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

TYMOFIY MYLOVANOV AND THOMAS TRÖGER

Abstract. We study the problem of mechanism design by an informed party in private value environments with transferrable utility. We show that, at the interim stage, it is optimal for the proposing party to offer a mechanism that would also be optimal for her ex ante, that is, before she has obtained her private information.

1. Introduction

The optimal design of contracts and institutions in the presence of privately informed market participants is central to economics. For the environments with transferable utility, a rich theory has emerged (see, e.g., Krishna (2002) and Milgrom (2004)) with applications including auctions, procurement, public good provision, organizational contract design, legislative bargaining, collusion, labor contracts, etc. Transferability of utility makes many problems tractable, allows for a clean welfare analysis, and has proved very rich in terms of explanatory power.

A restriction in much of this theory is that a contract or a mechanism is either designed by a third party, e.g., a fictitious benevolent planner, or is proposed by a party who has no private information. As such, the theory is not applicable to environments in which contracts or institutions arise endogenously as a choice of privately informed market participants.

One difficulty that arises if a contract is proposed by a privately informed party is that a mechanism proposal might reveal some of the proposer's private information to the other participants and the value of the mechanism is now determined endogenously in equilibrium. In addition, there is a time inconsistency problem: the preferences of the proposing party over the mechanisms might depend non-trivially on her information and differ from those before her information is realized. Consequently, the standard maximization approach of mechanism design is not applicable.

The seminal references on the mechanism design problem by an informed principal are Myerson (1983) and Maskin and Tirole (1990, 1992). Myerson's characterization is, however, complex and excludes standard trading environments with continuous types and transferrable utility. Maskin and Tirole's characterization, on the other hand, requires quite specific structure.

In this paper, we provide a simple and general answer that applies to arbitrary environments with transferrable utility and private values, that is, environments in which each agent's private information is not directly payoff-relevant to the other agents. Examples include any markets for private or public goods in which each agent is privately informed about her production costs or her willingness-to-pay, legislative bargaining with private values and transferable utility, and bargaining over bets by speculators with heterogenous private beliefs.

We show that, at the interim stage, it is optimal for the proposing agent to offer a mechanism that would also be optimal for her ex ante, that is, before she has obtained her private information.[1] Our result provides a justification for considering mechanisms that are ex-ante optimal for an agent who has full bargaining power at the interim stage.

The significance of this result is three-fold. First, it connects the informed principal problem to the standard approach in mechanism design that is used to characterize ex-ante optimal mechanisms. Second, despite the time inconsistency problem and the issue of information leakage at the proposal stage, in equilibrium a privately informed principal picks up the entire expected surplus that were available to her ex-ante before her information is realized. Finally, 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.

Prima facie the result about the equivalence between ex-ante and interim selected mechanisms might appear almost trivial. For an environment with private values and transferrable utility, Maskin and Tirole (1990) show that there exists a fully separating equilibrium in which each type of the principal offers a different mechanism, and that this equilibrium attains the same expected payoff payoff as an ex-ante optimal mechanism. Hence, neither time inconsistency nor information leakage are of essence in this environment.

Intuitively, in private value environments, the principal's private information is not directly relevant for the other parties and it is hard to see why revealing this information should affect the equilibrium outcome. The case in point is a bargaining environment over one unit of good (Myerson and Satterthwaite 1983). The optimal mechanism for the seller whose cost is commonly known is a posted price. The ex-ante optimal mechanism maximizing her expected payoff over costs is a collection of these optimal posted prices. As the agent is unconcerned about the principal's cost, the price is the only relevant parameter and the ex-ante optimal mechanism is dominant strategy incentive compatible and individually rational. In a separating equilibrium, a privately informed principal reveals her information through choice of the posted price but this is inconsequential (Yilankaya 1999). Similar arguments have been made for other private value environments (Tan 1996, Balestrieri 2008, Skreta 2009, Mylovanov and Tröger 2008).

This intuition, however, is incomplete and relies on the assumption about the nature of a disagreement outcome that realizes if parties fail to agree on a mechanism. As an illustration, we study the informed principal problem in the basic bargaining environment of Myerson and Satterthwaite, with a minor modification that the disagreement outcome is a lottery that allocates the good to the agent with probability $\alpha \in (0,1)$ and to the principal otherwise.[2] (If $\alpha \in \{0,1\}$, we have the classic case in which the principal is either a seller or a buyer of the good and the optimal mechanism is a posted price.)

Surprisingly, in this environment the ex-ante optimal mechanism selected by the informed principal is not a posted price. Instead, it is a collection of a participation fee for the

[1] Our solution concept for the mechanism-selection game is perfect-Bayesian equilibrium that has a strong "neologism-proofness" property (Mylovanov and Tröger forthcoming). Neologism-proofness was introduced by Farrell (1993); it is also related to the concept of perfect sequential equilibrium by Grossman and Perry (1986). Neologism-proof equilibria are stronger than the intuitive criterion (Cho and Kreps 1987) and undefeated equilibrium Mailath, Okuno-Fujiwara, and Postlewaite (1993).

[2] An equivalent interpretation is that the good is divisible and α and $1-\alpha$ are initial endowments.

agent, a buy-out option for the principal, and a resale stage.[3] The mechanism is incentive compatible in dominant strategies but only Bayesian individually rational. In equilibrium, the agent is kept in the dark about the principal's type and is uncertain, at the moment of accepting the mechanism, whether the principal will exercise her buy-out option. Concealing her information allows the principal to decrease the agent's information rents: individual rationality constraint for agent types is violated for some (different) principal types, but is satisfied in expectation over the principal's types.

2. Model

Consider players $i = 0, \ldots, n$ 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 principal and the agents, $A = \{0, \ldots, n\}$, 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 $t_0, \ldots, t_n$ are realizations of stochastically independent Borel probability measures $p_0, \ldots, p_n$ 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 $\mathbf{x} = (x_1, \ldots, x_n)$, and $v_0, \ldots, v_n$ 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.

[3]The structure of this mechanism is interesting for independent reasons. For example, the usual distortion of no trade is not a part of this mechanism. Instead, there is always trade, but sometimes the realized gains from trade are negative. In addition, the optimal contract combines different features observed in practice. For instance, in application of bargaining between a downsizing firm (principal) and a worker (agent) about new level of employment, the optimal mechanism can be interpreted as consisting of a fixed decrease in wage, a golden parachute that can be triggered by the firm, buy-out options that can be triggered by employees, and renegotiation over wage.

