Discussion Paper Series – CRC TR 224

Discussion Paper No. 727
Project A 03

# Neutral Optimum in Private-Values Settings

Tymofiy Mylovanov[1]
Thomas Tröger[2]

February 2026

[1]Department of Economics, University of Pittsburgh. Email: mylovanov@gmail.com.
[2]Department of Economics, University of Mannheim. Email: troeger@unimannheim.de.

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

# Neutral optimum in private-values settings[*]

Tymofiy Mylovanov[†]and Thomas Tröger[‡]

January 23, 2026

**Abstract**

We show that in informed-principal settings with generalized private values any neutral optimum (Myerson, 1983) is strongly neologism-proof (Mylovanov and Tröger, 2012) and hence is a strong unconstrained Pareto optimum in the setting of Maskin and Tirole (1990). Thus, in any setting with a unique strongly neologism-proof solution this concept is equivalent to neutral optimum. We rely on the unifying concept of neo-optimum that we develop in the companion paper Mylovanov and Tröger (2026). The main step is to prove that any neo-optimum is strongly neologism-proof.

## 1 Introduction

Myerson (1983) introduced neutral optimum as a solution to the problem of mechanism-design by an informed principal. Myerson proves the existence of a neutral optimum in arbitrary informed-principal settings, subject to technical assumptions. Neutral optimum is a refinement of perfect-Bayesian

[*] Support by the German Research Foundation (DFG) through CRC TR 224 (Project B01) is gratefully acknowledged. Support by the National Science Foundation through grant 0922365, as well as the hospitality of the Study Center in Gerzensee, Switzerland, are also gratefully acknowledged.

This paper has a companion paper Mylovanov and Tröger (2026), an earlier version of which contained the results of both papers.

[†] Department of Economics, University of Pittsburgh. Email: mylovanov@gmail.com.

[‡] Department of Economics, University of Mannheim. Email: troeger@uni-mannheim.de.

equilibrium in the signaling game where the principal is a sender who proposes a mechanism in which the principal as well as agents submit messages. The refinement is motivated by the standard mechanism-design doctrine that the principal can set a focal point for the agents' behavior. Two properties are central. First, relative to a neutral optimum, there exists no other direct revelation mechanism that would be strictly preferred by the principal independently of her private information. This stands in contrast to all signaling refinements that are consistent with strategic stability, such as the intuitive criterion. Secondly, neutral optimum balances the conflicts of interest between different private-information types of the principal.

Unfortunately, neutral optima are notoriously different to compute. Thus, with few exceptions (e.g., Severinov (2008)), the subsequent literature has ignored Myerson's elegant solution und has instead turned attention to classical signaling refinements or to other solution concepts, each of which was invented for a specific class of informed-principal settings.

Initiated by Maskin and Tirole (1990), a part of the informed-principal literature has considered settings with "private values" (e.g., Myerson (1985); Maskin and Tirole (1990); Tan (1996); Yilankaya (1999); Skreta (2009); Mylovanov and Tröger (2014); Wagner, Mylovanov, and Tröger (2015)). Here, the principal is privately informed about her goals, that is, she "has private information that is not directly payoff relevant to the agents, but may influence her design" (Mylovanov and Tröger, 2012). An example would be a seller with private information about her opportunity cost of selling who designs a profit-maximizing sales procedure. Private information is assumed to be stochastically independent across players. The solution concepts proposed in this context, strong unconstrained Pareto optimum (SUPO) by Maskin and Tirole (1990) and its generalization, strongly neologism-proof allocations by Mylovanov and Tröger (2012, 2014) are more intuitive than neutral optimum. They allow an economic interpretation where different private-information types of the principal resolve their conflicts of interest like traders in a competitive market. However, SUPO and strong neologism-proofness have so far remained disconnected from Myerson's (1983) approach.

We show that, in the generalized-private-values settings for which Mylovanov and Tröger (2012) show the existence of a strongly neologism-proof allocation, any neutral optimum is strongly neologism-proof. This result resolves a question that has remained open essentially since the publication of Maskin and Tirole (1990). An immediate implication is that in any setting with a unique strongly neologism-proof solution, this concept is equivalent to neutral optimum. Another important implication is that in quasilinear private-values environments (Mylovanov and Tröger, 2014), any neutral optimum is ex-ante optimal.

Towards establishing the connection between neutral optimum and strong neologism-proofness, there are three steps. First, a neutral optimum is a neo-optimum in any (private or non-private values) setting. This result is taken from our companion paper Mylovanov and Tröger (2026), where we introduce the concept of a neo-optimum. Neo-optimum inherits many of the properties of neutral optimum, but is easier to handle. Second, in private-values settings neo-optimum is equivalent to the novel concept of interim optimum by Koessler and Skreta (2023) if we translate their concept into the private-values context. Third, in private-values settings that satisfy the separability condition from Mylovanov and Tröger (2012), interim optimum is equivalent to strong neologism-proofness. In summary, we have the following picture in private-values settings:

$$\text{Neutral optimum} \quad \overset{Lemma1}{\Longrightarrow} \quad \text{Neo-optimum} \quad \underset{Lemma3}{\overset{Lemma2}{\Longleftrightarrow}} \quad \text{Interim-optimum}$$

$$\underset{Lemma5}{\overset{Lemma4}{\Longleftrightarrow}} \quad \text{Strongly neologism-proof allocation.}$$

Section 2 reviews private-values settings and introduces the core concepts. In Section 3 we prove our results. Section 4 summarizes the main conclusions.

# 2 Settings with private-values and solution concepts

As in Mylovanov and Tröger (2012), we consider the interaction of a principal (player 0) and $n \geq 1$ agents $i = 1, \ldots, n$. The players must collectively choose an outcome from a compact metric space of *basic outcomes* $Z$. Every player $i$ has a *type* $t_i$ that belongs to a finite type space $T_i$. The principal's type $t_0$ will often be denoted $t$ for short. The principal's type space $T_0$ will often be denoted $T$ for short. A *type profile* is any $\mathbf{t} \in \mathbf{T} = T \times \cdots \times T_n$. Sometimes we use the notation $\mathbf{T} = T_i \times \mathbf{T}_{-i}$, $\mathbf{t} = (t_i, \mathbf{t}_{-i})$, or $\mathbf{t} = (t_0, t_i, \mathbf{t}_{-0,i})$. Player $i$'s payoff function,

$$u_i : \ Z \times \mathbf{T} \to \mathrm{I\!R},$$

is assumed to be continuous (note that the continuity assumption is void if $Z$ is finite).

An *outcome* is a probability measure over basic outcomes; let $\mathcal{Z}$ denote the set of outcomes. We identify any $z \in Z$ with the point distribution that puts probability 1 on the point $z$; hence, $Z \subseteq \mathcal{Z}$. We extend the definition of $u_i$ to $\mathcal{Z} \times \mathbf{T}$ via the statistical expectation.

Some outcome $z_0 \in \mathcal{Z}$ is designated as the *disagreement outcome*. Every player's payoff from the disagreement outcome is normalized to 0 (for every profile of other players' types).

The interaction is described by a *mechanism-selection game*. First, for each player $i$ nature chooses a type. Let $p_i(t_i) > 0$ denote the probability of type $t_i \in T_i$. We assume that types are stochastically independent. Each player privately observes her type $t_i$. Second, the principal offers a *mechanism* $M$, which is a finite perfect-recall game form with players $N \cup \{0\}$ and with outcomes in $\mathcal{Z}$. Third, the players[1] decide simultaneously whether or not to accept $M$. If $M$ is accepted unanimously, then each player chooses a plan of actions in $M$, and the outcome specified by $M$ is implemented. If at least one player rejects $M$, then the disagreement outcome $z_0$ is implemented. In Mylovanov and Tröger (2012), we define perfect-Bayesian equilibrium for this game.

Consider now the continuation game that begins after the principal has proposed some arbitrary mechanism $M$. Based on observing $M$, the agents may change their belief about the principal's type, so that the belief at the beginning of the continuation game may be different from $p_0$.

An *allocation* is a function

$$\rho : \quad \mathbf{T} \rightarrow \mathcal{Z}$$

that assigns an outcome $\rho(\mathbf{t})$ to every type profile $\mathbf{t} \in \mathbf{T}$. Thus, an allocation may describe the outcome of any continuation game as a function of the type profile. Here we can also view the entire mechanism-selection game as a continuation game. Hence, an allocation can also describe the outcome of the entire mechanism-selection game.

Let $\rho$ denote the allocation induced by equilibrium play in a continuation game that begins with any belief $b \in B$ about the principal's type, where $B$ denotes the set of probability distributions on $T$.

[1] We are assuming the principal can reject her own mechanism. This will not play any significant role, but it corrects an imprecision in the game description in Mylovanov and Tröger (2012).

The expected payoff of type $t_i$ of player $i$ if she imitates type $\hat{t}_i$ in the continuation game is

$$U_i^{\rho,b}(\hat{t}_i, t_i) \quad = \quad \sum_{\mathbf{t}_{-i}\in\mathbf{T}_{-i}} u_i(\rho(\hat{t}_i, \mathbf{t}_{-i}), (t_i, \mathbf{t}_{-i}))\ \mathbf{q}_{-i}(\mathbf{t}_{-i}),$$

where $\mathbf{q}_{-i}(\mathbf{t}_{-i}) = b(t_0) \cdot p_1(t_1) \cdots p_{i-1}(t_{i-1}) \cdot p_{i+1}(t_{i+1}) \cdots p_n(t_n)$ if $i \neq 0$, and $\mathbf{q}_{-0}(\mathbf{t}_{-0}) = p_1(t_1) \cdot \cdots \cdot p_n(t_n)$.

The expected payoff of type $t_i$ of player $i$ from allocation $\rho$ is

$$U_i^{\rho,b}(t_i) \quad = \quad U_i^{\rho,b}(t_i, t_i).$$

**Definition 1.** *Given any belief $b \in B$, an allocation $\rho$ is called* $b$-feasible *if no type of any player has an incentive to reject $\rho$ or to imitate a different type: for all $i$,*

$$U_i^{\rho,b}(t_i) \quad \geq \quad U_i^{\rho,b}(\hat{t}_i, t_i) \text{ for all } t_i, \hat{t}_i, \tag{1}$$

$$U_i^{\rho,b}(t_i) \quad \geq \quad 0 \text{ for all } t_i. \tag{2}$$

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

Our crucial assumption for the rest of the paper is that the agents' payoff functions are independent of the principal's type.

