL: THE QUASI-LINEAR PRIVATE-VALUES CASE

TYMOFIY MYLOVANOV AND THOMAS TRÖGER

Abstract. We show that, in environments with independent private values and transferable utility, a privately informed principal can implement a contract that is ex-ante optimal for her. As an application, we consider a bilateral exchange environment (Myerson and Satterthwaite, 1983) in which the principal is one of the traders. If the property rights over the good are dispersed among the traders, the principal will implement a contract in which she is almost surely better off than if there were no uncertainty about her information. The optimal contract is a combination of a participation fee, a buyout option for the principal, and a resale stage with posted prices and, hence, is a generalization of the posted price that would be optimal if the principal's valuation were commonly known. We also provide a condition under which the principal implements the same contract regardless of whether the agents know her information or not.

Contents

Date: June 26, 2013.

Mylovanov: University of Pennsylvania, Department of Economics, 3718 Locust Walk, Philadelphia, PA 19104, USA, mylovanov@gmail.com. Troeger: University of Mannheim, Department of Economics, L7, 3-5 68131 Mannheim, Germany, troeger@uni-mannheim.de. We are deeply grateful to Simon Board, Yeon-Koo Che, Pierre Fleckinger, Marina Halac, Sergei Izmalkov, Philippe Jehiel, Stephan Lauermann, David Martimort, Eric Maskin, Benny Moldovanu, Moritz Meyer-ter-Vehn, Roger Myerson, Vasiliki Skreta, Sergei Severinov, James Shummer, Bruno Strulovici, Rakesh Vohra, Okan Yilankaya, and Charles Zheng as well as the audiences at University of Bonn, University of Essex, Harvard University, the Hebrew University of Jerusalem, University of Leicester, New Economic School, Northwestern University, Paris School of Economics, University of Pennsylvania, University of Pittsburgh, University of Rochester, Royal Holloway University of London, University of Southern California, Tel-Aviv University, Toulouse School of Economics, the ESSET in Gerzensee, the World Congress of Game Theory Society in Istanbul, and the Canadian Economic Theory Conference in Toronto for the very helpful comments. 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."

1. Introduction

The optimal design of contracts and institutions in the presence of privately informed market participants is central to economics, with applications including auctions, procurement, public good provision, organizational contract design, legislative bargaining, etc.. A restriction in much of this theory is that a contract or a mechanism is designed by a party who has no private information. As such, the theory is not applicable to a large set of environments in which contracts or institutions arise endogenously as a choice of privately informed agents such as in, e.g., collusion, resale, contract renegotiation, bargaining over arbitration procedures, design of international agreements, etc..

In this paper, we study the model in which the contract is selected by one of the privately informed market participants (principal) subject to the agreement by the others (agents). This is a signaling game in which the choice of a contract is a signal about the principal's private information (her type). The value for the principal of any contract is determined endogenously depending on the agents' belief and the continuation play assigned in equilibrium to the contract. Unlike in standard signaling games, the principal's strategy space is enormous and does not possess a nice structure. So far, except for several restricted environments, the informed principal games have been intractable.

We provide a solution to the informed-principal problem in the classic environment with independent private values and monetary transfers. Since the principal has private information, she might want to conceal this information from the agents at the moment of contract selection and become a player in the game induced by the contract she selects. If this occurs, the agents' belief about the principal's type translates into their belief about the principal's actions in this game and affect the agents' calculation of the optimal strategy. This is so even though the values are private and the agents' payoffs do not directly depend on the belief about the principal's information.

Nevertheless, the literature has identified a number of independent-private-values environments in which the principal cannot benefit from concealing her information and any equilibrium of the informed-principal game is outcome-equivalent to a collection of contracts that would be optimal if the type of the principal were commonly known (Myerson 1985, Maskin and Tirole 1990, Tan 1996, Yilankaya 1999, Balestrieri 2008, Skreta 2009). In these environments, the outcomes of the contracts that would be selected by the principal in three different informational environments, before the principal learns her information, if she is privately informed, and if her information is common knowledge among the agents, coincide:[1]

Environments identified in the literature:

Ex-ante contract $\Leftrightarrow$ Interim (informed-principal) contract $\Leftrightarrow$ Ex-post contract

and the informed-principal problem is trivial.

In general, however, this equivalence fails (see Fleckinger 2007 and Section 7 in this paper). Intuitively, the agents' uncertainty about the principal type might allow the principal to extract a higher surplus from the agents through relaxing their individual-rationality constraints.[2]

In the environments in which the principal can benefit from concealing her information from the agents, the informed-principal problem becomes complex and so far has been intractable. The difficulty is that at the moment of contract selection the principal would like to implement a contract that maximizes her payoff given the realized type. Contracts, however, will differ in how they distribute the surplus across the principal types and hence different types would like to choose different contracts. Yet, if different contracts are selected in equilibrium, the principal's type will be revealed to the agents and the principal won't be able to benefit from the agents' uncertainty about her type.

Our main result (Proposition 2) is the existence of the pooling equilibrium in an informed-principal game that implements an ex-ante optimal contract: All principal

[1]In all three environments, the contract is executed after the principal and the agents learn their types. Hence, the contract can condition both on the reports of the principal and the agents. The only difference between the environments is the principal's and the agents' beliefs at the moment of contract selection. The ex-ante contract is selected when both the agents and the principal share prior beliefs about the principal's type, the interim contract is selected when the principal has privately learned her type and the agents hold prior beliefs about the principal's type, and the ex-post contract is selected when the principal's type is commonly known.

[2]Manelli and Vincent (2010) and Gershkov, Goeree, Kushnir, Moldovanu, and Shi (forthcoming) show that in linear independent private values environments Bayesian incentive compatible allocations can be implemented in dominant strategies; this implies that the uncertainty about the principal's type cannot benefit her through the channel of relaxing the incentive compatibility constraints.

types select the same contract that maximizes the principal's weighted expected payoff, where the weights are given by the agents' prior belief about the distribution of principal types:

This paper $->$ all independent-private-values environments with transfers:

$$\text{Ex-ante contract} \Leftrightarrow \text{Interim (informed-principal) contract} \not\Leftrightarrow \text{Ex-post contract}$$

Hence, in environments in which the principal can benefit from keeping her information private, different principal types can jointly realize the entire available surplus even though they have conflicting preferences at the moment of contract selection. The ex-ante optimality result means that it is possible to simultaneously (i) maximize the expected payoff of all principal types and (ii) prevent each type of the principal from deviating to contracts that favor this type at the expense of destroying (disproportionally more) surplus of the other types. (The ex-ante optimality result, in general, fails in the environments without transfers; see Section 3 in the working paper version of this paper.)

This equilibrium outcome can be obtained by maximization of the principal's ex-ante expected payoff subject to interim incentive compatibility and individual rationality constraints. Hence, the ex-ante optimality result reduces the analysis of a complex informed-principal signaling game to a linear maximization program that can be solved using standard methods; it connects the informed-principal problem to the standard mechanism design approach, overcoming intractability issues.

The ex-ante optimality result can be easily applied to diverse environments such as, e.g., public good provision, multiunit or multigood auctions, collusion, legislative bargaining and voting, speculative trade, assignment problems, matching with transferable utility, etc.

We also obtain a general sufficient condition under which that the privacy of the principal's information is irrelevant: monotonicity of each agent's payoff in his *type* for any outcome. This result generalizes and clarifies the logic underlying the existing results in the literature (Myerson 1985, Maskin and Tirole 1990, Tan 1996, Yilankaya 1999, Balestrieri 2008, Skreta 2009). The irrelevance of the principal's private information in linear environments with monotonic payoffs shows that the informed-principal problem can indeed be ignored in many applications.

Nevertheless, the monotonicity of payoffs is not generic and the privacy of the principal's information will, in general, affect the outcome. As an application of our ex-ante optimality result, we focus on a linear environment with two outcomes in which each player is privately informed about their preferences over the outcomes. This environment has multiple interpretations, including bilateral trade, monopoly with an outside option for the buyer, trade in differentiated goods, and renegotiation of a labor contract, among others. In this environment, the principal can benefit from the privacy of her type and will implement a contract in which interim she is almost surely better off than if her type is commonly known if and only if the players' payoffs are non-monotonic.

For concreteness, we focus on the bilateral exchange interpretation of the environment in which the players are traders (Myerson and Satterthwaite (1983)) and one of the traders is designated as the principal.[3] In this application, the payoffs are non-monotonic if and only if the property rights over the good are dispersed among the traders.

We also discuss an indirect implementation of the ex-ante optimal contract for bilateral exchange environments. Its structure is novel and of independent interest. The contract can be implemented via a combination of a participation fee for the agent, a buyout option for the principal, and a resale stage with posted prices. In the first stage, the agent pays the participation fee and the good is tentatively allocated to the agent. In the second stage, the principal decides whether to exercise a buyout option, in which case the good becomes tentatively allocated to the principal. In the third stage, given the tentative allocation of property rights, the principal makes a take-it-or-leave-it offer to the agent to sell or buy the good. The first two stages consolidate the originally dispersed property rights over the good and allocate them either to the principal or the agent, determining whether the principal becomes the seller or the buyer in the third stage.

An important literature is devoted to showing optimality of posted prices in bilateral exchange environments (Riley and Zeckhauser 1983, Myerson 1985, Williams 1987, Yilankaya 1999). Our results indicate that uncertainty about the principal's valuation may generate more complex and interesting contracts. These contracts, nevertheless, are a generalization of a posted price contract that would be optimal in the environments with the extreme property rights allocation in which either the principal or the agent own the good.

1.1. **Related literature on private values.** The problem of mechanism-selection by an informed principal was introduced by Myerson (1983). He uses an axiomatic approach to define a solution and shows that it is always consistent with sequential equilibrium play in a mechanism-selection game. Unfortunately, the characterization in Myerson is abstract. Furthermore, his environment excludes the quasilinear environments studied here.

Maskin and Tirole (1990) consider mechanism-selection by an informed principal in a class of environments with independent private values under a number of specific structural assumptions. The focus of Maskin and Tirole (1990) is on risk-averse players. In their model, if players are risk-neutral so that utility is fully transferable, a privately informed principal uses the same mechanism as when her information is

[3]This environment is equivalent to a partnership dissolution problem (Cramton, Gibbons, and Klemperer 1987) in which one of the parties selects a dissolution mechanism subject to the approval of the mechanism by the other party. Cramton, Gibbons, and Klemperer (1987) have focused on conditions for ex-post efficient implementation. The informed principal, however, will maximize the expected revenue and will distort the allocation from the efficient one to minimize the information rents she has to leave to the agent.

public and the equivalence between the ex-ante optimal and the separating outcomes attains. Similar results are obtained in Myerson (1985), Tan (1996), Yilankaya (1999), Balestrieri (2008), and Skreta (2009). In the application studied in this paper, this "irrelevance" result is non-generic. In bilateral exchange environment, it holds if and only if the property rights are extreme, i.e., one of the parties owns the entire good. For non-extreme allocation of property rights, the principal is almost surely better off in the ex-ante optimal allocation than if her type is commonly known. In the literature, the first example with private values and transferable utility in which there exists a mechanism in which the principal can gain from privacy of her information is due to Fleckinger (2007).

1.2. **Relationship to Mylovanov and Tröger (2012).** Characterizing the entire set of equilibria in the informed principal-game is typically infeasible due to the complexity of the strategy space of the principal. The standard approach in the literature, instead, is to identify certain properties of a mechanism that ensure that it can be supported as an equilibrium outcome. The relevant property for the environments with private values is strong neologism-proofness: All equilibrium outcomes of the informed principal game found in the literature in these environments are strongly neologism-proof. An equilibrium outcome is strongly neologism-proof if no type of the principal can gain from proposing an alternative mechanism that is incentive compatible and individually rational given any belief about the principal that puts probability 0 on types that would strictly lose from proposing the alternative. Strong neologism-proofness is a generalization of the concept of "strongly unconstrained Pareto optimal" (SUPO) allocations of Maskin and Tirole (1990) and is related to some of the concepts in Myerson (1983) (Mylovanov and Tröger (2012)).

In Mylovanov and Tröger (2012), we establish existence of a strongly neologism-proof allocation in a large class of private value environments. The existence is obtained by generalizing the approach by Maskin and Tirole (1990) of considering a fictitious slack-exchange economy in which each type of the principal is a trader in the agents' participation and incentive compatibility constraints. This approach provides an abstract characterization in terms of incentive constraints or convex polyhedra that is *computationally* convenient, but that entails *no economic intuition* that would be useful for concrete applications. In particular, it is not helpful towards obtaining the ex-ante optimality result.[4] In this paper, we derive an envelope characterization of the strongly neologism-proof allocations in quasilinear environments. This characterization is obtained using methods different from the slack-exchange approach. The ex-ante optimality result is an implication of this characterization. The characterization is as follows. Let T_0 denote the principal's type space and let $U_0^\rho(t_0)$ denote the expected payoff of any principal type t_0 in any allocation ρ. Let p_0 be the prior belief about the principal's type. An incentive-feasible allocation ρ is strongly

[4]Furthermore, this approach does not immediately apply to the environments considered in this paper because the space of monetary transfers is unbounded.

neologism-proof if and only if

$$\eta(q_0) \quad \le \quad \int_{T_0} U_0^{\rho}(t_0)\mathrm{d}q_0(t_0) \quad \text{for all } q_0 \text{ absolutely continuous rel. to } p_0, \qquad (1)$$

where $\eta(q_0)$ is the principal's ex-ante optimal payoff given the belief q_0 about the principal, i.e., the maximal expected payoff on the set of allocations that are incentive-feasible with respect to q_0 and prior beliefs about the agents' types.

This is a rather restrictive condition that requires the principal's expected payoff in the allocation corresponding to ρ, when weighed according to q_0, to be not less than the total expected surplus available to the principal if q_0 reflects the agents' belief about the principal, and this condition must hold for all q_0 that are absolutely continuous relative to the prior belief about the principal p_0.

Condition (1) simplifies the expression of strong neologism-proofness considerably: instead of having to compare the principal's payoff in different allocations separately for each of her types, it is sufficient to consider her ex-ante expected payoff in different allocations.

1.3. **Other related literature.** Mechanism-selection by an informed principal in environments with common values was first considered by Myerson (1985) and Maskin and Tirole (1992). In environments with correlated types and a single agent, Cella (2008) shows that the principal benefits from the privacy of her information and Skreta (2009) discuses the optimal disclosure policy for the principal. With correlated types and multiple agents, Severinov (2008) provides an intriguing construction that allows the informed principal to extract the entire surplus. Balkenborg and Makris (2010) look at common value environments and provide a novel characterization of a solution to the informed principal problem. Izmalkov and Balestrieri (2012) study the problem of the informed principal in an environment with horizontally differentiated goods, where the principal is privately informed about the characteristic of the good. Halac (2012) considers optimal relational contracts in a repeated setting where the principal has persistent private information about her outside option. Nishimura (2012) analyzes properties of scoring procurement auctions in an independent private value environment with multidimensional quality and a privately informed buyer. An informed principal problem arises in Francetich and Troyan (2012) who study endogenous collusion agreements in auctions with interdependent values.

Finally, there exists a separate literature that studies the informed-principal problem in moral-hazard environments, rather than in adverse-selection environments considered here (see, for example, Beaudry (1994), Jost (1996), Bond and Gresik (1997), Mezzetti and Tsoulouhas (2000), Chade and Silvers (2002), and Kaya (2010)).

Structure of the paper. In Section 2, we introduce the basic concepts of our model. Section 3 characterizes strongly neologism-proof allocations. The ex-ante optimality result is presented in Section 4. In Section 5, we characterize ex-ante optimal allocations in linear-utility environments. Section 6 provides a condition under which the outcome is independent of whether the agents know the principal's information. The characterization of ex-ante optimal allocations is applied in Section

7 to a class of bilateral-trade environments. Proof details and some additional results are relegated to the Appendix.

2. Model

2.1. **Environment.** Consider players $i = 0, \dots, n$ who have to collectively choose from a space of basic outcomes

$$Z = A \times \mathbb{R}^n,$$

where the measurable space A represents a set of verifiable collective actions, and $\mathbb{R}^n$ is the set of vectors of agents' payments. For example, in an environment where the collective action is the allocation of a single unit of a private good among the players, $A = \{0, \dots, n\}$, indicating who obtains the good.

Every player i has a type $t_i \in T_i$ that captures her private information. A player's type space T_i may be any compact metric space. The product of players' type spaces is denoted $\mathbf{T} = T_0 \times \cdots \times T_n$. The types $t_0, \dots, t_n$ are realizations of stochastically independent Borel probability measures $p_0, \dots, p_n$ with $\text{supp}(p_i) = T_i$ for all i. The probability of any Borel set $B \subseteq T_i$ of player-i types is denoted $p_i(B)$.

Player i's payoff function is denoted

$$u_i : Z \times T_i \to \mathbb{R}.$$

We consider private-value environments with quasi-linear payoff functions,

$$\begin{aligned} u_0(a, \mathbf{x}, t_0) &= v_0(a, t_0) + x_1 + \cdots + x_n, \\ u_i(a, \mathbf{x}, t_i) &= v_i(a, t_i) - x_i, \end{aligned}$$

where $\mathbf{x} = (x_1, \dots, x_n)$, and $v_0, \dots, v_n$ are called *valuation functions*. We assume that the family of functions $(v_i(a, \cdot))_{a \in A}$ is equi-continuous for all i (observe that this assumption is void if type spaces are finite).

The players' interaction results in an outcome that is a probability measure on the set of basic outcomes; the set of outcomes is denoted

$$\mathcal{Z} = \mathcal{A} \times \mathbb{R}^n,$$

where $\mathcal{A}$ denotes the set of probability measures on A, and $\mathbb{R}^n$ is the vector of the agents' expected payments.

If the players cannot agree on an outcome, some exogenously given disagreement outcome $\underline{z}$ obtains. The disagreement outcome $\underline{z} = (\underline{\alpha}, 0, \dots, 0)$ for some (possibly random) collective action $\underline{\alpha} \in \mathcal{A}$. We normalize the valuation functions such that each player's expected valuation from the disagreement outcome equals 0, that is, $\int_A v_i(a, t_i) \mathrm{d}\underline{\alpha}(a) = 0$ for all i and t_i. This is *without loss of generality* since we can always subtract the disagreement payoffs from the payoffs from the basic outcomes.