If the players cannot agree on an outcome, some exogenously given disagreement outcome z_0 obtains. The disagreement outcome $z_0 = (\alpha_0, 0, \dots, 0)$ for some (possibly random) collective action $\alpha_0 \in \mathcal{A}$. We normalize the valuation functions such that each player's expected valuation from the disagreement outcome equals 0, that is, $\int_A v_i(a, t_i) \mathrm{d}\alpha_0(a) = 0$ for all i and t_i.

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

$$u_i(\zeta, t_i) \;=\; \int_A v_i(a, t_i) \mathrm{d}\alpha(a) - x_i \;\stackrel{\text{def}}{=}\; \int_Z u_i(z, t_i) \mathrm{d}\zeta(z),$$

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

A complete type-dependent description of the result of the players' interaction is given by an *allocation*, which is 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).

Given the presence of private information, incentive and participation constraints will play a major role in our analysis. Here, 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. To take account of this possibility, we work in the following with an arbitrary belief q_0 that is absolutely continuous relative to p_0.

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

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

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 (1) and the q_0-participation constraints (2) are satisfied,

$$\forall t_i, \hat{t}_i \in T_i : \qquad U_i^{\rho, q_0}(t_i) \geq U_i^{\rho, q_0}(\hat{t}_i, t_i), \tag{1}$$

$$\forall t_i \in T_i : \qquad U_i^{\rho, q_0}(t_i) \geq 0. \tag{2}$$

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 $t_0 \in \mathrm{supp}(q_0)$.

We are interested in the problem of optimally selecting a mechanism in the absence of a disinterested outsider. Rather, one of the players is designated as the proposer of the mechanism. We will assume from now on that the proposer is player 0. We call her the principal; the other players are called agents.

In earlier work (Mylovanov and Tröger forthcoming), we have introduced strongly neologism-proof allocations as a solution to the principal's mechanism-selection problem in environments with generalized private values, compact outcome spaces, and finite type spaces. Adapted to quasi-linear environments with arbitrary type spaces, the definition reads as follows.[4]

Definition 1. *An allocation ρ is called* 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.*

The idea behind neologism-proofness is that any observed deviation from an equilibrium mechanism should be accompanied by a "credible" belief. We call a belief credible if none of the principal-types who would suffer from the deviation is believed to make it. An allocation ρ fails strong neologism-proofness if an allocation ρ' together with a credible belief q_0 can be found such that a q_0-positive mass of types (hence, p_0-positive mass of types) strictly prefers ρ' to ρ, that is, if ρ is q_0-dominated by ρ'.

In Mylovanov and Tröger (forthcoming), we show for the finite-type-space environments considered there that any strongly neologism-proof allocation is a perfect-Bayesian equilibrium outcome of a mechanism-selection game in which any finite game form with perfect recall may be proposed as a mechanism. In environments with infinite type spaces, there is no "natural" set of feasible mechanisms, nor is there an obvious choice for the definition of equilibrium.[5] On the other hand, the basic idea behind our arguments in the earlier paper (Mylovanov and Tröger, forthcoming, proof of Proposition 1) is simple and appears quite general. Therefore, we see no point in extending our earlier definition of the mechanism-selection game and will make no attempt at doing so.

It will be shown 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

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

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.

Any solution with $q_0 = p_0$ is called an *ex-ante optimal* allocation.

3. Results

3.1. **Characterization.** In this section we present a characterization of strong neologism-proofness in quasi-linear environments. Strong neologism-proofness requires, for all beliefs

[4]The corresponding definition in Mylovanov and Tröger (forthcoming) includes provisions about "happy types" who obtain the highest feasible payoff. In quasilinear environments, there are no happy types because payments can be arbitrarily high.

[5]Any definition would have to deal with exempting deviations to mechanisms such that in the resulting continuation games no equilibrium exists, or in which the belief-equilibrium correspondence does not have the requisite continuity properties. See Zheng (2002) for an approach in a related context.

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. In particular, by setting $q_0 = p_0$ it follows that any strongly neologism-proof allocation is ex-ante optimal; this result is most convenient in environments with continuous type spaces where the ex-ante optimal payoffs are unique.

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

$$\eta(q_0) \;\leq\; \int_{T_0} U_0^{\rho}(t_0)\,dq_0(t_0) \quad \textit{for all beliefs } q_0. \tag{3}$$

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 (3) 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 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 types 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

with 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.

Proof. "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 (3) fails, that is

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

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

$$\int_{T_0} U_0^{\rho'}(t_0)\mathrm{d}q_0(t_0) - \int_{T_0} U_0^{\rho}(t_0)\mathrm{d}q_0(t_0) \stackrel{\text{def}}{=} \epsilon > 0. \tag{4}$$

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

$$\begin{aligned} x_1''(\mathbf{t}) &= x_1'(\mathbf{t}) - (U_0^{\rho}(t_0) - U_0^{\rho'}(t_0) + \epsilon). \\ x_i''(\mathbf{t}) &= x_i'(\mathbf{t}), \quad i = 2, \ldots, n. \end{aligned} \tag{5}$$

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

$$\begin{aligned} U_1^{\rho'',q_0}(\hat{t}_1, t_1) &= \int_{T_{-1}} \int_A v_1(a, t_1)\mathrm{d}\alpha'(\hat{t}_1, \mathbf{t}_{-1})(a)\mathrm{d}\mathbf{q}_{-1}(\mathbf{t}_{-1}) - \int_{T_{-1}} x_1''(\hat{t}_1, \mathbf{t}_{-1})\mathrm{d}\mathbf{q}_{-1}(\mathbf{t}_{-1}) \\ &\stackrel{(5)}{=} U_1^{\rho',q_0}(\hat{t}_1, t_1) + \int_{T_0} (U_0^{\rho}(t_0) - U_0^{\rho'}(t_0))\mathrm{d}q_0(t_0) + \epsilon \\ &\stackrel{(4)}{=} U_1^{\rho',q_0}(\hat{t}_1, t_1). \end{aligned}$$

For all $t_0 \in T_0$,

$$U_0^{\rho''}(t_0) - U_0^{\rho}(t_0) \stackrel{(5)}{=} U_0^{\rho'}(t_0) + (U_0^{\rho}(t_0) - U_0^{\rho'}(t_0) + \epsilon) - U_0^{\rho}(t_0) = \epsilon. \tag{6}$$

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 $t_0 \in \mathrm{supp}(r_0)$,

$$U_0^{\sigma}(t_0) \;\geq\; U_0^{\rho}(t_0) + \frac{1}{2}\epsilon. \tag{7}$$

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

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

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

$$\forall\ t_0, t_0' \in T_0 : \quad \text{if } |t_0 - t_0'| < \delta' \text{ then } |U_0^\rho(t_0) - U_0^\rho(t_0')| < \frac{\epsilon}{8}. \tag{9}$$