**Definition 2.** *An environment features* generalized private values *if, for all agents $i = 1, \ldots, n$,*

$$u_i(z, (t, \mathbf{t}_{-0})) = u_i(z, (t', \mathbf{t}_{-0})) \overset{def}{=} u_i(z, \mathbf{t}_{-0}) \quad \text{for all } \ z,\ t, t',\ \mathbf{t}_{-0}.$$

This assumption implies a very particular structure for the sets of feasible allocations. To describe it we need additional notation.

A *sub-allocation* is a function

$$\alpha : \quad \mathbf{T}_{-0} \to \mathcal{Z}.$$

The set of sub-allocations is a convex subset $Q$ of the linear space of maps from $\mathbf{T}_{-0}$ into the space of signed Borel measures on $Z$.

Any allocation $\rho$ corresponds to a family $(\rho_t)_{t\in T}$ of sub-allocations $\rho_t$, via the equation $\rho(t, t_{-0}) = \rho_t(t_{-0})$. In such a situation, we call $\rho_t$ the *allocation of type $t$*.

We can use any family $(\rho_t)_{t\in T}$ together with any belief $b \in B$ to define an averaged sub-allocation via the convex combination

$$\rho_b = \sum_{t\in T} b(t)\rho_t.$$

To make this more explicit, note that, for any $\mathbf{t_{-0}}$, the outcome $\rho_b(\mathbf{t_{-0}}) \in \mathcal{Z}$ assigns to any Borel set $S \subseteq Z$ the probability

$$\rho_b(\mathbf{t}_{-0})(S) = \sum_{t\in T} b(t)\rho_t(\mathbf{t}_{-0})(S).$$

A crucial implication of the generalized-private-values assumption is that agents care only about the averaged sub-allocation, and are otherwise indifferent concerning which type of the principal implements what. Thus, the agents' incentive and participation constraints can be written in terms of the averaged sub-allocation. Formally, given any sub-allocation $\alpha$, agent $i$, and types $t_i$ and $\hat{t}_i$, define

$$U_i^\alpha(\hat{t}_i, t_i) \;\; = \sum_{\mathbf{t}_{-i}\in\mathbf{T}_{-i}} u_i(\alpha(\hat{t}_i, \mathbf{t}_{-i-0}), (t_i, \mathbf{t}_{-i-0}))\; \mathbf{q}_{-i-0}(\mathbf{t}_{-i-0}).$$

Let $A$ denote the set of sub-allocations that satisfy the agents' incentive and participation constraints, that is, $\alpha \in A$ if and only if, for all $i \geq 1$,

$$\begin{aligned} U_i^\alpha(t_i) &\geq U_i^\alpha(\hat{t}_i, t_i) \text{ for all } t_i,\ \hat{t}_i \neq t_i, \\ U_i^\alpha(t_i) &\geq 0 \text{ for all } t_i. \end{aligned}$$

Using this definition, the condition that the constraints (1) and (2) are satisfied for all agents $i = 1, \ldots, n$ can be expressed as the statement

$$\rho_b \in A.$$

We call this condition "Agents' Feasibility" (AF).

We will also reformulate the incentive and participation constraints for the principal. Given any sub-allocation $\alpha$, the expected payoff of any type $t \in T$ of the principal is denoted

$$\Pi(\alpha)(t) \;\; = \sum_{\mathbf{t}_{-i}\in\mathbf{T}_{-i}} u_0(\alpha(\mathbf{t}_{-0}), (t, \mathbf{t}_{-0}))\; \mathbf{q}_{-0}(\mathbf{t}_{-0}).$$

Note that $\Pi$ is a linear mapping from $Q$ into $\mathrm{I\!R}^T$. For any allocation $\rho$ and types $\hat{t}$ and $t$,

$$U_0^\rho(\hat{t}, t) \ = \ \Pi(\rho_{\hat{t}})(t).$$

Thus, the constraints (1) for $i = 0$ can alternatively be expressed as $\Pi(\rho_t)(t) \geq \Pi(\rho_{t'})(t)$ for all $t, t' \in T$. We call these conditions "Principal's Incentive Compatibility" (PIC).

Thus, the constraints (2) for $i = 0$ can alternatively be expressed as $\Pi(\rho_t)(t) \geq 0$ for all $t \in T$. We call these conditions "Principal's Individual Rationality" (PIR).

In summary, given any belief $b \in B$, an allocation $\rho$ (or a family $(\rho_t)_{t \in T}$ in $Q$) is $b$-feasible if and only if AF holds for $(\rho_t)_{t \in T}$ together with the belief $b$, and PIC and PIR hold. We then also say that the payoff vector $U \in \mathrm{I\!R}^T$, $U(t) = \Pi(\rho_t)(t)$, is $b$-feasible.

### Strong neologism-proofness

Given any allocations $\rho$ and $\rho'$, the set of principal-types that are strictly better off in $\rho$ is denoted

$$S(\rho, \rho') \ = \ \{t \in T \mid U_0^\rho(t) > U_0^{\rho'}(t)\}.$$

The set of types who in $\rho$ obtain the highest feasible payoff is denoted

$$H(\rho) \ = \ \left\{ t \in T \mid U_0^\rho(t) = \sum_{\mathbf{t}_{-0} \in \mathbf{T}_{-0}} \max_{z \in Z} u_0(z, t, \mathbf{t}_{-0}) \mathbf{q}(\mathbf{t}_{-0}) \right\}.$$

Given any allocations $\rho$ and $\rho'$, Mylovanov and Tröger (2012) say that a belief $b \in B$ is *credible* for $\rho'$ relative to $\rho$ if $b$ puts zero probability mass on principal-types who are strictly better off in $\rho$ or who already enjoy the highest feasible payoff in $\rho$ that is,

$$(S(\rho, \rho') \cup H(\rho)) \cap \mathrm{supp}(b) = \emptyset,$$

where $\mathrm{supp}(b)$ denotes the support of $b$.

**Definition 3.** *An allocation $\rho$ is called* strongly neologism-proof *if $\rho$ is $p_0$-feasible and $S(\rho', \rho) \cap supp(b') = \emptyset$ for any belief $b'$ together with a $b'$-feasible allocation $\rho'$ such that $b'$ is credible for $\rho'$ relative to $\rho$.*

Mylovanov and Tröger (2012) prove existence and show that the concept generalizes strong unconstrained Pareto optimum (SUPO) of Maskin and Tirole (1990). While being technically convenient, a problem with strong neologism-proofness is that the underlying concept of credibility of beliefs is somewhat arbitrary and may be seen as too permissive. The concept of a neologism that was put forward by Farrell (1993) captures a more natural idea of credibility.

### Neo-optimum

Consider a vector $V \in \mathbb{R}^T$. We will interpret $V$ as a (fictitious, possibly non-feasible) payoff vector for the types of the principal. A belief $b$ together with a $b$-feasible allocation $\rho'$ is a *neologism* for $V$ (at the belief $p_0$) if $U_0^{\rho'}(\check{t}) > V(\check{t})$ for some $\check{t} \in T$, and the following conditions hold for all $t \in T$:

$$\text{if } U_0^{\rho'}(t) > V(t) \text{ then } b(t)p_0(t') \geq b(t')p_0(t) \text{ for all } t' \in T, \tag{3}$$

$$\text{if } U_0^{\rho'}(t) < V(t) \text{ then } b(t) = 0. \tag{4}$$

A neologism that can be seen as a plausible deviation relative to a given (not necessarily feasible) principal-payoff-vector $V$. The deviation must be feasible at a belief $b$ that is Bayes-consistent with the prior belief $p_0$, given that all principal types choose optimally whether or not to deviate, and the deviation must be profitable for at least one principal type $\check{t}$. Note that $b$ retains the relative likelihood across types that strictly gain (use (3) with switched roles of $t$ and $t'$, yielding $b(t)p_0(t') = b(t')p_0(t)$), and can shift belief probability mass from indifferent types to strictly gaining types. The concept is inspired by Farrell (1993). Note, however, that our definition, in contrast to Farrell's, refers to an arbitrary vector $V$ that may not be feasible.

**Definition 4.** *A vector $V \in \mathbb{R}^T$ is* neologism-proof *if no neologism exists for $V$ at $p_0$.*

Consistent with the naming, the definition is less restrictive than the concept of strong neologism-proofness, as long as no principal type obtains the highest feasible payoff. In our companion paper, we propose a concept which in general non-private-value settings is even weaker than neologism-proofness: we extend consideration to all allocations that yield payoff vectors that are *above limits of* neologism-proof payoff vectors.

**Definition 5.** *A $p_0$-feasible allocation $\rho$ is a* neo-optimum *if there exists a sequence $(V^n)_{n=1,2,\ldots}$ of neologism-proof payoff vectors such that*

$$U_0^\rho(t) \geq \lim_n V^n(t) \quad \text{for all } t \in T.$$

As we demonstrate in our companion paper Mylovanov and Tröger (2026), this concept is of central importance to informed-principal problems because it unifies a variety of other concepts that have been proposed in the literature, and it exists quite generally, including in settings with non-private-values. In particular, in our companion paper we show the following.

**Lemma 1.** *Consider any Bayesian incentive problem. Any neutral optimum is a neo-optimum with $p_0$ being the uniform distribution.*

The uniform-distribution assumption concerning $p_0$ is not essential. We make it in order to adapt the framework to Myerson (1983), where the setup is such that the prior is always uniform.

An immediate conclusion from Lemma 1 and Myerson's existence result for neutral optima is that a neo-optimum exists in any Bayesian incentive problem.

### Interim-optimum

As an intermediate step towards proving our main result, the equivalence of strong neologism-poofness and neo-optimum, we employ yet another solution concept that "fits in between" the other two concepts (cf. the diagram in the introduction). The concept was invented by Koessler and Skreta (2023) in a non-private-values context of information design.

**Definition 6.** *An allocation $\rho$ is* interim-optimal *if (i) $\rho$ is $p_0$-feasible and (ii) there does not exist a belief $b'$ together with a $b'$-feasible allocation $\rho'$ such that supp$(b') \subseteq S(\rho', \rho)$.*

## 3 Results

Interim-optimality is easily seen to be at least as strong as neo-optimum (and the result has nothing to do with private values). For clarity we repeat this result from our companion paper Mylovanov and Tröger (2026).

**Lemma 2.** *Any interim optimal allocation is a neo-optimum.*

*Proof.* Consider any interim optimal allocation $\rho$. Then, for all $\epsilon > 0$, no neologism exists for $U_0^\rho + \epsilon$ at the belief $p_0$. Thus, $U_0^\rho$ is a limit of neologism-proof payoff vectors, showing that $\rho$ is a neo-optimum. $\square$

The following result, the reverse of Lemma 2, is by far the harded piece of work.

