

Discussion Paper Series – CRC TR 224

Discussion Paper No. 643

Project B 01

Neo-Optimum: A Unifying Solution to the Informed-Principal Problem

Tymofiy Mylovanov¹

Thomas Tröger²

January 2026(First Version : February 2025)

¹Department of Economics, University of Pittsburgh. Email: mylovanov@gmail.com.

²Department 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.

Neo-Optimum: A Unifying Solution to the Informed-Principal Problem*

Tymofiy Mylovanov[†] and Thomas Tröger[‡]

January 22, 2026

Abstract

A mechanism proposal by a privately informed principal is a signal. The agents' belief updating endogenizes their incentives in the mechanism, implying that such design problems cannot be solved via optimizing subject to incentive constraints. We propose a solution, neo-optimum, that can be interpreted as principal-preferred perfect-Bayesian equilibrium. Its neologism-based definition allows an intuitive computation, as we demonstrate in several applications. Any Myerson neutral optimum is a neo-optimum, implying that a neo-optimum exists generally. Neo-optimum unifies the other known solution approaches in the informed-principal literature.

1 Introduction

Mechanism design—the theory of designing rules of interaction that provide incentives to reveal private information towards maximizing the principal's

*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.

Compared to an earlier version, we have moved the development of the private-values results to a different paper. Here, we only report the main result.

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

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

goal—is a cornerstone of economics, with many applications ranging from procurement and the regulation of firms (Laffont and Tirole, 1993) to the design of public institutions (Laffont, 2000), from macroeconomics (Kocherlakota, 2006) to testing and social distancing in pandemics (Tröger, 2025).

The standard approach to solving a mechanism-design problem relies on the revelation principle, which implies that any mechanism induces a direct revelation mechanism, thus transforming the design problem into an optimization subject to incentive constraints. As first recognized by Myerson (1983) and Maskin and Tirole (1990, 1992), this approach fails if the principal is privately informed about her own goals or about anything that concerns the incentives of the agents who participate in the mechanism. The proposal of a mechanism then is a signal in a signaling game that leads to an endogenous updated belief about the principal’s private-information type. The incentives of the agents in the mechanism depend on the updated belief. Thus, the answer to the question *which* direct revelation mechanism is induced by a given mechanism proposal becomes endogenous.

An example of a mechanism designer is a seller who designs a profit-maximizing sales contract while being privately informed about the quality of her good. If, say, the contract includes a warranty then a buyer may be willing to pay a higher price not only because she gets the warranty, but also because she is triggered into believing in a higher quality of the good.

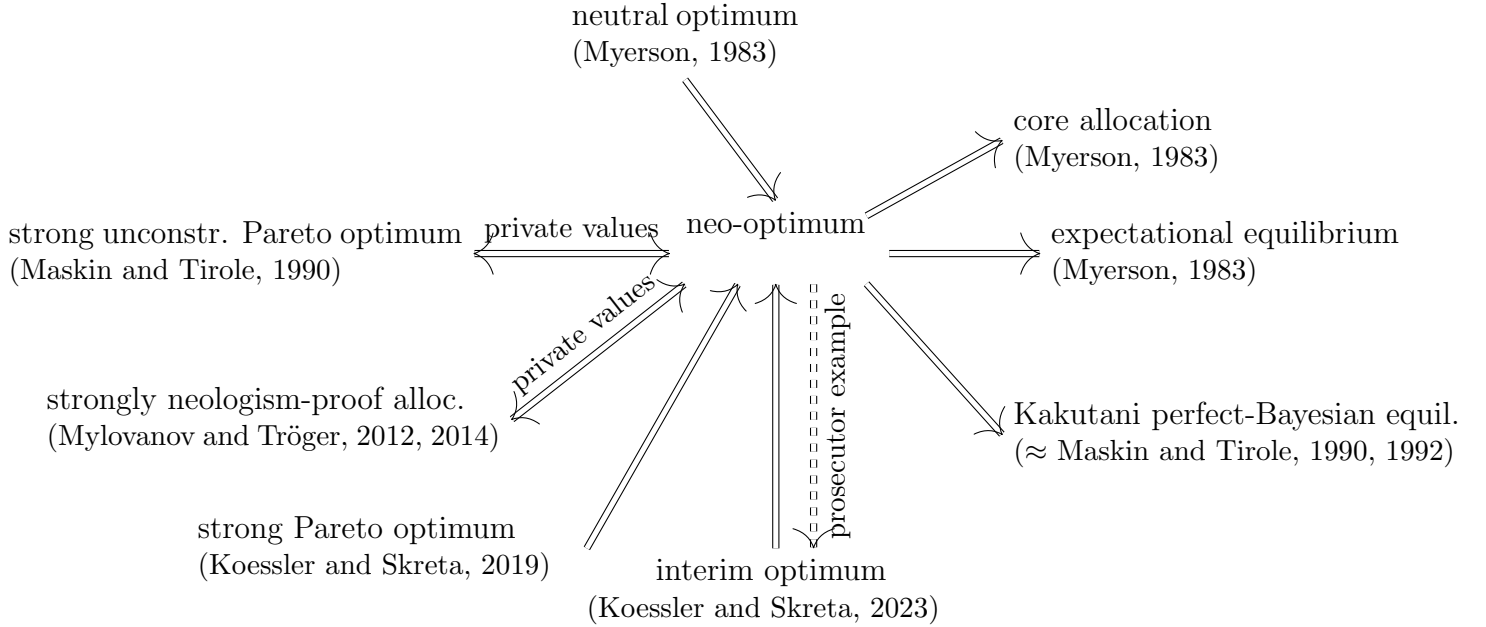
Modeling the principal as just a sender in a signaling game would, however, neglect the mechanism-design doctrine according to which the principal can select an equilibrium for the interaction. It is, for example, implausible that the principal offers a mechanism that, at the prior belief, yields a smaller payoff to all her private-information types than another mechanism. Yet, offering the low-payoff mechanism can be consistent with the logic of perfect-Bayesian equilibrium in signaling games because offering any alternative mechanism may trigger a “pessimistic” off-path belief that makes it unattractive.

The purpose of the paper is to provide a refinement of perfect-Bayesian equilibrium, neo-optimum, that is particularly suitable for informed-principal settings because it can be seen as “sender-preferred equilibrium”.

We show that neo-optimum exists broadly and connects the known solution approaches to informed-principal problems. Our existence result is not based on any specific assumptions concerning the payoff structure (such as, for instance, a single-crossing property), but follows from very general topological and convexity properties of feasibility sets. This is important.

In many informed-principal settings, due to the large signal space no single-crossing condition is satisfied.

The following diagram indicates how neo-optimum is connected via implication and equivalence relations to other solution concepts in the informed-principal literature.



Beyond finding neo-optima, our methods can be used to reconstruct all (including non-refined) perfect-Bayesian equilibria in virtually all existing papers in the informed-principal literature.

Neo-optima are always on the weak Pareto frontier of the set of perfect-Bayesian equilibria. Thus the predictions of neo-optimum are generally different from refinements such as the intuitive criterion (Cho and Kreps, 1987) that are consistent with Kohlberg-Mertens stability. But neo-optimum is also generally different from the principal's ex-ante-optimal perfect-Bayesian equilibrium. The ex-ante criterion has the obvious limitation that rescaling the payoff function of some type of the principal—although being strategically irrelevant—can change the prediction. Neo-optimum is invariant with respect to payoff rescalings. Nevertheless, it turns out that in many settings the neo-optima are also ex-ante optimal for the principal.

To capture the broadest range of settings, we introduce a reduced-form

description of informed-principal problems that focusses on the principal's payoff. A *payoff vector* refers to a payoff for each private-information type of the principal. For any (prior or updated) belief about the principal's type, a set of *feasible* payoff vectors is given. That's all.

In any particular application, the set of feasible payoff vectors at any belief will be determined by the details of the interaction. In a typical application, mechanisms are finite game forms with outcomes from a given set, and feasibility is defined via incentive constraints, as in Myerson (1983). For example, a seller may be able to choose from a set of sales contracts, and each continuation equilibrium in the interaction following the proposal of a sales contract induces a particular feasible payoff vector. The reduced form specifies the feasibility sets for *all* beliefs because, a priori, the principal's proposal may trigger any updated belief.

While a reduced-form description is possible for any signaling game,¹ the feasibility structure of informed-principal problems has a very useful additional property, *composition-closedness*, that allows us build our entire analysis around the feasible belief-payoff-vector pairs. Composition-closedness can be seen as a reduced-form version of the inscrutability principle in Myerson (1983). To define, consider a finite family of belief-payoff-vector pairs. Imagine that each type of principal makes a payoff-maximizing choice from this family. Suppose there exists a prior belief such that the belief component of each element of the family is consistent with Bayesian updating given the different types' choices. Then this prior belief together with the maximum-payoff vector resulting from the different principal types' choices is called a *composition*. Composition-closedness requires that any composition of feasible belief-payoff-vector pairs is feasible.² The justification for composition-closedness is that for any finite family of mechanisms there exists a "grand" mechanism in which the principal gives herself the option to select a mechanism from the family.

Both established approaches to defining perfect-Bayesian equilibria for

¹E.g., Mailath (1987).

²As for a concrete example, suppose the principal can have two types, called horizontal and vertical so that payoff vectors are represented in a two-dimensional coordinate system. Suppose the payoff vector $(1, 0)$ is feasible at the belief b^h that puts probability 1 on the horizontal type and the payoff vector $(0, 2)$ is feasible at the belief b^v that puts probability 1 on the vertical type. Then, for any belief b , the belief-pay-vector pair $(b, (1, 2))$ is a composition of $(b^h, (1, 0))$ and $(b^v, (0, 2))$. Thus, composition-closedness requires that the payoff vector $(1, 2)$ is feasible at all beliefs.

informed-principal signaling games, by Myerson (1983) and by Maskin and Tirole (1990, 1992), fit into our model. While the two approaches differ subtly with respect to the set of allowed mechanisms, in either approach, composition-closedness implies that focussing on fully pooling equilibria is without loss of generality, that is, all types of the principal propose the same mechanism on the equilibrium path.

Let us explain neo-optimum in more detail. A crucial ingredient is the concept of a neologism that is well-established in signaling games (Farrell, 1993). A neologism is a Bayes-consistent and feasible deviation relative to a given (not necessarily feasible) belief-payoff-vector pair. That is, the deviating payoff vector comes together with a deviating belief that puts probability 0 on types who would be harmed by the deviation, retains the relative likelihood across types that strictly gain, and can shift belief probability mass from indifferent types to strictly gaining types.