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

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 $x_i'''(\cdot)$ $(i = 1, \dots, n)$ 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 $t_0 \in D_k$,[6]

$$u_0(\rho'''(\mathbf{t}), t_0) = \frac{1}{q_0(D_k)} \int_{D_k} u_0(\rho''(t_0', \mathbf{t}_{-0}), t_0) \mathrm{d}q_0(t_0'). \tag{10}$$

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

$$\begin{aligned} U_0^{\rho'''}(t_0) &= \int_{\mathbf{T}_{-0}} u_0(\rho'''(\mathbf{t}), t_0) \mathrm{d}\mathbf{p}(t_{-0}) \\ &\overset{(10)}{=} \frac{1}{q_0(D_k)} \int_{D_k} \int_{\mathbf{T}_{-0}} u_0(\rho''(t_0', \mathbf{t}_{-0}), t_0) \mathrm{d}\mathbf{p}(t_{-0}) \mathrm{d}q_0(t_0') \\ &\overset{(8)}{>} \frac{1}{q_0(D_k)} \int_{D_k} \int_{\mathbf{T}_{-0}} (u_0(\rho''(t_0', \mathbf{t}_{-0}), t_0') - \frac{\epsilon}{8}) \mathrm{d}\mathbf{p}(t_{-0}) \mathrm{d}q_0(t_0') \\ &= \frac{1}{q_0(D_k)} \int_{D_k} (U_0^{\rho''}(t_0') - \frac{\epsilon}{8}) \mathrm{d}q_0(t_0') \\ &\overset{(6)}{=} \frac{1}{q_0(D_k)} \int_{D_k} (U_0^{\rho}(t_0') + \frac{7}{8}\epsilon) \mathrm{d}q_0(t_0') \\ &\overset{(9)}{>} \frac{1}{q_0(D_k)} \int_{D_k} (U_0^{\rho}(t_0) + \frac{3}{4}\epsilon) \mathrm{d}q_0(t_0') \\ &= U_0^\rho(t_0) + \frac{3}{4}\epsilon \quad \text{for all } t_0 \in \hat{T}_0. \end{aligned}$$

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.

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

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,

$$r_0'(\hat{T}_0') = 1, \quad \text{where } \hat{T}_0' = B''\cup(\hat{T}_0\setminus(D'\cup D'')). \tag{11}$$

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 $y_i(\cdot)$ $(i=1,\dots,n)$ 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

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

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

First we show that

$$\sigma'\in\mathcal{I}(r_0'). \tag{12}$$

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

$$\begin{aligned}
U_i^{\sigma',r_0'}(\hat{t}_i,t_i) &= \int_{T_{-0-i}}\int_{\hat{T}_0} u_i(\sigma'(\hat{t}_i,t_{-i}),t_i)\mathrm{d}r_0'(t_0)\mathrm{d}p_{-0-i}(t_{-0-i})\\
&= \int_{T_{-0-i}}\int_{\hat{T}_0\setminus(D'\cup D'')} u_i(\sigma^*(\hat{t}_i,t_{-i}),t_i)\mathrm{d}r_0(t_0)\mathrm{d}p_{-0-i}(t_{-0-i})\\
&\quad + \int_{T_{-0-i}}\int_{B''} u_i(\sigma'(\hat{t}_i,t_{-i}),t_i)\mathrm{d}r_0'(t_0)\mathrm{d}p_{-0-i}(t_{-0-i}).
\end{aligned} \tag{13}$$

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 (13) yields

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

This implies (12) because $\sigma^*\in\mathcal{I}(r_0)$.

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

$$U_0^{\sigma'}(t_0) - U_0^{\rho}(t_0) > \frac{\epsilon}{2}. \tag{14}$$

First consider $t_0 \in \hat{T}_0 \setminus (D' \cup D'')$. Then $U_0^{\sigma'}(t_0) = U_0^{\sigma^*}(t_0)$, so (14) is immediate from $\sigma^* \in \mathcal{E}$ and from $\hat{T}_0' \subseteq \hat{T}_0$.

For all $t_0 \in B''$, (14) 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, (7) holds for all $t_0 \in T_0 \setminus B^*$. By continuity of $U_0^{\sigma}(\cdot)$, (7) extends to all $t_0 \in \text{supp}(r_0)$. In particular, the principal's participation constraint is satisfied for all types in $\text{supp}(r_0)$. By construction, the same holds for all types not in $\text{supp}(r_0)$. Hence, σ is r_0-feasible. This completes the proof.

An important corollary from the proof is the following.

Corollary 1. *If a p_0-feasible allocation is not strongly neologism-proof, then it is* strictly *r_0-dominated for some belief r_0.*

In particular, the set of strongly neologism-proof principal-payoff vectors is always closed.

3.2. **Existence.** In this section, we present an existence result for strongly neologism-proof allocations in environments with finite type spaces. We assume here that the space of collective actions A is a compact metric space and make the technical assumption of separability that was introduced in our earlier paper (Mylovanov and Tröger forthcoming). Observe that the outcome space itself is not compact because there is no bound on payments.

Proposition 2. *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 relies on two lemmas that allow us to reduce the problem to one with a finite bound on payments; then we can apply the general existence result in (Mylovanov and Tröger forthcoming). 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

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

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

Then there exists a number λ such that, 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 λ.

Proof. Let $\overline{v}$ denote an upper bound for the absolute value of the valuation of any action for any type of any player.

By (2), 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,

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

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

Turning to interim expected payoffs,

$$|U_i^{\rho,q_0}(t_i, t_i') - U_i^{\rho,q_0}(t_i, t_i)| \leq \max_{a \in A} |v_i(a, t_i') - v_i(a, t_i)| \leq 2\overline{v}. \tag{15}$$

Hence,

$$U_i^{\rho,q_0}(t_i) \leq U_i^{\rho,q_0}(t_i, t_i') + 2\overline{v} \overset{(1)}{\leq} U_i^{\rho,q_0}(t_i') + 2\overline{v}.$$

Thus,

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

Because any player's interim payment can differ from her interim payoff by at most $\overline{v}$, we can set $\lambda = (n+4)\overline{v}$. 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 $\underline{\mathcal{L}} = \mathbb{R}^{|T_0|+\dots+|T_n|}$ 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 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 $||\mathbf{x}(\cdot)|| \leq \kappa ||\underline{\mathbf{x}}(\cdot)||$.*

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 2.