**Lemma 3.** *Any neo-optimum is interim optimal.*

To show this (for details see the Appendix), we start with a $p_0$-feasible allocation $\rho$ that is not interim optimal and show that it is not a neo-optimum. Let $U = U_0^\rho$ denote the corresponding payoff vector.

By assumption, there exists a belief $b'' \in B$ and a $b''$-feasible allocation $\rho''$ such that in the "deviation payoff vector" $U'' = U_0^{\rho''}$ all types in the support of $b''$ are strictly better off than in $U$. In general, $\rho''$ is not a neologism for $U$ because the relative probabilities of different types in the support of $b''$ are unrestricted, and types outside the support may also be better off in $U''$ than in $U$. The idea behind our proof is to apply a sequence a "surgeries" in which we apply further changes to $b''$ and $\rho''$ such that eventually a neologism for $U$ is obtained.

Because in $U''$ all types in the support of $b''$ are also strictly better off than in the payoff vectors in a neighborhood of $U$ and below, the surgery constructions extend to the existence of neologisms for all such payoff vectors, implying that $U$ is not a neo-optimum.

Starting with $b''$ and the allocation $\rho''$, the basic idea behind our surgeries is that we build a new belief $b'$ together with a new allocation$\rho'$ such that

$$\sum_{t\in T} b'(t)\rho'_t = \sum_{t\in T} b''(t)\rho''_t,$$

where each sub-allocation $\rho'_t$ will be a convex combination of various types' allocations in $\rho''$. By construction, the new sub-allocations belong to $Q$, and AF remains true for $\rho'$ together with $b'$.

If some type $t$'s new allocation $\rho'_t$ arises from a convex combination involving some type $\check{t}$'s old allocation $\rho''_{\check{t}}$, then we say that type $t$ obtains a chunk of type $\check{t}$'s allocation. Note that in this process a corresponding piece of probability mass from $b''(\check{t})$ must be moved into $b'(t)$ so that AF remains true.

The possibility of surgeries yields considerable freedom to construct new allocations $\rho'$, but care is needed to guarantee that PIC remains true so that $\rho'$ is $b'$-feasible. Several observations are helpful towards verifying PIC: first, if a type does not gain from choosing some other types' allocations, then she also cannot gain from any convex combination of these sub-allocations; second, if a type does not gain from choosing another type's allocation, then any convex combination of her own and that type's allocation is still at least as good for her as that type's allocation; third, if a type strictly loses from choosing another type's allocation, then this remains true for any perturbation of her original sub-allocation.

As a first surgery, the belief is kept fixed and each type outside the support of $b''$ gets restricted to choose either the disagreement outcome (to keep PIR in place) or her most preferred sub-allocation among the allocations of types in the support of $b''$. This will keep PIC in place and can only lead to a reduction of utility for the types outside the support.

If after this operation there exists a type $t$ outside the support who still obtains more than her $U$ utility, then, as a second surgery, we move some probability mass to her from her most preferred type in the initial support. AF and PIC (and PIR) are still in place, but now we have included $t$ into the support. In this way, we obtain a deviation $(b', U')$ such that in $U'$ all types in the support of $b'$ are strictly better off than in $U$, and all types outside the support are weakly better off in $U$ than in $U'$.

If the support of $b'$ contains a single type, we have obtained a neologism and are done. If it contains two types, say $t$ and $\check{t}$, the remaining surgeries are still comparatively easy. It is useful to introduce auxiliary variables that capture probabilities relative to the prior; we call the numbers $b'(t)/p_0(t)$ and $b'(\check{t})/p_0(\check{t})$ the $r$-values of the types $t$ and, resp., $\check{t}$ at the belief $b'$. If both types have the same $r$-value, then $(b', U')$ is a neologism for $(p_0, U)$ and we are done.

Otherwise one type, say $t$, has a smaller $r$ value than the other type, $\check{t}$. Now imagine that we change the deviation continuously, by moving an ever larger chunk of the allocation of type $\check{t}$, and a corresponding piece of belief probability mass, to type $t$. Along the way, any other type (i.e., the types outside the support of $b'$) always chooses her most preferred sub-allocation among the current allocations of the types $t$ and $\check{t}$ (or chooses the disagreement outcome if that is better). In this process, the $r$-value of type $t$ increases while the $r$-value of type $\check{t}$ decreases, and the utility of type $t$ can drop. AF, PIC, and PIR remain intact.

This process is continued until one of two things happens. Either both types' $r$ values are equalized, or the utility of type $t$ drops to her $U$ utility. In both cases we have arrived at a neologism and are done.

The general argument, where the type space (and thus the support of $b'$) can have any cardinality, is very much more complicated. The main reasons for the complications are that the number of incentive constraints in PIC increases fast (quadratically) with the cardinality of the type space, and that we have to find a deviation that equalizes the $r$-values across a potentially large number of types. These complications may have contributed to the fact that the underlying puzzle—the relation between neutral optimum and private-values solution concepts—has remained open essentially since the start of the informed-principal literature in the 1980s. In the following we provide a roadmap through the general argument.

The key to the general argument is the introduction of a special class of deviations. A feasible pair $(b', U')$ is a *deviation if at least one type has utility $> U$, all types not in $\mathrm{supp}(b')$ have utility $\leq U$ and each of them obtains the same sub-allocation as one of the types in $\mathrm{supp}(b')$, all types in $\mathrm{supp}(b')$ have utility $\geq U$, and the $U$-utility types $t \in \mathrm{supp}(b')$ have $r$-values $\leq r^*$, where $r^*$ is defined as the "target value" of $r$ that would be reached if all $r$-values of types with $> U$ utility were equal, that is

$$\sum_{U'(t)>U(t)} \left(r^* - r_{b'}(t)\right) p_0(t) = 0, \tag{5}$$

where $r_{b'}(t) = b'(t)/p_0(t)$ denotes the $r$-value of any type $t$ at the belief $b'$. (Note that $r^*$ is defined separately for each *deviation.)

Not all *deviations are allowed deviations in the definition of interim-optimality because some types in the support of $b'$ can have utility equal to $U$. However, a *deviation, with no $U$-utility type in the support of $b'$, exists by the first and second surgery arguments above.

If a *deviation is such that the $r$-values of all $> U$-utility types are equalized then, by construction, the *deviation is a neologism for $(p_0, U)$ and we are done.

Rather than explicitly describing the sequence of surgeries to be applied to the initial *deviation, we cut through to the end by considering a *deviation with the "right" properties.

Consider the *deviations that have a minimal cardinality of the support of $b'$ among all *deviations. Among these, consider the *deviations that have a maximum number of $U$-utility types in the support. Among these, we consider a *deviation that has a maximum number of types with $r$-value equal to $r^*$.

We claim that any such *deviation $(b', U')$ has the desired neologism properties. Suppose otherwise.

Let $r_0^*$ denote the value of $r^*$ for $(b', U')$. Then there exists a type $t^1$ in the support of $b'$ with $> U$ utility and an $r$-value below $r_0^*$. Let $T^{\leq}$ denote the set of $> U$-utility types with $r$-values $\leq r_0^*$. Let $T^{>}$ denote the $> U$-utility types with $r$-values $> r_0^*$.

Starting with $(b', U')$, we now consider the problem of maximizing the $r$-value of type $t^1$ via surgery subject to constraints. We consider surgery that concerns the types in $T^{\leq} \cup T^{>}$, while the other types in the support of $b'$ keep their sub-allocations, and each type outside the support of $b'$ chooses her best available sub-allocation among those of the types in the support of $b'$.

Using the numeration from the proof for reference, the constraints are that (10) all types in $T^{>}$ keep their sub-allocations, (11) each type in $T^{\leq}$ obtains a convex combination of the allocations of the types in $T^{\leq} \cup T^{>}$, (12) the $r$-value of type $t^1$ remains $\leq r_0^*$, (13) the $r$-values of the types in $T^{\leq} \setminus \{t^1\}$ remain the same as at the belief $b'$, (14) the $r$-values of the types $T^{>}$ remain $\geq r_0^*$, (15) each type in $T^{\leq}$ weakly prefers her new sub-allocation to the (old and new) allocation of each of the types in $\mathrm{supp}(b') \setminus (T^{\leq} \cup T^{>})$, (16) each type in $T^{\leq}$ weakly prefers her new sub-allocation to the new allocations of the types in $T^{\leq}$, and to the allocations of the types in $T^{>}$, and (17) the utility each type in $T^{\leq}$ does not fall below her $U$-utility.

We will show that at a solution to the maximization problem, denoted $(\hat{b}, \hat{u})$, the constraints (12), (14), (15), and (17) are not binding. This will allow us to increase the solution value via a perturbation that satisfies all constraints, and thus obtain a contradiction.

By construction, the solution $(\hat{b}, \hat{u})$ is a *deviation, where $\hat{b}$ has the same (minimum cardinality) support as $b'$. By the assumed maximality of the number of $U$-utility types in the support, at the optimum $(\hat{b}, \hat{u})$, the utility of no type in $T^{\leq}$ has dropped to her $U$-utility, that is, the constraints (17) are not binding. Note also that the $r^*$-value for $(\hat{b}, \hat{u})$ is still equal to $r_0^*$.

At $\hat{b}$, the $r$-value of type $t^1$ must still be strictly below $r_0^*$, and the $r$-values of the types in $T^{>}$ must still be strictly above $r_0^*$ because the number of types with $r$-values equal to $r^*$ was assumed to be already maximal at $(b', U')$, and by constraint (13) any type who before the optimization had an $r$ value equal to $r^*$ keeps it. Thus, the constraints (12) and (14) are not binding.

Suppose a constraint (15) is binding, that is, some type $t^i \in T^{\leq}$ is indifferent to a type $\mathring{t}$ that belongs to the support of $\hat{b}$ and who obtains her $U$ utility. Then we can do a surgery where all the probability mass and allocation of $\mathring{t}$ is moved to type $t^i$, yielding a new *deviation where the type $\mathring{t}$ does not belong to the belief support anymore, but this contradicts the minimality of the support of $\hat{b}$ among all *deviations.

Now we describe the perturbation of $(\hat{b}, \hat{U})$. Only the allocations of the types in a subset of $T^{\leq}$ are changed. In the subset we include all types with sub-allocations that type $t^1$ likes as well as her own allocation, and then include all types that any type included in the first round is indifferent to, and so on, until all indifferences in $T^{\leq}$ are exhausted. We denote this subset (which can be the singleton $\{t^1\}$) by $T^{\leq}_{=}$.