A player's (expected) payoff from any outcome $(\alpha, \mathbf{x}) \in \mathcal{Z}$ is denoted

$$u_i(\alpha, \mathbf{x}, t_i) = \int_A v_i(a, t_i) \mathrm{d}\alpha(a) - x_i,$$

where $x_0 = -x_1 - \cdots - x_n$.

An *allocation* is a complete type-dependent description of the result of the players' interaction; it is described by a map

$$\rho(\cdot) = (\alpha(\cdot), \mathbf{x}(\cdot)) : \ \mathbf{T} \to \mathcal{Z}$$

such that payments are uniformly bounded (that is, $\sup_{\mathbf{t}\in\mathbf{T}} ||\mathbf{x}(\mathbf{t})|| < \infty$, to guarantee integrability) and such that the appropriate measurability restrictions are satisfied (that is, for any measurable set $B \subseteq A$, the map $\mathbf{T} \to \mathbb{R}$, $\mathbf{t} \mapsto \alpha(\mathbf{t})(B)$ is Borel measurable, and $\mathbf{x}(\cdot)$ is Borel measurable).

An allocation describes the outcome of the informed principal problem as a function of the type profile. Alternatively, an allocation ρ can be interpreted as a *direct mechanism*, where the players $i = 0, \ldots, n$ simultaneously announce types $\hat{t}_i$ (=messages), and the outcome $\rho(\hat{t}_0, \ldots, \hat{t}_n)$ is implemented.

Our analysis focuses on the outcome of the informed principal problem as captured by allocations and abstracts from the specific mechanisms chosen by the principal. An allocation can be implemented by multiple mechanisms and some allocations can be implemented both in pooling equilibria in which all types of the principal offer the same mechanisms and in separating or semi-separating equilibria in which different types of the principal offer distinct mechanisms.

2.2. **Linear-utility environments.** A common assumption in the literature is that each player's valuation function depends linearly on her type. We say that the environment has *linear utilities* if (i) the set of basic collective actions is finite ($A = \{1, \ldots, |A|\}$), (ii) each player's type space is an interval ($T_i = [\underline{t}_i, \overline{t}_i]$), (iii) each player's valuation function $v_i(a, t_i)$ is an affine function of t_i, for all $a \in A$ (that is, there exist numbers s_i^a and c_i^a such that $v_i(a, t_i) = s_i^a t_i + c_i^a$), (iv) there exists of a strictly positive and continuous density f_i for each player's type distribution p_i (and we use F_i to denote the c.d.f.), (v) the disagreement outcome $(\underline{\alpha}, 0, \ldots, 0)$ is such that, for all i, $\int_A s_i^a \mathrm{d}\underline{\alpha}(a) = 0$ and $\int_A c_i^a \mathrm{d}\underline{\alpha}(a) = 0$, and (vi)

$$\forall i \geq 1 \ \exists a_i, b_i \in A : \ s_i^{a_i} \neq s_i^{b_i}. \tag{2}$$

Observe that (v) is not a substantial restriction, but simply expresses that disagreement payoffs are normalized to 0, and (vi) restricts attention to players $i \geq 1$ whose preferences over outcomes actually depend on their private information.

Linear-utility environments provide useful models for many applications, including bilateral exchange, single and multi-unit auctions, procurement, public good provision, non-linear pricing, franchise, legislative bargaining, and assignment problems with transferable utility.[5]

[5]For some recent papers using linear environments see, e.g., Che and Kim (2006), Ledyard and Palfrey (2007), Eliaz and Spiegler (2007), Hafalir and Krishna (2008), Pavlov (2008), Figueroa and Skreta (2009), Garratt, Tröger, and Zheng (2009), Celik (2009), Kirkegaard (2009), Lebrun (2009), Manelli and Vincent (2010), Krähmer (2012), and Gershkov, Goeree, Kushnir, Moldovanu, and Shi (forthcoming).

2.3. **Strongly neologism-proof allocation and equilibrium.** One of the players is designated as the proposer of the allocation. We will assume from now on that the proposer is player 0. We call her the principal; the other players are called agents.

Given the presence of private information, incentive and participation constraints will play a major role in our analysis. Expected payoffs are computed with respect to the prior beliefs $p_1, \dots, p_n$ about the agents' types. However, during the interaction the agents may update their belief about the principal's type, away from the prior p_0. Let q_0 denote a Borel probability measure on T_0 that represents the agents' belief about the principal's type. For our purposes it is enough to work with a belief q_0 that is either absolutely continuous relative to p_0 or is a point belief.

Given an allocation ρ and a belief q_0, the expected payoff of type t_i of player i if she announces type $\hat{t}_i$ is denoted[6]

$$U_i^{\rho,q_0}(\hat{t}_i, t_i) \;=\; \int_{\mathbf{T}_{-i}} u_i(\rho(\hat{t}_i, \mathbf{t}_{-i}), t_i)\mathrm{d}\mathbf{q}_{-i}(\mathbf{t}_{-i}),$$

where $\mathbf{q}_{-i}$ denotes the product measure obtained from deleting dimension i of $q_0, p_1, \dots, p_n$. The expected payoff of type t_i of player i from allocation ρ is

$$U_i^{\rho,q_0}(t_i) \;=\; U_i^{\rho,q_0}(t_i, t_i).$$

We will use the shortcut $U_0^{\rho}(t_0) = U_0^{\rho,q_0}(t_0)$, which is justified by the fact that the principal's expected payoff is independent of q_0.

An allocation ρ is called q_0*-feasible* if, for all players i, the q_0-incentive constraints (3) and the q_0-participation constraints (4) are satisfied,

$$\forall t_i, \hat{t}_i \in T_i: \qquad U_i^{\rho,q_0}(t_i) \geq U_i^{\rho,q_0}(\hat{t}_i, t_i), \tag{3}$$

$$\forall t_i \in T_i: \qquad U_i^{\rho,q_0}(t_i) \geq 0. \tag{4}$$

Given allocations ρ and ρ' and a belief q_0, we say that ρ is q_0*-dominated* by ρ' if ρ' is q_0-feasible and

$$\forall t_0 \in \mathrm{supp}(q_0): \qquad U_0^{\rho'}(t_0) \geq U_0^{\rho}(t_0),$$

$$\exists B \subseteq \mathrm{supp}(q_0),\ q_0(B) > 0\ \forall t_0 \in B: \qquad U_0^{\rho'}(t_0) > U_0^{\rho}(t_0).$$

The domination is *strict* if "$>$" holds for all $t_0 \in \mathrm{supp}(q_0)$.

Characterizing the entire set of equilibria in the informed principal game is typically infeasible due to the complexity of the strategy space of the principal. The standard approach in the literature, instead, is to identify certain properties of a mechanism that ensure that it can be supported as an equilibrium outcome. The relevant property for the environments with private values is strong neologism-proofness. Strong neologism-proofness is a generalization of the concept of "strongly unconstrained Pareto optimal" (SUPO) allocations of Maskin and Tirole (1990) and is related to some of the concepts in Myerson (1983) (Mylovanov and Tröger (2012)).

[6]Here and throughout the paper the principal's beliefs about the agents are held constant and are, consequently, suppressed in the notation.

Definition 1. *An allocation ρ is* strongly neologism-proof *if (i) ρ is p_0-feasible and (ii) ρ is not q_0-dominated for any belief q_0 that is absolutely continuous relative to p_0.*

2.4. **Perfect Bayesian equilibrium.** The following simple argument from Mylovanov and Tröger (2012) shows that such allocations are consistent with equilibrium play in a non-cooperative mechanism-selection game. Consider the principal's choice between either obtaining the payoff from a given strongly neologism-proof allocation or proposing any alternative mechanism. Suppose that some types of the principal propose the alternative mechanism. By Bayesian rationality, this mechanism implements an allocation that is incentive-feasible given a belief that puts probability 0 on the set of types that would strictly lose from proposing the alternative. By definition of strong neologism-proofness, then, no type of the principal has a strict incentive to propose the alternative.[7] Hence, by proposing the strongly neologism-proof allocation as a direct mechanism the principal can solve her mechanism-selection problem.

If types are finite, we consider a mechanism-selection game in which any finite game form with perfect recall may be proposed as a mechanism (cf. Myerson (1983), Maskin and Tirole (1990)); with non-finite type spaces, the game interpretation is informal as there is no "natural" set of feasible mechanisms, nor is there an obvious choice for the definition of equilibrium.

2.5. **Ex-ante optimal allocations.** A core point of our paper will be that strong neologism-proofness is closely related to the ex-ante optimality of an allocation. For any belief q_0, the problem of maximizing the principal's q_0-ex-ante expected payoff across all allocations that are q_0-feasible is

$$\max_{\rho \ q_0\text{-feasible}} \int_{T_0} U_0^\rho(t_0) \mathrm{d}q_0(t_0). \tag{5}$$

Let $\eta(q_0)$ denote the supremum value of the problem. In general, a maximum may fail to exist. This may be because arbitrarily high payoffs can be achieved ($\eta(q_0) = \infty$), or because the supremum cannot be achieved exactly.

Definition 2. *An allocation ρ is* ex-ante optimal *if it solves problem (5) with $q_0 = p_0$.*

Note that any ex-ante optimal allocation is p_0-feasible and, in particular, satisfies the principal's incentive constraints.

3. Characterization of strong neologism-proofness

The main result in this section is a characterization of strong neologism-proofness in quasi-linear environments. We show that strong neologism-proofness requires, for all beliefs q_0 that are absolutely continuous with respect to the prior p_0, that the principal's highest possible q_0-ex-ante expected payoff cannot exceed the q_0-ex-ante expectation of the vector of her strongly neologism-proof payoffs. This envelope characterization greatly simplifies the expression of strong neologism-proofness; it plays a central role in our analysis.

[7] Myerson (1983, Theorem 2) uses a related argument to show that his concept of a strong solution is consistent with equilibrium play.

Proposition 1. *A p_0-feasible allocation ρ is strongly neologism-proof if and only if*

$$\eta(q_0) \quad \leq \quad \int_{T_0} U_0^{\rho}(t_0)\, dq_0(t_0) \quad \text{for all } q_0 \text{ absolutely continuous rel. to } p_0. \qquad (6)$$

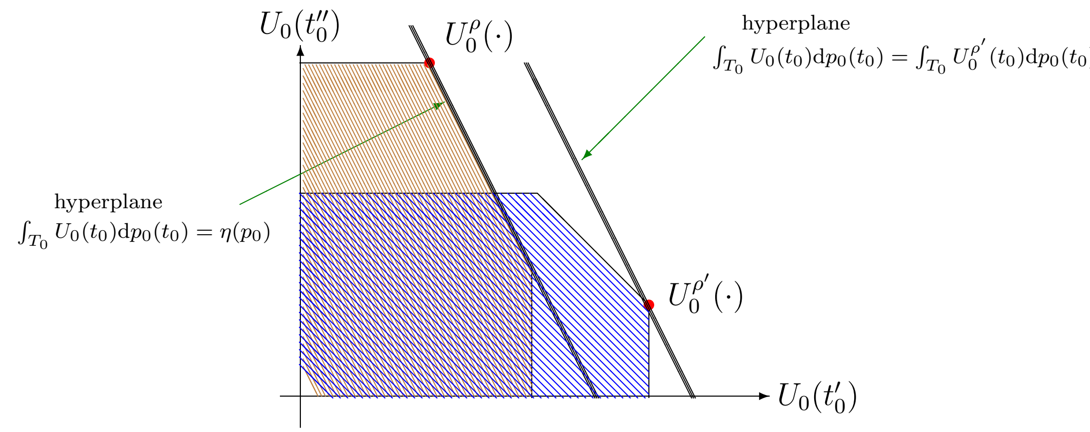

FIGURE 1. Illustration of condition (6) for $T_0 = \{t_0', t_0''\}$. Let ρ and ρ' be two strongly neologism-proof allocations and p_0 and p_0' be the corresponding prior beliefs. The brown and the blue areas are the regions of incentive feasible principal-type payoff vectors for prior beliefs p_0 and p_0' respectively. By (6),

$$\int_{T_0} U_0^{\rho'}(t_0)\mathrm{d}p_0(t_0) \;\; \geq \;\; \int_{T_0} U_0^{\rho}(t_0)\mathrm{d}p_0(t_0) = \eta(p_0),$$
$$\int_{T_0} U_0^{\rho}(t_0)\mathrm{d}p_0'(t_0) \;\; \geq \;\; \int_{T_0} U_0^{\rho'}(t_0)\mathrm{d}p_0'(t_0) = \eta(p_0').$$

We prove the "if" part by showing the counterfactual, which is simple: an allocation that q_0-dominates ρ also yields a strictly higher q_0-ex-ante-expected payoff, and $\eta(q_0)$ is, by definition, not smaller than this payoff.

To prove "only if", we again show the counterfactual. That is, we suppose that, given a strongly neologism-proof allocation ρ, there exists a belief q_0 such that (6) fails. By definition of $\eta(q_0)$, there exists a q_0-feasible allocation ρ' with a strictly higher q_0-ex-ante-expected payoff than ρ. Starting with ρ', by redistributing payments between principal-types we can construct an allocation ρ'' such that each principal-type is strictly better off than in ρ. This may lead, however, to a violation of a principal-type's incentive constraint in ρ''. The remaining, more difficult, part of the proof consists in resurrecting the principal's incentive constraints.

We find a belief r_0 and an allocation σ that r_0-dominates ρ, thereby showing that ρ is not strongly neologism-proof. Starting with the belief q_0 and the allocation ρ'', this can be imagined as being achieved by altering the allocation and the belief multiple times in a procedure that ends with r_0 and σ after finitely many steps.

In environments with finite type spaces, the procedure can be imagined as follows. Suppose ρ'' violates the incentive constraint of some principal-type. We may restrict attention here to types in the support of q_0 (all other types may be assumed to announce whatever type is optimal among the type announcements in the support of q_0). Alter ρ'' by giving the type with the violated constraint a different allocation: the average over what she had and what she is attracted to. Alter q_0 by adding to her previous probability the probability of the type that she was attracted to, and assign this type probability 0. From the viewpoint of the agents (i.e., in expectation over the principal's types), the new allocation together with the new belief is indistinguishable from the old one together with the old belief. Moreover, the new belief has a smaller support. Repeating this procedure leads to smaller and smaller supports, until incentive compatibility is satisfied.

The procedure is more complicated in environments with non-finite type spaces. First, we partition the principal's type space into a finite number of small cells such that when we replace in each cell the allocation by its average across the cell, then the new allocation ρ''' is q_0-almost surely better than ρ. The crucial property of the new allocation is that, in the direct-mechanism interpretation, there exist only *finitely* many essentially different announcements of principal-types. In summary, ρ''' belongs to the set $\mathcal{E}$ of all allocations that (i) have this finiteness property, and (ii) are r_0-almost surely better for the principal than ρ, where (iii) r_0 is any belief such that the agents' r_0-incentive and participation constraints are satisfied (while the principal's constraints are not necessarily satisfied). We consider an allocation σ^* in $\mathcal{E}$ that is minimal with respect to the finiteness property (that is, it is not possible to further reduce the number of essentially different principal-type announcements without violating (ii) or (iii)). Using the averaging idea from the finite-type world, we show that σ^* satisfies the principal's incentive constraints r_0-almost surely. Hence, we can construct an r_0-feasible allocation σ by altering σ^* on an r_0-probability-0 set. Using continuity and the fact that property (ii) holds for σ^*, we conclude that ρ is r_0-dominated by σ.

The complete proof is in the appendix.

Remark 1. *In the appendix (Section 9.3), we show that a strongly neologism-proof allocation exists under weak technical assumptions if the type spaces are finite, and exists in any linear environment.*

These existence results build on the existence result proved in Mylovanov and Tröger (2012). Stochastic independence between the principal's and the agents' types is required for these existence results. Stochastic independence between the agents' types can be dropped if the type spaces are finite and it is not required for the proof of Proposition 1.

4. Ex-ante optimality of strongly neologism-proof allocations

If we set $q_0 = p_0$ in Proposition 1, we can conclude

Proposition 2. *Any strongly neologism-proof allocation is ex-ante optimal.*

The significance of this result is that it connects the complex informed principal problem to the standard mechanism design approach that can be used to characterize ex-ante optimal mechanisms making the informed principal problem tractable in applications.

This result is most convenient in environments where the ex-ante optimal payoffs are unique (such as many environments with continuous type spaces): in such environments there is an essentially unique candidate for a strongly neologism-proof allocation.

Proposition 2 also implies that the issue of the principal's information leakage through the choice of the mechanism imposes no cost on the principal in terms of the total surplus she realizes in equilibrium: Different principal types, despite their conflict of preference about how to allocate the available surplus, can coordinate on a mechanism that maximizes their ex-ante expected total surplus.

A further implication of this result is that in the environments in which the principal learns her type over time, the principal is indifferent about whether to write an ex-ante (long-term) contract or offer a (short-term) contract after her information is realized; this might explain why sometimes we do not observe complete long-term contracts.

5. Ex-ante optimality in linear-utility environments

In this section, we provide a characterization of ex-ante optimality in linear-utility environments. Importantly, in contrast to the main characterization result in Ledyard and Palfrey (2007), we do not restrict attention to environments with monotonic payoffs (cf. Definition 4 below). Allowing non-monotonicities is crucial for many of the applications because it amounts to allowing arbitrary disagreement outcomes. Technically, the challenge arising from non-monotonicities is that the participation constraint is not necessarily binding for the lowest agent type.[8]

In a linear-utility environment, the payoff-relevant aspects of the collective-action choice are captured by the set

$$\mathcal{V} \;=\; \{(\hat{s}_0,\ldots,\hat{s}_n,\hat{c}_0,\ldots,\hat{c}_n)\in \mathbb{R}^{2n+2} \mid \exists \alpha\in\mathcal{A}\ \forall i:\ \hat{s}_i=\int_A s_i^a \mathrm{d}\alpha(a),\ \hat{c}_i=\int_A c_i^a \mathrm{d}\alpha(a)\}.$$

Therefore, we can think of an allocation as directly determining a vector $(s_0(\mathbf{t}),\ldots,c_n(\mathbf{t}))\in\mathcal{V}$ for any type profile $\mathbf{t}$. Also, instead of determining payments we can think of an allocation as directly determining the players' utilities $u_0(\mathbf{t}),\ldots,u_n(\mathbf{t})$ (a player's payment is then given by $s_i(\mathbf{t})t_i + c_i(\mathbf{t}) - u_i(\mathbf{t})$).