Consider any sequence of payment bounds (λ_l) such that $\lambda_l \to \infty$. From Mylovanov and Tröger, (forthcoming), forthcoming), 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 1, ρ^* is strictly q_0-dominated by some allocation ρ', for some belief q_0.

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

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.

4. Application: Bargaining

There are two players $i = 0, 1$ who have to allocate one unit of divisible private good. The set of verifiable collective actions is

$$A = [0, 1] \cup \{\alpha_0\},$$

where $a \in [0, 1]$ indicates the amount of good allocated to the agent and $a = \alpha_0$ is the default allocation that results if the players fail to agree on an outcome, i.e., the disagreement outcome is $z_0 = (\alpha_0, 0, 0)$. The players payoff from the disagreement outcome are $u_1(z_0, t_1) = \gamma t_1$ and $u_0(z_0, t_0) \leq (1 - \gamma)t_0$ for some $\gamma \in [0, 1]$.

The disagreement outcome could have multiple interpretations. The value γ could represent the default property rights (e.g., grandfathered allocation of pollution rights in a model of international trade of pollution permits or shares in a company in a model of partnership dissolution), the expected outcome of a court, arbitration, or another dispute resolution procedure, the amount of good to be delivered to the agent that was agreed upon ex-ante, the agent's probability of acquiring the good outside of the relationship with the principal.

The type spaces are $T_0 = T_1 = [0, 1]$. The principal's (marginal) value distribution F_0 is assumed to be atomless. The agent's (marginal) value distribution is assumed to have strictly increasing and continuous buyer- and seller virtual valuation functions $\psi_b(t_1) = t_1 - \frac{1-F_1(t_1)}{f_1(t_1)}$ and $\psi_s(t_1) = t_1 + \frac{F_1(t_1)}{f_1(t_1)}$.

We characterize the allocations that maximize the ex-ante expected utility of player 0. The objective is

$$\int_0^1 U_0(t_0)\mathrm{d}F_0(t_0) \;=\; -U_1(0) + \int_0^1\int_0^1 (\psi_b(t_1) - t_0)x(t_0, t_1)\mathrm{d}F_1(t_1)\mathrm{d}F_0(t_0),$$

where ψ_b denotes the virtual valuation function of player 1, $x(t_0, t_1) \in [-a, 1 - a]$ denotes the (expected) quantity sold to player 1, and $U_1(0)$ denotes the expected utility of type 0 of player 1 (in expectation over player 0's type).

The problem is to maximize the objective subject to the constraints

$$\begin{aligned}
&U_1(0) + \int_0^{t_1}\int_0^1 (x(t_0, s) - \hat{u}_1'(s))\mathrm{d}F_0(t_0)\mathrm{d}s \geq 0 \quad \forall t_1 \in T_1 \quad (PC),\\
&x(t_0, t_1) \text{ is weakly increasing in } t_1 \quad \forall t_0 \in T_0 \quad (MC),\\
&x(t_0, 0) \geq 0, \quad x(t_0, 1) \leq 1 \quad \forall t_0 \in T_0. \quad (CC)
\end{aligned}$$

Observe that the agent's monotonicity constraint MC is required *not* in expectation over t_0, but separately for each t_0, which is justified by Manelli and Vincent (2010) and Gershkov et al (2012). There is an analogous monotonicity constraint for the principal which we relax, hoping that it is automatically satisfied. The last set of constraints (CC) expresses the capacity restriction. The variables over which we are maximizing are the number $U_1(0)$ and the x-function.

We can define the convex set

$$\Omega \;=\; \{(U_1(0), x(\cdot,\cdot)) \mid (MC),\ (CC)\}$$

Define a function G from Ω into the space $\mathcal{C}([0,1])$ of continuous real-valued functions on $[0,1]$ as follows,

$$G(U_1(0), x)\ (t_1) \;=\; U_1(0) + \int_0^{t_1}\int_0^1 (x(t_0,s) - \hat{u}_1'(s))\mathrm{d}F_0(t_0)\mathrm{d}s \quad (t_1 \in [0,1]).$$

Then we can express our maximization problem as

$$\begin{aligned} \max_{(U_1(0),x)\in\Omega} \quad & -U_1(0) + \int_0^1\int_0^1 (\psi_b(t_1) - t_0)x(t_0,t_1)\mathrm{d}F_1(t_1)\mathrm{d}F_0(t_0) \\ \text{s.t.} \quad & G(U_1(0), x) \geq 0. \end{aligned}$$

Now we apply Theorem 1, p. 217 and Theorem 2, p. 221 in Luenberger (1969) to get necessary and sufficient conditions for a solution.

Observe that the dual to $\mathcal{C}([0,1])$ is the space $NBV[0,1]$ of normalized functions of bounded variation on $[0,1]$ (see Luenberger, p. 115), and the positive cone in $NBV[0,1]$ that corresponds to the standard positive cone in $\mathcal{C}([0,1])$ is the set of weakly increasing right-continuous functions on $[0,1]$ (cf. Luenberger, p. 215).

From Luenberger, p. 217, if $(U_1(0), x) = (U_1(0)^*, x^*)$ solves the original optimization problem then there exists a weakly increasing right-continuous function z^* on $[0,1]$ such that $(U_1(0), x) = (U_1(0)^*, x^*)$ solves[7]

$$\begin{aligned} \max_{(U_1(0),x)\in\Omega} \quad & -U_1(0) + \int_0^1\int_0^1 (\psi_b(t_1) - t_0)x(t_0,t_1)\mathrm{d}F_1(t_1)\mathrm{d}F_0(t_0) \\ & + \int_0^1 (U_1(0) + \int_0^{t_1}\int_0^1 (x(t_0,s) - \hat{u}_1'(s))\mathrm{d}F_0(t_0)\mathrm{d}s)\mathrm{d}z^*(t_1) \end{aligned}$$

and

$$\int_0^1 (U_1(0)^* + \int_0^{t_1}\int_0^1 (x^*(t_0,s) - \hat{u}_1'(s))\mathrm{d}F_0(t_0)\mathrm{d}s)\mathrm{d}z^*(t_1) \;=\; 0. \quad (*)$$

(Here, z^* is the "Lagrange multiplier".)

Luenberger also states that the value reached at the maximum of the new maximization problem equals the value reached at the optimum of the original problem.

Towards solving the problem, observe that $z^*(1) = 1$ so that $U_1(0)$ cancels out in the above objective (otherwise the problem has no solution, a contradiction to Luenberger's Theorem).

[7]Our integration areas are always closed intervals.