As a perturbing surgery, the allocation of each type in $T^{\leq}_{=}$ is now changed such that a small fraction of her new allocation comes from her respective most preferred type in $T^{>}$. The fraction will be the same for all types in $T^{\leq}_{=}$, implying that incentive compatibility relative to each other and to the types in $T^{>}$ remains intact. By construction, there are no indifferences from types in $T^{\leq}_{=}$ to types in $T^{\leq} \setminus T^{\leq}_{=}$ if the perturbation is small.

Due to the new allocation chunks and corresponding probability masses, the types in $T^{\leq}_{=}$ will now have increased $r$-values. For all types except $t^1$, the $r$ values must be brought back to their previous levels to satisfy constraint (13).

To this end, we consider a directed graph with nodes $T^{\leq}_{=}$ where each edge corresponds to an indifference. We select a tree with root $t^1$ in $T^{\leq}_{=}$. All the direct predecessor types of the tree's end nodes get chunks of the end node's allocations and corresponding probability masses such that the end nodes are back to their correct $r$ values. These corrections are iterated backwards through the tree. Due to the indifferences along the way, the involved types keep their utility levels. Eventually only type $t^1$ gains probability mass, yielding the desired contradiction.

The following result is immediate from the definitions (and has nothing to do with private values.

**Lemma 4.** *Any strongly neologism-proof allocation is interim optimal.*

The reverse implication is the last piece needed. Here we need a property from Mylovanov and Tröger (2012).

**Definition 7.** *A setting with generalized private values is called* separable *if there exists a sub-allocation $\alpha$ for which the defining inequalities of the set $A$ are all satisfied strictly.*

**Lemma 5.** *In any separable generalized-private-values setting, any interim optimal allocation is strongly neologism-proof.*

To prove this (for details see the appendix), we consider an interim optimal allocation $\rho$ and suppose it is not strongly neologism-proof.

Let $U_0^\rho$ denote the payoff vector implemented by $\rho$. By assumption, there exists a belief $b'$ and a $b'$-feasible allocation $\rho'$ such that, for all $t \in \text{supp}(b')$, we have that (i) $U_0^{\rho'}(t) \geq U_0^\rho(t)$ with strict inequality for at least one type $t'$, and (ii) $U_0^{\rho'}(t)$ is below the maximum feasible payoff.

We now do surgery in order to find a deviation as required in the definition of interim-optimality. By the separability assumption, there exists an allocation such that all agents' constraints are satisfied strictly; we perturb the allocation $\rho'$ by having type $t'$ offer this "separating" allocation with a small probability and simultaneously slightly increasing the probability mass for type $t'$. She will still be strictly better off than in $\rho$. Given the new allocation, the agents' constraints are satisfied strictly. Thus, we can again perturb it without violating the agents' constraints; we do this by giving all types in $\text{supp}(b')$ their "maximum feasible payoff" with a small probability. Now all types in the belief support are strictly better off than in $\rho$, but PIC may not hold anymore. It can be restored by further surgery, using a trick from Mylovanov and Tröger (2014). If one type in the belief support is attracted to the allocation of another type in the belief support, then we let the first type offer the average allocation of what both types used to offer, and move all the probability mass from the second type to the first type. This procedure continues until incentive compatibility is satisfied for the types in the (remaining) support. Let the types outside the support choose their optimum among the allocations of the types in the support. Then we have a deviation as considered in the definition of interim-optimality.

# 4 Summary

He we sum up the main implications of the above lemmata.

**Corollary 1.** *Consider a separable generalized-private-values environment in the sense of Mylovanov and Tröger (2012). Then strong neologism-proofness, interim optimality and neo-optimum are equivalent.*

**Corollary 2.** *In any separable generalized-private-values environment that is also a Bayesian incentive problem, any neutral optimum is strongly neologism-proof.*

Since strong neologism-proofness often yields sharp properties related to competitive equilibria (Maskin and Tirole, 1990)—such as ex-ante optimality in quasi-linear settings (Mylovanov and Tröger, 2014)—the same properties apply to any neutral optimum.

From the existence result in Mylovanov and Tröger (2012), we can also conclude that a neo-optimum exists broadly in private-value settings, including settings that do not satisfy the finiteness properties of Bayesian incentive problems as defined by Myerson (1983).

# Appendix A: omitted proofs

*Proof of Lemma 3.* Consider a neo-optimum $\nu$ that implements a payoff vector $U = U_0^\nu$.

Throughout the proof, we will call a pair $(b, V)$ *feasible* if $b \in B$ and $V = U_0^\rho$ for some $b$-feasible allocation $\rho$.

Also, let $\zeta_0$ denote the sub-allocation that always implements the outside option, that is $\zeta_0(\mathbf{t}_{-0}) = z_0$ for all $\mathbf{t}_{-0} \in \mathbf{T}_{-0}$.

Suppose that $\nu$ is not interim-optimal, that is, some $(b', U')$ is feasible such that $U'(t) > U(t)$ for all $t \in \text{supp}(b')$. The pair $(b', U')$ has the same property relative to all payoff vectors in a neighborhood of $U$ and to everything below. Thus, to obtain a contradiction it is sufficient to show that a neologism exists for $U$.

Define $r_{b'}(t) = b'(t)/p_0(t)$ for all $t \in T$.

Call a feasible pair $(b', U')$ a *deviation if there exists an allocation $\rho'$ such that $U'(t) = \Pi(\rho'_t)(t)$ for all $t$, $U'(t) \leq U(t)$ for all $t \in T \setminus \text{supp}(b')$,

$$\{\rho'_t \mid t \in T\} \subseteq \{\rho'_t \mid t \in \text{supp}(b')\} \cup \{\zeta_0\},$$

$U'(t) \geq U(t)$ for all $t \in \text{supp}(b')$ with a strict inequality for at least one type, and any type $t \in \text{supp}(b')$ with $U'(t) = U(t)$ satisfies $r_{b'}(t) \leq r^*$, where we define $r^*$ via the equation

$$r^* \cdot \sum_{U'(t)>U(t)} p_0(t) + \sum_{U'(t)=U(t)} b'(t) = 1. \tag{6}$$

Note that we can also write this equation in the form (5).

As a first step, we show that a *deviation exists. Because $\nu$ is not interim-optimal, a feasible pair $(b'', U'')$ exists such that $U''(t) > U(t)$ for all $t \in \mathrm{supp}(b'')$. Let $\rho''$ denote a corresponding allocation. That is,

$$U''(t) = \Pi(\rho''_t)(t) \geq \Pi(\rho''_{\check{t}})(t) \tag{7}$$

for all $t, \check{t} \in T$, and

$$\sum_{t' \in T} b''(t')\rho''_{t'} \in A. \tag{8}$$

For all $t \in T \setminus \mathrm{supp}(b'')$, select a

$$v(t) \in \arg \max_{\check{t} \in \mathrm{supp}(b'')} \Pi(\rho''_{\check{t}})(t), \tag{9}$$

and let $v(t) = t$ for all $t \in \mathrm{supp}(b'')$. For all $t \in T$, define $\rho'_t = \rho''_{v(t)}$ if $\Pi(\rho''_{v(t)})(t) \geq 0$ and otherwise define $\rho'_t = \zeta_0$. Let $U'(t) = \Pi(\rho'_t)(t)$ for all $t \in T$.

In other words, every type outside $\mathrm{supp}(b'')$ gets restricted to their respective best allocation of a type in $\mathrm{supp}(b'')$, while the types in $\mathrm{supp}(b'')$ keep their allocations. The only exception from this rule are types outside $\mathrm{supp}(b'')$ who are better off with the outside option.

The construction has a second part in which we move from the belief $b''$ to a new belief $b'$. If a type $t$ outside $\mathrm{supp}(b'')$ has utility $U'(t) > U(t)$, then we move a bit of probability to her from the type $v(t)$. This includes $t$ into the support $\mathrm{supp}(b')$, without affecting AF because the types $t$ and $v(t)$ get the same allocation.

We will now describe the second part of the construction more formally. For any $t' \in \mathrm{supp}(b'')$, let $w(t')$ denote the number of types outside $\mathrm{supp}(b'')$ which choose the allocation of type $t'$ and still get more than their $U$ utility. That is,

$$w(t') = |\{t \in T \setminus \mathrm{supp}(b'') \mid v(t) = t',\ U'(t) > U(t)\}|.$$

Given any $\epsilon > 0$, define for all $t \in T$,

$$b'(t) = \begin{cases} 0 & \text{if } t \notin \text{supp}(b'') \ \text{ and } \ U'(t) \leq U(t), \\ \epsilon & \text{if } t \notin \text{supp}(b'') \ \text{ and } \ U'(t) > U(t), \\ b''(t) - \epsilon w(t) & \text{if } t \in \text{supp}(b''), \end{cases}$$

where $\epsilon$ is chosen so small that $b'(t) > 0$ for all $t \in \text{supp}(b'')$.

By construction, PIR holds for $\rho'$, that is, $\Pi(\rho'_t)(t) \geq 0$ for all $t \in T$.

To see that PIC holds as well, consider any $t, t' \in T$. If $\Pi(\rho''_{v(t')})(t') < 0$ then $\rho'_{t'} = \zeta_0$ so that obviously $\Pi(\rho'_t)(t) \geq \Pi(\rho'_{t'})(t)$. If $\Pi(\rho''_{v(t')})(t') \geq 0$, then

$$\Pi(\rho'_t)(t) \geq \Pi(\rho''_{v(t)})(t) \geq \Pi(\rho''_{v(t')})(t) = \Pi(\rho'_{t'})(t),$$

where in the cases with $t \in \text{supp}(b')$ the second of the above inequalities follows from (7) with $t = v(t)$ and $\check{t} = v(t')$, and in the other cases this inequality follows from (9).

To show that AF holds for the allocation $\rho'$ together with the belief $b'$, recall (8) and note that

$$\sum_{t' \in T} b'(t')\rho'_{t'} = \sum_{t' \in T} b''(t')\rho''_{t'}.$$

We conclude that $(b', U')$ is a feasible pair. Moreover, by construction,

$$\text{supp}(b') = \{t \in T \mid U'(t) > U(t)\}.$$

Thus, $(b', U')$ is a *deviation.