[8]Formally, this is equivalent to allowing linear type-dependent disagreement payoffs. Jullien (2000) analyzes mechanism design with linear and non-linear type-dependent disagreement payoffs, but in his model there is only one agent and the principal has no private information.

We also use the following shortcuts

$$\begin{aligned}\overline{s}_i(t_i) &= \int_{T_{-i}} s_i(\mathbf{t})\mathrm{d}F_{-i}(\mathbf{t}_{-i}),\\ \overline{c}_i(t_i) &= \int_{T_{-i}} c_i(\mathbf{t})\mathrm{d}F_{-i}(\mathbf{t}_{-i}),\\ \overline{u}_i(t_i) &= \int_{T_{-i}} u_i(\mathbf{t})\mathrm{d}F_{-i}(\mathbf{t}_{-i}).\end{aligned}$$

For all $i \geq 1$ and c.d.f.s $z_i^*(\cdot)$ on T_i, define the virtual valuation function

$$\psi_i^{z_i^*}(t_i) \;=\; t_i - \frac{z_i^*(t_i) - F(t_i)}{f_i(t_i)} \quad (t_i \in T_i).$$

The *ironed virtual valuation* $\overline{\psi}_i^{z_i^*}$ is defined as follows.[9] Let

$$H_i(q) \;=\; \int_0^q \psi_i^{z_i^*}(F_i^{-1}(r))\mathrm{d}r \quad (q \in [0,1]).$$

Let $\overline{H}_i$ denote the convex hull of H_i. Because $\overline{H}_i$ is convex, its derivative exists Lebesgue-a.e. and is weakly increasing; let $\overline{H}_i'$ be a weakly increasing extension to $[0,1]$ and define

$$\overline{\psi}_i^{z_i^*}(t_i) \;=\; \overline{H}_i'(F_i(t_i)).$$

One can think of $\overline{\psi}_i^{z_i^*}(\cdot)$ as constructed by ironing the non-monotonicities of $\psi_i^{z_i^*}(\cdot)$.

We characterize ex-ante optimality in terms of virtual-surplus maximization. For all $v = (\hat{s}_0, \dots, \hat{c}_n) \in \mathcal{V}$ and $\mathbf{t} \in \mathbf{T}$, define the virtual surplus function

$$V^{z_1^*,\dots,z_n^*}(v,\mathbf{t}) \;=\; \hat{s}_0 t_0 + \hat{c}_0 + \sum_{i=1}^n \hat{s}_i \overline{\psi}_i^{z_i^*}(t_i) + \hat{c}_i. \tag{7}$$

Here is the existence and characterization result.[10]

Proposition 3. *In any linear-utility environment, an ex-ante optimal allocation exists. An allocation $u_0(\mathbf{t}),\dots,u_n(\mathbf{t}),s_0(\mathbf{t}),\dots,c_n(\mathbf{t})$ is ex-ante optimal if and only if there exist c.d.f.s z_i^* on T_i $(i = 1,\dots,n)$ such that the following conditions hold:*

$$\forall i \geq 1,\ t_i \in \mathit{supp}(z_i^*):\ \overline{u}_i(t_i) = 0, \tag{8}$$

$$(s_0(\mathbf{t}),\dots,c_n(\mathbf{t})) \in \arg\max_{v\in\mathcal{V}} V^{z_1^*,\dots,z_n^*}(v,\mathbf{t}), \quad \textit{a.e. } \mathbf{t}, \tag{9}$$

$$\overline{s}_i(\cdot) \textit{ is weakly increasing for all } i \geq 0, \tag{10}$$

$$\overline{u}_i(t_i) = \overline{u}_i(\underline{t}_i) + \int_{\underline{t}_i}^{t_i} \overline{s}_i(y)dy \textit{ for all } i \geq 0, t_i \in T_i. \tag{11}$$

$$\overline{u}_i(t_i) \geq 0 \textit{ for all } i \geq 0, t_i \in T_i. \tag{12}$$

[9] The construction follows Myerson (1981), who considered the case $z_i^*(t_i) = 1$.

[10] The if-and-only-if part of this result holds for arbitrary c.d.f.s F_0 on T_0; the assumptions concerning smoothness of F_0 are not needed.

The core part of the conditions is the virtual-surplus maximization (9). If this maximization problem has a unique solution, then $s_i(\mathbf{t})$ is automatically weakly increasing in t_i, for any t_{-i}; in general, however, (10) is an independent condition. The Lagrange multiplier functions z_i^* indicate which agent types' participation conditions have bite; condition (8) requires that z_i^* puts all its mass on types for which the participation constraint is binding. The envelope condition (11) requires that payments are chosen such that all players' incentive constraints are satisfied. The participation constraints are (12).

The proof of Proposition 3 begins with the observation that the solutions to the principal's F_0-ex-ante problem are precisely the solutions to the principal's relaxed F_0-ex-ante problem in which the principal's incentive and participation constraints are satisfied (cf. Proposition 9 and Corollary 2). In order to characterize the solutions to the relaxed problem, we take a Lagrangian approach. The crucial insight is to take a Lagrangian approach only with respect to the agents' participation constraints (30), and *not* with respect to the monotonicity constraints (28). The monotonicity constraints are treated with a generalization of the ironing techniques of Myerson (1981). The details are in the appendix.

6. Irrelevance of privacy of the principal's information

Proposition 1 also implies that the question of whether or not the principal benefits from the uncertainty about her information or, equivalently, offers an allocation that differs from what she would if her information were commonly known ("best separable allocation") boils down to the question of whether or not a best separable allocation is ex-ante optimal for various beliefs about the principal's type, as stated in the proposition below.

Definition 3. *An allocation ρ is* best separable *if, for all point beliefs q_0, ρ is q_0-feasible and ρ is not q_0-dominated.*

Observe that the concept of a best separable allocation[11] is entirely independent of the agent's prior belief p_0. The principal would optimally propose a best separable allocation if her type were commonly known, that is, if the agents did have a point belief about the principal's type. Equivalently, a best separable allocation will be selected if the principal is restricted to offer a mechanism in which she is not a player herself.[12,13]

[11]Maskin and Tirole (1990) use the term *full-information optimal allocation* instead.

[12]Zheng (2002) calls such mechanisms "transparent".

[13]In independent-private-values environments, a best separable allocation which is not p_0-dominated is a strong solution (Myerson 1983). By definition of strong neologism proofness, a strongly neologism-proof allocation cannot be dominated by a best separable allocation, while a best separable allocation ρ is a strong solution if it is strongly neologism-proof. Moreover, a best separable allocation is the only candidate for a strong solution. If both a strongly neologism-proof allocation and a strong solution exist, then they lead to the same principal-type payoff vector.

Proposition 4. *A best separable allocation is strongly neologism-proof if and only if it solves problem (5) for all q_0 that are absolutely continuous relative to p_0.*

Proof. "if" is immediate from Proposition 1. To see "only if", consider a best separable allocation ρ that is strongly neologism-proof. As a best separable allocation, ρ is q_0-feasible for *all* beliefs q_0. Hence, it solves problem (5) by Proposition 1. *QED*

Proposition 4 can also be used to understand when restrictions on the class of mechanisms available to the principal, often made in applied models, are with loss of generality. For instance, a best separable allocation will be selected if the principal is restricted to offer a mechanism in which she is not a player herself; if a best separable allocation is dominated given the prior or some other belief, it is not strongly neologism-proof. Similarly, if an equilibrium allocation in a semi-separating or a pooling equilibrium of an informed principal game with a restricted set of mechanisms is dominated given some belief, e.g., the belief that put the entire mass on the set of separating types, it is not strongly neologism-proof.

Mylovanov and Tröger (2012) show that a strongly neologism-proof allocation exists in independent-private-values environments that satisfy a separability condition. In the appendix, we show that a strongly neologism-proof allocation exists in quasilinear environments if the type spaces are finite or the environment is linear. In all these environments, a best separable allocation is a strong solution if and only if it is strongly neologism-proof.[14]

Our next results describe a sufficient condition under which the solution of the informed principal problem is independent of the amount of the agents' uncertainty about the principal's type. It is the following monotonicity condition.

Definition 4. *In linear environments, the players' payoffs are monotonic in types if and only if*[15]

$$s_i^a \geq 0 \quad \text{for all } i,\ a. \tag{13}$$

A best separable allocation exists in any linear environment. Moreover, if payoffs are monotonic, then such an allocation satisfies the conditions given in Proposition 3 when F_0 is any point belief. Here, for all agents i, z_i^* puts all its weight on the lowest type $\underline{t}_i$. The crucial observation is that the best allocation in fact satisfies the conditions given in Proposition 3 for arbitrary beliefs F_0. Hence, using footnote 10 and Proposition 4, we obtain the following result.

Proposition 5. *Consider a linear environment with monotonic payoffs. Then any best separable allocation is ex-ante optimal and strongly-neologism proof.*

[14]Myerson (1983, Theorem 2) proves that in any environment with finite type spaces and a finite outcome space, a strong solution is a perfect Bayesian equilibrium outcome of an informed-principal game where any finite simultaneous-move game form is a feasible mechanism.

[15]Assuming weakly increasing payoff functions is without loss of generality. If the payoff of some player i is weakly decreasing for all actions, we can redefine her types as $\hat{t}_i = -t_i$ and obtain a weakly increasing payoff function.

The interpretation of Proposition 5 is simple: if payoffs are monotonic, then the privacy of the principal's private information is irrelevant for mechanism design—she will offer the same mechanism as when her type is commonly known, and she does not gain from being a player in her own mechanism.

Much of the intuition for the monotonicity condition can be gained from the analysis of a bilateral trade environment in which a single unit is traded (Myerson and Satterthwaite, 1983). The optimal mechanism for the seller (principal) if her type (=cost) is commonly known is a posted price. Hence, the best separable allocation arises from a collection of these optimal posted prices. Under a regularity condition on the distribution of the agent's valuation that requires monotonicity of the virtual valuation function, a best separable allocation arises from pointwise maximization of the virtual surplus function. Any ex-ante optimal allocation is a best separable allocation because it still arises from pointwise maximization of the virtual surplus function (e.g., Ledyard and Palfrey, 2007). The reason this works is that, whatever type the principal has, the agent's equilibrium payoff is always increasing in the agent's type. Hence, it is always the same lowest-valuation type of the agent for whom the participation constraint is binding. In other words, any best separable allocation is ex-ante optimal because the agent's payoff is monotonic in her type.

The irrelevance result extends to environments in which the regularity condition fails because the virtual surplus function is additively separable in types of different players and multiplicatively separable in the agent's virtual utility $\psi_i(t_i)$ and the marginal value of action s_i^a. Consequently, the objective in (9) is additively separable in the agents' and the principal's types and the type-wise maximizer of (9) is not affected by the belief about the principal's type.

The irrelevance result has been previously observed in a number of environments (Myerson 1985, Maskin and Tirole 1990, Tan 1996, Yilankaya 1999, Balestrieri 2008, Skreta 2009).[16] Proposition 5 illustrates the unifying feature underlying these results.

7. Application: Bilateral trade

In this section, we impose more structure on linear environments and use Proposition 3 to study properties of ex-ante optimal allocations.

We consider a linear-utility environment with one agent ($n = 1$) and two outcomes ($A = \{0, 1\}$). The type spaces are $T_0 = T_1 = [0, 1]$. We assume that the agent's type distribution F_1 has strictly increasing virtual valuation functions $\psi^b(t_1) = t_1 - (1 - F_1(t_1))/f_1(t_1)$ and $\psi^s(t_1) = t_1 + F_1(t_1)/f_1(t_1)$.

Any probability distribution on $A = \{0, 1\}$ can be described by the probability $\alpha \in [0, 1]$ of action 1. Let $0 < \underline{\alpha} < 1$ denote the probability of action 1 upon disagreement. The disagreement payoffs are normalized to 0 and player i's ($i = 0, 1$) valuation function is given by $v_i(a, t_i) = s_i^a t_i$, where

$$s_0^a = \mathbf{1}_{a=0} - (1 - \underline{\alpha}), \quad s_1^a = \mathbf{1}_{a=1} - \underline{\alpha}.$$

[16] A related result is Nishimura (2012) in an environment with generalized private values (Mylovanov and Tröger 2012).

Observe that $0 < \underline{\alpha} < 1$ implies that s_i^0 and s_i^1 have different signs and, therefore, the payoffs are non-monotonic in type.

Applications. *Bilateral exchange.* Our primary application is bilateral exchange: the set of collective actions $A = \{0, 1\}$ indicates who gets assigned one unit of an indivisible good. A player's type represents her valuation of the good, the disagreement outcome is the good is assigned to the agent with probability $\underline{\alpha}$ and to the principal with probability $1 - \underline{\alpha}$, and payoffs are written such that each player's payoff from the disagreement outcome is normalized to 0.

Procurement. Fleckinger (2007) considers a non-linear procurement environment with countervailing incentives (Lewis and Sappington 1989, Jehiel, Moldovanu, and Stacchetti 1999, Jullien 2000). Our model captures its linear version. The agent can produce a good. The fixed costs of production are decreasing in the agent's type, while the marginal costs are increasing in the agent's type. We can express the fixed and the marginal costs respectively as $C_0 - \underline{\alpha} t_1$ and qt_1, where $q \in [0, 1]$ is the amount of production, the agent's total cost can be written as $C_0 + (q - \underline{\alpha})t_1$. Since we allow for randomization on $A = \{0, 1\}$, the agent's payoff is equivalent to that in the above model up to constant C_0. The principal's valuation function is equal to qt_0. While in this setting the principal's payoff is different from that in bilateral exchange, the characterization of ex-ante optimal allocation in Proposition 6 continues to hold.

Renegotiation of a labor contract. The principal is a company and the agent is a union. The disagreement outcome $\underline{\alpha}$ represents the amount of labor union members are required to supply to the firm per existing contract. There is an exogenous shock to the economy affecting the value of labor for the firm and the union; the realized values of labor are the parties private information. The total amount of labor is normalized to 1, and the transfer represents a net change in wage over the current contract.

Priceline. The principal is a seller and the agent is a buyer. The seller is a monopolist who sells hotel nights. There are two locations. Action $a \in \{0, 1\}$ indicates the location of the hotel given to the agent. The types capture relative valuations of the two locations. If the agent rejects the contract, he goes to the market in which he expects to find a hotel in the first location with probability $\underline{\alpha}$.

In what follows, we use the partnership dissolution interpretation of the model. That is, the environment is the standard two-party divisible-good exchange setting of Myerson and Satterthwaite (1983) under the assumption that one party is designated as the principal and, as in Cramton, Gibbons, and Klemperer (1987), the disagreement outcome reflects the share of the good owned by each party (the disagreement outcome may also include a side payment which we normalize to 0).[17] We allow for non-extreme property rights.

The following result describes the unique ex-ante optimal allocation, which by Proposition 2 is an equilibrium outcome of the mechanism-selection game.

Proposition 6. *Consider the bilateral-trade environment with non-extreme property rights. There exists an a.e. unique ex-ante optimal allocation* $\rho(\cdot) = (\alpha(\cdot), \mathbf{x}(\cdot))$,

$$\alpha(t_0, t_1) \quad = \quad \begin{cases} 0 & \text{if } t_0 < t_0^*, \ \psi^s(t_1) < t_0, \\ 1 & \text{if } t_0 < t_0^*, \ \psi^s(t_1) > t_0, \\ 0 & \text{if } t_0 > t_0^*, \ \psi^b(t_1) < t_0, \\ 1 & \text{if } t_0 > t_0^*, \ \psi^b(t_1) > t_0, \end{cases}$$

where $t_0^* = F_0^{-1}(\underline{\alpha})$ *and* $\mathbf{x}(\cdot)$ *is chosen such that* ρ *is* F_0*-feasible, and such that the participation constraints of the agent-types in the interval* $[(\psi^s)^{-1}(t_0^*), (\psi^b)^{-1}(t_0^*)]$ *are satisfied with equality.*

The proof of this result consists of a computation that uses the conditions provided in Proposition 3; the details are in the Appendix.

Observe that in the ex-ante optimal allocation there is trade with probability 1 and the allocation is deterministic. The outcome is sometimes less efficient than the disagreement outcome and the entire good is sometimes allocated to the party with a lower valuation.[18]

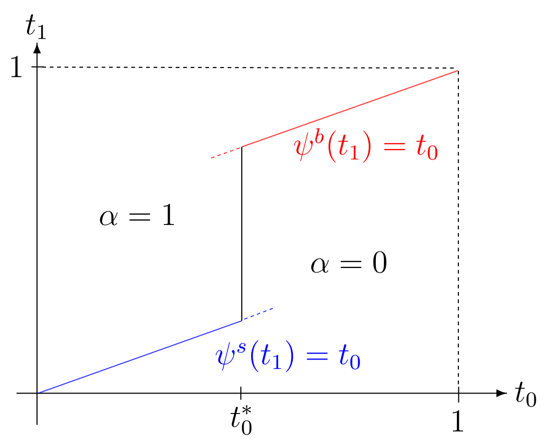

FIGURE 2. The ex-ante optimal allocation in a bilateral-trade environment.

[17]Cramton, Gibbons, and Klemperer (1987) shows that dispersed property rights might allow implementing an ex-post efficient allocation. The informed principal, however, will find it optimal to distort the allocation away from the efficient one in order to extract higher rents from the agent.

[18]Figueroa and Skreta (2009) present an environment with type-dependent outside options in which the optimal mechanism includes overselling. This type of inefficiency is caused by the structure of the outside option designed by the principal; there is no uncertainty about the principal's valuation in their model.

We compare the ex-ante optimal allocation to the best-separable allocation — the allocation that the principal would optimally propose if her type were commonly known. Using methods very similar to those used in the proof of Proposition 6, we obtain the following result.