If no neologism exists for some payoff vector together with the prior belief, the payoff vector is *neologism-proof*. A payoff vector is a *neo-optimum* if (i) it is feasible at the prior belief and (ii) there exists a payoff vector *below* it that is a *limit* of neologism-proof payoff vectors.

Many established signaling-game refinements, including the intuitive criterion, rely on ideas related to neologisms. Farrell (1993) recognized that asking for a prior-belief feasible *and* neologism-proof payoff vector is generally too much; non-existence occurs in very simple signaling games. The subsequent literature on signaling games responded by arguing in favor of belief-payoff-vector deviations different from Farrell’s neologisms (e.g., Rabin (1990), Mailath, Okuno-Fujiwara, and Postlewaite (1993), Reny (2025), Zeng (2025)). Neo-optimum is fundamentally different: it insists on neologism-proofness as the relevant criterion, but disentangles it from feasibility. Rather, neo-optimum is content with being a limit of neologism-proof payoff vectors, or with being at least as good as such a limit for all types of the principal.

As a concrete example we consider a Spence (1973) job-market setting. It belongs to the class of informed-principal problems analyzed in Maskin and Tirole (1992). The principal is a worker who is privately informed about her productivity type, which can be high or low. To determine her task level and wage, she proposes a mechanism to an employer. One payoff vector that is feasible at all beliefs is the least-cost-separating one. It arises from both types getting a wage equal to their respective productivity types, the low-productivity type getting the lowest possible task level, and the high-productivity type getting the task level at which the low type is indifferent

between the two types' outcomes. Another prominent payoff vector, feasible at the prior belief, arises from both types getting the lowest task level and the wage equal to average productivity; we call this the best-pooling payoff vector. It is then easy to see that the unique neo-optimum is the least-cost-separating payoff vector or the best-pooling payoff vector, whichever is preferred by the high-productivity type.³

As for the perfect-Bayesian equilibria in the Spence example, we recall that the least-cost-separating payoff vector is a neo-optimum if the prior belief puts a sufficiently high weight on the low-productivity type. From this it is immediate that the least-cost-separating payoff vector, being feasible at all beliefs, is a perfect-Bayesian equilibrium at any interior prior belief. Generalizing this logic to other settings recovers the perfect-Bayesian equilibria constructed in Maskin and Tirole (1992) and a number of subsequent papers (e.g., Koessler and Skreta (2016); Nishimura (2022); Balzer (2017); Dosis (2022); Zhao (2023))

We show that any Myerson (1983) neutral optimum is a neo-optimum. From Myerson's existence result, this implies that a neo-optimum exists in any Bayesian incentive problem as defined by Myerson. Neo-optimum is stronger than Myerson's other solution concepts, core and expectational equilibrium. Whenever it yields a unique prediction it identifies the unique neutral optimum. But neo-optimum is much more intuitive and easier to handle than neutral optimum. In particular, neo-optimum avoids any reference to Myerson's extension axiom that connects solutions across different 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. 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

³Undefeated equilibrium (Mailath, Okuno-Fujiwara, and Postlewaite, 1993) and most persuasive equilibrium (Zeng, 2025) yield an analogous prediction in the classical job-market signaling game with two types where signals are not mechanisms but just education levels.

Mylovanov and Tröger (2012, 2014) have so far remained disconnected from Myerson’s (1983) approach.

In a companion paper, we show that, in the generalized-private-values settings for which Mylovanov and Tröger (2012) show the existence of a strongly neologism-proof allocation, this solution concept is in fact *equivalent* to neo-optimum. In particular, in these settings any neutral optimum is strongly neologism-proof. This result resolves a question that has remained open essentially since the literature started. Another implication is that in quasilinear private-values environments (Mylovanov and Tröger, 2014), any neo-optimum, and thus any neutral optimum, is ex-ante optimal.

Koessler and Skreta (2019) consider an informed principal who can, partially or fully, “certify” her type. The model fits into our framework. Koessler and Skreta (2019) propose a solution concept, strong Pareto optimum (SPO), and show that any prior-feasible SPO is a perfect-Bayesian equilibrium and is ex-ante optimal. But a prior-feasible SPO exists only in settings with a sufficiently rich certifiability structure. Using neo-optimum instead of SPO as a solution concept in their setting, existence is generally guaranteed and the qualitative results of Koessler and Skreta (2019) remain largely intact.

Settings in which the principal is an information designer also fit into our framework. To analyze informed-principal information design, Koessler and Skreta (2023) introduce a new refinement of perfect-Bayesian equilibrium, interim optimality, and provide existence and characterization results. We show that any interim optimum is a neo-optimum. In the different variants of the introductory prosecutor-judge example in Koessler and Skreta (2023) the reverse is also true, that is, any neo-optimum is interim-optimal. Thus, the arguments given by Koessler and Skreta (2023) in favor of interim-optimum as a solution concept—predictive power and robustness—apply similarly to neo-optimum.

Balkenborg and Makris (2015) consider a common-value setup similar to Maskin and Tirole (1992), but in contrast to the latter focus on an equilibrium refinement called *assured allocation* that in two-type settings is the unique Myerson (1983) core allocation and hence is equivalent to neo-optimum and neutral optimum. However, in some settings the assured allocation is dominated by a stochastic allocation and is not a core allocation, implying that it is not a neo-optimum.

Section 2 introduces the model, including the concept of Kakutani perfect-Bayesian equilibrium. Section 3 introduces the main concept, neo-optimum. At the end of this section, we also show how other perfect-Bayesian equi-

librium constructions in the informed-principal literature can be recovered using the concept of neo-optimum. In Section 4 we compare neo-optimum to the solution concepts in Myerson (1983). In Section 5, we discuss various applications and compare neo-optimum to other solution concepts. Some proofs are relegated to the appendix.

2 Model

2.1 Informed-Principal settings

The *principal* is a privately informed entity. Let T denote the set of the principal's feasible private-information types. For technical simplicity, we assume that T is finite.⁴ The set of probability distributions with support in T is denoted B . A *belief* about the principal's type is a $b \in B$, where $b(t)$ denotes the likelihood assigned to type t .

The principal's *payoff* is represented by a *vector* $U \in \mathbb{R}^T$, where $U(t)$ for all $t \in T$ denotes the principal's payoff when she has the type t .

An *informed-principal setting* is characterized by the *set of feasible belief-payoff-vector pairs*

$$K \subseteq B \times \mathbb{R}^T.$$

For any belief $b \in B$, we say that a payoff vector U is *feasible at b* or *b -feasible* if $(b, U) \in K$.⁵ We assume that K is topologically closed and, for each $b \in B$, the set of b -feasible payoff vectors is non-empty.

When convenient, we will use the following language. We say that a payoff vector V is *above* a payoff vector U (or U is *below* V), written $V \geq U$, if $V(t) \geq U(t)$ for all $t \in T$. In other words, V is above U if and only if all principal types weakly prefer V to U . (We say that V *dominates* U if V is above U and $V \neq U$.)

Applications

As a first example, following Maskin and Tirole (1992) and inspired by Spence (1973), let the principal be a worker who proposes a mechanism for deter-

⁴We see no obstacle against extending our central concept, neo-optimum, to continuous-type settings.

⁵Given any set $M \subseteq B \times \mathbb{R}^T$ and any $b \in B$, we will also use the notation $M(b) = \{U \mid (b, U) \in M\}$. For example, $K(b)$ is the set of b -feasible payoff vectors.

mining her wage w and task (or effort) level e to a potential employer. The principal has one of two productivity (or output) types, $T = \{\theta_L, \theta_H\}$. Any belief b can be identified with the probability of the high type θ_H , that is, $b \in B = [0, 1]$. Let $w - e/\theta$ denote the principal's payoff if she has the type θ , works at level e , and gets the wage w . This specification reflects that, when working at any task level e , if the worker's productivity is high then her cost of working is smaller than if her productivity is low. The employer then obtains the output θ and pays the wage w , so she obtains the payoff $\theta - w$ (in particular, the output is independent of the task level e). A mechanism is a game form in which the principal and the employer play, and each end node is a task-level-wage pair $(e, w) \in [0, \infty) \times [0, \theta_H]$. The mechanism is played if the employer accepts it; if she rejects, both the worker and the employer get the payoff 0. By the revelation principle, given any belief $b \in B$, the mechanism implements a task-level-wage pair⁶ (e_L, w_L) for type θ_L and a task-level-wage pair (e_H, w_H) for type θ_H such that incentive compatibility is satisfied,

$$w_L - e_L/\theta_L \geq w_H - e_H/\theta_L \quad \text{and} \quad w_H - e_H/\theta_H \geq w_L - e_L/\theta_H, \quad (1)$$

and the employer's participation constraint is satisfied,

$$b(\theta_H - w_H) + (1 - b)(\theta_L - w_L) \geq 0. \quad (2)$$

Thus,

$$K = \{(b, (w_L - \frac{e_L}{\theta_L}, w_H - \frac{e_H}{\theta_H})) \mid (1), (2)\}.$$

After standard manipulations, we obtain the following characterization,

$$\begin{aligned} K = \bigcup_{\bar{e} \geq 0} \bigcup_{b \in B} \{b\} \times \text{conv} \bigg\{ & (\theta_L, \theta_H - \frac{\theta_L(\theta_H - \theta_L)}{\theta_H}), \\ & (b\theta_H + (1 - b)\theta_L, b\theta_H + (1 - b)\theta_L), \\ & (0, 0), (-\frac{\bar{e}}{\theta_L}, -\frac{\bar{e}}{\theta_H}), (-\frac{\bar{e}}{\theta_L}, \theta_H - \theta_L - \frac{\bar{e}}{\theta_H}) \bigg\}, \end{aligned}$$

⁶Given the linearity of the payoff functions, probability distributions over task-level-wage pairs need not be considered. Also, without loss of generality the mechanism is never rejected.