The objective of the above problem can be rewritten (by changing the order of integration), so that we have to

$$\max_{(U_1(0),x)\in\Omega} \quad \int_0^1 \int_0^1 \left(t_1 - \frac{z^*(t_1) - F_1(t_1)}{f_1(t_1)} - t_0 \right) x(t_0, t_1) \mathrm{d}F_1(t_1) \mathrm{d}F_0(t_0). \quad (**)$$

(Here, $U_1(0)$ does not occur anymore; it is chosen such that at least one agent's participation constraint is binding.)

Now it seems to me we have to solve (**) for any z^*, and then determine z^* such that the corresponding solution x^* satisfies the condition (*). Then Luenberger, p. 221, shows that in fact we do have a solution to the original problem.

Solving (**) for a given z^* is not difficult. Fix any t_0 and

$$\begin{aligned} \max_{x(t_0,\cdot)} \quad & \int_0^1 h(t_0, t_1) x(t_0, t_1) \mathrm{d}F_1(t_1) \quad (***) \\ \text{s.t.} \quad & x(t_0, t_1) \text{ is weakly increasing in } t_1, \\ & x(t_0, 0) \geq 0, \quad x(t_0, 1) \leq 1, \end{aligned}$$

where we use the shortcut

$$h(t_0, t_1) \;=\; t_1 - \frac{z^*(t_1) - F_1(t_1)}{f_1(t_1)} - t_0.$$

Define

$$H(t_0, t_1) \;=\; \int_0^{t_1} f_1(s) h(t_0, s) \mathrm{d}s.$$

Using integration by parts,the objective of (***) can be written as

$$-\int_0^1 H(t_0, t_1) \mathrm{d}x(t_0, t_1) + H(t_0, 1) x(t_0, 1).$$

Holding $x(t_0, 1)$ fixed first, one sees that the integral is minimized if $\mathrm{d}x$ puts all weight where H is minimal;then the value of the objective is

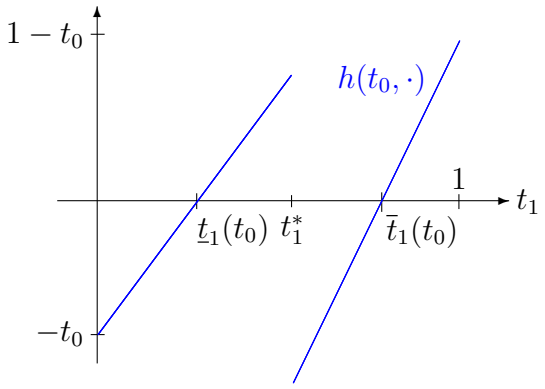
$$-\min_{t_1} H(t_0, t_1)(x(t_0, 1) - (-a)) + H(t_0, 1) x(t_0, 1). \tag{16}$$

We will solve problem (***) specifically for the cases in which z^* puts all weight on a single point $0 < t_1^* < 1$, that is,

$$z^*(t_1) = \mathbf{1}_{t_1 \geq t_1^*} \quad \text{for all } t_1.$$

Hence,

$$h(t_0, t_1) \;=\; \begin{cases} \psi_s(t_1) - t_0 & \text{if } t_1 < t_1^*, \\ \psi_b(t_1) - t_0 & \text{if } t_1 \geq t_1^*. \end{cases}$$



For later use, observe that

$$\begin{aligned} H(t_0,1) &= \int_0^1 \left(t_1 + \frac{F_1(t_1)}{f_1(t_1)}\right) f_1(t_1)\mathrm{d}t_1 - (1-t_1^*) - t_0 \\ &= \int_0^1 (t_1 F_1(t_1))'\mathrm{d}t_1 - (1-t_1^*) - t_0 \\ &= t_1^* - t_0 \end{aligned}$$

W.l.o.g. (for atomless F_0) we restrict, from now on, attention to types t_0 such that $0 < t_0 < 1$.

Observe that for all $t_1 \approx 1 > t_0$,

$$h(t_0,t_1) \geq t_1 - \frac{1-F_1(t_1)}{f_1(t_1)} - t_0 \approx t_1 - t_0 > 0.$$

Hence, $H(t_0,t_1)$ is strictly increasing in t_1 if $t_1 \approx 1$. Hence,

$$H(t_0,1) \;>\; \min_{t_1} H(t_0,t_1).$$

Hence, maximizing (16) implies

$$x(t_0,1) \;=\; 1-a. \tag{17}$$

Now, as said above, we have to find z^* (that is, t_1^*) such that the corresponding solution x^* satisfies the condition (*). Then Luenberger, p. 221, shows that in fact we do have a solution to the original problem.

For any $t_0 \in (0,1)$, let

$$\underline{t}_1(t_0) = \psi_s^{-1}(t_0) \in (0,t_0), \quad \bar{t}_1(t_0) = \psi_b^{-1}(t_0) \in (t_0,1).$$

Let t_0^* denote an a-quantile of F_0, that is, $F_0(t_0^*) = a$. By the intermediate value theorem, there exists

$$t_1^* \;\in\; (\underline{t}_1(t_0^*), \bar{t}_1(t_0^*)) \tag{18}$$

such that

$$\int_{\underline{t}_1(t_0^*)}^{t_1^*} \underbrace{(\psi_s(t_1) - t_0^*)}_{\geq 0}\,\mathrm{d}F_1(t_1) \;=\; -\int_{t_1^*}^{\bar{t}_1(t_0^*)} \underbrace{(\psi_b(t_1) - t_0^*)}_{\leq 0}\,\mathrm{d}F_1(t_1).$$

Observe that, for all t_0,[8]

$$h(t_0,t_1) \quad \begin{cases} <0 & \text{if } t_1<\underline{t}_1(t_0),\ t_1<t_1^*,\\ >0 & \text{if } \underline{t}_1(t_0)<t_1<t_1^*,\\ <0 & \text{if } t_1<\bar{t}_1(t_0),\ t_1>t_1^*,\\ >0 & \text{if } t_1>\bar{t}_1(t_0),\ t_1>t_1^*. \end{cases}$$

Consider first types t_0 such that $\underline{t}_1(t_0)<t_1^*<\bar{t}_1(t_0)$. Then $H(t_0,t_1)$ is decreasing in t_1 up to the point $\underline{t}_1(t_0)$, then increasing up to t_1^*, then decreasing up to $\bar{t}_1(t_0)$, then increasing up to the point 1. Thus, if t_0 is such that