Proposition 7. *Consider the bilateral-trade environment with non-extreme property rights. There exists an a.e. unique best-separable allocation* $\rho(\cdot) = (\alpha(\cdot), \mathbf{x}(\cdot))$,

$$\alpha(t_0, t_1) = \begin{cases} 0 & \text{if } \psi^s(t_1) < t_0, \\ 1 & \text{if } \psi^b(t_1) > t_0, \\ \underline{\alpha} & \text{otherwise,} \end{cases}$$

and $\mathbf{x}(\cdot)$ *is chosen such that, for all* $t_0 \in T_0$*, if the agent believes in type* t_0*, then* ρ *is incentive-feasible and the participation constraints of the agent-types in the interval* $[(\psi^s)^{-1}(t_0), (\psi^b)^{-1}(t_0)]$ *are satisfied with equality.*

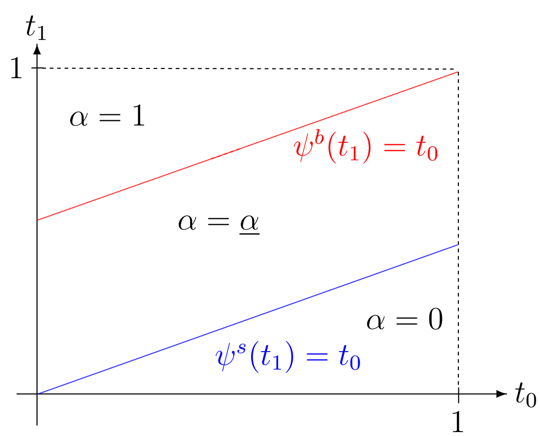

FIGURE 3. The best-separable allocation in a bilateral-trade environment.

Hence, in the best-separable allocation, in contrast to the ex-ante optimal allocation, each type of the principal fails to trade with the agent with a positive probability ($= F_1((\psi^b)^{-1}(t_0)) - F_1((\psi^s)^{-1}(t_0))$) and when the trade occurs it increases efficiency relative to the disagreement outcome. Because the ex-ante optimal allocation is strongly neologism-proof, each type of the principal is at least as well off as in the best-separable allocation. In fact, due to the additional volume of trade in the ex-ante optimal allocation relative to the best-separable allocation, the envelope formula (11) implies that the difference $\overline{u}_0(t_0) - \overline{u}_0(t_0^*)$ between the expected utilities of type t_0^* and any other type t_0 is larger for the ex-ante optimal allocation than for the best-separable allocation. Therefore:

Corollary 1. *In the bilateral-trade environment with non-extreme property rights, all types* $t_0 \neq t_0^*$ *of the principal are strictly better off in the ex-ante optimal allocation than in the best-separable allocation.*

Thus, the principal can use the privacy of her information in order to increase her payoff.

Yilankaya (1999) shows that, if the default allocation of the property rights is extreme ($\underline{\alpha} = 0$ or $\underline{\alpha} = 1$), then the uncertainty of the principal's valuation plays no role and she will implement a best-separable allocation by making, e.g., a posted price offer.

The best-separable allocation described in Proposition 7 can be implemented by using, for each type t_0, a bid price of $(\psi^s)^{-1}(t_0)$ and an ask price of $(\psi^b)^{-1}(t_0)$.

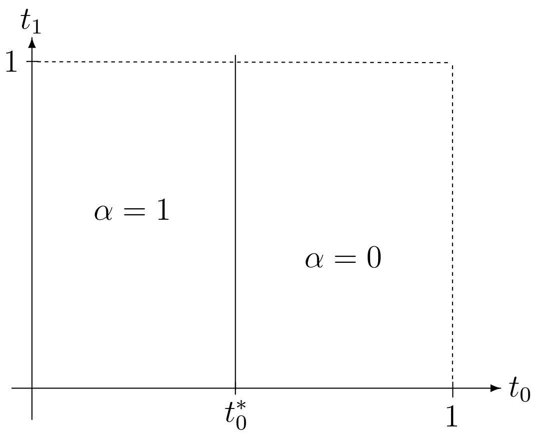

FIGURE 4. The outcome of the second stage in the three stage mechanism implementing the ex-ante optimal allocation in a bilateral-trade environment.

In contrast, the ex-ante optimal allocation described in Proposition 6 is implemented by a multi-stage mechanism involving a combination of a participation fee for the agent, a buyout option for the principal, and a resale stage with posted prices: In the first stage, the agent pays the participation fee and the good is tentatively allocated to the agent. In the second stage, the principal decides whether to exercise a buyout option, in which case the good becomes tentatively allocated to the principal; this option will be exercised by the types $t_0 > t_0^*$ of the principal. In the third stage, given the tentative allocation of the good, the principal makes a take-it-or-leave-it fixed-price offer to the agent to sell or buy the good. Hence, the first two stages consolidate the originally dispersed property rights to the good and allocate the good either to the principal or the agent, determining whether the principal becomes the seller or the buyer in the third stage. This mechanism is a generalization of the bid and ask price mechanism that implements the best separable allocation as well as a generalization of a posted price mechanism that would be optimal in the environments with the extreme property rights allocation in which either the principal or the agent own the good (Williams 1987, Yilankaya 1999).

An important lesson from the bilateral-trade application is that the ex-ante optimal mechanism differs from the mechanism the principal would offer if her valuation were commonly known. The intuition for why the principal strictly gains from the privacy of her information is as follows. A low-valuation principal will set low prices. Hence, when dealing with a low-valuation principal, many agent-types will get the good. Normalizing disagreement payoffs to 0, this implies that the agent's payoff will be increasing over a relatively large range of her type space, implying that the agent's participation constraint will be binding for relatively low agent types. Vice versa, when dealing with a high-valuation principal, the agent's participation constraint will be binding for relatively high types. In summary, the agent's participation constraint will be binding for different types, depending on which principal type the agent is dealing with.

In an ex-ante optimal allocation, the agent's participation (and incentive) constraints are only required to hold in expectation over the principal's types. As a result, in the ex-ante optimal allocation the principal can extract more rents than if her valuation is commonly known. In the multi-stage mechanism implementing the ex-ante optimal allocation, at the moment of accepting the mechanism and paying the participation fee, the agent is kept in the dark about the principal's type and is uncertain whether the principal will exercise her buy-out option. The agent's participation constraint can be violated conditional on a particular type of the principal, but is satisfied in expectation over the principal's types.

8. Conclusions

This paper considers the informed principal problem in the environments with independent private values and transferable utility. Our main result is that an informed principal can implement an allocation that maximizes her ex-ante expected payoff. This result holds for the environments much more general than those considered in the literature so far and clarifies which results from the existing literature extend to more general environments. It also reduces the analysis of a complex informed-principal signaling game to a linear maximization program that can be solved using standard methods and, thus, it connects the informed-principal problem to the standard mechanism design approach.

An important feature of the model that differentiates it from most of the literature is that the principal can benefit from concealing her private information from the agents at the moment of contract selection. As an illustration, we consider a bilateral exchange environment in which the principal is one of the traders. We show that if the property rights over the good are dispersed among the traders, then the principal will implement a contract in which she is almost surely better off than if her type is commonly known. The optimal contract is a combination of a participation fee, a buyout option for the principal, and a resale stage with posted prices and, hence, is a generalization of the posted price that would be optimal if the principal's valuation were commonly known.

The ex-ante optimality results relies on the assumptions of private values and quasi-linear preferences; it is easy to construct examples where this result fails if either of the assumptions is relaxed. It would be interesting to explore whether our basic approach - reformulating the informed principal problem as an appropriate optimization problem - extends to other settings, including those with interdependent values, non-quasilinear preferences, and moral hazard.

9. Appendix

9.1. **Proof of Proposition 1.** "if" Suppose that ρ is not strongly neologism-proof. Then there exists a belief q_0 and an allocation ρ' that q_0-dominates ρ. We obtain a contradiction because

$$\eta(q_0) \geq \int_{T_0} U_0^{\rho'}(t_0)\mathrm{d}q_0(t_0) > \int_{T_0} U_0^{\rho}(t_0)\mathrm{d}q_0(t_0).$$

"only if". Consider a strongly neologism-proof allocation $\rho = (\alpha(\cdot), x_1(\cdot), \ldots, x_n(\cdot))$.

Suppose there exists a belief q_0 such that (6) fails, that is

$$\eta(q_0) \quad > \quad \int_{T_0} U_0^{\rho}(t_0)\mathrm{d}q_0(t_0).$$

By definition of $\eta(q_0)$, there exists a q_0-feasible allocation $\rho' = (\alpha'(\cdot), x_1'(\cdot), \ldots, x_n'(\cdot))$ such that

$$\int_{T_0} U_0^{\rho'}(t_0)\mathrm{d}q_0(t_0) - \int_{T_0} U_0^{\rho}(t_0)\mathrm{d}q_0(t_0) \stackrel{\text{def}}{=} \epsilon > 0. \tag{14}$$

Let $\rho'' = (\alpha'(\cdot), x_1''(\cdots), \ldots, x_n''(\cdot))$, where

$$\begin{aligned} x_1''(\mathbf{t}) &= x_1'(\mathbf{t}) - (U_0^{\rho}(t_0) - U_0^{\rho'}(t_0) + \epsilon). \\ x_i''(\mathbf{t}) &= x_i'(\mathbf{t}), \quad i = 2, \ldots, n. \end{aligned} \tag{15}$$

Then ρ'' satisfies the q_0-incentive and participation constraints for all $i \notin \{0, 1\}$. Also, ρ'' satisfies the q_0-incentive and participation constraints for $i = 1$ because

$$\begin{aligned} U_1^{\rho'',q_0}(\hat{t}_1, t_1) &= \int_{T_{-1}} \int_A v_1(a, t_1)\mathrm{d}\alpha'(\hat{t}_1, \mathbf{t}_{-1})(a)\mathrm{d}\mathbf{q}_{-1}(\mathbf{t}_{-1}) - \int_{T_{-1}} x_1''(\hat{t}_1, \mathbf{t}_{-1})\mathrm{d}\mathbf{q}_{-1}(\mathbf{t}_{-1}) \\ &\stackrel{(15)}{=} U_1^{\rho',q_0}(\hat{t}_1, t_1) + \int_{T_0} (U_0^{\rho}(t_0) - U_0^{\rho'}(t_0))\mathrm{d}q_0(t_0) + \epsilon \\ &\stackrel{(14)}{=} U_1^{\rho',q_0}(\hat{t}_1, t_1). \end{aligned}$$

For all $t_0 \in T_0$,

$$U_0^{\rho''}(t_0) - U_0^{\rho}(t_0) \stackrel{(15)}{=} U_0^{\rho'}(t_0) + (U_0^{\rho}(t_0) - U_0^{\rho'}(t_0) + \epsilon) - U_0^{\rho}(t_0) = \epsilon. \tag{16}$$

In other words, in ρ'' every type of the principal is—by the amount ϵ—better off than in ρ. In particular, ρ'' satisfies the participation constraints for $i = 0$. However, ρ'' may violate a incentive constraint for $i = 0$.

To complete the proof, we show that there exists a belief r_0 and an r_0-feasible allocation σ such that, for all $t_0 \in \mathrm{supp}(r_0)$,

$$U_0^{\sigma}(t_0) \quad \geq \quad U_0^{\rho}(t_0) + \frac{1}{2}\epsilon. \tag{17}$$

It follows that ρ is r_0-dominated by σ; this contradicts the strong neologism-proofness of ρ.

Because v_0 is equi-continuous and T_0 is compact, there exists $\delta > 0$ such that

$$\forall\ t_0, t_0' \in T_0,\ z \in Z: \quad \text{if } |t_0 - t_0'| < \delta \text{ then } |u_0(z, t_0) - u_0(z, t_0')| < \frac{\epsilon}{8}. \tag{18}$$

Similarly, because ρ is p_0-feasible, U_0^ρ is uniformly continuous. Hence, there exists $\delta' > 0$ such that

$$\forall\ t_0, t_0' \in T_0: \quad \text{if } |t_0 - t_0'| < \delta' \text{ then } |U_0^\rho(t_0) - U_0^\rho(t_0')| < \frac{\epsilon}{8}. \tag{19}$$

By compactness of T_0, there exists a finite partition $\hat{D}_1, \ldots, \hat{D}_{\hat{k}}$ of T_0 such that $\text{diam}(\hat{D}_k) < \min\{\delta, \delta'\}$ for all $k = 1, \ldots, \hat{k}$. By dropping any cell $\hat{D}_k$ with $q_0(\hat{D}_k) = 0$, we obtain a partition $D_1, \ldots, D_{\overline{k}}$ of some set $\hat{T}_0 \subseteq T_0$, where $q_0(\hat{T}_0) = 1$ and $q_0(D_k) > 0$ for all $k = 1, \ldots, \overline{k}$.

We construct an allocation $\rho''' = (\alpha'''(\cdot), \mathbf{x}'''(\cdot))$ from ρ'' as follows. Given any $\mathbf{t} \in \mathbf{T}$ with $t_0 \in D_k$ for some k, we define $\alpha'''(\mathbf{t})$, and $x_i'''(\cdot)$ $(i = 1, \ldots, n)$ by taking the average over all types in D_k. That is,

$$\begin{aligned}
\alpha'''(\mathbf{t})(B) &= \frac{1}{q_0(D_k)} \int_{D_k} \alpha'(t_0', t_{-0})(B) \mathrm{d}q_0(t_0') \quad \text{for all measurable sets } B \subseteq A, \\
x_i'''(\mathbf{t}) &= \frac{1}{q_0(D_k)} \int_{D_k} x_i''(t_0', t_{-0}) \mathrm{d}q_0(t_0').
\end{aligned}$$

Given any $t_0 \in T_0 \setminus \hat{T}_0$, let $\hat{t}_0 \in \hat{T}_0$ be an announcement that is optimal for t_0 among all announcements in $\hat{T}_0$ in the direct-mechanism interpretation of ρ'''; define $\rho'''(t_0, \mathbf{t}_{-0}) = \rho'''(\hat{t}_0, \mathbf{t}_{-0})$ for all $\mathbf{t}_{-0} \in \mathbf{T}_{-0}$. (By construction of ρ''', there are at most $\overline{k}$ essentially different announcements, so that an optimal one exists.)

By Fubini's Theorem for transition probabilities, for all k and $t_0 \in D_k$,[19]

$$u_0(\rho'''(\mathbf{t}), t_0) = \frac{1}{q_0(D_k)} \int_{D_k} u_0(\rho''(t_0', \mathbf{t}_{-0}), t_0) \mathrm{d}q_0(t_0'). \tag{20}$$

[19]See, e.g., Bauer, Probability Theory, Ch. 36.

Hence, letting $\mathbf{p}$ denote the product measure of $p_1, \ldots, p_n$,

$$
\begin{aligned}
U_0^{\rho'''}(t_0) &= \int_{\mathbf{T}_{-0}} u_0(\rho'''(\mathbf{t}), t_0) \mathrm{d}\mathbf{p}(t_{-0}) \\
&\stackrel{(20)}{=} \frac{1}{q_0(D_k)} \int_{D_k} \int_{\mathbf{T}_{-0}} u_0(\rho''(t_0', \mathbf{t}_{-0}), t_0) \mathrm{d}\mathbf{p}(t_{-0}) \mathrm{d}q_0(t_0') \\
&\stackrel{(18)}{>} \frac{1}{q_0(D_k)} \int_{D_k} \int_{\mathbf{T}_{-0}} (u_0(\rho''(t_0', \mathbf{t}_{-0}), t_0') - \frac{\epsilon}{8}) \mathrm{d}\mathbf{p}(t_{-0}) \mathrm{d}q_0(t_0') \\
&= \frac{1}{q_0(D_k)} \int_{D_k} (U_0^{\rho''}(t_0') - \frac{\epsilon}{8}) \mathrm{d}q_0(t_0') \\
&\stackrel{(16)}{=} \frac{1}{q_0(D_k)} \int_{D_k} (U_0^{\rho}(t_0') + \frac{7}{8}\epsilon) \mathrm{d}q_0(t_0') \\
&\stackrel{(19)}{>} \frac{1}{q_0(D_k)} \int_{D_k} (U_0^{\rho}(t_0) + \frac{3}{4}\epsilon) \mathrm{d}q_0(t_0') \\
&= U_0^{\rho}(t_0) + \frac{3}{4}\epsilon \quad \text{for all } t_0 \in \hat{T}_0.
\end{aligned}
$$

Let $\mathcal{I}(q_0)$ denote the set of allocations that satisfy the agents' (but not necessarily the principal's) q_0-incentive and participation constraints.

We show that $\rho''' \in \mathcal{I}(q_0)$. To see this, consider any $i = 1, \ldots, n$ and $\hat{t}_i, t_i \in T_i$. Then

$$
\begin{aligned}
U_i^{\rho''', q_0}(\hat{t}_i, t_i) &= \int_{T_{-0-i}} \int_{T_0} u_i(\rho'''(\hat{t}_i, t_{-i}), t_i) \mathrm{d}q_0(t_0) \mathrm{d}p_{-0-i}(t_{-0-i}) \\
&= \int_{T_{-0-i}} \sum_k \int_{D_k} u_i(\rho'''(\hat{t}_i, t_{-i}), t_i) \mathrm{d}q_0(t_0) \mathrm{d}p_{-0-i}(t_{-0-i}) \\
&= \int_{T_{-0-i}} \sum_k q_0(D_k) u_i(\rho'''(\hat{t}_i, t_{-i-0}, t_{0k}), t_i) \mathrm{d}p_{-0-i}(t_{-0-i}),
\end{aligned}
$$

where we have selected any $t_{0k} \in D_k$ for all k. Applying Fubini's Theorem for transition probabilities, we conclude that

$$
\begin{aligned}
U_i^{\rho''', q_0}(\hat{t}_i, t_i) &= \int_{T_{-0-i}} \sum_k \int_{D_k} u_i(\rho''(\hat{t}_i, t_{-i-0}, t_0'), t_i) \mathrm{d}q_0(t_0') \mathrm{d}p_{-0-i}(t_{-0-i}) \\
&= \int_{T_{-0-i}} \int_{T_0} u_i(\rho''(\hat{t}_i, t_{-i-0}, t_0'), t_i) \mathrm{d}q_0(t_0') \mathrm{d}p_{-0-i}(t_{-0-i}) \\
&= U_i^{\rho'', q_0}(\hat{t}_i, t_i).
\end{aligned}
$$

Hence, $\rho''' \in \mathcal{I}(q_0)$ because $\rho'' \in \mathcal{I}(q_0)$.