Let $D_2$ denote the set of *deviations $(b', U')$ such that $|\text{supp}(b')|$ is minimal among all *deviations. Let $D_1$ denote the set of *deviations $(b', U')$ in $D_2$ such that $|\{t \in \text{supp}(b') \mid U'(t) = U(t)\}|$ is maximal among all *deviations in $D_2$. Let $D_0$ denote the set of *deviations $(b', U')$ in $D_1$ such that $|\{t \in \text{supp}(b') \mid U'(t) > U(t), r_{b'}(t) = r^*\}|$ is maximal among all *deviations in $D_1$.

In the following, we consider a *deviation $(b', U') \in D_0$. Let $\rho'$ denote an allocation for this *deviation. Let $r^*_0$ denote the $r^*$-value for this *deviation.

It remains to show that $(b', U')$ is a neologism for $(p_0, U)$. To prove this, we have to show that all types $t$ with $U'(t) > U(t)$ satisfy $r_{b'}(t) = r^*_0$.

Suppose otherwise. Then there exists a type $t^1$ with $U'(t^1) > U(t^1)$ and $r_{b'}(t^1) < r^*_0$ as well as a type $t$ with $U'(t) > U(t)$ and $r_{b'}(t) > r^*_0$. Let $t^1, \ldots, t^n$ denote the types with $U'(t^i) > U(t^i)$ and $r_{b'}(t^i) \leq r^*_0$. Let $t^{n+1}, \ldots, t^{n+m}$ denote the types with $U'(t^i) > U(t^i)$ and $r_{b'}(t^i) > r^*_0$.

Now consider the following problem, where $y$ stands for the probability mass assigned to type $t^1$, and $x_{k,i}$ stands for the fraction of the allocation of type $t^k$ that is reassigned to type $t^i$.

$$
\begin{aligned}
\max_{\substack{y,\ (x_{k,i})_{k=1,\dots,n+m,} \\ i=1,\dots,n+m}} \quad & y, \\
\text{s.t.} \quad & x_{k,i} = \mathbf{1}_{k=i} \text{ for all } k \text{ and all } i \geq n+1, && (10) \\
& x_{k,i} \geq 0 \text{ for all } k \text{ and all } i \leq n, \\
& \sum_{k=1}^{n+m} x_{k,i} = 1 \text{ for all } i \leq n, && (11) \\
& y \leq p_0(t^1) r_0^*, && (12) \\
& b'(t^k) = x_{k,1} y + \sum_{i=2}^{n} x_{k,i} b'(t^i) \quad \text{for all } k \leq n, && (13) \\
& b'(t^k) - x_{k,1} y - \sum_{i=2}^{n} x_{k,i} b'(t^i) \geq p_0(t^k) r_0^* \quad \text{for all } k > n, && (14) \\
& \sum_{k=1}^{n+m} x_{k,i} \Pi(\rho'_k)(t^i) \geq \Pi(\rho'_{\check{t}})(t^i) \text{ for all } \check{t} \in \operatorname{supp}(b') \setminus \{t^1, \dots, t^{n+m}\}, \\
& && (15) \\
& \sum_{k=1}^{n+m} (x_{k,i} - x_{k,j}) \Pi(\rho'_k)(t^i) \geq 0 \text{ for all } i \leq n \text{ and all } j, && (16) \\
& \sum_{k=1}^{n+m} x_{k,i} \Pi(\rho'_k)(t^i) \geq U(t^i) \text{ for all } i \leq n. && (17)
\end{aligned}
$$

Note that all constraints are satisfied at the point $y = b'(t^1)$ and $x_{k,i} = \mathbf{1}_{k=i}$ for all $k$ and $i$. Thus, the feasibility set is non-empty and the solution value—which exists by the extreme-value theorem of Weierstrass—is $\geq b'(t^1)$.

Given any solution $\hat{y}$, $(\hat{x}_{k,i})$, define an allocation $\hat{\rho}$ as follows:

$$\hat{\rho}_{t^i} = \sum_{k=1}^{n+m} \hat{x}_{k,i} \rho'_{t^k} \quad \text{for all } i = 1, \dots, n+m;$$

$\hat{\rho}_t = \rho'_t$ for all $t \in \operatorname{supp}(b') \setminus \{t^1, \dots, t^{n+m}\}$; $\hat{\rho}_t = \zeta_0$ for all $t \in T \setminus \operatorname{supp}(b')$ with $\max_{\check{t} \in \operatorname{supp}(b')} \Pi(\hat{\rho}_{\check{t}})(t) < 0$, and $\hat{\rho}_t = \hat{\rho}_{\hat{v}(t)}$ for all $t \in T \setminus \operatorname{supp}(b')$ with

$\max_{\tilde{t} \in \mathrm{supp}(b')} \Pi(\hat{\rho}_{\tilde{t}})(t) \geq 0$, where we choose any

$$\hat{v}(t) \in \arg \max_{\tilde{t} \in \mathrm{supp}(b')} \Pi(\hat{\rho}_{\tilde{t}})(t).$$

This construction already implies that $\hat{\rho}$ satisfies the PIR conditions for all $t \in T \setminus \mathrm{supp}(b')$.

Define a utility vector $\hat{U}$ via $\hat{U}(t) = \Pi(\hat{\rho}_t)(t)$ for all $t$. Define a belief $\hat{b}$ as follows: $\hat{b}(t^1) = \hat{y}$; $\hat{b}(t^k) = b'(t^k)$ for $k = 2, \ldots, n$;

$$\hat{b}(t^k) = b'(t^k) - \hat{x}_{k,1}\hat{y} - \sum_{i=2}^{n} \hat{x}_{k,i} b'(t^i) \quad \text{for all } k = n+1, \ldots, n+m;$$

$\hat{b}(t) = b'(t)$ for all $t \in T \setminus \{t^1, \ldots, t^{n+m}\}$. Note that $\hat{b} \in B$ because

$$\begin{aligned}
\sum_{t \in T} \hat{b}(t) &= \sum_{t \in \mathrm{supp}(b') \setminus \{t^1, \ldots, t^{n+m}\}} b'(t) + \hat{y} + \sum_{k=2}^{n} b'(t^k) \\
&\quad + \sum_{k=n+1}^{n+m} \left( b'(t^k) - \hat{x}_{k,1}\hat{y} - \sum_{i=2}^{n} \hat{x}_{k,i} b'(t^i) \right) \\
&= \sum_{t \in \mathrm{supp}(b') \setminus \{t^1\}} b'(t) + \left( 1 - \sum_{k=n+1}^{n+m} \hat{x}_{k,1} \right) \hat{y} - \sum_{i=2}^{n} \sum_{k=n+1}^{n+m} \hat{x}_{k,i} b'(t^i)
\end{aligned}$$

and, recalling $b' \in B$ and using constraint (11), the above chain continues as

$$\begin{aligned}
&= 1 - b'(t^1) + \sum_{k=1}^{n} \hat{x}_{k,1}\hat{y} - \sum_{i=2}^{n} \left( 1 - \sum_{k=1}^{n} \hat{x}_{k,i} \right) b'(t^i) \\
&= 1 - \sum_{i=1}^{n} b'(t^i) + \sum_{k=1}^{n} \hat{x}_{k,1}\hat{y} + \sum_{i=2}^{n} \sum_{k=1}^{n} \hat{x}_{k,i} b'(t^i) \; = \; 1,
\end{aligned}$$

where the last equation follows from the formula that is obtained by summing the constraints (13) across all $k \leq n$. Next we show that $(\hat{b}, \hat{U})$ is feasible. To verify condition AF for $(\hat{b}, \hat{U})$, note that

$$\sum_{t \in T \setminus \{t^1, \ldots, t^{n+m}\}} \hat{b}(t) \hat{\rho}_t \;=\; \sum_{t \in T \setminus \{t^1, \ldots, t^{n+m}\}} b'(t) \rho'_t$$

and

$$\begin{aligned}\sum_{i=1}^{n+m}\hat{b}(t^i)\hat{\rho}_{t^i} &= \hat{y}\hat{\rho}_{t^1}+\sum_{i=2}^{n}b'(t^i)\hat{\rho}_{t^i}+\sum_{k=n+1}^{n+m}\hat{b}(t^k)\rho'_{t^k}\\ &= \hat{y}\sum_{k=1}^{n+m}\hat{x}_{k,1}\rho'_{t^k}+\sum_{i=2}^{n}b'(t^i)\sum_{k=1}^{n+m}\hat{x}_{k,i}\rho'_{t^k}\\ &\quad +\sum_{k=n+1}^{n+m}\left(b'(t^k)-\hat{x}_{k,1}\hat{y}-\sum_{i=2}^{n}\hat{x}_{k,i}b'(t^i)\right)\rho'_{t^k}\\ &= \sum_{k=1}^{n}\left(\hat{y}\hat{x}_{k,1}+\sum_{i=2}^{n}b'(t^i)\hat{x}_{k,i}\right)\rho'_{t^k}+\sum_{k=n+1}^{n+m}b'(t^k)\rho'_{t^k}\\ &\overset{(13)}{=} \textstyle\sum_{k=1}^{n+m} \quad b'(t^k)\rho'_{t^k}.\end{aligned}$$

We also have to verify PIC. For any $t \in T \setminus \mathrm{supp}(b')$ and $t' \in \mathrm{supp}(b')$,

$$\Pi(\hat{\rho}_t)(t) \geq \Pi(\hat{\rho}_{\hat{v}(t)})(t) \geq \Pi(\hat{\rho}_{t'})(t),$$

by definition of $\hat{v}(t)$. Similarly, for any $t, t' \in T \setminus \mathrm{supp}(b')$ with $\hat{\rho}_{t'} \neq \zeta_0$,

$$\Pi(\hat{\rho}_t)(t) \geq \Pi(\hat{\rho}_{\hat{v}(t)})(t) \geq \Pi(\hat{\rho}_{v(t')})(t) = \Pi(\hat{\rho}_{t'})(t).$$

Moreover, for any $t, t' \in T \setminus \mathrm{supp}(b')$ with $\hat{\rho}_{t'} = \zeta_0$,

$$\Pi(\hat{\rho}_t)(t) \geq 0 = \Pi(\hat{\rho}_{t'})(t).$$

For all $t \in \mathrm{supp}(b') \setminus \{t^1, \ldots, t^n\}$ and all $t' \in T$,

$$\Pi(\hat{\rho}_t)(t) = \Pi(\rho'_t)(t) = \max_{\tilde{t} \in T} \Pi(\rho'_{\tilde{t}})(t) \geq \Pi(\hat{\rho}_{t'})(t).$$

For all $t \in \{t^1, \ldots, t^n\}$ and all $t' \in \{t^1, \ldots, t^{n+m}\}$, constraint (16) directly implies $\Pi(\hat{\rho}_t)(t) \geq \Pi(\hat{\rho}_{t'})(t)$.