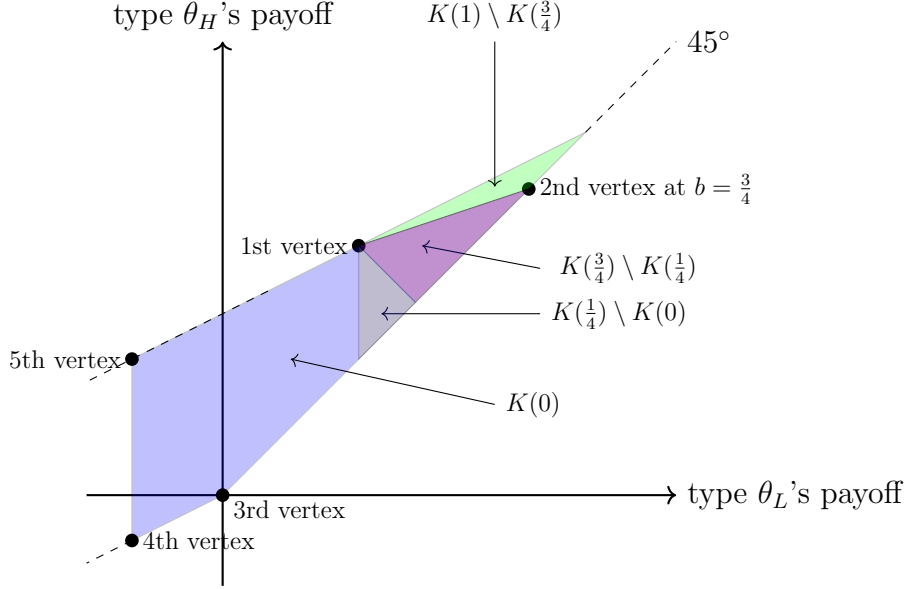


Figure 1: The sets of feasible payoff vectors at different beliefs b in the Spence-job-market example for generic values of θ_L and θ_H . The vertices refer to the points that span the convex hull for a particular value of \bar{e} in the definition of K . The higher b , the larger the feasibility set so that $K(0) \subset K(1/4) \subset K(3/4) \subset K(1)$.

where $\text{conv}\{\dots\}$ is the convex hull of five “vertex” payoff vectors: the first is called least-cost separating, the second arises from both types pooling at task level 0 and getting a wage equal to the expected productivity, the third arises from both types pooling at task level 0 and getting the wage 0, the fourth is the payoff vector where both types choose a task level \bar{e} and get the wage 0, and the fifth the payoff vector where the low type chooses a task level \bar{e} and gets the wage 0 while the high type chooses the task level such that the low type is made indifferent to choosing the high type’s task level and getting the wage θ_H . The union of the convex hulls across all $\bar{e} \geq 0$ captures what is feasible at the belief b .

The projections of K onto the space of payoff vectors for several beliefs b are represented graphically in Figure 1.

Secondly, going beyond a particular example, we would like to emphasize Myerson’s (1983) “Bayesian incentive problems” as a broad framework that is covered by our model. The framework captures mechanism design problems with adverse selection and moral hazard, finitely many agents, arbitrary

finite outcome spaces and arbitrary payoff functions.⁷ A feasibility set K is naturally associated to any Bayesian incentive problem Γ , as follows. For any belief $b \in B$, the set $K(b)$ is determined via the incentive constraints in Γ when b is interpreted as a normalized likelihood vector in the sense of Myerson (1983, Section 5), and in the incentive constraints (1983, (3.1)) of all players except the principal (i.e., player 1) there is an additional factor $b(t_1)$, where t_1 is the first component of the summation variable t_{-i} . (An “incentive compatible mechanism” in the terminology of Myerson (1983) leads to a payoff vector that is b -feasible with b the uniform distribution on T .)

As a third application, our model covers the information-design framework of Bergemann and Morris (2019). Suppose the information designer is one of the players, say player 1 (i.e., $v = u_1$ in the terminology of Bergemann and Morris). As in the previous example, any belief b is interpreted as a normalized likelihood vector in the sense of Myerson (1983, Section 5). For any belief $b \in B$, the set $K(b)$ is determined via the obedience constraints (Definition 1 in Bergemann and Morris (2019)), augmented with an additional factor $b(t_1)$ for all $i \neq 1$.⁸ A special case of this framework is Koessler and Skreta (2023), who assume that the principal is fully informed about the state of the world and the other players have no prior information.⁹

Compositions and inscrutibility

Next we introduce our central structural property for feasibility sets, Assumption 1. This is motivated by a reduced-form generalization of Myerson’s (1983) inscrutability principle. The fundamental concept, *composition*, is best understood via a thought experiment. Suppose that the principal had to choose a belief-payoff-vector pair from a given family. Depending on her type, she cares about the corresponding dimension in each payoff vector. A *composition* is a belief-payoff-vector pair that arises from a payoff-maximizing choice of each type, and the choice decisions of the various principal types are Bayes-consistent.

⁷The Spence example with the particular cost structure described above becomes a Bayesian incentive problem if we cut the outcome space at some highest feasible task level. Then any outcome can be represented as a probability distribution over four outcomes, combined of the highest/lowest task/wage levels.

⁸Correcting a typo in Bergemann and Morris (2019), the summation variables in their formula (2) should be a_{-i} and t_{-i} , where the latter has t_1 as its first coordinate.

⁹Koessler and Skreta (2023) sketch a more general model in Section 8 of their paper.

A belief-payoff-vector pair $(b, U) \in B \times \mathbb{R}^T$ is the

$$\text{composition of } (b_k, U_k)_{k=1, \dots, \bar{k}} \quad (\bar{k} \geq 1, (b_k, U_k) \in B \times \mathbb{R}^T)$$

if, for all $t \in T$,

$$U(t) = \max_{k=1, \dots, \bar{k}} U_k(t) \quad (3)$$

and if there exist (choice-probability) functions $c_k : T \rightarrow [0, 1]$ ($k = 1, \dots, \bar{k}$) with $\sum_{k=1}^{\bar{k}} c_k(t) = 1$ ($t \in T$) such that

$$\text{if } c_k(t) > 0, \text{ then } U_k(t) \geq U_l(t) \text{ for all } l \neq k, \quad (4)$$

and such that, for all k , Bayes' rule is satisfied,

$$c_k(t'')b(t'')b_k(t') = c_k(t')b(t')b_k(t'') \quad \text{for all } t', t'' \in T. \quad (5)$$

To digest the notation, suppose that each type t chooses among the payoff vectors $U_1, \dots, U_{\bar{k}}$, where type t 's probability of choosing any U_k is denoted $c_k(t)$. Then (3) expresses that U is the payoff vector that results from optimal choice, (4) expresses that each type chooses optimally, and (5) expresses that the belief b_k that is formed upon observing the choice k is consistent with the grand belief b and the choice probability distributions c_k . The fact that (5) captures Bayes rule is easiest to see if all numbers involved are strictly positive, implying

$$\frac{b_k(t')}{b_k(t'')} = \frac{c_k(t')b(t')}{c_k(t'')b(t'')}. \quad (6)$$

The right-hand side in this equation expresses the relative probability of entering the interaction and choosing k , across types t' and t'' , taking the initial belief b into account; the equality with the left-hand side expresses that the belief that is formed upon observing k is consistent with the actual relative choice probabilities.

A special case covered by (5) is that some k ends up being never chosen. Formally, this case occurs if $c_k(t)b(t) = 0$ for all t . For such k , the conditions (5) are void because both sides are equal to 0. In all other k , we can find a type t'' such that $c_k(t'')b(t'') \neq 0$. Then the conditions (5) imply that $b_k(t') = 0$ for any t' with $b(t')c_k(t') = 0$ and $b_k(t') > 0$ for any t' with

$b(t')c_k(t') > 0$. Thus, in (6) either the numerators on both sides are strictly positive or the numerators on both sides are equal to zero, and the same is true for the denominators.

For any finite family of belief-payoff-vector pairs in K , let \overline{K} denote the set of compositions of elements of K . Throughout the paper, we will maintain the following assumption.

Assumption 1. *The feasibility set K is composition-closed, that is, $\overline{K} = K$.*

In all applications that we have described, the feasibility set K is composition-closed. In the setting of Myerson (1983), the argument leading to composition-closedness is called the inscrutability principle. Intuitively, the argument is as follows. For each k , by definition of the feasibility set, there exists a direct mechanism M_k that induces the payoff vector U_k at the belief b_k . Now consider the indirect mechanism \hat{M} that gives the principal the option to select any of the direct mechanisms $M_1, \dots, M_{\overline{k}}$ for play. The indirect mechanism then has an equilibrium in which any type t selects any M_k with probability $c_k(t)$, and in M_k the truth-telling and obedient equilibrium is played because the belief b_k prevails at the start of M_k .

2.2 Perfect-Bayesian equilibrium

In the previous section, we posited that the principal chooses among belief-payoff-vector pairs. That was a preparatory step towards conceptualizing the idea that the principal is a sender in a signaling game where the signals are mechanisms.

Which mechanisms should be allowed as signals for the principal in the mechanism-selection game? The standard literature on mechanism design where the principal has no private information evokes the revelation principle and so justifies the focus on direct revelation mechanisms. But now, for any belief b , any continuation equilibrium in any mechanism corresponds to a different direct mechanism, and the belief b is endogenous. The role of a mechanism as a signal generally depends on all its continuation equilibria for all possible beliefs. Here it can matter which equilibrium concept is used for continuation equilibria; this is particularly relevant if sequential mechanisms are allowed. Also, should one allow mechanisms such that a continuation equilibrium exists for some beliefs about the principal and not for others? It is not obvious how the set of possible mechanisms and continuation equilibria can be specified such that it does not appear restrictive and still a perfect

Bayesian equilibrium exists in the signaling game where the principal is the sender and signals are mechanisms.

In the literature, there are two different approaches to specifying the principal’s set of mechanisms. First, Myerson (1983), given any Bayesian incentive problem, defines “generalized mechanisms” that allow for arbitrary finite message spaces for all players. The mechanism may reveal information about these messages privately to each player, to influence her private action, and some public outcome is implemented. The equilibrium concept for continuation equilibria in generalized mechanisms is Nash equilibrium (in particular, generalized mechanisms are treated as simultaneous-move game forms).

The second approach, which can be traced back to Maskin and Tirole (1990, 1992), is to avoid the explicit specification of the set of feasible mechanisms, and instead restrict the continuation-equilibrium payoff properties of mechanisms. Given any (in whatever way specified) mechanism μ , we define a set of belief-payoff-vector pairs M^μ , as follows: $(b, U) \in M^\mu$ if and only if the payoff vector U can be induced by (in whatever way specified) continuation-equilibrium play of μ at the belief b . A set of belief-payoff-vectors $M \subseteq B \times \mathbb{R}^T$ is called a *Kakutani set*¹⁰ if, for all $b \in B$, the set of payoff vectors $M(b) = \{U \mid (b, U) \in M\}$ is non-empty and convex, and the set M is compact (hence, the correspondence $b \mapsto M(b)$ is upper hemicontinuous). Rather than explicitly describing which μ ’s are feasible, Maskin and Tirole (1990, 1992) assume that only such μ are feasible where M^μ is a Kakutani set.