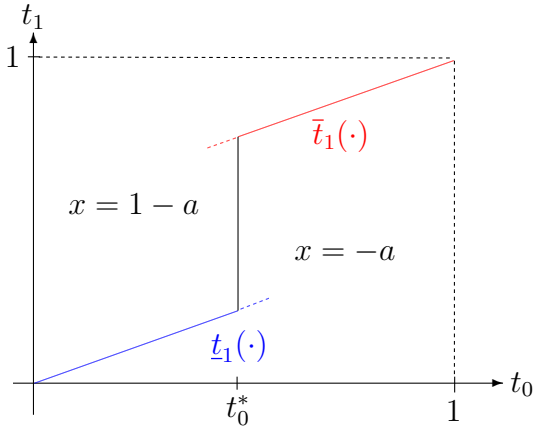
$$\int_{\underline{t}_1(t_0)}^{t_1^*} \underbrace{(\psi_s(t_1)-t_0)}_{\geq 0}\,\mathrm{d}F_1(t_1) \quad > \quad -\int_{t_1^*}^{\bar{t}_1(t_0)} \underbrace{(\psi_b(t_1)-t_0)}_{\leq 0}\,\mathrm{d}F_1(t_1),$$

then $H(t_0,t_1)$ is minimized at $t_1=\underline{t}_1(t_0)$, and is minimized at $t_1=\bar{t}_1(t_0)$ if the reverse inequality holds.

By construction, we have "=" if $t_0=t_0^*$ and ">" if $t_0<t_0^*$. Hence, the optimal x satisfies

$$x(t_0,t_1)=x^*(t_0,t_1) \quad = \quad \begin{cases} -a & \text{if } t_0<t_0^*,\ t_1<\underline{t}_1(t_0),\\ 1-a & \text{if } t_0<t_0^*,\ t_1>\underline{t}_1(t_0),\\ -a & \text{if } t_0>t_0^*,\ t_1<\bar{t}_1(t_0),\\ 1-a & \text{if } t_0>t_0^*,\ t_1>\bar{t}_1(t_0). \end{cases}$$

This formula extends types t_0 so large that (i) $\underline{t}_1(t_0)\geq t_1^*$ or (ii) $\bar{t}_1(t_0)\leq t_1^*$. For types (i), the function $H(t_0,t_1)$ is decreasing in t_1 up to the point $\bar{t}_1(t_0)$, then increasing; hence, it is minimized at $t_1=\bar{t}_1(t_0)$. For types (ii), the function $H(t_0,t_1)$ is decreasing in t_1 up to the point $\underline{t}_1(t_0)$, then increasing; hence, it is minimized at $t_1=\underline{t}_1(t_0)$.



Using (18), we find

$$\int_0^1 x^*(t_0,t_1^*)\mathrm{d}F_0(t_0) \quad = \quad (1-a)F_0(t_0^*)-a(1-F_0(t_0^*))=0.$$

Thus, from the envelope formula and because $t_1\mapsto\int x^*(t_0,t_1)\mathrm{d}F_0(t_0)$ is a weakly increasing function, the agent's expected payoff $U_1(t_1)$ is minimized at $t_1=t_1^*$, that is, $U_1(t_1^*)=0$ if $U_1(0)=U_1(0)^*$ is chosen optimally.

[8]It is sufficient to describe h up to a set of points of Lebesgue measure 0.

This implies (*). Hence, using Luenberger's theorem, x^* together with $U_1(0)^*$ solves the original maximization problem. Finally, note that $x^*(t_0, t_1)$ is decreasing in t_0, confirming incentive compatibility for the principal.

Next we turn to the principal's ex-ante expected utility

$$\eta(F_0) \;=\; \int (-\min_{t_1} H(t_0,t_1) + H(t_0,1)(1-a))\mathrm{d}F_0(t_0).$$

Consider first cases $t_0 < t_0^*$. Then

$$\begin{aligned}
-\min_{t_1} H(t_0,t_1) + H(t_0,1)(1-a) &= -H(t_0,\underline{t}_1(t_0)) + (t_1^* - t_0)(1-a)\\
&= \int_0^{\underline{t}_1(t_0)} (t_0 - \psi_s(t_1))\mathrm{d}F_1(t_1) + (t_1^* - t_0)(1-a)\\
&= t_0 F_1(\underline{t}_1(t_0)) - \int_0^{\underline{t}_1(t_0)} \psi_s(t_1)\mathrm{d}F_1(t_1)\\
&\quad +(t_1^* - t_0)(1-a).
\end{aligned}$$

Similarly, if $t_0 > t_0^*$, then

$$\begin{aligned}
-\min_{t_1} H(t_0,t_1) + H(t_0,1)(1-a) &= -H(t_0,\bar{t}_1(t_0)) + H(t_0,1)(1-a)\\
&= \int_{\bar{t}_1(t_0)}^1 (\psi_b(t_1) - t_0)\mathrm{d}F_1(t_1) - (t_1^* - t_0)a\\
&= -t_0(1 - F_1(\bar{t}_1(t_0))) + \int_{\bar{t}_1(t_0)}^1 \psi_b(t_1)\mathrm{d}F_1(t_1)\\
&\quad -(t_1^* - t_0)a\\
&= t_0 F_1(\bar{t}_1(t_0)) + \int_{\bar{t}_1(t_0)}^1 \psi_b(t_1)\mathrm{d}F_1(t_1)\\
&\quad -t_1^* a - (1-a)t_0.
\end{aligned}$$

Hence, using that $F_0(t_0^*) = a$,

$$\begin{aligned}
\eta(F_0) \;=\; &\int_0^{t_0^*} t_0 F_1(\underline{t}_1(t_0))\mathrm{d}F_0(t_0) - \int_0^{t_0^*}\int_0^{\underline{t}_1(t_0)} \psi_s(t_1)\mathrm{d}F_1(t_1)\mathrm{d}F_0(t_0)\\
&+ \int_{t_0^*}^1 t_0 F_1(\bar{t}_1(t_0))\mathrm{d}F_0(t_0) + \int_{t_0^*}^1\int_{\bar{t}_1(t_0)}^1 \psi_b(t_1)\mathrm{d}F_1(t_1)\mathrm{d}F_0(t_0)\\
&-(1-a)\int_0^1 t_0\mathrm{d}F_0(t_0). \qquad (19)
\end{aligned}$$