Given ρ''' and any $t_0 \in T_0$, let

$$
D^{\rho'''}(t_0) = \{t_0' \in T_0 \mid \forall t_{-0} : \ \rho'''(t_0', t_{-0}) = \rho'''(t_0, t_{-0})\}.
$$

By construction, the set

$$\mathcal{D}^{\rho'''} = \{D^{\rho'''}(t_0) \mid t_0 \in T_0\}$$

is a finite partition of T_0 (with at most $\overline{k}$ cells).

In summary, $\rho''' \in \mathcal{E}$, where we define

$$\begin{aligned}\mathcal{E} = \{\sigma \mid & |\mathcal{D}^\sigma| < \infty,\\ & \exists r_0 : \ \sigma \in \mathcal{I}(r_0), \ \exists \hat{T}_0 : \ r_0(\hat{T}_0) = 1,\\ & \forall t_0 \in \hat{T}_0 : U_0^\sigma(t_0) - U_0^\rho(t_0) > \frac{\epsilon}{2},\\ & \forall t_0 \in T_0 \setminus \hat{T}_0, t_0' \in T_0 : \ U_0^\sigma(t_0) \geq U_0^\sigma(t_0', t_0),\\ & \forall t_0 \in \hat{T}_0 : \ \hat{T}_0 \cap \arg\max_{t_0' \in T_0} U_0^\sigma(t_0', t_0) \neq \emptyset\}.\end{aligned}$$

Because $\mathcal{E} \neq \emptyset$, there exists $\sigma^* \in \mathcal{E}$ with minimal $|\mathcal{D}^{\sigma^*}|$. Let r_0 denote a corresponding belief and let $\hat{T}_0$ a corresponding probability-1 set.

Let B^* denote the set of principal-types for which an incentive constraint is violated in σ^*. Then $B^* \subseteq \hat{T}_0$ because $\sigma^* \in \mathcal{E}$. We will show that $r_0(B^*) = 0$.

Suppose that $r_0(B^*) > 0$. We will show that this contradicts the minimality of $|\mathcal{D}^{\sigma^*}|$.

Because $|\mathcal{D}^{\sigma^*}| < \infty$, there exists $D' \in \mathcal{D}^{\sigma^*}$ such that $r_0(B^* \cap D') > 0$.

By violation of the incentive constraint, there exists $D'' \in \mathcal{D}^{\sigma^*} \setminus \{D'\}$ such that

$$r_0(B'') > 0, \quad \text{where } B'' = \{t_0 \in B^* \cap D' \mid U_0^{\sigma^*}(\hat{t}_0, t_0) > U_0^{\sigma^*}(t_0) \text{ if } \hat{t}_0 \in D''\}.$$

We construct a new belief r_0' by

$$r_0'(B) = r_0(B \cap B'')\frac{r_0(D' \cup D'')}{r_0(B'')} + r_0(B \setminus \{D' \cup D''\}) \quad \text{for any Borel set } B \subseteq T_0.$$

Clearly, r_0' is absolutely continuous relative to r_0 (hence, relative to p_0). Also,

$$r_0'(\hat{T}_0') = 1, \quad \text{where } \hat{T}_0' = B'' \cup (\hat{T}_0 \setminus (D' \cup D'')). \tag{21}$$

We construct an allocation $\sigma' = (\beta(\cdot), \mathbf{y}(\cdot))$ from $\sigma^* = (\beta^*(\cdot), \mathbf{y}^*(\cdot))$ as follows.

Given any $\mathbf{t} \in \mathbf{T}$ with $t_0 \in B''$, we define $\beta(\mathbf{t})$, and $y_i(\cdot)$ $(i = 1, \ldots, n)$ by taking the average over all types in $D' \cup D''$. That is, for all measurable sets $B \subseteq A$,

$$\begin{aligned}\beta(\mathbf{t})(B) &= \frac{r_0(D')}{r_0(D' \cup D'')}\beta^*(t_0', t_{-0})(B) + \frac{r_0(D'')}{r_0(D' \cup D'')}\beta^*(t_0'', t_{-0})(B),\\ y_i(\mathbf{t}) &= \frac{r_0(D')}{r_0(D' \cup D'')}y_i^*(t_0', t_{-0}) + \frac{r_0(D'')}{r_0(D' \cup D'')}y_i^*(t_0'', t_{-0}),\end{aligned}$$

where we have picked any $t_0' \in D'$ and $t_0'' \in D''$.

Given any $\mathbf{t} \in \mathbf{T}$ with $t_0 \in \hat{T}_0 \setminus (D' \cup D'')$, we define $\sigma'(\mathbf{t}) = \sigma^*(\mathbf{t})$. For all $\mathbf{t} \in \mathbf{T}$ with $t_0 \notin \hat{T}_0'$, define $\sigma'(\mathbf{t})$ by letting type t_0 announce, in the direct-mechanism interpretation of σ', whatever type she finds optimal in $\hat{T}_0'$. Then

$$|\mathcal{D}^{\sigma'}| \leq |\mathcal{D}^{\sigma^*} \setminus \{D', D''\}| + 1 < |\mathcal{D}^{\sigma^*}|.$$

We will show now that $\sigma' \in \mathcal{E}$, yielding a contradiction to the minimality of $|\mathcal{D}^{\sigma^*}|$.

First we show that

$$\sigma' \in \mathcal{I}(r_0'). \tag{22}$$

Consider any $i = 1, \ldots, n$ and $\hat{t}_i, t_i \in T_i$. Then

$$\begin{aligned} U_i^{\sigma', r_0'}(\hat{t}_i, t_i) &= \int_{T_{-0-i}} \int_{\hat{T}_0} u_i(\sigma'(\hat{t}_i, t_{-i}), t_i) \mathrm{d}r_0'(t_0) \mathrm{d}p_{-0-i}(t_{-0-i}) \\ &= \int_{T_{-0-i}} \int_{\hat{T}_0 \setminus (D' \cup D'')} u_i(\sigma^*(\hat{t}_i, t_{-i}), t_i) \mathrm{d}r_0(t_0) \mathrm{d}p_{-0-i}(t_{-0-i}) \\ &\quad + \int_{T_{-0-i}} \int_{B''} u_i(\sigma'(\hat{t}_i, t_{-i}), t_i) \mathrm{d}r_0'(t_0) \mathrm{d}p_{-0-i}(t_{-0-i}). \end{aligned} \tag{23}$$

Picking any $\check{t}_0 \in B''$, and applying Fubini's theorem for transition probabilities,

$$\begin{aligned} \int_{B''} u_i(\sigma'(\hat{t}_i, t_{-i}), t_i) \mathrm{d}r_0'(t_0) &= u_i(\sigma'(\hat{t}_i, \check{t}_0, t_{-0-i}), t_i) r_0'(B'') \\ &= \left(\frac{r_0(D')}{r_0(D' \cup D'')} u_i(\sigma^*(\hat{t}_i, t_0', t_{-0-i}), t_i) + \frac{r_0(D'')}{r_0(D' \cup D'')} u_i(\sigma^*(\hat{t}_i, t_0'', t_{-0-i}), t_i) \right) r_0'(B'') \\ &= r_0(D') u_i(\sigma^*(\hat{t}_i, t_0', t_{-0-i}), t_i) + r_0(D'') u_i(\sigma^*(\hat{t}_i, t_0'', t_{-0-i}), t_i) \\ &= \int_{D' \cup D''} u_i(\sigma'(\hat{t}_i, t_{-i}), t_i) \mathrm{d}r_0(t_0). \end{aligned}$$

Plugging this into (23) yields

$$U_i^{\sigma', r_0'}(\hat{t}_i, t_i) = U_i^{\sigma^*, r_0}(\hat{t}_i, t_i).$$

This implies (22) because $\sigma^* \in \mathcal{I}(r_0)$.

Next we show that, for all $t_0 \in \hat{T}_0'$,

$$U_0^{\sigma'}(t_0) - U_0^{\rho}(t_0) > \frac{\epsilon}{2}. \tag{24}$$

First consider $t_0 \in \hat{T}_0 \setminus (D' \cup D'')$. Then $U_0^{\sigma'}(t_0) = U_0^{\sigma^*}(t_0)$, so (24) is immediate from $\sigma^* \in \mathcal{E}$ and from $\hat{T}_0' \subseteq \hat{T}_0$.

For all $t_0 \in B''$, (24) holds because

$$U_0^{\sigma'}(t_0) = \frac{r_0(D')}{r_0(D' \cup D'')} U_0^{\sigma^*}(t_0) + \frac{r_0(D'')}{r_0(D' \cup D'')} U_0^{\sigma^*}(t_0'', t_0) > U_0^{\sigma^*}(t_0).$$

This completes the proof that $\sigma' \in \mathcal{E}$, thereby contradicting the minimality of $|\mathcal{D}^{\sigma^*}|$.

We conclude that $r_0(B^*) = 0$.

Given any $\mathbf{t} \in \mathbf{T}$ with $t_0 \notin B^*$, we define $\sigma(\mathbf{t}) = \sigma^*(\mathbf{t})$. For all $\mathbf{t} \in \mathbf{T}$ with $t_0 \in B^*$, we define $\sigma(\mathbf{t})$ by letting type t_0 announce, in the direct-mechanism interpretation of σ^*, whatever type she finds optimal in $T_0 \setminus B^*$, or assign the disagreement outcome if t_0 prefers that.

By construction, the principal's incentive constraints are satisfied for σ. Also, the agents' r_0-incentive and participation constraints are satisfied because $\sigma(\mathbf{t})$ equals

$\sigma^*(\mathbf{t})$ for a r_0-probability-1 set of principal-types, and because these constraints are satisfied for σ^*.

By construction, (17) holds for all $t_0 \in T_0 \setminus B^*$. By continuity of $U_0^\sigma(\cdot)$, (17) extends to all $t_0 \in \mathrm{supp}(r_0)$. In particular, the principal's participation constraint is satisfied for all types in $\mathrm{supp}(r_0)$. By construction, the same holds for all types not in $\mathrm{supp}(r_0)$. Hence, σ is r_0-feasible. This completes the proof. *QED*

9.2. **Some additional implications of Proposition 1 .** At several points in our analysis, it is useful to consider a simpler variant of the principal's ex-ante problem. An allocation ρ is called q_0-*agent-feasible* if, for all agents $i \geq 1$, the q_0-incentive constraints (3) and the q_0-participation constraints (4) are satisfied. That is, in an agent-feasible allocation the principal's incentive and participation constraints may be violated. Let $\tilde{\eta}(q_0)$ denote the supremum value of the principal's *relaxed* q_0-*ex-ante problem*

$$\max_{\rho \ q_0\text{-agent feasible}} \int_{T_0} U_0^\rho(t_0) \mathrm{d}q_0(t_0). \tag{25}$$

Technically, the relaxed ex-ante problem is often easier to solve than the standard ex-ante problem. Obviously, $\tilde{\eta}(q_0) \geq \eta(q_0)$.

From the proof of Proposition 1 it is clear that the characterization continues to hold when all ex-ante optimizations are replaced by relaxed ex-ante optimizations.

Corollary 2. *A* p_0-*feasible allocation* ρ *is strongly neologism-proof if and only if*

$$\tilde{\eta}(q_0) \quad \leq \quad \int_{T_0} U_0^\rho(t_0) dq_0(t_0) \quad \text{for all } q_0 \text{ absolutely continuous rel. to } p_0. \tag{26}$$

Thus, $\eta(p_0) = \tilde{\eta}(p_0)$ *in any environment in which a strongly neologism-proof allocation exists.*

Another corollary provides a sufficient condition for an allocation to be strongly neologism-proof; it follows from the arguments in the proof of Proposition 1. This result is useful towards characterizing ex-ante optimal allocations in linear environments (Proposition 3).

Corollary 3. *If a* p_0-*feasible allocation is not strongly neologism-proof, then it is* strictly q_0-*dominated for some belief* q_0 *that is absolutely continuous with respect to* p_0.

This corollary implies that the set of strongly neologism-proof principal-payoff vectors is always closed and is helpful in proving the existence of strongly neologism-proof allocations in environments with finite type spaces (Proposition 8).

9.3. **Existence of strongly neologism-proof allocations.** In this section, we use the results of Section 9.2 together with the existence result from Mylovanov and Tröger (2012) in order to show that a strongly neologism-proof allocation exists in any environment with finite type spaces that satisfies weak technical assumptions, and in any linear-utility environment.

9.3.1. *Existence in environments with finite type spaces.* In Mylovanov and Tröger (2012) we prove the existence of a strongly neologism-proof allocation in environments with finite type spaces under otherwise rather weak assumptions.[20] We now extend the existence result to quasi-linear environments in which the set of collective actions, A, is compact. We make the assumption of separability that was introduced in Mylovanov and Tröger (2012); it requires that there exists an allocation such that the incentive and participation constraints of all types of all agents are satisfied as strict inequalities.

Proposition 8. *Suppose that the type spaces $T_0, \ldots, T_n$ are finite, that A is a compact metric space, the valuation functions $v_0, \ldots, v_n$ are continuous, and separability holds. Then a strongly neologism-proof allocation exists.*

The proof has the following steps. First, we provide an upper bound λ for the absolute value of the interim expected payment of any type of any player in any incentive-feasible allocation. Then we show that there exists a number κ such that any scheme of interim expected payments that can occur in a q_0-feasible allocation at all can also be obtained from a payment scheme that involves payments at most κ times as large (in absolute value) as the largest interim expected payment of any type of any player. We approximate the outcome space of the quasilinear environment with a sequence of outcome spaces with larger and larger finite bounds on payments. These environments have compact outcome spaces, so that strongly neologism-proof allocations exist by Mylovanov and Tröger (2012). Moreover, we can assume that payments in these allocations are bounded by $\kappa\lambda$. Hence, the sequence of strongly neologism-proof allocations has a convergent subsequence. Using Corollary 3, we show that the subsequence limit is strongly neologism-proof in the quasilinear environment.

The proof or Proposition 8 relies on two lemmas. Given any allocation $\rho(\cdot) = (\alpha(\cdot), \mathbf{x}(\cdot))$ and any belief q_0 about the principal's type, the interim expected payment function of any player i is denoted

$$\underline{x}_i^{\rho,q_0}(t_i) \quad = \quad \int_{\mathbf{T}_{-i}} x_i(t_i, \mathbf{t}_{-i}) \mathrm{d}\mathbf{q}_{-i}(\mathbf{t}_{-i}).$$

Lemma 1. *Suppose that A is a compact metric space, and the valuation functions $v_0, \ldots, v_n$ are continuous.*

[20]As observed by Maskin and Tirole (1990), once can consider a fictitious economy in which the principal-types trade amounts of slack allowed for the various constraints. Any competitive equilibrium in this fictitious economy corresponds to an allocation that is a strongly neologism-proof equilibrium allocation of the mechanism-selection game. The proof in Mylovanov and Tröger (2012) employs this idea and, consequently, the existence proofs in this paper are indirectly based on that approach.

Then, for all beliefs q_0, in any q_0-feasible allocation, the absolute value of the interim expected payment of any type of any player is smaller than

$$\lambda \;=\; (n+4)\max_{i,a,t_i}|v_i(a,t_i)|.$$

Proof. Let $\overline{v} = \max_{i,a,t_i}|v_i(a,t_i)|$ denote an upper bound for the absolute value of the valuation of any action for any type of any player.

By (4), each player's q_0-ex-ante expected payoff is bounded below by 0. On the other hand, the sum of the players' q_0-ex-ante expected payoffs is bounded above by $(n+1)\overline{v}$ because payments cancel. Hence,

$$0 \le \int_{T_i} U_i^{\rho,q_0}(t_i)\mathrm{d}q_i(t_i) \le (n+1)\overline{v} \quad \text{for all } i,$$

where we define $q_i = p_i$ for all $i = 1,\dots,n$.

Turning to interim expected payoffs,

$$|U_i^{\rho,q_0}(t_i,t_i') - U_i^{\rho,q_0}(t_i,t_i)| \le \max_{a\in A}|v_i(a,t_i') - v_i(a,t_i)| \le 2\overline{v}. \tag{27}$$

Hence,

$$U_i^{\rho,q_0}(t_i) \le U_i^{\rho,q_0}(t_i,t_i') + 2\overline{v} \overset{(3)}{\le} U_i^{\rho,q_0}(t_i') + 2\overline{v}.$$

Thus,

$$U_i^{\rho,q_0}(t_i) \le \int_{T_i} U_i^{\rho,q_0}(t_i')\mathrm{d}q_i(t_i') + 2\overline{v} \le (n+3)\overline{v}.$$

Because any player's interim payment can differ from her interim payoff by at most $\overline{v}$, we obtain the desired bound. This completes the proof.

With finite type spaces, both the space of payment schemes $\mathcal{L} = \mathbb{R}^{|\mathbf{T}|n}$ and the space of interim expected payment schemes $\underline{\mathcal{L}} = \mathbb{R}^{|T_0|+\cdots+|T_n|}$ are finite-dimensional vector spaces. Endow both spaces with the max-norm. We define the linear map

$$\phi^{q_0} : \mathcal{L} \to \underline{\mathcal{L}},\ \mathbf{x}(\cdot) \mapsto (\underline{x}_0^{\rho,q_0}(\cdot),\dots,\underline{x}_n^{\rho,q_n}(\cdot)).$$

The following lemma says that there exists a number κ such that any scheme of interim expected payments that can occur in a q_0-feasible allocation at all can also be obtained from a payment scheme that involves payments at most κ times as large (in absolute value) as the largest interim expected payment of any type of any player.

Lemma 2. *Suppose that $T_0,\dots,T_n$ are finite. Consider any belief q_0. There exists a number κ such that, for every $\underline{\mathbf{x}}(\cdot) \in \underline{\mathcal{L}}$, there exists $\mathbf{x}(\cdot) \in \mathcal{L}$ such that $\phi^{q_0}(\mathbf{x}(\cdot)) = \underline{\mathbf{x}}(\cdot)$ and $||\mathbf{x}(\cdot)|| \le \kappa||\underline{\mathbf{x}}(\cdot)||$.*

Proof. The set $\phi^{q_0}(\mathcal{L})$ is a finite-dimensional vector space, hence a Banach space (with the norm induced by the max-norm in $\underline{\mathcal{L}}$), and ϕ^{q_0} maps onto that space. Hence, the claim is immediate from the open mapping theorem in functional analysis.