Myerson’s (1983) approach allows some mechanisms that are not allowed by Maskin and Tirole (1990) because, for some generalized mechanisms M and beliefs b , the set $M(b)$ is a non-singleton set of isolated points, hence non-convex. Maskin and Tirole’s (1990, 1992) approach, on the other hand, is not restricted to Myerson’s Bayesian incentive problems, and mechanisms outside the set of Myerson’s “generalized mechanisms” can be allowed.

According to both approaches, in a perfect-Bayesian equilibrium (or, “expectational equilibrium”, in Myerson’s framework) of the mechanism-selection game, we can assume without loss of generality from the point of view of the principal’s equilibrium payoff vector that all types of the principal pool at the same mechanism.

The possibility of pooling follows from composition-closedness. To see

¹⁰The terminology is adapted from Pęski (2022).

this, let b^* denote the (interior) prior belief about the principal.¹¹ Let $\mu_1, \dots, \mu_{\bar{k}}$ denote the list of mechanisms that are chosen with positive probability by at least one type of principal. In a perfect-Bayesian equilibrium, each type of principal chooses an optimal mechanism from the list, and the beliefs b_k are consistent with these choices in the sense of Bayes' rule. For any k , let U_k denote the payoff vector that is induced by the continuation-equilibrium play of μ_k . Let $c_k(t)$ denote the probability that any type t chooses the mechanism μ_k . By composition-closedness (with $b = b^*$), the payoff vector $U = \max_k U_k$ is b^* -feasible. Thus, without loss of generality all types pool at a mechanism that induces U at the belief b^* .

According to Myerson (1983), given any Bayesian incentive problem, in an *expectational equilibrium* all types of the principal pool at an incentive-compatible direct revelation mechanism such that for any generalized mechanism μ' there exists a belief b' and a continuation equilibrium in μ' at belief b' such that no type of principal gains from deviating to μ' .

The following alternative concept follows the spirit of Maskin and Tirole (1990, 1992). Given any informed-principal setting K and any interior prior belief b^* , a payoff vector U is a *Kakutani perfect-Bayesian equilibrium* if U is b^* -feasible, and for any Kakutani set $M' \subseteq K$ there exists $(b', U') \in M'$ such that U' is below U .

Note that the definition of Kakutani perfect-Bayesian equilibrium considers each Kakutani set to be a feasible signal. We cannot exclude the possibility that the definition is unreasonably restrictive in some specific settings, but we are not aware of any such setting. Also, this possibility has no bearing on our main concept, neo-optimum, which is not based on Kakutani sets.

The following observation is often useful for bounding the set of Kakutani perfect-Bayesian equilibria from below; the lemma's conclusion is immediate from the assumption because $M' = B \times \{\underline{U}\}$ is a Kakutani set.

Lemma 1. *If there exists a payoff vector \underline{U} that is feasible at all beliefs, then \underline{U} is below all Kakutani Perfect-Bayesian equilibria.*

To illustrate the concept of Kakutani perfect-Bayesian equilibrium, consider again our Spence-job-market example. The least-cost separating payoff

¹¹Myerson does not need to specify a prior belief as it is implicitly built into the definition of a Bayesian incentive problem. Translated to our terminology, Myerson's prior is the uniform distribution on T .

vector, $U^{lcs} = (\theta_L, \theta_H - \theta_L(\theta_H - \theta_L)/\theta_H)$, illustrated as the 1st vertex in Figure 1, is b^* -feasible for any b^* . Any Kakutani set M leaves a trace $\cup_{b \in B} M(b)$ in the space of payoff vectors. By definition, the trace includes a 0-feasible payoff vector U' , and from our earlier characterization of the 0-feasibility set in the Spence example it follows that $U'(t) \leq U^{lcs}(t)$ for all $t \in T$. Thus, using the “pessimistic” belief $b' = 0$ we see that U^{lcs} is a Kakutani perfect-Bayesian equilibrium. On the other hand, because U^{lcs} is feasible at all beliefs, Lemma 1 implies that any Kakutani perfect-Bayesian equilibrium U is above U^{lcs} . It follows that the set of Kakutani perfect-Bayesian equilibria is

$$\{U \mid U \text{ is } b^*\text{-feasible, } \forall t \in T : U(t) \geq U^{lcs}(t)\}.$$

For later use, we note the following.

Lemma 2. *In any informed-principal setting with any interior prior, the set of Kakutani Perfect-Bayesian equilibria is closed.*

Proof. Denote the feasibility set by K and the prior by b^* . Consider a sequence of equilibria $(U_n) \rightarrow U^*$. Then U_n is b^* -feasible for all n . Because the set of b^* -feasible payoff vectors, $K(b^*)$, is closed, it follows that U^* is b^* -feasible.

Consider any Kakutani set $M' \subseteq K$. Using the equilibrium property along the sequence, there exists a sequence $(b'_n, U'_n) \in M'$ such that $U'_n(t) \leq U_n(t)$ for all n and t . Because M' is compact, there exists $(b', U') \in M'$ such that $(b'_n, U'_n) \rightarrow (b', U')$ along some subsequence. Thus, $U'(t) \leq U^*(t)$ for all t , proving that U^* is a Kakutani Perfect-Bayesian equilibrium. \square

3 Neo-optimum

In many informed-principal settings (such as the Spence-job-market example above with b^* being sufficiently close to 1), multiple Kakutani perfect-Bayesian equilibria exist. Similarly, Myerson (1983) observes that multiple expectational equilibria exist in many Bayesian incentive problems.

The main goal of our paper is to show how to select perfect-Bayesian equilibria that are “sender preferred” in an intuitive sense. Importantly, our refinement will be invariant to scaling the utility of each sender type. Thus, the refinement is a-priori unrelated to ex-ante optimality for the sender. Our refinement provides a unified perspective of the informed-principal literature, and opens the door to solving new problems.

We start from a concept inspired by Farrell (1993). Consider a feasibility set K , an interior belief b , and a payoff-vector $U \in \mathbb{R}^{|T|}$. A belief-payoff-vector pair $(\hat{b}, \hat{U}) \in K$ is a *neologism* for (b, U) ¹² if $\hat{U}(\tilde{t}) > U(\tilde{t})$ for some $\tilde{t} \in T$, and the following conditions hold for all $t \in T$:

$$\text{if } \hat{U}(t) > U(t) \text{ then } \hat{b}(t)b(t') \geq \hat{b}(t')b(t) \text{ for all } t' \in T, \quad (7)$$

$$\text{if } \hat{U}(t) < U(t) \text{ then } \hat{b}(t) = 0. \quad (8)$$

Intuitively, a neologism that can be seen as a Bayes-consistent and feasible deviation relative to a given (not necessarily feasible) belief-payoff-vector pair. The deviating payoff vector \hat{U} comes together with a belief that puts probability 0 on types who would be harmed by the deviation (see (8)), retains the relative likelihood across types that strictly gain (see (7) with switched roles of t and t' , yielding $b_1(t)b(t') = b_1(t')b(t)$), and can shift belief probability mass from indifferent types to strictly gaining types (see (7) with t' such that $\hat{U}(t') = U(t')$).

A payoff vector U is *b-neologism-proof* if no neologism exists for (b, U) . As implicitly suggested by the principal's "speeches" proposed in Myerson (1983), an ideal solution for the principal would be a payoff vector that is feasible at the prior belief b^* , and is b^* -neologism-proof. In many settings, however, such a payoff vector does not exist. Below we will review this well-known fact in our Spence example; see Farrell (1993) for a different example in an elementary signaling game.

In the spirit of the literature following Farrell (1993), one may respond to the non-existence problem by restricting the set of neologisms that are considered legitimate. We follow an alternative approach: we select the b^* -feasible payoff vectors that are *above limits of* b^* -neologism-proof payoff vectors. Note that the selected payoff vector itself may not be b^* -neologism-proof. Here is the definition.

Given any interior prior belief b^* , a payoff vector U is a *b^* -neo-optimum* if U is b^* -feasible and there exists a payoff vector $V \leq U$ such that V is a limit of b^* -neologism-proof payoff vectors.

The justification of neo-optimum is indirect. If we pose this in terms of a speech of the principal, she would say: if some types of myself did renounce part of their payoff (so that the payoff vector V remains), then no prior-belief-neologism would break arbitrarily close approximations of this payoff

¹²Sometimes we say instead that (\hat{b}, \hat{U}) is a neologism for U at b .

vector; thus, my actual payoff vector U should also be considered a justified selection.

Our first remark is that all neo-optima lie on the principal's weak Pareto frontier. This is immediate from the definition of neo-optimum.

Remark 1. *Let b^* denote an interior belief. Consider a payoff vector U such that $U(t) < \hat{U}(t)$ for all $t \in T$ for some b^* -feasible \hat{U} . Then U is not a b^* -neo-optimum.*

Another important remark is that the notion of neo-optimum is independent of each principal type's utility scale: if a positive affine transformation is applied to some type's utility, then the set of neo-optima remains unchanged. This reveals a fundamental difference to the notion of the principal's *ex-ante optimum*, which by definition is any payoff vector that maximizes $\sum_t b^*(t)U(t)$ among all b^* -feasible payoff vectors U .

In the rest of the paper, we will show that neo-optima are refinements of perfect-Bayesian equilibria, exist broadly, often lead to a unique prediction, and provide a unification of the various solution concepts that have been proposed in the informed-principal literature.

In a given application, and given any prior belief b^* , one can find the neo-optima via the following intuitive steps. First, characterize the set $P(b^*) \subseteq \mathbb{R}^T$ of b^* -neologism-proof payoff vectors (independently of their feasibility) by checking any possible neologism at b^* . Second, form the topological closure of $P(b^*)$ and include all vectors above elements of the closed set. Third, intersect the resulting set with the set of b^* -feasible payoff vectors to obtain the set of neo-optima.