This then also implies that for all $t \in \{t^1, \ldots, t^n\}$ and all $t' \in T \setminus \mathrm{supp}(b')$, $\Pi(\hat{\rho}_t)(t) \geq \Pi(\hat{\rho}_{\hat{v}(t')})(t) = \Pi(\hat{\rho}_{t'})(t)$.

For all $t \in \{t^1, \ldots, t^n\}$ and all $t' \in \mathrm{supp}(b') \setminus \{t^1, \ldots, t^{n+m}\}$, constraint (15) directly implies $\Pi(\hat{\rho}_t)(t) \geq \Pi(\hat{\rho}_{t'})(t)$. This completes the proof of PIC.

Next we show that $(\hat{b}, \hat{U})$ is a *-deviation.

For all $i = 1, \ldots, n$, we have $\hat{U}(t^i) \geq U(t^i)$ by constraint (17). (In particular, $\hat{\rho}$ satisfies the PIR conditions for all these types $t = t^i$).

For all $t \in \mathrm{supp}(b') \setminus \{t^1, \ldots, t^{n+m}\}$, we have $\hat{U}(t) = U'(t) = U(t)$ by construction. (In particular, $\hat{\rho}$ satisfies the PIR conditions for all these types $t$). For all $i = n+1, \ldots, n+m$, we have $\hat{U}(t^i) = U'(t^i) > U(t^i)$ by construction. (In particular, $\hat{\rho}$ satisfies the PIR conditions for all these types $t = t^i$). For all $t \in T \setminus \mathrm{supp}(b')$,

$$\hat{U}(t) = \Pi(\hat{\rho}_{\hat{v}(t)}) \leq \max_{\check{t} \in \mathrm{supp}(b')} \Pi(\rho'_{\check{t}})(t) \leq \Pi(\rho'_t)(t) = U'(t) \leq U(t).$$

In particular, for any type $t \in T$, if $\hat{U}(t) > U(t)$ then $U'(t) > U(t)$, and all types $t \in T$ with $\hat{U}(t) = U(t)$ satisfy $r_{\hat{b}}(t) \leq r_0^*$. Thus, to complete the proof that $(\hat{b}, \hat{U})$ is a *-deviation, it remains to show that the $r^*$ value for $(\hat{b}, \hat{U})$ satisfies $r^* \geq r_0^*$.

Using the definition (6),

$$\begin{aligned} 1 &= r_0^* \cdot \sum_{U'(t) > U(t)} p_0(t) + \sum_{U'(t) = U(t)} b'(t) \\ &\geq r_0^* \cdot \sum_{\hat{U}(t) > U(t)} p_0(t) + \sum_{\substack{U'(t) > U(t), \\ \hat{U}(t) = U(t)}} \hat{b}(t) + \sum_{U'(t) = U(t)} b'(t) \\ &= r_0^* \cdot \sum_{\hat{U}(t) > U(t)} p_0(t) + \sum_{\hat{U}(t) = U(t)} \hat{b}(t), \end{aligned}$$

implying that $r^* \geq r_0^*$.

Next we show that the constraints (12), (14), (15), and (17) are not binding at the solution $\hat{y}$, $(\hat{x}_{k,i})$.

Note that $(\hat{b}, \hat{U}) \in D_2$ because $\mathrm{supp}(\hat{b}) = \mathrm{supp}(b')$. Moreover, because any type $t \in \mathrm{supp}(\hat{b})$ with $U'(t) = U(t)$ also satisfies $\hat{U}(t) = U(t)$, we even have $(\hat{b}, \hat{U}) \in D_1$. Thus, $\hat{U}(t^i) > U(t^i)$ for all $i = 1, \ldots, n$, implying that the constraints (17) are not binding.

As a consequence, $r^* = r_0^*$.

By construction, any type $t \in T$ with $r_{b'}(t^1) = r_0^*$ also satisfies $r_{\hat{b}}(t^1) = r_0^*$. Thus, we even have $(\hat{b}, \hat{U}) \in D_0$, implying that the constraints (12) and (14) are not binding.

To show that the constraints (15) are not binding, we suppose that

$$\hat{U}(t^i) = \Pi(\hat{\rho}_{\mathring{t}})(t^i)$$

for some $i \leq n$ and some $\mathring{t} \in \operatorname{supp}(b') \setminus \{t^1, \dots, t^{n+m}\}$ and derive a contradiction. Define an allocation $\mathring{\rho}$ as follows. Let

$$\mathring{\rho}_{t^i} = \frac{\hat{b}(t^i)}{\hat{b}(t^i) + \hat{b}(\mathring{t})}\hat{\rho}_{t^i} + \frac{\hat{b}(\mathring{t})}{\hat{b}(t^i) + \hat{b}(\mathring{t})}\hat{\rho}_{\mathring{t}}; \tag{18}$$

let $\mathring{\rho}_t = \hat{\rho}_t$ for all $t \in \operatorname{supp}(b') \setminus \{t^i, \mathring{t}\}$; for all $t \in \{\mathring{t}\} \cup T \setminus \operatorname{supp}(b')$ with $\max_{\check{t} \in \operatorname{supp}(b') \setminus \{\mathring{t}\}} \Pi(\mathring{\rho}_{\check{t}})(t) \geq 0$, let $\mathring{\rho}_t = \mathring{\rho}_{\mathring{v}(t)}$, where we choose any

$$\mathring{v}(t) \in \arg \max_{\check{t} \in \operatorname{supp}(b') \setminus \{\mathring{t}\}} \Pi(\mathring{\rho}_{\check{t}})(t);$$

for all $t \in \{\mathring{t}\} \cup T \setminus \operatorname{supp}(b')$ with $\max_{\check{t} \in \operatorname{supp}(b') \setminus \{\mathring{t}\}} \Pi(\mathring{\rho}_{\check{t}})(t) < 0$, let $\mathring{\rho}_t = \zeta_0$.

Define a belief $\mathring{b}$ as follows. Let

$$\mathring{b}(t^i) = \hat{b}(t^i) + \hat{b}(\mathring{t}); \tag{19}$$

let $\mathring{b}(\mathring{t}) = 0$; let $\mathring{b}(t) = \hat{b}(t)$ for all $\in T \setminus \{t^i, \mathring{t}\}$.

Define a payoff vector $\mathring{U}$ via $\mathring{U}(t) = \Pi(\mathring{\rho}_t)(t)$ for all $t \in T$.

By the supposed indifference, $\mathring{U}(t^i) = \hat{U}(t^i)$. Moverover, the set of alternative sub-allocations to choose from (beyond the disagreement outcome) has shrunk:

$$\{\mathring{\rho}_{t'} | t' \in T \setminus \{t^i\}\} \subseteq \{\hat{\rho}_{t'} | t' \in T \setminus \{t^i\}\} \cup \{\zeta_0\}$$

Thus, because $(\hat{\rho}_t)_{t \in T}$ satisfies PIC, the allocation $\mathring{\rho}$ also satisfies the PIC conditions for $t = t^i$ and all $t' \in T$. The same holds for $t \in \operatorname{supp}(b') \setminus \{t^i, \mathring{t}\}$ and all $t' \neq t^i$ because $\mathring{\rho}_t = \hat{\rho}_i$.

By construction, the allocation $\mathring{\rho}$ satisfies PIR. It also satisfies the PIC conditions for all $t \in \operatorname{supp}(b') \setminus \{t^i, \mathring{t}\}$ and for $t' = t^i$ because

$$\Pi(\mathring{\rho}_{t^i})(t) \overset{(18)}{\leq} \max\{\Pi(\hat{\rho}_{t^i})(t), \Pi(\hat{\rho}_{\mathring{t}})(t)\} \leq \Pi(\hat{\rho}_t)(t) = \Pi(\mathring{\rho}_t)(t).$$

Finally, the allocation $\mathring{\rho}$ satisfies the PIC conditions for all $t \in \{\mathring{t}\} \cup T \setminus \operatorname{supp}(b')$ and all $t'$ due to the definition of $\mathring{v}(t)$. This completes the verification of PIC.

To verify AF for $(\mathring{\rho}_t)_{t \in T}$ together with $\mathring{b}$, note that

$$\sum_{t \in T \setminus \{t^i, \mathring{t}\}} \mathring{b}(t)\mathring{\rho}_t \quad = \quad \sum_{t \in T \setminus \{t^i, \mathring{t}\}} \hat{b}(t)\hat{\rho}_t,$$

and

$$\mathring{b}(t^i)\hat{\rho}_{t^i} + \mathring{b}(\mathring{t})\hat{\rho}_{\mathring{t}} \;\; = \;\; \hat{b}(t^i)\hat{\rho}_{t^i} + \hat{b}(\mathring{t})\hat{\rho}_{\mathring{t}}$$

by (18) and (19). Thus, $(\mathring{b}, \mathring{U})$ is feasible.

Also note that $\operatorname{supp}(\mathring{b}) = \operatorname{supp}(b') \setminus \{\mathring{t}\}$.

To obtain a contradiction it remains to verify that $(\mathring{b}, \mathring{U})$ is a *deviation because then $(b', U') \notin D_2$.

By construction, $\mathring{U}(t) = \hat{U}(t)$ for all $t \in \operatorname{supp}(\mathring{b})$ and $\mathring{U}(t) \leq \hat{U}(t)$ for all $t \in T \setminus \operatorname{supp}(\mathring{b})$. Thus, using the definition (6),

$$\begin{aligned} 1 &= r_0^* \cdot \sum_{\hat{U}(t) > U(t)} p_0(t) + \sum_{\hat{U}(t) = U(t)} \hat{b}(t) \\ &= r_0^* \cdot \sum_{\mathring{U}(t) > U(t)} p_0(t) + \sum_{\mathring{U}(t) = U(t)} \mathring{b}(t) \; + \hat{b}(\mathring{t}), \end{aligned}$$

implying that the $r^*$ value for $(\mathring{b}, \mathring{U})$ satisfies $r^* > r_0^*$. Thus, $(\mathring{b}, \mathring{U})$ is a *deviation.

To obtain the final contradiction, we will now define a perturbation of the presumed max-solution that satisfies all constraints and increases the solution value.