It is convenient to rearrange the above formula by first changing the order of integration, then using a transformation of variable, and finally using integration by parts. Specifically,

$$\begin{aligned}
\int_0^{t_0^*}\int_0^{\underline{t}_1(t_0)}\psi_s(t_1)\mathrm{d}F_1(t_1)\mathrm{d}F_0(t_0) &= \int_0^{\underline{t}_1(t_0^*)}\int_{\psi_s(t_1)}^{t_0^*}\mathrm{d}F_0(t_0)\psi_s(t_1)\mathrm{d}F_1(t_1)\\
&= \int_0^{\underline{t}_1(t_0^*)}(a-F_0(\psi_s(t_1)))\psi_s(t_1)f_1(t_1)\mathrm{d}t_1\\
&\overset{\psi_s(t_1)=t_0,\ t_1=\underline{t}_1(t_0)}{=} \int_0^{t_0^*}(a-F_0(t_0))t_0 f_1(\underline{t}_1(t_0))\mathrm{d}\underline{t}_1(t_0)\\
&= \int_0^{t_0^*}(a-F_0(t_0))t_0\mathrm{d}F_1(\underline{t}_1(t_0))\\
&= -\int_0^{t_0^*}(a-F_0(t_0)-f_0(t_0)t_0)F_1(\underline{t}_1(t_0))\mathrm{d}t_0\\
&= -a\int_0^{t_0^*}F_1(\underline{t}_1(t_0))\mathrm{d}t_0+\int_0^{t_0^*}t_0F_1(\underline{t}_1(t_0))\mathrm{d}F_0(t_0)\\
&\quad +\int_0^{t_0^*}F_1(\underline{t}_1(t_0))F_0(t_0)\mathrm{d}t_0. \qquad (20)
\end{aligned}$$

An analogous computation yields

$$\begin{aligned}
\int_{t_0^*}^1\int_{\bar{t}_1(t_0)}^1\psi_b(t_1)\mathrm{d}F_1(t_1)\mathrm{d}F_0(t_0) &= a\int_{t_0^*}^1 F_1(\bar{t}_1(t_0))\mathrm{d}t_0-\int_{t_0^*}^1 t_0F_1(\bar{t}_1(t_0))\mathrm{d}F_0(t_0)\\
&\quad -\int_{t_0^*}^1 F_1(\bar{t}_1(t_0))F_0(t_0)\mathrm{d}t_0+(1-a). \qquad (21)
\end{aligned}$$

Plugging (20) and (21) into (19) yields

$$\begin{aligned}
\eta(F_0) &= a\int_0^{t_0^*}F_1(\underline{t}_1(t_0))\mathrm{d}t_0-\int_0^{t_0^*}F_1(\underline{t}_1(t_0))F_0(t_0)\mathrm{d}t_0\\
&\quad +a\int_{t_0^*}^1 F_1(\bar{t}_1(t_0))\mathrm{d}t_0-\int_{t_0^*}^1 F_1(\bar{t}_1(t_0))F_0(t_0)\mathrm{d}t_0+(1-a)\\
&\quad -(1-a)\int_0^1 t_0\mathrm{d}F_0(t_0). \qquad (22)
\end{aligned}$$

There is an alternative way to express the principal's ex-ante expected utility,

$$\eta(F_0) = \int_0^1 U_0(t_0)\mathrm{d}F_0(t_0),$$

where $U_0(t_0)$ denotes the expected utility of a given type t_0.

Using the envelope formula, for all $t_0 < t_0^*$,

$$\begin{aligned} U_0(t_0) &= U_0(t_0^*) - \int_{t_0}^{t_0^*} \int_0^1 (-x^*(s,t_1))\mathrm{d}F_1(t_1)\mathrm{d}s \\ &= U_0(t_0^*) - \int_{t_0}^{t_0^*} (F_1(\underline{t}_1(s)) - (1-a))\mathrm{d}s \\ &= U_0(t_0^*) - \int_{t_0}^{t_0^*} F_1(\underline{t}_1(s))\mathrm{d}s + (t_0^* - t_0)(1-a). \end{aligned} \tag{23}$$

Similarly, for all $t_0 > t_0^*$,

$$\begin{aligned} U_0(t_0) &= U_0(t_0^*) + \int_{t_0^*}^{t_0} \int_0^1 (-x^*(s,t_1))\mathrm{d}F_1(t_1)\mathrm{d}s \\ &= U_0(t_0^*) + \int_{t_0^*}^{t_0} (F_1(\bar{t}_1(s)) - (1-a))\mathrm{d}s \\ &= U_0(t_0^*) + \int_{t_0^*}^{t_0} F_1(\bar{t}_1(s))\mathrm{d}s + (t_0^* - t_0)(1-a). \end{aligned} \tag{24}$$

Hence,

$$\begin{aligned} \eta(F_0) &= U_0(t_0^*) - \int_0^{t_0^*} \int_{t_0}^{t_0^*} F_1(\underline{t}_1(s))\mathrm{d}s\mathrm{d}F_0(t_0) \\ &\quad + \int_{t_0^*}^1 \int_{t_0^*}^{t_0} F_1(\bar{t}_1(s))\mathrm{d}s\mathrm{d}F_0(t_0) - (1-a)\int_0^1 t_0\mathrm{d}F_0(t_0) + t_0^*(1-a). \end{aligned}$$

Using integration by parts,

$$\begin{aligned} \eta(F_0) &= U_0(t_0^*) - \int_0^{t_0^*} F_1(\underline{t}_1(t_0))F_0(t_0)\mathrm{d}t_0 \\ &\quad - \int_{t_0^*}^1 F_1(\bar{t}_1(t_0))F_0(t_0)\mathrm{d}t_0 + \int_{t_0^*}^1 F_1(\bar{t}_1(s))\mathrm{d}s \\ &\quad -(1-a)\int_0^1 t_0\mathrm{d}F_0(t_0) + t_0^*(1-a). \end{aligned} \tag{25}$$

Comparing (25) to (22) yields that

$$U_0(t_0^*) = a\int_0^{t_0^*} F_1(\underline{t}_1(t_0))\mathrm{d}t_0 - (1-a)\int_{t_0^*}^1 F_1(\bar{t}_1(t_0))\mathrm{d}t_0 + (1-a) - t_0^*(1-a).$$

Plugging this into (23), we find

$$\begin{aligned} U_0^*(t_0) &= a\int_0^{t_0} F_1(\underline{t}_1(s))\mathrm{d}s - (1-a)\int_{t_0}^{t_0^*} F_1(\underline{t}_1(s))\mathrm{d}s - (1-a)\int_{t_0^*}^1 F_1(\bar{t}_1(s))\mathrm{d}s \\ &\quad + (1-t_0)(1-a) \quad \text{for all } t_0 < t_0^*. \end{aligned} \tag{26}$$