Examples

Consider the Spence setting with a prior belief $b^* > (\theta_H - \theta_L)/\theta_H$. This inequality guarantees that the high type strictly prefers the best pooling equilibrium (which is represented as the 2nd vertex in Figure 1 at $b = b^*$),

$$U^{pool*} = (b^*\theta_H + (1 - b^*)\theta_L, b^*\theta_H + (1 - b^*)\theta_L),$$

over the least-cost separating payoff vector, U^{lcs} . Note that in this case there is a multiplicity of Kakutani perfect-Bayesian equilibria. From Remark 1 it is immediate that U^{pool*} is the unique neo-optimum candidate. In order to

prove that U^{pool*} is a neo-optimum, we compute the set $P(b^*)$ of b^* -neologism-proof payoff vectors. First, we show that only payoff vectors weakly above the least-cost separating one can be neologism-proof:

$$P(b^*) \subseteq \{U \mid U \geq U^{lcs}\}.$$

To see this, note that $(\hat{b}, U^{lcs}) \in K$ for all $\hat{b} \in B$. Thus, for any U such that $U(\theta_L) < U^{lcs}(\theta_L)$ and $U(\theta_H) \geq U^{lcs}(\theta_H)$, the belief-payoff vector pair $(0, U^{lcs})$ is a neologism for (b^*, U) . Vice versa, for any U such that $U(\theta_L) \geq U^{lcs}(\theta_L)$ and $U(\theta_H) < U^{lcs}(\theta_H)$, the belief-payoff vector pair $(1, U^{lcs})$ is a neologism for (b^*, U) . Lastly, for any U such that $U(\theta_L) < U^{lcs}(\theta_L)$ and $U(\theta_H) < U^{lcs}(\theta_H)$, the belief-payoff vector pair (b^*, U^{lcs}) is a neologism for (b^*, U) .

Second, no payoff vector that is dominated by the best pooling one can be neologism-proof.

$$P(b^*) \cap \{U \mid U \leq U^{pool*}, U \neq U^{pool*}\} = \emptyset.$$

This follows from using (b^*, U^{pool*}) as a neologism.

Third, consider the segment between the least-cost separating payoff vector and (θ_H, θ_H) , the best pooling payoff vector at the belief 1. We define the “lower right” of this segment as the set of payoff vectors U such that $U(\theta_L) \geq \hat{U}(\theta_L)$ and $U(\theta_H) < \hat{U}(\theta_H)$ for some \hat{U} in the segment. Given any such U , the pair $(1, \hat{U})$ is a neologism for (b^*, U) . Thus,

$$P(b^*) \cap (\text{lower right of segment between } U^{lcs} \text{ and } (\theta_H, \theta_H)) = \emptyset.$$

The three restrictions that we have described characterize $P(b^*)$; the set is sketched in Figure 2.

It is apparent that no b^* -feasible and b^* -neologism payoff vector exists. It is also apparent that U^{pool*} is the unique payoff vector that is above a point in the topological closure of $P(b^*)$. Thus, U^{pool*} is the unique neo-optimum.

In the case $b^* = (\theta_H - \theta_L)/\theta_H$, the set of neo-optima equals the segment spanned by U^{lcs} and U^{pool*} .

In cases with a prior belief $b^* < (\theta_H - \theta_L)/\theta_H$, the payoff vector U^{lcs} is the unique Kakutani perfect-Bayesian equilibrium and thus is the unique neo-optimum; this is implied by Proposition 1 below.

Note that neo-optimum in general differs from ex-ante optimum. If the prior belief b^* is slightly below $(\theta_H - \theta_L)/\theta_H$, then the payoff vector U^{lcs} is

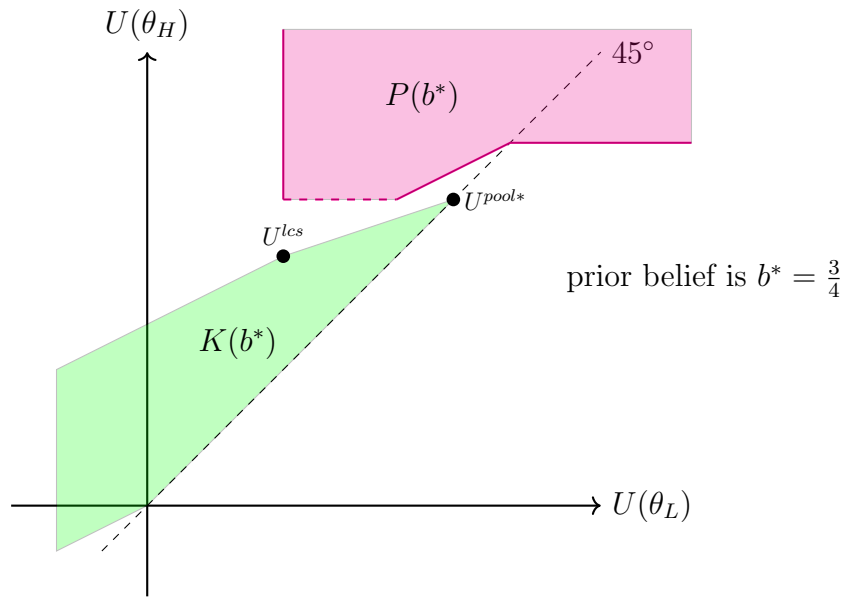


Figure 2: In the Spence-job-market example, at the prior belief $b^* = 3/4$, the set of feasible payoff vectors $K(b^*)$ has an empty intersection with $P(b^*)$, the set of payoff vectors U such that (b^*, U) is neologism-proof (the set extends infinitely to the upper right). The unique neo-optimum is U^{pool*} .

the unique neo-optimum and the payoff vector U^{pool*} is the unique ex-ante optimum.¹³

A class of settings in which neo-optimum is trivially unique and identical to ex-ante optimum is characterized by the property that $U(t) = U(t')$ for all $(b, U) \in K$ and all $t, t' \in T$. Such settings are considered in Koessler and Skreta (2016):¹⁴ different types of the principal have identical goals and represent different information about the agents' payoffs.

In Section 5 we present several other cases of mechanism-design by a privately informed seller. These examples reveal some aspects that are not present in our Spence example. In particular, one example features b -feasibility sets that are not nested across different beliefs b , and the other example has multiple Kakutani perfect-Bayesian equilibria such that different types of the principal prefer different equilibria, yet there is a unique neo-optimum.

Our next result shows that neo-optimum is a refinement of perfect-Bayesian equilibrium.

Proposition 1. *Let b^* denote an interior belief. Any b^* -neo-optimum is a Kakutani Perfect-Bayesian equilibrium.*

Towards proving this, the main technical hurdle is Lemma 3 below. (This is also the essential step towards establishing many of the equilibria in the informed-principal literature.) It says that from any given finite list of Kakutani sets belief-payoff-vector pairs can be selected to form a composition that entails any given belief.

Lemma 3. *Let \mathcal{P} be a finite set of Kakutani sets. Let $b \in B$.*

Then there exists a composition $(b, U^) \in B \times \mathbb{R}^T$ of some $(b_P^*, U_P^*)_{P \in \mathcal{P}}$, where $(b_P^*, U_P^*) \in P$ for all $P \in \mathcal{P}$.*

The proof works as follows. For each $n = 1, 2, \dots$, we define a function Ψ_n that continuously maps each list of payoff vectors that can occur in the

¹³Recall that we have defined ex-ante optimum via maximizing the principal's ex-ante expected payoff across all b^* -feasible payoff vectors. A generally different point is obtained if the maximization is restricted to the Kakutani perfect-Bayesian equilibria. This "ex-ante-optimal equilibrium" is identical to neo-optimum in the Spence example for all $b^* \neq (\theta_H - \theta_L)/\theta_H$. For an example where ex-ante-optimal equilibrium and neo-optimum are different for an open set of prior beliefs, consider the discussion at the end of the bilateral-trade example in Section 5.

¹⁴See also Izmalkov and Balestrieri (2012).

signals in \mathcal{P} to a vector of smoothed optimal choice probabilities, that is, every type chooses every $P \in \mathcal{P}$ with a positive probability, and as n tends to infinity, almost all weight is put on utility-maximizing P s. Requiring strictly positive choice probabilities is a form of trembling which guarantees that by Bayesian updating from b a unique belief about the sender's type is assigned to each signal.

We have another correspondence given by the signals themselves. Suppose a belief is given for each signal in \mathcal{P} . We consider the correspondence that assigns to any such list of beliefs a set of lists of payoff vectors by applying the signals to the beliefs. Combining this correspondence with Ψ_n , we obtain a correspondence that has a fixed point by Kakutani's Theorem.¹⁵

A fixed point consists of, for each signal, a belief and a payoff vector that belongs to the signal at this belief such that the beliefs are consistent with the sender's smoothed optimal choice.

By taking n to infinity we consider a sequence of fixed points with trembling probabilities tending to 0. By choosing an appropriate subsequence, we can guarantee that the sequence converges. In the limit, there is no trembling restriction so that the beliefs are fully consistent with the sender's optimal choice among the signals. Because the signals are compact, each limit payoff vector belongs to the respective signal at the limit belief. By construction, the maximum of the limit payoff vectors together with belief b is the composition of the list of limit payoff vectors together with the limit beliefs. The details of the proof are in the Appendix.

Proof of Proposition 1. Consider a b^* -feasible and b^* -neologism-proof payoff vector U . It is sufficient to show that U is a Kakutani perfect-Bayesian equilibrium. By Lemma 2, the conclusion then extends to limit points. By definition of equilibrium, it then extends to points above.

Consider any Kakutani set M in K . By Lemma 3, there exists a composition (b^*, \cdot) of some $(b_1, V) \in M$ and some element of $B \times \{U\}$.

In particular, $(b_1, V) \in K$. Let $c_1(t)$ denote the choice probabilities of M in the composition.

Suppose that $V \not\leq U$. For any $t \in T$ with $V(t) > U(t)$, we have $c_1(t) = 1$ by definition of a composition. Applying (5) with $t'' = t$, (7) follows.

¹⁵The detour through introducing trembles is needed to guarantee that Ψ_n is single-valued, and thus the combined correspondence is convex-valued, a prerequisite for Kakutani's Theorem.

By assumption, $V(\hat{t}) > U(\hat{t})$ for some $\hat{t} \in T = \text{supp}(b^*)$, implying $c_1(\hat{t}) = 1$.

Now consider any $t \in T$ with $V(t) < U(t)$. Then $c_1(t) = 0$. Because b is interior, $b(\hat{t}) > 0$. Applying (5) with $t' = t$ and $t'' = \hat{t}$ now yields (8).

Thus, (b_1, V) is a neologism for (b^*, U) . But this contradicts the assumption on U . Hence, $V \leq U$, as was to be shown. \square

As an immediate corollary to Proposition 1, we have a way of establishing other (and, in many settings, all) Kakutani perfect-Bayesian equilibria. Virtually all known equilibria in the informed-principal literature can be reconstructed in this way.