Denote $T^{\leq} = \{t^1, \ldots, t^n\}$. Given the allocation $\hat{\rho}$, we say that $(v_1, \ldots, v_l)$ (where $l \geq 1$) is a *chain-indifference path* in $T^{\leq}$ if $v_1, \ldots, v_l \in T^{\leq}$ and $\Pi(\hat{\rho}_{v_{i+1}})(v_i) = \Pi(\hat{\rho}_{v_i})(v_i)$ for all $i < l$.

Let $T^{\leq}_{=}$ denote the types $t \in T^{\leq}$ such that a chain-indifference path $(v_1, \ldots, v_l)$ exists with $v_1 = t^1$ and $v_l = t$.

An *indifference graph* $(T^{\leq}_{=}, g)$ is defined as a directed graph such that (i) the set of nodes equals $T^{\leq}_{=}$ and (ii) $\Pi(\hat{\rho}_{t'})(t) = \Pi(\hat{\rho}_t)(t)$ for each edge $(t, t') \in g$.

By definition of $T^{\leq}_{=}$, there exists an indifference graph such that, for all $t' \in T^{\leq}_{=}$, there exists a (chain-indifference) path from $t^1$ to $t'$. Requiring this property, let $(T^{\leq}_{=}, g)$ denote an indifference graph with a minimal number of edges.

Then $(T^{\leq}_{=}, g)$ is a tree with root $t^1$; that is, no edge points to $t^1$, and there exists a unique path from $t^1$ to each node in $T^{\leq}_{=}$. (To see the uniqueness statement, suppose that paths $p_1$ and $p_2$ lead to the same node, and $(t^{1'}, t'') \in p_1$, $(t^{2'}, t'') \in p_2$ with $t^{1'} \neq t^{2'}$ are edges where the two paths join. Then

$(T^{\leq}_{=}, g \setminus \{(t^{1'}, t'')\})$ is an indifference graph will a smaller number of edges in which still there exists a path from $t^1$ to any other node—contradiction.)

For each $t \in T^{\leq}_{=}$, let the index of a "most preferred type" among those with $r > r^*$ (recall that the constraints (14) are not binding) be denoted

$$\iota(t) \in \arg \max_{j \in \{n+1, \ldots, n+m\}} \Pi(\hat{\rho}_{t^j})(t).$$

For all $i$ with $t^i \in T^{\leq}_{=}$, define

$$\sigma(i) = \{j \in \{1, \ldots, n\} \mid t^j \text{ is a direct successor of } t^i \text{ in } (T^{\leq}_{=}, g)\}.$$

Note that $\sigma(i) = \emptyset$ means that $t^i$ is an end node in $(T^{\leq}_{=}, g)$. For each $j \neq 1$ with $t^j \in T^{\leq}_{=}$, let $\sigma^{-1}(j)$ denote the index of the direct predecessor of $t^j$ in $(T^{\leq}_{=}, g)$.

Fix any $0 < \epsilon < 1$. The following definition works recursively from the end nodes backwards through the tree. Define

$$\omega_j \;=\; \frac{b'(t^j)}{b'(t^{\sigma^{-1}(j)})} \frac{\epsilon + (1-\epsilon) \sum_{k \in \sigma(j)} \omega_k}{(1-\epsilon)} \quad \text{for all } j \notin \sigma(1) \text{ with } t^j \in T^{\leq}_{=}$$

and

$$z_j \;=\; b'(t^j) \frac{\epsilon + (1-\epsilon) \sum_{k \in \sigma(j)} \omega_k}{(1-\epsilon)} \quad \text{for all } j \in \sigma(1).$$

Define

$$\mathring{y} \;=\; \frac{1}{1-\epsilon} \hat{y} + \sum_{j \in \sigma(1)} z_j \tag{20}$$

and

$$\omega_j \;=\; \frac{z_j}{\mathring{y}} \quad \text{for all } j \in \sigma(1).$$

Thus, replacing $z_j = \omega_j \mathring{y}$ in (20) and solving for $\mathring{y}$, we find that

$$\mathring{y} \;=\; \frac{\hat{y}}{(1 - \sum_{j \in \sigma(1)} \omega_j)(1-\epsilon)}. \tag{21}$$

For all $i$ with $t^i \in T^{\le}_{=}$ and all $k = 1, \ldots, n+m$, define

$$\mathring{x}_{k,i} \;\; = \;\; \left( \hat{x}_{k,i} \left( 1 - \sum_{j \in \sigma(i)} \omega_j \right) + \sum_{j \in \sigma(i)} \hat{x}_{k,j} \omega_j \right) (1 - \epsilon) + \mathbf{1}_{k=\iota(t^i)} \epsilon. \quad (22)$$

For all $i \le n$ with $t^i \notin T^{\le}_{=}$, and all $i = n+1, \ldots, n+m$ and all $k = 1, \ldots, n+m$, define $\mathring{x}_{k,i} = \mathbf{1}_{k=i}$.

First note that

$$0 < \omega_j \to_{\epsilon \to 0} 0 \quad \text{for all } j \ne 1 \text{ with } t^j \in T^{\le}_{=}.$$

(This is seen recursively, arguing backwards from the end nodes in $(T^{\le}_{=}, g)$.)

Thus, through choosing $\epsilon$ sufficiently close to 0, we can guarantee that $\sum_{j \in \sigma(i)} \omega_j$ is close to 0 for all $i = 1, \ldots, n$, implying

$$\mathring{x}_{k,i} \ge 0$$

and, using (21),

$$\mathring{y} > \hat{y}.$$

In particular, once we show that $\mathring{y}, (\mathring{x}_{k,i})$ satisfies all remaining constraints of our max-problem, then we have a contradiction to the assumption that $\hat{y}, (\hat{x}_{k,i})$ is a solution.

First of all, recall that the constraints (12), (14), (15), and (17) are not binding at the solution $\hat{y}$, $(\hat{x}_{k,i})$.

Thus, because $\mathring{x}_{k,i} \to \hat{x}_{k,i}$ and $\mathring{y} \to \hat{y}$ as $\epsilon \to 0$, the constraints (12), (14), (15), and (17) are also strictly satisfied at $\mathring{y}, (\mathring{x}_{k,i})$, assuming $\epsilon$ is sufficiently close to 0.

Define the auxiliary variables $\hat{b}(t^1) = \hat{y}$ and $\hat{b}(t^i) = b'(t^i)$ for all $i = 2, \ldots, n$. Defining the column vectors $b'^n = (b'(t^1), \ldots, b'(t^n))^T$ and $\hat{b}^n = (\hat{b}(t^1), \ldots, \hat{b}(t^n))^T$ and the square matrix $\hat{X} = (\hat{x}_{k,i})_{k \le n,\ i \le n}$, constraint (13) reads

$$b'^n = \hat{X} \hat{b}^n. \quad (23)$$

Defining the square matrices $\mathring{X} = (\mathring{x}_{k,i})_{k \le n,\ i \le n}$ and $H = (h_{j,i})_{j \le n,\ i \le n}$ via $h_{j,i} = \mathbf{1}_{j=i}$ if $t^i \notin T^{\le}_{=}$, and

$$h_{i,i} = \left( 1 - \sum_{j \in \sigma(i)} \omega_j \right) (1 - \epsilon), \quad h_{j,i} = \omega_j (1 - \epsilon) \text{ for all } j \in \sigma(i), \quad h_{j,i} = 0 \text{ otherwise,}$$

if $t^i \in T^{\leqq}$, definition (22) implies the matrix-product equation

$$\mathring{X} = \hat{X} H. \tag{24}$$

Now define the auxiliary variables $\mathring{b}(t^1) = \mathring{y}$ and $\mathring{b}(t^i) = b'(t^i)$ for all $i = 2, \ldots, n$. Defining the column vector $\mathring{b}^n = (\mathring{b}(t^1), \ldots, \mathring{b}(t^n))^T$, the definition of the $\omega_j$ variables implies that

$$\mathring{b}(t^{\sigma^{-1}(j)})\omega_j(1-\epsilon) + \mathring{b}(t^j)(1 - \sum_{k \in \sigma(j)} \omega_k)(1-\epsilon) = \hat{b}(t^j) \ \text{ for all } j \neq 1 \text{ with } t^j \in T^{\leqq},$$

and (21) implies

$$\mathring{b}(t^1)(1 - \sum_{k \in \sigma(1)} \omega_k)(1 - \epsilon) = \hat{b}(t^1).$$

In matrix notation,

$$H\mathring{b}^n = \hat{b}^n.$$

Together with (23) and (24) this implies

$$b'^n = \mathring{X}\mathring{b}^n.$$

That is, constraint (13) holds for $\mathring{y}, (\mathring{x}_{k,i})$.

That constraint (11) holds for $(\mathring{x}_{k,i})$ is seen by summing (22) across all $k = 1, \ldots, n+m$ and noting that (11) holds for $(\hat{x}_{k,i})$.

It remains to verify (16) for $(\mathring{x}_{k,i})$, that is, for all $i \leq n$ and all $j$,

$$\sum_{k=1}^{n+m} (\mathring{x}_{k,i} - \mathring{x}_{k,j}) \Pi(\rho'_k)(t^i) \geq 0. \tag{25}$$

Consider any $i$ with $t^i \notin T^{\leqq}$ and any $j > n$, or $j \leq n$ with $t^j \notin T^{\leqq}$. Then (25) is immediate because $\mathring{x}_{k,i} = \hat{x}_{k,i}$ and $\mathring{x}_{k,j} = \hat{x}_{k,j}$ and (16) holds for $(\hat{x}_{k,i})$.

Consider any $i$ with $t^i \notin T^{\leqq}$ and any $j$ with $t^j \in T^{\leqq}$. Then (25) follows from (22) because

$$\begin{aligned}
&\sum_{k=1}^{n+m}(\mathring{x}_{k,i}-\mathring{x}_{k,j})\Pi(\rho'_k)(t^i)\\
=&\sum_{k=1}^{n+m}(\hat{x}_{k,i}-\mathring{x}_{k,j})\Pi(\rho'_k)(t^i)\\
=&(1-\sum_{l\in\sigma(j)}\omega_l)(1-\epsilon)\sum_{k=1}^{n+m}(\hat{x}_{k,i}-\hat{x}_{k,j})\Pi(\rho'_k)(t^i)\\
&+\sum_{l\in\sigma(j)}\omega_l(1-\epsilon)\sum_{k=1}^{n+m}(\hat{x}_{k,i}-\hat{x}_{k,l})\Pi(\rho'_k)(t^i)+\epsilon\sum_{k=1}^{n+m}(\hat{x}_{k,i}-\hat{x}_{k,\iota(t^j)})\Pi(\rho'_k)(t^i)\\
\geq&0,
\end{aligned}$$

where the inequality follows because (16) holds for $(\hat{x}_{k,i})$.