Proof of Proposition 8. Consider any sequence of payment bounds (λ_l) such that $\lambda_l \to \infty$. From Mylovanov and Tröger (2012), for each l, there exists an allocation

ρ_l that is strongly neologism-proof in the environment with payment bound λ_l. By Lemma 2 and Lemma 1 (with $q_0 = p_0$), w.l.o.g., all these allocations use payments that are bounded by the same number $\kappa\lambda$. Hence, the sequence of payment schemes in the sequence ρ_l is bounded in the max-norm. Hence, there exists a convergent subsequence with limit ρ^* (in the dimension of the probability measures on collective actions, the convergence is meant as a weak convergence).

As a limit of p_0-feasible allocations, ρ^* is p_0-feasible. Suppose that ρ^* is not strongly neologism-proof. By Corollary 3, ρ^* is strictly q_0-dominated by some allocation ρ', for some belief q_0.

If l is sufficiently large, then ρ' is a feasible allocation in the environment with payment bound λ_l (w.l.o.g. by Lemma 2 and Lemma 1).

Moreover, if l is sufficiently large, then ρ_l is strictly q_0-dominated by ρ' because ρ_l approximates ρ^*. This contradicts the fact that ρ_l is strongly neologism-proof in the environment with payment bound λ_l. *QED*

9.3.2. *Existence in linear-utility environments.* The following result extends existence to environments with continuous type spaces. The proof relies on Proposition 8, but it is not a simple extension and is the most involved result in this paper.

Proposition 9. *A strongly neologism-proof allocation exists in any linear-utility environment.*

To prove this, we use the version of our characterization result that refers to the relaxed ex-ante problems (Corollary 2). We use the finite-type existence result (Proposition 8) and consider a continuous-type limit.

Using standard envelope arguments (e.g., Mas-Colell, Whinston, and Green (1995), Chapter 23), the principal's relaxed F_0-ex-ante problem can be written as

$$\max_{u_0(\mathbf{t}),\ldots,u_n(\mathbf{t}),(s_0(\mathbf{t}),\ldots,c_n(\mathbf{t}))\in\mathcal{V}} \int_T u_0(\mathbf{t})\mathrm{d}F(\mathbf{t}),$$

$$\text{s.t.} \quad \overline{s}_i(\cdot) \text{ weakly increasing for all } i \geq 1, \tag{28}$$

$$\overline{u}_i(t_i) = \overline{u}_i(\underline{t}_i) + \int_{\underline{t}_i}^{t_i} \overline{s}_i(y)\mathrm{d}y \text{ for all } i \geq 1, t_i \in T_i, \tag{29}$$

$$\overline{u}_i(t_i) \geq 0 \text{ for all } i \geq 1, t_i \in T_i, \tag{30}$$

$$\int_T \left(\sum_{i=0}^{n} s_i(\mathbf{t})t_i + c_i(\mathbf{t}) - u_i(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t}) = 0, \tag{31}$$

Notice that we require the budget to be balanced ex-ante (31). By Börgers and Norman (2009), this is equivalent to an ex-post budget balance condition.

Our proof techniques rely on the linearity of the environment. The linearity of the environment implies the linearity of the envelope formula (29) in $\overline{s}_i$. This linearity, together with the linearity of the budget balance equation (31) allows us to get well-behaved limits when we consider the relevant convergent subsequences.

The proof of Proposition 9 begins with the observation that an ex-ante optimizing principal who implements any vector $(\hat{s}_0, \ldots, \hat{s}_n)$ will combine this with a vector $(\hat{c}_0, \ldots, \hat{c}_n)$ that has a minimal sum $\sum_i \hat{c}_i$, so that she can charge the largest payments. Hence, we can work with a simplified $\mathcal{V}$ in which $(\hat{s}_0, \ldots, \hat{s}_n)$ uniquely determines $(\hat{c}_0, \ldots, \hat{c}_n)$, and we can ignore the vector $(\hat{c}_0, \ldots, \hat{c}_n)$ in the following.

We define a sequence $m = 1, 2, \ldots$ of finer and finer finite-type approximations of the linear-utility environment. Each of these environments m can be shown to be separable, so that a strongly neologism-proof allocation ρ^m exists by Proposition 8. We use the notation $s_i^m(\cdot)$ and $u_i^m(\cdot)$ to refer to the components of ρ^m. We extend each allocation ρ^m to the original continuous type spaces by letting the intermediate types make optimal type announcements in the direct-mechanism interpretation. For each player i, the sequence of interim-averages $(\overline{s}_i^m(\cdot))_{m=1,2,\ldots}$ has an almost-everywhere convergent subsequence by Helly's selection theorem (let $\hat{s}_i(\cdot)$ denote the limit), and the interim-averages $(\overline{u}_i^m(\cdot))_{m=1,2,\ldots}$ have a uniformly convergent subsequence by Arzela-Ascoli's theorem (let $\hat{u}_i(\cdot)$ denote the limit). The sequence $(s_i^m(\cdot))_{m=1,2,\ldots}$ has a weakly convergent subsequence by Alaoglu's theorem. For the weak limits $s_i^*(\cdot)$ $(i = 0, \ldots, n)$ one can compute the interim averages $\overline{s}_i^*(\cdot)$. The crucial step is to show that $\overline{s}_i^*(\cdot) = \hat{s}_i(\cdot)$. Once we have that, we know that $\overline{s}_i^*(\cdot)$ is weakly increasing and we can use the envelope theorem to define the interim averages $\overline{u}_i^*(\cdot)$ and a corresponding limit allocation ρ^*. Then one shows that $\overline{u}_i^*(\cdot) = \hat{u}_i(\cdot)$. This implies that the monotonicity conditions (28), the participation constraints (30), and the budget balance condition (31) hold in the limit ρ^*.

To verify the condition of Corollary 2, we suppose that (26) fails. Thus, there exists a belief G_0 absolutely continuous relative to F_0, and a G_0-feasible allocation ρ' with a higher G_0-ex-ante expected payoff for the principal than ρ^*. We consider the sequence of finite-type-spaces environments $m = 1, 2, \ldots$ with beliefs G_0^m that approximate G_0.

For each m, we partition the space of continuous type profiles into cells that correspond to the discrete type profiles in the environment m. We construct an allocation ρ'^m by taking the average of ρ' in each cell, and by adding correction terms to the payments so that ρ'^m satisfies the agents' (not necessarily the principal's) incentive and participation constraints, as well as the budget balance condition, with respect to the belief G_0^m. We show that the correction terms vanish as $m \to \infty$. Thus, if m is large, then the G_0^m-ex-ante expectation of ρ'^m is larger than the G_0^m-ex-ante expectation of ρ^m. This contradicts the fact that ρ^m is strongly neologism-proof for all m and the proof is complete.

Proof of Proposition 9. For a probability-1 set of type profiles $\mathbf{t}$, a solution to the unconstrained F_0-ex-ante problem will implement an outcome that puts probability 0 on any action $a \in A$ such that

$$\exists b \in A,\ (s_0^a, \ldots, s_n^a) = (s_0^b, \ldots, s_n^b),\ \sum_{i=0}^{n} c_i^a < \sum_{i=0}^{n} c_i^b. \tag{32}$$

(Otherwise, the principal could always implement b instead of a and extract larger payments from the agents.)

Moreover, if instead $\sum_{i=0}^n c_i^a = \sum_{i=0}^n c_i^b$ in (32), then either a or b may be used without changing the interim expected utility $\overline{u}_i(t_i)$ of any type t_i of any player i.

Hence, without loss of generality, we may assume A is such that, for all $a \in A$, the vector $s^a = (s_0^a, \ldots, s_n^a)$ uniquely determines the vector $c^a = (c_0^a, \ldots, c_n^a) = \Phi(s^a)$. Extending Φ linearly to the convex hull $\mathcal{S}$ of $\{s^a | a \in A\}$, we have

$$\mathcal{V} \;=\; \{(\hat{s}, \hat{c}) \in \mathbb{R}^{2n+2} \mid \hat{s} \in \mathcal{S}, \hat{c} = \Phi(\hat{s})\}.$$

For all players $i = 0, \ldots, n$, naturals $m = 1, 2, \ldots$, and $k = 1, \ldots, m$, let $C_i^m(k) = [F_i^{-1}((k-1)/m), F_i^{-1}(k/m))$ and $t_i^{m,k} = E_{F_i}[t_i \mid t_i \in C_i^m(k)]$. Define the finite type space

$$T_i^m = \{t_i^{m,1}, \ldots, t_i^{m,m}\}$$

and let F_i^m be the c.d.f. for the uniform distribution on T_i^m.

In the following, we will use the quantile functions $F_i^{-1}(q_i) = \min\{t_i \in T_i \mid F_i(t_i) \geq q_i\}$ ($q_i \in [0,1]$); define $(F_i^m)^{-1}$ analogously. Let $F^{-1}(q) = (F_0^{-1}(q_0), \ldots, F_n^{-1}(q_n))$ for all $q = (q_0, \ldots, q_n) \in [0,1]^{n+1}$; define $(F^m)^{-1}$ analogously. Then

$$\mid F_i^{-1}(q_i) - (F_i^m)^{-1}(q_i) \mid \;\leq \frac{1}{m \cdot \min_{t_i \in T_i} f_i(t_i)} \;\stackrel{\text{def}}{=}\; \delta_i^m. \tag{33}$$

Next we show that each of the discrete environments just defined is separable in the sense of Mylovanov and Tröger (2012). For all $i \geq 1$, $t_i \in T_i^m$, and $a, b \in A$, define

$$p_i^{a,b}(t_i) \;=\; \begin{cases} \frac{t_i - \underline{t}_i}{\overline{t}_i - \underline{t}_i} & \text{if } s_i^a \geq s_i^b, \\ \frac{\overline{t}_i - t_i}{\overline{t}_i - \underline{t}_i} & \text{otherwise.} \end{cases}$$

Define a function $\alpha(\mathbf{t})$ for all $\mathbf{t} \in \mathbf{T}^m$ by the following randomization: select any number $i \in \{1, \ldots, n\}$ with equal probability ($= 1/n$), then choose action a_i with probability $(1/n)\sum_{j=1}^n p_j^{a_i,b_i}(t_j)$ and choose b_i with the remaining probability, where we use the notation a_i, b_i from (2).

By construction, $p_j^{a_i,b_i}(t_j)$ is strictly increasing in t_j if agent j weakly prefers a_i to b_i, and is strictly decreasing if agent j weakly prefers b_i to a_i. Hence, for any agent i, type $t_i \in T_i^m$, and t_{-i}, as t_i increases, the randomized action $\alpha(\mathbf{t})$ shifts probability from less preferred actions to more preferred actions. Thus, using (2), the function

$$\hat{s}_i(\mathbf{t}) \;=\; \int_A s_i^a \mathrm{d}\alpha(\mathbf{t})(a)$$

is strictly increasing in t_i, for all $\mathbf{t}_{-i}$. Hence, we can define payments such that all agents' incentive constraints are satisfied with strict inequality. By adding constant payments we can guarantee that, in addition, all agents' participation constraints are satisfied with strict inequality, showing separability.

Because A is finite, it is trivially compact and the valuation functions are continuous. Hence, by Proposition 8, for each of the discrete-type-space environments constructed above ($m = 1, 2, \ldots$), there exists a strongly neologism-proof allocation

$$\rho^m(\mathbf{t}) \;=\; (u_0^m(\mathbf{t}), \ldots, u_n^m(\mathbf{t}), s_0^m(\mathbf{t}), \ldots, c_n^m(\mathbf{t}))$$

that is defined for all $\mathbf{t} \in \mathbf{T}^m = T_0^m \times \cdots \times T_n^m$.

We extend ρ^m to all $\mathbf{t} \in \mathbf{T}$ by assuming that, in the direct-mechanism interpretation of ρ^m, any type $t_i \in (t_i^{m,k}, t_i^{m,k+1})$ makes an optimal type announcement from the set $\{t_i^{m,k}, t_i^{m,k+1}\}$, any type $t_i > t_i^{m,m}$ announces the type $t_i^{m,m}$, and any type $t_i < t_i^{m,1}$ announces the type $t_i^{m,1}$.

Then the functions

$$\overline{s}_i^m(t_i) \;=\; \int_{T_{-i}} s_i^m(\mathbf{t})\mathrm{d}F_{-i}^m(\mathbf{t}_{-i}), \quad (i \geq 0,\; t_i \in T_i)$$

are weakly increasing on T_i. Moreover, defining

$$\overline{u}_i^m(t_i) \;=\; \int_{T_{-i}} u_i^m(\mathbf{t})\mathrm{d}F_{-i}^m(\mathbf{t}_{-i}),$$

the envelope formula holds on T_i, that is,

$$\overline{u}_i^m(t_i) \;=\; \overline{u}_i^m(\underline{t}_i) + \int_{\underline{t}_i}^{t_i} \overline{s}_i^m(y)\mathrm{d}y \quad \text{for all } i \geq 0,\; t_i \in T_i. \tag{34}$$

Observe that this formula includes the principal $i = 0$.

From (34), for all m, i,

$$|\overline{u}_i^m(t_i) - \overline{u}_i^m(t_i')| \leq \max_a |s_i^a| \cdot |t_i - t_i'| \quad (t_i, t_i' \in T_i).$$

Hence, the family of functions $(\overline{u}_i^m)_{m=1,2,\ldots}$ is equicontinuous. Moreover, by Lemma 1, it is uniformly bounded. Hence, by Arzela and Ascoli's Theorem, there exists a subsequence m' such that

$$\max_{t_i \in T_i} |\overline{u}_i^{m'}(t_i) - \hat{u}_i^*(t_i)| \to 0 \tag{35}$$

for some continuous function $\hat{u}_i^*$; i.e., the subsequence converges uniformly.

For all $i \geq 0$, the composite function $s_i^m \circ (F^m)^{-1}$ belongs to $L_2([0,1]^{n+1})$. The sequence $(s_i^m \circ (F^m)^{-1})_{m=1,2\ldots}$ is $||\cdot||_2$ -bounded (for instance, $\max_{a\in A} |s_i^a|$ is a bound). Hence, by Alaoglu's Theorem, there exists a subsequence m' such that

$$s_i^{m'} \circ (F^{m'})^{-1} \to_{\text{weakly}} h_i^* \tag{36}$$

for some $h_i^* \in L_2([0,1]^{n+1})$. Define

$$\overline{h}_i^*(q_i) \;=\; \int_{[0,1]^n} h_i^*(q)\mathrm{d}q_{-i}.$$

Define

$$s_i^*(\mathbf{t}) \;=\; h_i^*(F_0(t_0), \ldots, F_n(t_n)), \quad (\mathbf{t} \in T).$$

Define

$$\overline{s}_i^*(t_i) \;=\; \int_{T_{-i}} s_i^*(\mathbf{t})\mathrm{d}F_{-i}(\mathbf{t}_{-i}).$$

(At this point, is it not yet clear whether $\overline{s}_i^*$ is a weakly increasing function.)

Note that

$$\overline{s}_i^*(F_i^{-1}(q_i)) \;=\; \overline{h}_i^*(q_i). \tag{37}$$

Because the functions $\overline{s}_i^m$ are weakly increasing, Helly's selection theorem implies the existence of a subsequence m' such that

$$\overline{s}_i^{m'}(t_i) \to \hat{s}_i^*(t_i) \quad \text{Lebesgue-a.e. } t_i \in T_i. \tag{38}$$

for some $\hat{s}_i^* \in L_2(T_i)$. This convergence translates into the quantile space:[21]

$$\overline{s}_i^{m'}(F_i^{-1}(q_i)) \to \hat{s}_i^*(F_i^{-1}(q_i)) \quad \text{Lebesgue-a.e. } q_i \in [0,1]. \tag{39}$$

Because the functions $\overline{s}_i^m \circ F_i^m$ are weakly increasing, Helly's selection theorem implies the existence of a subsequence m' such that

$$\overline{s}_i^{m'}((F_i^{m'})^{-1}(q_i)) \to \hat{h}_i^*(q_i) \quad \text{Lebesgue-a.e. } q_i \in [0,1]. \tag{40}$$

for some $\hat{h}_i^* \in L_2([0,1])$.

From now on we will work with a subsequence m' such that (35), (36), (38), and (40) hold.

First we show that

$$\overline{s}_i^{m'} \circ (F_i^{m'})^{-1} \to_{\text{weakly}} \overline{h}_i^*. \tag{41}$$

To see this, notice that, for all $g \in L_2([0,1])$,

$$\begin{aligned}\int_0^1 \overline{s}_i^{m'}((F_i^{m'})^{-1}(q_i))g(q_i)\mathrm{d}q_i &= \int_{[0,1]^{n+1}} s_i^{m'} \circ (F^{m'})^{-1}(q)g(q_i)\mathrm{d}q \\ &\overset{(36)}{\to} \int_{[0,1]^{n+1}} h_i^*(q)g(q_i)\mathrm{d}q \\ &= \int_0^1 \overline{h}_i^*(q_i)g(q_i)\mathrm{d}q_i.\end{aligned}$$

Using (40) and (41),

$$\hat{h}_i^*(q_i) = \overline{h}_i^*(q_i) \quad \text{Lebesgue-a.e. } q_i \in [0,1]. \tag{42}$$

[21] In order to be able to move between quantile space and type space, it is important that an "Lebesgue-a.e. property" in one space translates into an "Lebesgue-a.e. property" in the other space. This follows from the assumption of positive densities. In particular, consider any set Lebesgue measurable set $X \subseteq T_i$ and $Q = \{q_i | F_i^{-1}(q_i) \in X\}$. Then

$$\Pr[Q] = \int_0^1 \mathbf{1}_{F_i^{-1}(q_i)\in X}\mathrm{d}q_i = \int_{T_i} \mathbf{1}_{t_i \in X}\mathrm{d}F_i(t_i) = \int_X f_i(t_i)\mathrm{d}t_i.$$

Hence, Q has Lebesgue-measure 0 if and only if X has Lebesgue-measure 0.