Corollary 1. *Consider any interior prior belief b^* and another interior belief b' . Any payoff vector that is b^* -feasible and is a b' -neo-optimum is a Kakutani perfect-Bayesian equilibrium at the prior belief b^* .*

Note that the result applies to the equilibria in Maskin and Tirole (1992) (see also the correcting formulation in Dosis (2022)) and the literature building on it: if the “Rothschild-Stiglitz-Wilson allocation” (which is feasible at all beliefs) is undominated for some interior belief b' , then its induced payoff vector is b' -neologism-proof and hence is a b' -neo-optimum, implying by Corollary 1 that it is a Kakutani perfect-Bayesian equilibrium at any interior prior belief b^* .

4 Neo-optimum versus Myerson’s solution concepts in Bayesian incentive problems

In this section, we use the simpler term neo-optimum instead of b^* -neo-optimum because we follow Myerson’s formulation of Bayesian incentive problems in which the prior belief b^* is the uniform distribution on T . We show that a neo-optimum exists in any Bayesian incentive problem and explain its connection to Myerson’s solution concepts neutral optimum, expectational equilibrium, and core.

Given any Bayesian incentive problem, Myerson (1983) identifies a set of payoff vectors that he calls *neutral optima*. He argues that neutral optima represent a “fair” compromise across all types of the principal. By verifying Myerson’s axioms in the proof below, we obtain the following result.

Proposition 2. *Consider any Bayesian incentive problem. Any neutral optimum is a neo-optimum.*

An immediate conclusion is that in the Spence example above with a highest feasible task level, the neutral optimum is generically unique and identical to neo-optimum for all prior beliefs $b^* \neq (\theta_H - \theta_L)/\theta_H$.

Myerson shows that a neutral optimum always exists. Thus, we have the following.

Corollary 2. *In any Bayesian incentive problem, a neo-optimum exists.*

This is a very broad existence result because Myerson allows for arbitrary outcome spaces, any number of agents, arbitrary payoff functions, and moral hazard; the essential restriction is that type and outcome spaces are assumed to be finite.

We emphasize that neo-optimum is not only much easier to handle than neutral optimum in applications, but neo-optimum is also conceptually simpler than neutral optimum because it avoids any reference to Myerson's (1983) Extension axiom, which relates properties of solutions across different Bayesian incentive problems.

While we will not dwell on Myerson's axioms, we still find it worthwhile to mention that the axioms are very useful for understanding why a prior-belief-feasible and neologism-proof payoff vector often fails to exist. The problem with neologism-proofness as a solution concept is that it violates two axioms, Openness and Domination. Indeed, as the Spence example above shows, a limit of a sequence of neologism-proof payoff vectors may fail to be neologism-proof, and a payoff vector above a neologism-proof payoff vector may not share the same property. The concept of a neo-optimum relaxes the ideal of a prior-belief-feasible and neologism-proof payoff vector just enough so that all of Myerson's axioms are satisfied, thus restoring existence.

Proof of Proposition 2. Let b^* denote the uniform distribution on T . We say that a payoff vector $U \in \mathbb{R}^T$ is *neo-blocked* if for some $\epsilon > 0$, a neologism exists for all (b^*, V) such that $V \leq U + \epsilon$. To complete the proof, it is sufficient to show that the concept of neo-blocking satisfies Myerson's four axioms.

The axioms Extension, Domination, and Openness are clear by construction. Consider the axiom Strong Solution. Let U be a strong solution. We have to show that U is not neo-blocked. For this it is sufficient that no

neologism exists for (b^*, U) . Suppose that (\hat{b}, \hat{U}) is a neologism. The main step is to show that (*) there exists a belief b' such that $(b^*, \max\{\hat{U}, U\})$ is a composition of (\hat{b}, \hat{U}) and (b', U) .

By definition of a strong solution, (b', U) is feasible. Given (*), composition-closedness then implies that $(b^*, \max\{\hat{U}, U\})$ is feasible. Because, by definition of a strong solution, U is undominated, $\max\{\hat{U}, U\} = U$, implying that $\hat{U} \leq U$, contradicting the fact that (\hat{b}, \hat{U}) is a neologism.

To show (*), take any \check{t} such that $\hat{U}(\check{t}) > U(\check{t})$ and define $\hat{d} = b^*(\check{t})/\hat{b}(\check{t})$. For all $t \in T$, define

$$c_1(t) = \frac{\hat{d} \hat{b}(t)}{b^*(t)}, \quad c_2(t) = 1 - c_1(t), \quad b'(t) = \frac{b^*(t)c_2(t)}{1 - \hat{d}}.$$

This together with the definition of a neologism implies $c_1(t) = 1$ for all t with $\hat{U}(t) > U(t)$, $0 \leq c_1(t) \leq 1$ for all t with $\hat{U}(t) = U(t)$, and $c_1(t) = 0$ for all t with $\hat{U}(t) < U(t)$. Next,

$$\sum_t b^*(t)c_2(t) = 1 - \sum_t b^*(t)c_1(t) = 1 - \sum_t \hat{d} \hat{b}(t) = 1 - \hat{d},$$

implying that b' is a probability distribution. Moreover, using the definitions above it is straightforward to verify that

$$c_1(t'')b^*(t'')\hat{b}(t') = c_1(t')b^*(t')\hat{b}(t'')$$

and

$$c_2(t'')b^*(t'')b'(t') = c_2(t')b^*(t')b'(t'')$$

for all $t', t'' \in T$. This completes the proof of (*). \square

The following result establishes that neo-optimum is an equilibrium refinement.

Proposition 3. *In any Bayesian incentive problem, any neo-optimum is an expectational equilibrium.*

Proof. Consider a payoff vector U such that no neologism exists for (b^*, U) , where b^* is the uniform distribution. Consider any generalized mechanism μ' as defined in Myerson (1983). It is sufficient to show that there exists a belief b' and a continuation equilibrium in μ' at belief b' such that no type of principal gains from deviating to μ' .

Consider a fictitious game in which the principal (with type distributed according to b^*) first chooses between getting the payoff vector U and the game ends, or deciding that μ' will be played. Because the fictitious game is finite, there exists a sequential equilibrium. Let b' denote a belief at the start of the continuation game μ' that is consistent with the sequential equilibrium. For all t , denote by $c_1(t)$ the probability that type t decides to play μ' , and denote by $V(t)$ her expected payoff in the continuation game μ' .

By definition of a sequential equilibrium, V is b' -feasible.

Suppose that $V \not\leq U$. By the sequential-equilibrium conditions, (b', V) is a neologism for (b^*, U) . But this contradicts the assumption on U . Hence, $V \leq U$, so that at belief b' no type of principal gains from choosing μ' . \square

To relate neo-optimum to Myerson's (1983) other solution concept, core, we need additional notation. We formulate the relation generally for all informed-principal settings, not restricted to Bayesian incentive problems. Given any interior belief b^* and a non-empty set $S \subseteq T$, let b^S denote the belief derived from the information that the type belongs to S , that is,

$$b^S(t) = 0 \text{ for all } t \notin S, \\ \text{and } b^S(t')b^*(t'') = b^S(t'')b^*(t') \text{ for all } t', t'' \in S.$$

A payoff vector U is called a b^* -core payoff vector if U is b^* -feasible and for any payoff vector V that dominates U there exists $S \supseteq \{t \in T \mid V(t) > U(t)\}$ such that V is not b^S -feasible. Myerson motivates the concept with ideas involving neologisms. While a core payoff vector always exists if b^* is the uniform distribution, it is not always an equilibrium. Next we show that neo-optimum is stronger than the core.

Proposition 4. *Let b^* denote an interior belief. Then any b^* -neo-optimum is a b^* -core payoff vector.*

Proof. Denote $D = \{t \in T \mid V(t) > U(t)\}$.

Suppose that U is b^* -feasible, but is not a core payoff vector, that is, there exists a payoff vector V that dominates U and V is b^S -feasible for all $S \supseteq D$.

Fix $\epsilon > 0$ such that $V(t) > U(t) + \epsilon$ for all $t \in D$.

Consider any $W \leq U + \epsilon$.

Note that $D \subseteq D^W := \{t \in T \mid V(t) > W(t)\}$. Thus, V is b^{D^W} -feasible, implying that (b^{D^W}, V) is a neologism for (b^*, W) .

We conclude that U is not a b^* -neo-optimum. \square

5 Neo-optimum in specific settings

5.1 Private-value settings

Mylovanov and Tröger (2012) establish a solution for a principal who “has private information that is not directly payoff relevant to the agents, but may influence her design”—the private-values case. The concept, strongly neologism-proof allocations, is a generalization of strong unconstrained Pareto optimum (SUPO) defined by Maskin and Tirole (1990). In our companion paper Mylovanov and Tröger (2025) we show it is equivalent to neo-optimum.

Proposition 5. *(Mylovanov and Tröger, 2025) Consider a separable generalized-private-values environment in the sense of Mylovanov and Tröger (2012). Consider any interior prior b^* . Then a payoff vector U is strongly neologism-proof if and only if U is a b^* -neo-optimum.*

Together with Proposition 2 this resolves a long-standing open question concerning the relation between the private-values solution concepts and neutral optimum: 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) together with the only-if part of Proposition 5, we can also conclude that a neo-optimum broadly exists in private-value settings, including settings that do not satisfy the finiteness properties of Bayesian incentive problems as defined by Myerson (1983).

As an intermediate step towards proving Proposition 5, we employ yet another solution concept. In a sense, this is the missing piece that allows to connect private-values settings to neutral optimum. The concept was invented by Koessler and Skreta (2023) in a non-private-values context of information design. Given any belief b^* , a payoff vector U is b^* -interim-optimal if (i) U is b^* -feasible and (ii) there does not exist a belief b together with a b -feasible payoff vector V such that $\text{supp}(b) \subseteq \{t \in T \mid V(t) > U(t)\}$. Interim-optimality is easily seen to be at least as strong as neo-optimum (and the result has nothing to do with private values):

Remark 2. *Consider any informed-principal setting. Given any interior belief b^* , any b^* -interim-optimal payoff vector is a b^* -neo-optimum.*

Proof. Consider any b^* -interim optimal payoff vector U . Then, for all $\epsilon > 0$, no neologism exists for $(b^*, U + \epsilon)$. Thus, U is a limit of b^* -neologism-proof payoff vectors, showing that it is a neo-optimum. \square

Interim-optimality—in contrast to neo-optimum—is not a generally applicable solution concept because existence may fail, as can be seen in our Spence example with $b^* > (\theta_H - \theta_L)/\theta_H$.