Consider any $i$ with $t^i \in T^{\leq}_{=}$ and $j \leq n$ with $t^j \notin T^{\leq}_{=}$. By definition of the indifference tree, (16) holds as a strict inequality for $(\hat{x}_{k,i})$. Thus, assuming that $\epsilon$ is sufficiently close to 0, (25) holds.

Consider any $i$ with $t^i \in T^{\leq}_{=}$. By definition of the indifference tree,

$$\sum_{k=1}^{n+m}\mathring{x}_{k,i}\Pi(\rho'_k)(t^i)=(1-\epsilon)\sum_{k=1}^{n+m}\hat{x}_{k,i}\Pi(\rho'_k)(t^i)+\epsilon\ \Pi(\rho'_{\iota(t^i)})(t^i). \tag{26}$$

Because (16) holds for $(\hat{x}_{k,i})$ with $j=\iota(t^i)$, we conclude that

$$\sum_{k=1}^{n+m}\mathring{x}_{k,i}\Pi(\rho'_k)(t^i)\geq\Pi(\rho'_{\iota(t^i)})(t^i).$$

Thus, for any $j>n$, using the definition of $\iota(t^i)$,

$$\sum_{k=1}^{n+m}\mathring{x}_{k,i}\Pi(\rho'_k)(t^i)\geq\Pi(\rho'_{t^j})(t^i),$$

implying (25).

Finally, consider any $i$ with $t^i \in T^{\leq}_{=}$ and any $j$ with $t^j \in T^{\leq}_{=}$. Applying (26), and applying it again with $i$ replaced by $j$, we find

$$\begin{aligned}
&\sum_{k=1}^{n+m} (\mathring{x}_{k,i} - \mathring{x}_{k,j})\,\Pi(\rho'_k)(t^i) \\
=&(1-\epsilon) \sum_{k=1}^{n+m} (\hat{x}_{k,i} - \hat{x}_{k,j})\,\Pi(\rho'_k)(t^i) + \epsilon \left( \Pi(\rho'_{\iota(t^i)})(t^i) - \Pi(\rho'_{\iota(t^j)})(t^i) \right) \\
\geq&0,
\end{aligned}$$

where the inequality follows because both terms are $\geq 0$—the left term because (16) holds for $(\hat{x}_{k,i})$ and the right term by definition of $\iota(t^i)$.

In summary, we have shown that $\mathring{y}, (\mathring{x}_{k,i})$ satisfies all constraints of the max problem and $\mathring{y} > \hat{y}$, contradicting the fact that $\hat{y}, (\hat{x}_{k,i})$ is a solution. □

*Proof of Lemma 5.* Let $\rho$ denote an interim optimal allocation and suppose that $\rho$ is not strongly neologism-proof. Then there exists a belief $b'$ and a $b'$-feasible allocation $\rho'$ such that $b'$ puts zero probability on all types that are strictly better off in $\rho$ than in $\rho'$ or that already obtain in $\rho$ the maximum feasible payoff. Moreover, there exists a type $t' \in \text{supp}(b')$ such that $U_0^{\rho'}(t') > U_0^{\rho}(t')$.

Let $\rho'_{b'} \in A$ denote the $b'$-averaged sub-allocation.

Fix a belief $\hat{b}$ such that, for all $t \neq t'$, $\hat{b}(t) = \delta b'(t)$ and

$$\hat{b}(t') = 1 - \delta + \delta b'(t')),$$

where $\delta < 1$ is chosen sufficiently close to 1 such that $\text{supp}(\hat{b}) = \text{supp}(b')$.

By separability, there exists a sub-allocation $e$ such that all agents' incentive and participation constraints in the definition of $A$ are satisfied strictly. Let $\hat{\rho}$ denote the allocation such that $\hat{\rho}_t = \rho'_t$ for all $t \neq t'$, and

$$\hat{\rho}_{t'} = \frac{\delta b'(t')}{\hat{b}(t')} \rho'_{t'} + \frac{1-\delta}{\hat{b}(t')} e.$$

At the $\hat{b}$-average of the allocation $\hat{\rho}$, the agents expect to obtain the allocation

$\rho'_{b'}$ with probability $\delta$, and the allocation $e$ with probability $1-\delta$. Formally,

$$\begin{aligned}\sum_{t\in T}\hat{b}(t)\hat{\rho}_t &= \hat{b}(t')\hat{\rho}_{t'}+\sum_{t\neq t'}\hat{b}(t)\hat{\rho}_t\\ &= \delta b'(t')\rho'_{t'}+(1-\delta)e+\sum_{t\neq t'}\delta b'(t)\hat{\rho}_t\\ &= \delta\sum_{t\in T}b'(t)\rho'_t+(1-\delta)e\\ &= \delta\rho'_{b'}+(1-\delta)e.\end{aligned}$$

Thus, at $\hat{\rho}$ the defining inequalities of the set $A$ are satisfied strictly.

Moreover, $U_0^{\hat{\rho}}(t)=U_0^{\rho'}(t)\geq U_0^{\rho}(t)$ for all $t\in\operatorname{supp}(\hat{b})\setminus\{t'\}$, and $U_0^{\hat{\rho}}(t')>U_0^{\rho}(t')$ assuming $\delta$ is sufficiently close to 1.

Let $\overline{e}$ denote any allocation where each principal type obtains their maximum feasible payoff. Also recall that at $\hat{\rho}$, none of the types in $\operatorname{supp}(b')$ obtains their maximum feasible payoff.

Thus, for all $0<\epsilon<1$, the allocation $\rho''=\epsilon\overline{e}+(1-\epsilon)\hat{\rho}$ is such that $U_0^{\rho''}(t)>U_0^{\rho}(t)$ for all $t\in\operatorname{supp}(b')$.

Moreover, if $\epsilon$ is sufficiently close to 0 then at the $\hat{b}$-average $\rho''_{\hat{b}}$, the defining inequalities of the set $A$ are still satisfied strictly.

We can further change $\rho''$ to an allocation $\rho'''$ by giving to each principal type outside the $\operatorname{supp}(b')$ their respective best sub-allocation among the types in $\operatorname{supp}(b')$, or assign the sub-allocation $\zeta_0$ that always implements the disagreement outcome. That is, $\rho'''_t=\rho''_t$ for all $t\in\operatorname{supp}(b')$, and

$$\rho'''_t\in\arg\max_{\alpha\in\{\rho''_{\hat{t}}|\hat{t}\in\operatorname{supp}(b')\}\cup\{\zeta_0\}}\Pi(\alpha)(t)\quad\text{for all } t\in T\setminus\operatorname{supp}(b').$$

In summary, we have shown that the pair $(\hat{b},\rho''')$ belongs to the set

$$\begin{aligned}R \;=\; \{(\mathring{b},\mathring{\rho})\mid\; &\mathring{b}\in B\\ &\mathring{\rho}_{\mathring{b}}\in A\\ &U_0^{\mathring{\rho}}(t)>U_0^{\rho}(t)\;\;\text{for all } t\in\operatorname{supp}(\mathring{b}),\\ &\mathring{\rho}_t\in\arg\max_{\alpha\in\{\mathring{\rho}_{\hat{t}}|\hat{t}\in\operatorname{supp}(\mathring{b})\}\cup\{\zeta_0\}}\Pi(\alpha)(t)\text{ for all } t\in T\setminus\operatorname{supp}(\mathring{b}).\}\end{aligned}$$

In particular, the set $R$ is non-empty.

Consider $(\mathring{b},\mathring{\rho})\in R$ with minimal support size $|\operatorname{supp}(\mathring{b})|$. We claim that $\mathring{\rho}$ satisfies PIC.

Suppose otherwise. Then there exist types $t', t'' \in \text{supp}(\mathring{b})$ such that $\Pi(\mathring{\rho}_{t''})(t') < \Pi(\mathring{\rho}_{t'})(t')$.

Define a new belief $\check{b} \in B$ as follows: $\check{b}(t') = \mathring{b}(t') + \mathring{b}(t'')$, $\check{b}(t'') = 0$, and $\check{b}(t) = \mathring{b}(t)$ for all $t \in T \setminus \{t', t''\}$.

Define a new allocation $\check{\rho}$ as follows:

$$\check{\rho}_{t'} = \frac{\mathring{b}(t')}{\check{b}(t')}\mathring{\rho}_{t'} + \frac{\mathring{b}(t'')}{\check{b}(t')}\mathring{\rho}_{t''},$$

$\check{\rho}_t = \mathring{\rho}_t$ for all $t \in \text{supp}(\check{b}) \setminus \{t'\}$, and

$$\check{\rho}_t \in \arg \max_{\alpha \in \{\check{\rho}_{\hat{t}} | \hat{t} \in \text{supp}(\check{b})\} \cup \{\zeta_0\}} \Pi(\alpha)(t) \text{ for all } t \in T \setminus \text{supp}(\check{b}).$$

By construction $(\check{b}, \check{\rho}) \in R$ and $|\text{supp}(\check{b})| = |\text{supp}(\mathring{b})| - 1$, contradicting the assumed minimality.

Thus $\mathring{\rho}$ satisfies PIC. Note that PIR holds by construction.

We conclude that $\mathring{\rho}$ is $\mathring{b}$-feasible and $\text{supp}(\mathring{b}) \subseteq S(\mathring{\rho}, \rho)$, contradicting the assumption that $\rho$ is terim optimal. □

# References

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

Koessler, F., and V. Skreta (2023): "Informed information design," *Journal of Political Economy*, 131(11), 3186–3232.

Maskin, E., and J. Tirole (1990): "The principal-agent relationship with an informed principal: The case of private values," *Econometrica*, 58(2), 379–409.

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

——— (1985): "Analysis of Two Bargaining Problems with Incomplete Information," in *Game Theoretic Models of Bargaining*, ed. by A. Roth, pp. 59–69. Cambridge University Press.

Mylovanov, T., and T. Tröger (2012): "Informed-principal problems in environments with generalized private values," *Theoretical Economics*, 7(3), 465–488.

——— (2014): "Mechanism Design by an Informed Principal: Private Values with Transferable Utility," *Review of Economic Studies*, 81(4), 1668–1707.

——— (2026): "Neo-optimum: a unifying solution to the informed-principal problem," *Unpublished manuscript.*

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.

Wagner, C., T. Mylovanov, and T. Tröger (2015): "Informed-principal problem with moral hazard, risk neutrality, and no limited liability," *Journal of Economic Theory*, 159(PA), 280–289.

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.