Similarly, using (24),

$$\begin{aligned} U_0^*(t_0) &= a\int_0^{t_0^*} F_1(\underline{t}_1(s))\mathrm{d}s + a\int_{t_0^*}^{t_0} F_1(\bar{t}_1(s))\mathrm{d}s - (1-a)\int_{t_0}^1 F_1(\bar{t}_1(s))\mathrm{d}s \\ &\quad + (1-t_0)(1-a) \quad \text{for all } t_0 > t_0^*. \end{aligned} \tag{27}$$

Now we show that $U_0^*(\cdot)$ is neologism-proof. Suppose otherwise. Then there exists a belief G_0 and a G_0-feasible payoff vector $U_0(\cdot)$ such that $U_0(\cdot)$ dominates $U_0^*(\cdot)$ (with positive probability under G_0). Thus,

$$\int_0^1 U_0^*(t_0)\mathrm{d}G_0(t_0) < \int_0^1 U_0(t_0)\mathrm{d}G_0(t_0) \leq \eta(G_0).$$

We obtain a contradiction, thus completing the proof of neologism-proofness, by showing that

$$\eta(G_0) \leq \int_0^1 U_0^*(t_0)\mathrm{d}G_0(t_0) \quad \text{for all } G_0. \tag{28}$$

Using (26) and (27), the claim (28) can be written as

$$\begin{aligned} \eta(G_0) \leq\ & a\int_0^{t_0^*}\int_0^{t_0} F_1(\underline{t}_1(s))\mathrm{d}s\mathrm{d}G_0(t_0) - (1-a)\int_0^{t_0^*}\int_{t_0}^{t_0^*} F_1(\underline{t}_1(s))\mathrm{d}s\mathrm{d}G_0(t_0) \\ & -(1-a)\int_0^{t_0^*}\int_{t_0^*}^1 F_1(\bar{t}_1(s))\mathrm{d}s\mathrm{d}G_0(t_0) + a\int_{t_0^*}^1\int_0^{t_0^*} F_1(\underline{t}_1(s))\mathrm{d}s\mathrm{d}G_0(t_0) \\ & +a\int_{t_0^*}^1\int_{t_0^*}^{t_0} F_1(\bar{t}_1(s))\mathrm{d}s\mathrm{d}G_0(t_0) - (1-a)\int_{t_0^*}^1\int_{t_0}^1 F_1(\bar{t}_1(s))\mathrm{d}s\mathrm{d}G_0(t_0) \\ & +(1-a) - (1-a)\int_0^1 t_0\mathrm{d}G_0(t_0). \end{aligned}$$

Using integration by parts, this can be rewritten as

$$\begin{aligned} \eta(G_0) \leq\ & -\int_0^{t_0^*} F_1(\underline{t}_1(t_0))G_0(t_0)\mathrm{d}t_0 - \int_{t_0^*}^1 F_1(\bar{t}_1(t_0))G_0(t_0)\mathrm{d}t_0 \\ & +a\int_0^{t_0^*} F_1(\underline{t}_1(t_0))\mathrm{d}t_0 + a\int_{t_0^*}^1 F_1(\bar{t}_1(t_0))\mathrm{d}t_0 + (1-a) - (1-a)\int_0^1 t_0\mathrm{d}G_0(t_0). \end{aligned}$$

Combining this with (22), and using the notation t_0^{**} for an a-quantile of G_0, shows that the claim (28) can be written as

$$\begin{aligned} & -\int_0^{t_0^{**}} F_1(\underline{t}_1(s))G_0(s)\mathrm{d}s - \int_{t_0^{**}}^1 F_1(\bar{t}_1(s))G_0(s)\mathrm{d}s \\ & +a\int_0^{t_0^{**}} F_1(\underline{t}_1(s))\mathrm{d}s + a\int_{t_0^{**}}^1 F_1(\bar{t}_1(s))\mathrm{d}s \\ \leq\ & -\int_0^{t_0^{*}} F_1(\underline{t}_1(s))G_0(s)\mathrm{d}s - \int_{t_0^{*}}^1 F_1(\bar{t}_1(s))G_0(s)\mathrm{d}s \\ & +a\int_0^{t_0^{*}} F_1(\underline{t}_1(s))\mathrm{d}s + a\int_{t_0^{*}}^1 F_1(\bar{t}_1(s))\mathrm{d}s. \end{aligned}$$

To verify this, consider the case $t_0^* \geq t_0^{**}$ ("$\leq$" is analogous). Then (28) holds because

$$0 \ \leq \ \int_{t_0^{**}}^{t_0^*} (\underbrace{F_1(\bar{t}_1(s)) - F_1(\underline{t}_1(s))}_{>0})(\underbrace{G_0(s) - a}_{\geq 0})\mathrm{d}s.$$

Observe that all the above computations assume that F_0 has no atom. This is all we have to consider. Starting with any prior F_0, we only need to consider G_0 that are absolutely continuous relative to F_0. Assuming that the prior F_0 has a density, we only need to consider G_0 that have a density as well.

References

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

Cho, I.-K., and D. M. Kreps (1987): "Signaling Games and Stable Equilibria," *The Quarterly Journal of Economics*, 102(2), 179–221.

Farrell, J. (1993): "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior*, 5, 514–531.

Gershkov, A., J. K. Goeree, A. Kushnir, B. Moldovanu, and X. Shi (2012): "On the equivalence of Bayesian and dominant strategy implementation," *mimeo*.

Grossman, S. J., and M. Perry (1986): "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39, 97–119.

Krishna, V. (2002): *Auction Theory*. Academic Press.

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

Mailath, G. J., M. Okuno-Fujiwara, and A. Postlewaite (1993): "Belief-Based Refinements in Signalling Games," *Journal of Economic Theory*, 60(2), 241–276.

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

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.

Milgrom, P. (2004): *Putting Auction Theory to Work*. Cambridge University Press.

Myerson, R. B. (1983): "Mechanism design by an informed principal," *Econometrica*, 51(6), 1767–1798.

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 (2008): "Optimal Auction Design and Irrelevance of Privacy of Information," *mimeo*.

——— (forthcoming): "Informed-principal problems in environments with generalized private values," *Theoretical Economics*.

Skreta, V. (2009): "On the informed seller problem: optimal information disclosure," *Review of Economic Design*, 15(1), 1–36.

Tan, G. (1996): "Optimal Procurement Mechanisms for an Informed Buyer," *Canadian Journal of Economics*, 29(3), 699–716.

Yilankaya, O. (1999): "A note on the seller's optimal mechanism in bilateral trade with two-sided incomplete information," *Journal of Economic Theory*, 87(1), 125–143.

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