To prove Proposition 5 in our companion paper, we first show that in private-values settings, interim-optimality and neo-optimality are in fact equivalent. Then we use the separability assumption in Mylovanov and Tröger (2012) to show that interim-optimality and strong neologism-proofness are equivalent.

5.2 Neo-optimum in settings with certification

Koessler and Skreta (2019) consider a seller-principal who proposes a mechanism to guide her interaction with a single buyer. Values are interdependent: each trader has private information concerning both traders' valuations. The seller can provide partial or complete evidence about (i.e., “certify”) her type. The model fits into our framework, with Lemma 1 in Koessler and Skreta (2019) describing the sets of feasible payoff vectors for all beliefs.

Koessler and Skreta (2019) propose a solution concept, strong Pareto optimum (SPO). They show that any prior-feasible SPO is an expectational equilibrium and is ex-ante optimal, but it exists only in settings with sufficient (e.g., full) certifiability. Using neo-optimum instead of SPO as a solution concept in their setting, existence is generally guaranteed and their qualitative results remain largely intact. In the following we sketch how.

Remark 3. *Let b^* denote an interior prior belief. Any b^* -feasible SPO as defined in Koessler and Skreta (2019) is a b^* -neo-optimum.*

Proof. Consider a b^* -feasible SPO (payoff/profit vector) V^* . We show that for any $\epsilon > 0$, the payoff vector $V^* + \epsilon$ is neologism-proof. Suppose there exists a neologism (b, U) for $(b^*, V^* + \epsilon)$. Define a payoff vector V via $V(t) = U(t)$ for all $t \in \text{supp}(b)$, and $V(t) = V^*(t)$ for all other t . Then (using the terminology of Koessler and Skreta (2019)) V is “buyer-feasible” at the belief

$\pi = b$. Moreover, $V(t) \geq V^*(t)$ for all $t \in T$, with strict inequality for all $t \in \text{supp}(b)$, contradicting the definition of an SPO payoff vector. We conclude that V^* is a limit of b^* -neologism-proof payoff vectors, and hence is a b^* -neo-optimum. \square

Thus, all the properties that we have established for neo-optima also hold for any prior-feasible SPO. Moreover, in all cases where there is a unique neo-optimum, it is identical to prior-feasible SPO if the latter exists.

Note that Koessler and Skreta (2019) includes two extreme cases, full certifiability and no certifiability (“soft information”). Importantly, SPO is a generally useful solution concept *only* under sufficient certifiability, while neo-optimum can always be used.

Remark 4. *Let b^* denote an interior prior belief. A b^* -feasible SPO as defined in Koessler and Skreta (2019) may not exist with soft information, but a b^* -neo-optimum always exists.*

Koessler and Skreta (2019) themselves remark on the non-existence problem. Interestingly, non-existence can also be seen from our Spence-job-market example, restricted via a (sufficiently high) highest feasible task level \bar{e} . Equivalently, we may assume that only two task levels, 0 and \bar{e} , are feasible. This setting is included as a soft-information case in Koessler and Skreta (2019). The worker-principal is the seller. Selling her labor with a certain probability p means working at the task level $e = p\bar{e}$. Thus, our Spence example with the prior b^* slightly below $(\theta_H - \theta_L)/\theta_H$ yields instances where no b^* -feasible SPO exists because it would be ex-ante optimal by the results in Koessler and Skreta (2019) and would be a b^* -neo-optimum by our Remark 3.

Neo-optimum, however, always exists: any setting considered in Koessler and Skreta (2019) is a Bayesian incentive problem, except that feasibility is defined without truth-telling constraints for the principal; as observed in Koessler and Skreta (2023), Myerson’s 1983 proof that a neutral optimum exists extends and, by the same logic as in our Proposition 2, any neutral optimum is a neo-optimum.

In the other extreme case, full certifiability, the main qualitative insight of Koessler and Skreta (2019) is that a prior-feasible SPO exists and is ex-ante optimal. But this is also true for neo-optimum.

Remark 5. *With full certifiability, any neo-optimum is ex-ante optimal.*

Indeed, in the proof of Proposition 3 in Koessler and Skreta (2019), the pair (π^0, \tilde{V}) is a neologism for (π^0, \hat{V}) because with full certifiability there are no incentive constraints for the seller.

5.3 Neo-optimum in information-design settings

Koessler and Skreta (2023) analyze information design by a privately informed designer. Their setting fits into our framework.¹⁶ Koessler and Skreta (2023) introduce a new solution concept, interim optimality, which we have generalized to arbitrary informed-principal settings in Section 5.1. The main results in Koessler and Skreta (2023) concern existence of interim-optimum in their setting, the proof that interim-optima are perfect-Bayesian equilibria, and characterization results in special settings.

From Remark 2 we know that any interim optimum is a neo-optimum. In the different variants of the introductory prosecutor-judge example in Koessler and Skreta (2023) the reverse is also true, that is, any neo-optimum is interim-optimal. The question to what extent this reverse implication holds in general information-design settings is left for future research.

Consider the three-actions variation of the prosecutor-judge setting in Koessler and Skreta (2023) with a belief that the defendant is guilty with a probability $b < 1/3$. A payoff vector $(U(t_G), U(t_I))$ is b -feasible if and only if

$$U(t) = 2\mu(a_2|t) + 3\mu(a_3|t) \quad \text{for } t \in \{t_G, t_I\},$$

where μ is a mechanism that satisfies the relevant obedience constraints, that is, μ is b -incentive compatible as defined in Koessler and Skreta (2023).

To prepare, we extend the arguments in Koessler and Skreta (2023) to show the following.

Lemma 4. *Consider a belief $0 < b < 1/3$ in the three-actions variation of the prosecutor-judge setting in Koessler and Skreta (2023). The set of b -feasible payoff vectors is the convex hull of the points $(0, 0)$, $(3, 0)$, $(3, \frac{3}{2}\frac{b}{1-b})$, and $(2, 4\frac{b}{1-b})$.*

See Figure 5 for an illustration of the b -feasibility set and the b' -feasibility set for some $0 < b < b' < 1/3$.

¹⁶Note a special feature of information-design settings: different types of the principal cannot imitate each other in a given direct mechanism. Thus, even in settings with state-independent preferences the feasibility of a payoff vector does not exclude the possibility that different types obtain different payoffs.

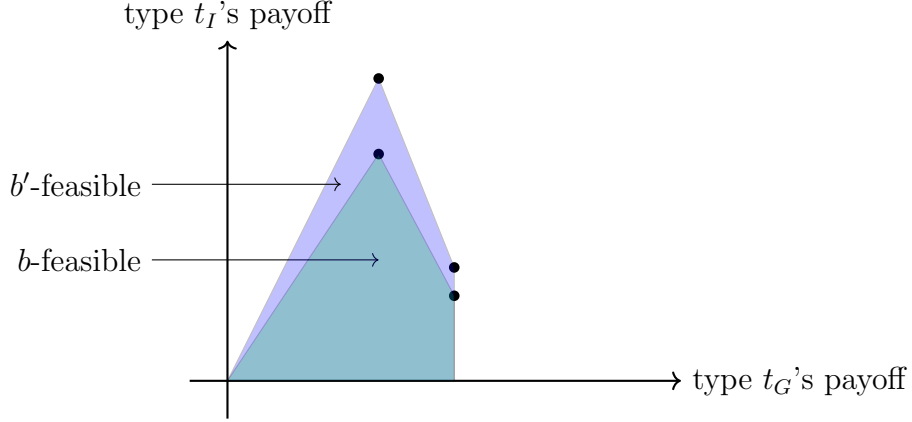


Figure 3: The sets of feasible payoff vectors at different beliefs $0 < b < b' < 1/3$ in the three-action prosecutor-judge example in Koessler and Skreta (2023). The thick vertices refer to the ex-ante optimal point $(2, 4 \frac{b}{1-b})$ (or, resp., $(2, 4 \frac{b'}{1-b'})$) and the point $(3, \frac{3}{2} \frac{b}{1-b})$ (or, resp., $(3, \frac{3}{2} \frac{b'}{1-b'})$) where the information designer splits the believed probability of the guilty type t_G into 0 and $2/3$.

Koessler and Skreta (2023) show that, in the three-action and four-action prosecutor-judge examples with a prior belief $b^* < 1/3$, a payoff vector U^* is interim-optimal if and only if

$$(*) \ U^* \text{ is } b^*\text{-feasible and } U^*(t_G) = 3.$$

We claim that condition $(*)$ is also necessary for U^* to be a b^* -neo-optimum.

Consider the three-action version and consider a b^* -neo-optimum U^* . By feasibility, $U^*(t_G) \leq 3$. Suppose that $U^*(t_G) < 3$.

Defining $V = (3, \frac{3}{2} \frac{b^*}{1-b^*})$, note first that (b^*, V) is a neologism for all $U \leq V$, $U \neq V$.

By Lemma 4, for any b^* -feasible U with $U(t_G) < 3$, there exists a b' -feasible V with $b' > b^*$, $V(t_G) > U(t_G)$, and $V(t_I) = U(t_I)$. Thus, (b', V) is a neologism for U .

These arguments contradict the fact that U^* is a b^* -neo-optimum. This completes the proof that U^* satisfies $(*)$ in the three-action version.

The arguments extend easily to the four-action version. Because the designer's payoff from action \bar{a}_0 is smaller than from action a_3 , the b -feasibility sets in the four-action version are more restricted than in the three-action version, for all b . But the points $(0, 0)$, $(3, \frac{3}{2} \frac{b}{1-b})$, and $(2, 4 \frac{b}{1-b})$ are still fea-

sible for all $b < 1/3$. Thus the above arguments showing the necessity of (*) extend to the four-action version.

In the two-action version of the prosecutor-judge example in Koessler and Skreta (2023), the equivalence between neo-optimum and interim-optimum follows with the help of Remark 1.

5.4 Bilateral trade with interdependent values

For another illustration of Kakutani PBE and neo-optimum, let the principal be a seller who, as in Dosis (2022), proposes a mechanism for determining the probability $q \in [0, 1]$ that she produces a unit of an indivisible good for a buyer, and let $p \in \mathbb{R}$ denote the buyer's payment. The principal has one of two (non-negative) cost types, $T = \{c_1, c_2\}$. Any belief b can be identified with the probability of the type c_2 , that is, $b \in B = [0, 1]$. Let $p - c_i q$ ($i = 1, 2$) denote the principal's payoff if she has the type c_i , sells with probability q , and receives the payment p ; the buyer then obtains the payoff $-p + v_i q$, where $v_1 > c_1$, $v_2 > c_2$, and $v_2 > v_1$. If the payments were bounded, the analysis would remain the same and the setting would be a Bayesian incentive problem, implying that a neo-optimum exists (cf. Corollary 2).