Let $\delta > 0$. For all m' large enough such that $\delta_i^{m'} < \delta$,[22]

$$\begin{aligned}
&\int_0^1 \left|\overline{s}_i^{m'}(F_i^{-1}(q_i)) - \overline{s}_i^{m'}((F_i^{m'})^{-1}(q_i))\right| \mathrm{d}q_i \qquad (43)\\
\overset{(33)}{\leq}\; &\int_0^1 \max\{\overline{s}_i^{m'}(F_i^{-1}(q_i)) - \overline{s}_i^{m'}((F_i)^{-1}(q_i)-\delta), s_i^{m'}((F_i)^{-1}(q_i)+\delta) - \overline{s}_i^{m'}(F_i^{-1}(q_i))\}\mathrm{d}q_i\\
\leq\; &\int_0^1 (\overline{s}_i^{m'}(F_i^{-1}(q_i)) - \overline{s}_i^{m'}((F_i)^{-1}(q_i)-\delta))\mathrm{d}q_i + \int_0^1 (s_i^{m'}((F_i)^{-1}(q_i)+\delta) - \overline{s}_i^{m'}(F_i^{-1}(q_i)))\mathrm{d}q_i\\
\overset{(39)}{\to}\; &\int_0^1 \left|\hat{s}_i^*(F_i^{-1}(q_i)) - \hat{s}_i^*((F_i)^{-1}(q_i)-\delta)\right| \mathrm{d}q_i + \left|\hat{s}_i^*(F_i^{-1}(q_i)+\delta) - \hat{s}_i^*((F_i)^{-1}(q_i))\right| \mathrm{d}q_i\\
=\; &\int_{T_i} \left|\hat{s}_i^*(t_i) - \hat{s}_i^*(t_i-\delta)\right| f_i(t_i)\mathrm{d}t_i + \int_{T_i} \left|\hat{s}_i^*(t_i+\delta) - \hat{s}_i^*(t_i)\right| f_i(t_i)\mathrm{d}t_i \quad \to 0 \quad \text{as } \delta \to 0.
\end{aligned}$$

(The limits in the last line have this reason: $\hat{s}_i^*$ is weakly increasing, thus is continuous Lebesgue-a.e., implying that the family of functions $k_\delta(t_i) = \hat{s}_i^*(t_i) - \hat{s}_i^*(t_i - \delta)$ converges to 0 Lebesgue-a.e. t_i as $\delta \to 0$.)

Taking the limit $m' \to \infty$ in (43), we conclude that

$$\hat{s}_i^*(F_i^{-1}(q_i)) = \hat{h}_i^*(q_i) \quad \text{Lebesgue-a.e. } q_i \in [0,1].$$

Combining this with (42), we conclude that

$$\hat{s}_i^*(F_i^{-1}(q_i)) = \overline{h}_i^*(q_i) \quad \text{Lebesgue-a.e. } q_i \in [0,1].$$

Transforming back into type space and using (37), we have

$$\hat{s}_i^*(t_i) = \overline{s}_i^*(t_i) \quad \text{Lebesgue-a.e. } t_i \in T_i. \qquad (44)$$

For any $\mathbf{t} \in \mathbf{T}$, define

$$(c_0^*(\mathbf{t}), \ldots, c_n^*(\mathbf{t})) \;=\; \Phi(s_0^*(\mathbf{t}), \ldots, s_n^*(\mathbf{t})).$$

Define $u_0^*(\mathbf{t}), \ldots, u_n^*(\mathbf{t})$ via payments such that

$$u_i^*(\mathbf{t}) \;=\; \hat{u}_i^*(\underline{t}_i) + \int_{\underline{t}_i}^{t_i} \overline{s}_i^*(y)\mathrm{d}y \quad \text{for all } i \geq 0,\ \mathbf{t} \in \mathbf{T}. \qquad (45)$$

This completes the definition of the allocation

$$\rho^* \;=\; (u_0^*(\cdot), \ldots, u_n^*(\cdot), s_0^*(\cdot), \ldots, c_n^*(\cdot)).$$

It remains to show that ρ^* is strongly neologism-proof.

First we show that ρ^* is F_0-feasible.

[22]Extend the functions $\overline{s}_i^{m'}$ and $\hat{s}_i^*$ constantly to the left and to the right of T_i in this computation.

Because all functions $\overline{s}_0^*, \ldots, \overline{s}_n^*$ are weakly increasing by (44), the envelope condition (45) guarantees that all players' incentive constraints are satisfied. Moreover,

$$\begin{aligned}\lim_{m'} \max_{t_i \in T_i} |\overline{u}_i^{m'}(t_i) - \overline{u}_i^*(t_i)| &= \lim_{m'} \max_{t_i \in T_i} \int_{\underline{t}_i}^{t_i} \mid \overline{s}_i^{m'}(y) - \overline{s}_i^*(y) \mid \mathrm{d}y \\ &\leq \lim_{m'} \int_{\underline{t}_i}^{\bar{t}_i} \mid \overline{s}_i^{m'}(y) - \overline{s}_i^*(y) \mid \mathrm{d}y \\ &\overset{(38),(44)}{=} 0. \end{aligned} \tag{46}$$

Hence, because, for all m and all players i, the allocation ρ^m satisfies the participation constraints for all types in T_i^m, the allocation ρ^* satisfies player i's participation constraints for all types in T_i.

Because weak convergence is preserved under each component of the affine map $\Phi = (\Phi_0, \ldots, \Phi_n)$, (36) implies that

$$\begin{aligned} c_i^{m'} \circ (F^{m'})^{-1} &= \Phi_i \circ \left(s_0^{m'} \circ (F^{m'})^{-1}, \ldots, s_n^{m'} \circ (F^{m'})^{-1} \right) \\ &\to_{\text{weakly}} \Phi_i \circ \left(s_0^* \circ F^{-1}, \ldots, s_n^* \circ F^{-1} \right) = c_i^* \circ F^{-1}. \end{aligned}$$

This allows us to verify the budget-balance condition (31) for ρ^*, as follows:

$$\begin{aligned} 0 &= \int_{\mathbf{T}} \left(\sum_{i=0}^{n} s_i^{m'}(\mathbf{t}) t_i + c_i^{m'}(\mathbf{t}) - u_i^{m'}(\mathbf{t}) \right) \mathrm{d}F^{m'}(\mathbf{t}) \\ &= \sum_{i=0}^{n} \left(\int_{[0,1]} \underbrace{\overline{s}_i^{m'}((F_i^{m'})^{-1}(q_i))}_{\to \overline{s}_i^*(F_i^{-1}(q_i))} \underbrace{(F_i^{m'})^{-1}(q_i)}_{\to F_i^{-1}(q_i)} \mathrm{d}q_i + \int_{[0,1]^{n+1}} c_i^{m'}((F^{m'})^{-1}(q)) \mathrm{d}q \right) \\ &\qquad - \sum_{i=0}^{n} \overline{u}_i^{m'}(t_i) \mathrm{d}F_i^{m'}(t_i) \\ &\overset{(46)}{\to} \sum_{i=0}^{n} \int_{T_i} (\overline{s}_i^*(t_i) t_i - \overline{u}_i^*(t_i) t_i) \mathrm{d}F_i(t_i) + \int_{[0,1]^{n+1}} c_i^*(F^{-1}(q)) \mathrm{d}q \\ &= \int_{T} \left(\sum_{i=0}^{n} s_i^*(\mathbf{t}) t_i + c_i^*(\mathbf{t}) - u_i^*(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t}). \end{aligned}$$

It remains to verify the condition stated in Proposition 1. Suppose it fails. Then there exists a belief G_0 absolutely continuous relative to F_0 and a G_0-feasible allocation $\rho' = (s_0'(\cdot), \ldots, u_n'(\cdot))$ with a higher G_0-ex-ante payoff for the principal than ρ^*.

Define G_0^m such that $\Pr_{G_0^m}[t_i^{m,k}] = G_0(F_0^{-1}(k/m)) - G_0(F_0^{-1}((k-1)/m))$. We will use the shortcuts $G = (G_0, F_{-0})$ and $G^m = (G_0^m, F_{-0}^m)$.

For all m, we define an allocation $\rho'^m = ({s'_0}^m(\cdot), \ldots, {u'_n}^m(\cdot))$: for all $\mathbf{t}^m = (t_0^{m,k_0}, \ldots, t_n^{m,k_n})$ ($k_i \in \{1, \ldots, m\}$),

$$\begin{aligned} {s'_i}^m(\mathbf{t}^m) &= E_G[s'_i(\mathbf{t}) \mid \forall i: \ t_i \in C_i^m(k_i)], \\ {c'_i}^m(\mathbf{t}^m) &= E_G[c'_i(\mathbf{t}) \mid \forall i: \ t_i \in C_i^m(k_i)], \\ {u'_i}^m(\mathbf{t}^m) &= E_G[u'_i(\mathbf{t}) \mid \forall i: \ t_i \in C_i^m(k_i)] + \mathbf{1}_{i=0}\epsilon_0^m - \mathbf{1}_{i\geq 1}\epsilon_i^m(t_i^{m,k}), \end{aligned}$$

where ϵ_0^m and $\epsilon_i^m(t_i^{m,k})$ are defined below.

For all $i \geq 1$, m, and k, let

$$\begin{aligned} \gamma_i^m(t_i^{m,k}) &= m\int_0^{(k-1)/m} \left(\int_{F_i^{-1}(q_i)}^{F_i^{-1}(q_i+\frac{1}{m})} \overline{s'_i}(y)\mathrm{d}y \right. \\ &\quad \left. - \left(F_i^{-1}\left(q_i + \frac{1}{m}\right) - F_i^{-1}(q_i) \right) m \int_{F_i^{-1}(q_i)}^{F_i^{-1}(q_i+\frac{1}{m})} \overline{s'_i}(y) f_i(y)\mathrm{d}y \right) \mathrm{d}q_i \end{aligned}$$

and

$$\epsilon_i^m(t_i^{m,k}) = \gamma_i^m(t_i^{m,k}) - \max_{k'} |\gamma_i^m(t_i^{m,k'})| \ \leq \ 0.$$

Then

$$\begin{aligned} |\gamma_i^m(t_i^{m,k})| &\leq m\int_0^1 \left| F_i^{-1}\left(q_i + \frac{1}{m}\right) - F_i^{-1}(q_i) \right| \\ &\quad \cdot \left| \frac{\int_{F_i^{-1}(q_i)}^{F_i^{-1}\left(q_i+\frac{1}{m}\right)} \overline{s'_i}(y)\mathrm{d}y}{F_i^{-1}\left(q_i+\frac{1}{m}\right) - F_i^{-1}(q_i)} - \frac{\int_{F_i^{-1}(q_i)}^{F_i^{-1}\left(q_i+\frac{1}{m}\right)} \overline{s'_i}(y) f_i(y)\mathrm{d}y}{1/m} \right| \mathrm{d}q_i \\ &= m\int_0^1 \left| F_i^{-1}\left(q_i + \frac{1}{m}\right) - F_i^{-1}(q_i) \right| \cdot |\sigma_1(q_i) - \sigma_2(q_i)| \, \mathrm{d}q_i, \end{aligned}$$

where $\sigma_1(q_i), \sigma_2(q_i) \in [\overline{s'_i}(q_i), \overline{s'_i}(q_i + \frac{1}{m})]$ because $\overline{s'_i}$ is weakly increasing. Therefore,

$$\begin{aligned} |\gamma_i^m(t_i^{m,k})| &\leq m\int_0^1 \left| F_i^{-1}\left(q_i + \frac{1}{m}\right) - F_i^{-1}(q_i) \right| \cdot \left(\overline{s'_i}\left(q_i + \frac{1}{m}\right) - \overline{s'_i}(q_i) \right) \mathrm{d}q_i \\ &\leq \frac{1}{\min_{t_i \in T_i} f_i(t_i)} \int_0^1 \left(\overline{s'_i}\left(q_i + \frac{1}{m}\right) - \overline{s'_i}(q_i) \right) \mathrm{d}q_i \ \to 0 \ \text{ as } m \to \infty. \end{aligned}$$

For all i and m, define a function $\phi_i^m : T_i \to T_i^m$ such that $\phi_i^m(t_i) = t_i^{m,k}$ for all $t_i \in C_i^m(k)$. Observe that $\phi_i^m(t_i) \to t_i$ as $m \to \infty$ for all t_i and hence

$$\epsilon_0^m \ \stackrel{\text{def}}{=} \ \int_T \sum_{i=0}^n s'_i(\mathbf{t})(t_i - \phi_i^m(t_i))\mathrm{d}G(\mathbf{t}) + \sum_{i=1}^n \sum_{k=1}^m \frac{\epsilon_i^m(t_i^{m,k})}{m} \quad \to 0.$$

This completes the definition of ρ'^m. By construction, for all $i \geq 1$ and $t_i^{m,k} \in T_i^m$,

$$\begin{aligned}
\overline{s'_i}^m(t_i^{m,k}) &\stackrel{\text{def}}{=} \int_{\mathbf{T}_{-i}} s'^m_i(t_i^{m,k}, \mathbf{t}_{-i})\mathrm{d}G_{-i}^m(\mathbf{t}_{-i}) \\
&= \frac{1}{\Pr_{F_i}(C_i^{k,m})} \int_{C_i^{m,k}} \int_{\mathbf{T}_{-i}} s'_i(\mathbf{t})\mathrm{d}G_{-i}(\mathbf{t}_{-i})\mathrm{d}F_i(t_i) \\
&= m \int_{C_i^{m,k}} \overline{s'_i}(t_i)\mathrm{d}F_i(t_i) \\
&= m \int_{F_i^{-1}((k-1)/m)}^{F_i^{-1}(k/m)} \overline{s'_i}(t_i) f_i(t_i)\mathrm{d}t_i \\
&= m \int_{(k-1)/m}^{k/m} \overline{s'_i}(F_i^{-1}(q_i))\mathrm{d}q_i,
\end{aligned}$$

and similar for $\overline{c'_i}^m$ and $\overline{u'_i}^m$. In particular, the agents' participation constraints are satisfied for ρ'^m. To verify the agents' incentive constraints, notice that, using the shortcut

$$\Delta = \epsilon_i^m(t_i^{m,k+1}) - \epsilon_i^m(t_i^{m,k}) = \gamma_i^m(t_i^{m,k+1}) - \gamma_i^m(t_i^{m,k}),$$

we have

$$\begin{aligned}
\overline{u'_i}^m(t_i^{m,k+1}) - \overline{u'_i}^m(t_i^{m,k}) &= m \int_{C_i^{m,k+1}} \overline{u'_i}(t_i)\mathrm{d}F_i(t_i) - m \int_{C_i^{m,k}} \overline{u'_i}(t_i)\mathrm{d}F_i(t_i) - \Delta \\
&= m \int_{k/m}^{(k+1)/m} \overline{u'_i}(F_i^{-1}(q_i))\mathrm{d}q_i - m \int_{(k-1)/m}^{k/m} \overline{u'_i}(F_i^{-1}(q_i))\mathrm{d}q_i - \Delta \\
&= m \int_{(k-1)/m}^{k/m} \left(\overline{u'_i}(F_i^{-1}(q_i + \frac{1}{m})) - \overline{u'_i}(F_i^{-1}(q_i)) \right) \mathrm{d}q_i - \Delta \\
&= m \int_{(k-1)/m}^{k/m} \int_{F_i^{-1}(q_i)}^{F_i^{-1}(q_i + \frac{1}{m})} \overline{s'_i}(y)\mathrm{d}y\mathrm{d}q_i - \Delta \\
&= m^2 \int_{(k-1)/m}^{k/m} (F_i^{-1}(q_i + \frac{1}{m}) - F_i^{-1}(q_i)) \int_{F_i^{-1}(q_i)}^{F_i^{-1}(q_i + \frac{1}{m})} \overline{s'_i}(y) f_i(y)\mathrm{d}y\mathrm{d}q_i.
\end{aligned}$$

Then, by the first mean value theorem for integration, there exists $\xi_i \in [(k-1)/m, k/m]$ such that

$$\begin{aligned} \ldots &= m^2 \int_{F_i^{-1}(\xi_i)}^{F_i^{-1}(\xi_i+\frac{1}{m})} \overline{s_i'}(y) f_i(y) \mathrm{d}y \int_{(k-1)/m}^{k/m} (F_i^{-1}(q_i + \frac{1}{m}) - F_i^{-1}(q_i)) \mathrm{d}q_i \\ &= \frac{1}{1/m} \int_{F_i^{-1}(\xi_i)}^{F_i^{-1}(\xi_i+\frac{1}{m})} \overline{s_i'}(y) f_i(y) \mathrm{d}y \cdot (t_i^{m,k+1} - t_i^{m,k}) \\ &= \frac{1}{1/m} \int_{\xi_i}^{\xi_i+\frac{1}{m}} \overline{s_i'}(F_i^{-1}(y)) \mathrm{d}y \cdot (t_i^{m,k+1} - t_i^{m,k}) \\ & \begin{cases} \geq \overline{s}_i^m(t_i^{m,k})(t_i^{m,k+1} - t_i^{m,k}) \\ \leq \overline{s}_i^m(t_i^{m,k+1})(t_i^{m,k+1} - t_i^{m,k}), \end{cases} \end{aligned}$$

showing incentive compatibility.

Moreover, due to the correcting term ϵ_0^m, the ex-ante budget balance condition for ρ' implies that the ex-ante budget balance condition holds for ρ'^m. Finally,

$$\max_{t_0 \in T_0^m} \mid \overline{u_0'}^m(t_0) - \overline{u_0'}(t_0) \mid \quad \leq |\epsilon_0^m| + 2 \sum_{i=1}^{m} \max_{k \in \{1,\ldots,m\}} |\gamma_i^m(t_i^{m,k})| \ \to 0 \quad \text{as } m \to \infty.$$

Hence, the principal's G_0^m-ex-ante payoff according to $(\rho')^m$ converges to the principal's G_0-ex-ante payoff according to ρ' as $m \to \infty$. This contradicts the fact that ρ^m is strongly neologism-proof for all m. *QED*

9.4. **Proof of Proposition 3.** From Proposition 9 and Corollary 2, the solutions to the principal's F_0-ex-ante problem are precisely the solutions to the principal's relaxed F_0-ex-ante problem in which the principal's incentive and participation constraints are satisfied.

Using (31) to rewrite the objective, the principal's relaxed F_0-ex-ante problem is to

$$\begin{aligned} \max_{u_1(\mathbf{t}),\ldots,u_n(\mathbf{t}),(s_0(\mathbf{t}),\ldots,c_n(\mathbf{t})) \in \mathcal{V}} \quad & \int_T \left(\sum_{i=0}^{n} s_i(\mathbf{t}) t_i + c_i(\mathbf{t}) - \sum_{i=1}^{n} u_i(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t}), \\ \text{s.t.} \quad & \overline{s}_i(\cdot) \text{ weakly increasing for all } i \geq 1, \\ & \overline{u}_i(t_i) = \overline{u}_i(\underline{t}_i) + \int_{\underline{t}_i}^{t_i} \overline{s}_i(y) \mathrm{d}y \text{ for all } i \geq 1, t_i \in T_i, \\ & \overline{u}_i(t_i) \geq 0 \text{ for all } i \geq 1, t_i \in T_i. \end{aligned}$$

Using the virtual valuation functions $\psi_i(t_i) = t_i - (1 - F_i(t_i))/f_i(t_i)$ and (29), the objective of this problem can be rewritten as

$$\int_T \left(s_0(\mathbf{t}) t_0 + c_0(\mathbf{t}) + \sum_{i=1}^{n} s_i(\mathbf{t}) \psi_i(t_i) + c_i(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t}) - \sum_{i=1}^{n} \overline{u}_i(\underline{t}_i),$$

Here, only the lowest-types' utilities $\underline{u}_i = \overline{u}_i(\underline{t}_i)$ occur. Thus we can define payments to satisfy (29) in a separate (second) step, and can simplify the maximization problem as follows:

$$\max_{(\underline{u}_1,\ldots,\underline{u}_n)\in\mathbf{R}^n,\ (s_0(\mathbf{t}),\ldots,c_n(\mathbf{t}))\in\mathcal{V}} \int_T \left(s_0(\mathbf{t})t_0 + c_0(\mathbf{t}) + \sum_{i=1}^n s_i(\mathbf{t})\psi_i(t_i) + c_i(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t}) - \sum_{i=1}^n \underline{u}_i,$$

$$\text{s.t.} \quad \overline{s}_i(\cdot) \text{ weakly increasing for all } i \geq 1, \tag{47}$$

$$\underline{u}_i + \int_{\underline{t}_i}^{t_i} \overline{s}_i(y)\mathrm{d}y \geq 0 \text{ for all } i \geq 1, t_i \in T_i.$$

Defining

$$\mathcal{M} \;=\; \{(s_0(\cdot),\ldots,c_n(\cdot)) \mid (47)\},$$

we have to

$$\max_{(\underline{u}_1,\ldots,\underline{u}_n)\in\mathbf{R}^n,\ (s_0(\cdot),\ldots,c_n(\cdot))\in\mathcal{M}} \int_T \left(s_0(\mathbf{t})t_0 + c_0(\mathbf{t}) + \sum_{i=1}^n s_i(\mathbf{t})\psi_i(t_i) + c_i(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t}) - \sum_{i=1}^n \underline{u}_i,$$

$$\text{s.t.} \quad \underline{u}_i + \int_{\underline{t}_i}^{t_i} \overline{s}_i(y)\mathrm{d}y \geq 0 \text{ for all } i \geq 1, t_i \in T_i.$$

Using the Lagrange approach (e.g., Luenberger (1969), Chapter 8), we have to

$$\max_{(\underline{u}_1,\ldots,\underline{u}_n)\in\mathbf{R}^n,\ (s_0(\cdot),\ldots,c_n(\cdot))\in\mathcal{M}} \int_T \left(s_0(\mathbf{t})t_0 + c_0(\mathbf{t}) + \sum_{i=1}^n s_i(\mathbf{t})\psi_i(t_i) + c_i(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t}) - \sum_{i=1}^n \underline{u}_i + \sum_{i=1}^n \int_{T_i} \left(\underline{u}_i + \int_{\underline{t}_i}^{t_i} \overline{s}_i(y)\mathrm{d}y \right) \mathrm{d}z_i^*(t_i),$$

where z_i^* $(i \geq 1)$ is a right-continuous and weakly increasing function on T_i such that

$$\sum_{i=1}^n \int_{T_i} \left(\underline{u}_i^* + \int_{\underline{t}_i^*}^{t_i} \overline{s}_i^*(y)\mathrm{d}y \right) \mathrm{d}z_i^*(t_i) \;=\; 0, \tag{48}$$

where $(\underline{u}_1^*,\ldots,\underline{u}_n^*, s_0^*(\cdot),\ldots,c_n^*(\cdot))$ denotes a solution to the maximization problem.

Because the solution value is the same as to the F_0-ex-ante optimization problem (e.g., Luenberger (1969), Chapter 8), we cannot reach arbitrarily high values, implying that $z_i^*(\bar{t}_i) = 1$ for all i (otherwise $\underline{i}$ could be chosen to achieve arbitrarily high values for the objective).

Hence, $\underline{u}_1, \ldots, \underline{u}_n$ cancel out and the objective becomes

$$\int_T \left(s_0(\mathbf{t})t_0 + c_0(\mathbf{t}) + \sum_{i=1}^n s_i(\mathbf{t})\psi_i(t_i) + c_i(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t})$$
$$+ \sum_{i=1}^n \int_{T_i} \int_{\underline{t}_i}^{t_i} \overline{s}_i(y)\mathrm{d}y\mathrm{d}z_i^*(t_i),$$

Using integration by parts, we can rewrite the objective as

$$\int_T \left(s_0(\mathbf{t})t_0 + c_0(\mathbf{t}) + \sum_{i=1}^n s_i(\mathbf{t})\psi_i^{z_i^*}(t_i) + c_i(\mathbf{t}) \right) \mathrm{d}F(\mathbf{t}),$$

By the arguments of Myerson (1981), maximization of this objective is equivalent to (9), provided that there exists a solution to (9) that belongs to $\mathcal{M}$. The existence of a solution that belongs to $\mathcal{M}$ is argued as follows. Observe that

$$(s_0^*(\mathbf{t}), \ldots, c_n^*(\mathbf{t})) \quad \in \quad \arg\max_{b \in \mathcal{V}} b \cdot d,$$

where

$$\begin{aligned} b &= (\hat{s}_0, \ldots, \hat{c}_n), \\ d &= (t_0, \overline{\psi}_1^*(t_1), \ldots, \overline{\psi}_n^*(t_n), 1, \ldots, 1). \end{aligned}$$

If t_i becomes larger, one component of d becomes larger, or d remains constant. Consider the problem to $\max_{b\in\mathcal{V}} b \cdot d^j$ for two vectors $d^1, d^2 \in \mathbb{R}^m$ with $d^2 = d^1 + (\delta, 0, \ldots, 0)$ for some $\delta > 0$. Let $b^j \in \arg\max_b b \cdot d^j$ for $j = 1, 2$. Then we claim that $b_1^2 \geq b_1^1$. To see this, consider $\hat{b} \in \mathcal{V}$ such that $\hat{b}_1 < b_1^1$. By optimality of b^1, we have $\hat{b} \cdot d^1 \leq b^1 \cdot d^1$, implying $\hat{b} \cdot d^2 < b^1 \cdot d^2$. Hence, $\hat{b} \notin \arg\max_b b \cdot d^2$, as was to be shown.

Finally, observe that (48) is equivalent to (8). Hence, a solution to the principal's relaxed F_0-ex-ante problem is characterized by the conditions (8), (9), (10) for $i \neq 0$, (11) for $i \neq 0$, and (12) for $i \neq 0$. The additional conditions (10), (11), and (12) for $i = 0$ are the principal's incentive and participation constraints. This completes the characterization.

9.5. **Proof of Proposition 6.** We make use of the conditions provided in Proposition 3. Observe that

$$V^{z_1^*,\ldots,z_n^*}(v, \mathbf{t}) \quad = \quad \overline{\psi}_1^{z_1^*}(t_1)\hat{s}_1 - t_0\hat{s}_0.$$

Condition (9) yields that $s_1(\mathbf{t}) = 1 - \underline{\alpha}$ if $\overline{\psi}_1^{z_1^*}(t_1) > t_0$ and $s_1(\mathbf{t}) = -\underline{\alpha}$ if $\overline{\psi}_1^{z_1^*}(t_1) < t_0$. Therefore,

$$\overline{s}_1(t_1) \quad = \quad F_0(\overline{\psi}_1^{z_1^*}(t_1)) - \underline{\alpha} \quad \text{a.e. } t_1. \tag{49}$$

Let $[\underline{y}_1, \overline{y}_1]$ denote the interval of types t_1 such that $\overline{u}_1(t_1) = 0$. By the monotonicity condition (10), if $t_1 > \overline{y}_1$, then $\overline{s}(t_1) \geq 0$. In fact, by definition of $\overline{y}_1$ and the envelope formula (11),

$$\overline{s}_1(t_1) > 0 \text{ if } t_1 > \overline{y}_1, \text{ and } \overline{s}_1(t_1) < 0 \text{ if } t_1 < \underline{y}_1. \tag{50}$$

First, we show that $\underline{y}_1 < \overline{y}_1$. Suppose that $\underline{y}_1 = \overline{y}_1$. Then $z_1^*(t_1) = \mathbf{1}_{t_1 \geq \underline{y}_1}$ by (8), implying that $\psi_1^{z_1^*}(t_1) = \psi^s(t_1)$ if $t_1 < \underline{y}_1$ and $\psi_1^{z_1^*}(t_1) = \psi^b(t_1)$ if $t_1 \geq \underline{y}_1$.

Suppose, furthermore, that $\underline{y}_1 = \overline{y}_1 = \underline{t}_1$. Then $\psi_1^{z_1^*} = \psi^b$ is strictly increasing. Hence, $\overline{\psi}_1^{z_1^*} = \psi^b$. Hence, for $t_1 \approx 0$, $\overline{\psi}_1^{z_1^*}(t_1) < 0$, implying $\overline{s}_1(t_1) = -\underline{\alpha} < 0$ by (49). That is, $\overline{u}_1$ is strictly decreasing at $t_1 \approx 0$ by (11), a contradiction to $\overline{u}_1(\underline{t}_1) = \overline{u}_1(\underline{y}_1) = 0$ and (12). For a similar reason, it cannot be that $\underline{y}_1 = \overline{y}_1 = \overline{t}_1$.

Thus, suppose that $\underline{y}_1 = \overline{y}_1 \in (\underline{t}_1, \overline{t}_1)$. Because $\psi_1^{z_1^*}$ jumps downwards at $\underline{y}_1$, ironing implies that, for all t_1 in an open neighborhood of $\underline{y}_1$, the function $\overline{\psi}_1^{z_1^*}(t_1)$ is constant. Hence, by (49) and (10), $\overline{s}_1(t_1) = \text{const.}$ for all t_1 in the open neighborhood, contradicting (50).

Hence, $\underline{y}_1 < \overline{y}_1$. Then $\overline{s}_1(t_1) = 0$ on $(\underline{y}_1, \overline{y}_1)$ by the envelope formula (11). Hence, using (49) and (50), for a.e. t_1,

$$\overline{\psi}_1^{z_1^*}(t_1) \quad \begin{cases} > F_0^{-1}(\underline{\alpha}_1) & \text{if } t_1 > \overline{y}_1, \\ = F_0^{-1}(\underline{\alpha}_1) & \text{if } t_1 \in [\underline{y}_1, \overline{y}_1], \\ < F_0^{-1}(\underline{\alpha}_1) & \text{if } t_1 < \underline{y}_1. \end{cases} \tag{51}$$

This extends to all t_1 because $\overline{\psi}_1^{z_1^*}$ is continuous. Notice that $\underline{y}_1 > \underline{t}_1$ because otherwise $\psi_1^{z_1^*}(\underline{t}_1) \geq \overline{\psi}_1^{z_1^*}(\underline{t}_1)$, but this is impossible because

$$\psi_1^{z_1^*}(\underline{t}_1) \leq \psi^s(\underline{t}_1) = \underline{t}_1 = 0 < F_0^{-1}(\underline{\alpha}_1) = \overline{\psi}_1^{z_1^*}(\underline{t}_1).$$

Similarly, $\overline{y}_1 < \overline{t}_1$.

From (8), $\psi_1^{z_1^*}(t_1) = \psi^s(t_1)$ for all $t_1 < \underline{y}_1$. Hence, because ψ^s is strictly increasing and using (51),

$$\overline{\psi}_1^{z_1^*}(t_1) \quad = \quad \psi^s(t_1) \quad \text{for all } t_1 < \underline{y}_1.$$

Similarly, because ψ^b is strictly increasing,

$$\overline{\psi}_1^{z_1^*}(t_1) \quad = \quad \psi^b(t_1) \quad \text{for all } t_1 > \overline{y}_1.$$

Hence, using (51) and the continuity of $\overline{\psi}_1^{z_1^*}$, $\psi^s(\underline{y}_1) = F_0^{-1}(\underline{\alpha})$, implying

$$\underline{y}_1 = (\psi^s)^{-1}(F_0^{-1}(\underline{\alpha})).$$

Similarly,

$$\overline{y}_1 = (\psi^b)^{-1}(F_0^{-1}(\underline{\alpha})).$$

This completes the proof.

References

Balestrieri, F. (2008): "A modified English auction for an informed buyer," mimeo.

Balkenborg, D., and M. Makris (2010): "An Undominated Mechanism for a Class of Informed Principal Problems with Common Values," *mimeo*.

Beaudry, P. (1994): "Why an Informed Principal May Leave Rents to an Agent," *International Economic Review*, 35(4), 821–32.

Bond, E. W., and T. A. Gresik (1997): "Competition between asymmetrically informed principals," *Economic Theory*, 10(2), 227–240.

Börgers, T., and P. Norman (2009): "A note on budget balance under interim participation constraints: the case of independent types," *Economic Theory*, 39(3), 477–489.

Celik, G. (2009): "Mechanism design with collusive supervision," *Journal of Economic Theory*, 144(1), 69–95.

Cella, M. (2008): "Informed principal with correlation," *Games and Economic Behavior*, 64(2), 433–456.

Chade, H., and R. Silvers (2002): "Informed principal, moral hazard, and the value of a more informative technology," *Economics Letters*, 74(3), 291–300.

Che, Y.-K., and J. Kim (2006): "Robustly Collusion-Proof Implementation," *Econometrica*, 74(4), 1063–1107.

Cramton, P., R. Gibbons, and P. Klemperer (1987): "Dissolving a Partnership Efficiently," *Econometrica*, 55(3), 615–32.

Eliaz, K., and R. Spiegler (2007): "A Mechanism-Design Approach to Speculative Trade," *Econometrica*, 75(3), 875–884.

Figueroa, N., and V. Skreta (2009): "The role of optimal threats in auction design," *Journal of Economic Theory*, 144(2), 884–897.

Fleckinger, P. (2007): "Informed principal and countervailing incentives," *Economics Letters*, 94(2), 240–244.

Francetich, A., and P. Troyan (2012): "Collusion Agreements in Auctions: Design and Execution by an Informed Principal," *mimeo*.

Garratt, R. J., T. Tröger, and C. Z. Zheng (2009): "Collusion via Resale," *Econometrica*, 77(4), 1095–1136.

Gershkov, A., J. K. Goeree, A. Kushnir, B. Moldovanu, and X. Shi (forthcoming): "On the equivalence of Bayesian and dominant strategy implementation," *Econometrica*.

Hafalir, I., and V. Krishna (2008): "Asymmetric Auctions with Resale," *American Economic Review*, 98(1), 87–112.

Halac, M. (2012): "Relational Contracts and the Value of Relationships," *American Economic Review*, 102(2), 750–79.

Izmalkov, S., and F. Balestrieri (2012): "The informed seller problem: The case of horizontal differentiation," *mimeo*.

Jehiel, P., B. Moldovanu, and E. Stacchetti (1999): "Multidimensional Mechanism Design for Auctions with Externalities," *Journal of Economic Theory*, 85(2), 258–293.

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.

Jullien, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93(1), 1–47.

Kaya, A. (2010): "When Does It Pay To Get Informed?," *International Economic Review*, 51(2), 533–551.

Kirkegaard, R. (2009): "Asymmetric first price auctions," *Journal of Economic Theory*, 144(4), 1617–1635.

Krähmer, D. (2012): "Auction design with endogenously correlated buyer types," *Journal of Economic Theory*, 147(1), 118–141.

Lebrun, B. (2009): "Auctions with almost homogeneous bidders," *Journal of Economic Theory*, 144(3), 1341–1351.

Ledyard, J. O., and T. R. Palfrey (2007): "A general characterization of interim efficient mechanisms for independent linear environments," *Journal of Economic Theory*, 133(1), 441–466.

Lewis, T. R., and D. E. M. Sappington (1989): "Countervailing incentives in agency problems," *Journal of Economic Theory*, 49(2), 294–313.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*. John Wiley and Sons, Inc., New York, NY.

Manelli, A. M., and D. R. Vincent (2010): "Bayesian and Dominant Strategy Implementation in the Independent Private Values Model," *Econometrica*, 78(6), 1905–1938.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*. Oxford University Press.

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. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

——— (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.

Myerson, R. B., and M. A. Satterthwaite (1983): "Efficient mechanisms for bilateral trading," *Journal of Economic Theory*, 29(2), 265–281.

Mylovanov, T., and T. Tröger (2012): "Informed-principal problems in environments with generalized private values," *Theoretical Economics*, 7, 465–488.

Nishimura, T. (2012): "Scoring Auction by an Informed Principal," *mimeo.*

Pavlov, G. (2008): "Auction design in the presence of collusion," *Theoretical Economics*, 3(3).

Riley, J., and R. J. Zeckhauser (1983): "Optimal selling strategies: when to haggle, when to hold firm," *Quarterly Journal of Economics*, 98, 267–289.

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.

Williams, S. R. (1987): "Efficient performance in two-agent bargaining," *Journal of economic theory*, 41(1), 154–172.

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.

Zheng, C. Z. (2002): "Optimal auction with resale," *Econometrica*, 70(6), 2197–2224.