A mechanism is a game form in which the seller and the buyer play, and each end node is a probability-payment pair $(q, p) \in [0, 1] \times \mathbb{R}$. The mechanism is played if the buyer accepts it. By the revelation principle, given any belief $b \in B$, the mechanism (or, more precisely, the action of proposing the mechanism) implements a probability-payment pair¹⁷ $(q_i(b), p_i(b))$ for each type c_i such that incentive compatibility is satisfied,

$$p_1 - c_1 q_1 \geq p_2 - c_1 q_2 \quad \text{and} \quad p_2 - c_2 q_2 \geq p_1 - c_2 q_1, \quad (9)$$

and the buyer's participation constraint is satisfied,

$$(1 - b)(v_1 q_1 - p_1) + b(v_2 q_2 - p_2) \geq 0. \quad (10)$$

Thus,

$$K = \{(b, (p_1 - c_1 q_1, p_2 - c_2 q_2)) \mid (9), (10)\}.$$

We now distinguish two cases. In the first, $c_1 < c_2$; this is consistent with the interpretation that the seller has a low-quality good if her type is c_1 and has

¹⁷Given the linearity of the payoff functions, probability distributions over probability-payment pairs need not be considered.

a high-quality good if her type is c_2 . In the second case, $c_1 > c_2$; recalling $v_1 < v_2$, this case can be interpreted in terms of fit: if the seller's type is c_1 then the good is a relatively better fit for the seller, and otherwise fits the buyer relatively better.

For any set of payoff vectors S , denote the payoff vectors that are below some payoff vector in S along a 45 degree ray by

$$\text{diagram}(S) = \{U \mid \exists x \geq 0 : U + (x, x) \in S\}.$$

As is clear from the constraints (9) and (10), if all payoff vectors in a set S are b -feasible for some $b \in B$, then all payoff vectors in $\text{diagram}(S)$ are b -feasible as well because the payments to all types of the seller may be changed by the same amount x .

Consider now the quality-uncertainty case $c_1 < c_2$. After standard manipulations of the constraints (9) and (10), we obtain the following characterization,

$$K = \bigcup_{b \in B} \{b\} \times \text{diagram}(\text{conv}\{((1-b)v_1 + bv_2 - c_1, (1-b)v_1 + bv_2 - c_2)), \\ ((1-b)(v_1 - c_1), (1-b)(v_1 - c_1))\}).$$

where $\text{conv}\{\dots\}$ denotes the line section between two “vertex” payoff vectors. The first of these arises from both types selling with probability 1, and the payment is equal to the buyer's expected valuation. The second vertex payoff vector arises from type c_1 selling with probability 1 and type c_2 selling with probability 0, where the payments are determined such that the incentive constraint of type c_1 (that is, the left constraint in (9)) and the agent's participation constraint (10) are binding (in particular, whenever $b < 1$ so that the buyer is not certain to face type c_2 , a seller of this type obtains a payment although she is not selling anything).

As illustrated in Figure 4, the bounding line sections tilt around a common point \underline{U} that is feasible at all beliefs. This point is characterized as follows. The payoff of a type- c_1 seller is maximized across all points in the 0-feasible set $K(0)$, that is, under the constraint that the buyer is sure to face type c_1 and accepts the mechanism. This is achieved via the type- c_1 seller selling the good for sure (i.e., $q_1 = 1$) and have the buyer pay $p_1 = v_1$. The vertical component of the tilting point can be computed by the type- c_2 seller

maximizing her payoff across all points in the 1-feasible set that satisfy the incentive constraint of type c_1 given her outcome (q_1, p_1) . This is achieved at the outcome (q_2, p_2) such that $-c_1 + v_1 = -q_2 c_1 + p_2$ and $q_2 v_2 - p_2 = 0$, that is,

$$q_2 = \frac{v_1 - c_1}{v_2 - c_1}, \quad p_2 = \frac{v_1 - c_1}{v_2 - c_1} v_2.$$

(In particular, $q_2 < 1$ so that the high-quality seller keeps the good with positive probability.)

By Lemma 1, any Kakutani PBE is above \underline{U} . Thus, whenever the 1st vertex does not dominate \underline{U} (that is, at all interior prior beliefs b^* sufficiently close to 0), the point \underline{U} is the unique Kakutani PBE which, by Proposition 1 then also is the unique neo-optimum. Using Corollary 1, we conclude that, for all interior priors, the set of Kakutani PBEs is equal to the points above \underline{U} that are feasible given the prior. Using Remark 1, the 1st vertex is the unique neo-optimum if it dominates \underline{U} strictly. In the non-generic case where a seller of type c_2 is indifferent between \underline{U} and the 1st vertex, the line section connecting these two points is the set of neo-optima.

Consider now the fit-uncertainty case $c_1 > c_2$. After standard manipulations of the constraints (9) and (10), we obtain the following characterization,

$$K = \bigcup_{b \in B} \{b\} \times \text{diagraph}(\text{conv}\{((1-b)v_1 + bv_2 - c_1, (1-b)v_1 + bv_2 - c_2), (b(v_2 - c_2), b(v_2 - c_2))\}).$$

The first “vertex” payoff vector argument in $\text{conv}\{\dots\}$ is the same as in the quality-uncertainty case (i.e., both seller types sell with probability 1 at a price equal to the buyer’s expected valuation). The second vertex payoff vector arises from type c_1 selling with probability 0 and type c_2 selling with probability 1, where the payments are determined such that the incentive constraint of type c_2 (that is, the right constraint in (9)) and the agent’s participation constraint (10) are binding (in particular, whenever $b > 0$ so that the buyer is not certain to face type c_1 , a seller of this type obtains a payment although she is not selling anything).

As illustrated in Figure 5, the b -feasibility sets are nested, with a larger probability b of the type c_2 leading to more feasible points. The largest point that is common to all b -feasibility sets, \underline{U} , arises from the 1st vertex at $b = 0$, that is, both seller types sell with probability 1 at the price v_1 .

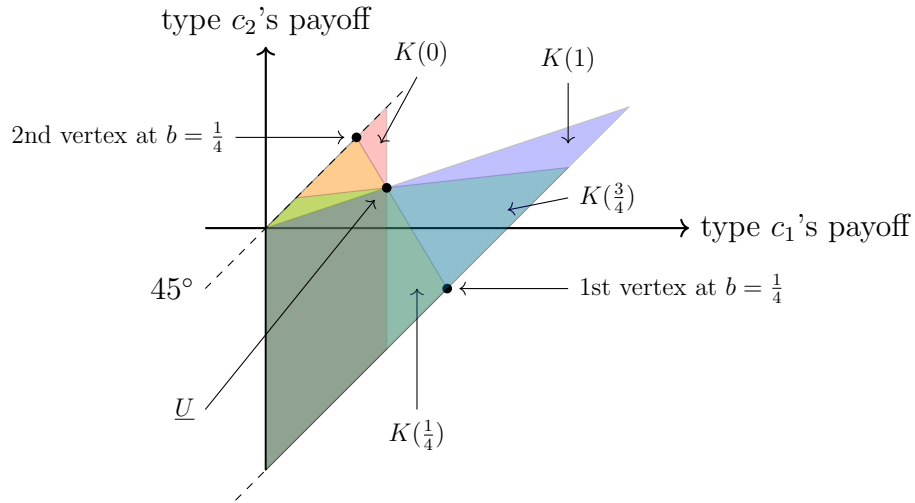


Figure 4: The sets of b -feasible payoff vectors at different beliefs b in the informed-seller example with quality-uncertainty (we omit the parts of the feasible sets that are to the left of the vertical axis). In contrast to the Spence example, the feasibility sets are not nested across different beliefs. Each b -feasibility set is bounded by the line section between the corresponding 1st and 2nd vertices, and includes all points below this line section along 45 degree rays. As b changes, the line sections tilt around a common point \underline{U} that is feasible at all beliefs. (The diagram corresponds to the case $c_1 = 1$, $c_2 = 3$, $v_1 = 2$, $v_2 = 4$.)

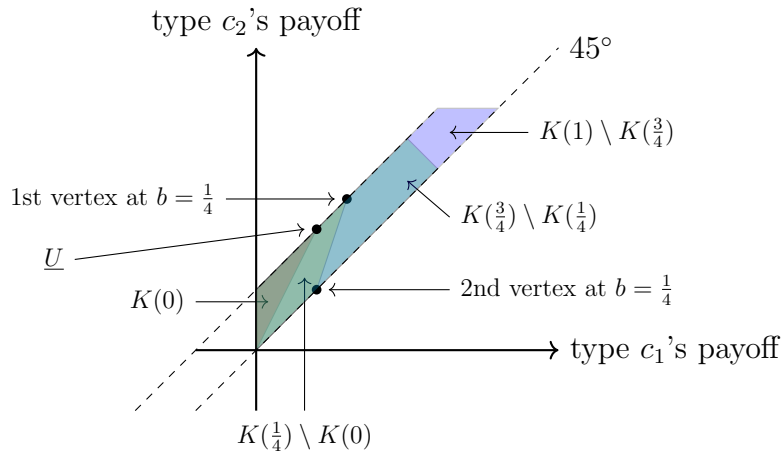


Figure 5: The sets of b -feasible payoff vectors at different beliefs b in the informed-seller example with fit-uncertainty (we omit the parts of the feasible sets that are to the left of the vertical axis). The feasibility sets are nested across different beliefs. Each b -feasibility set is bounded by the line section between the corresponding 1st and 2nd vertices, and includes all points below this line section along 45 degree rays. The largest payoff vector that is feasible at all beliefs is \underline{U} . (The diagram corresponds to the case $c_1 = 1$, $c_2 = 0$, $v_1 = 2$, $v_2 = 4$.)

Given any interior prior belief b^* , the set of Kakutani PBE is easily characterized. By Lemma 1, any Kakutani PBE is above \underline{U} . On the other hand, any b^* -feasible point above \underline{U} is a Kakutani PBE because the “pessimistic” off-path belief $b = 0$ can always be used.

Using Remark 1, the 1st vertex is the unique neo-optimum if it dominates all other b^* -feasible points strictly (which happens for all b^* sufficiently close to 0 and is the case if $b^* = 1/4$ in Figure 5).

For all b^* sufficiently close to 1 (such as in the case $b^* = 3/4$ in Figure 5), there exist multiple b^* -feasible points above \underline{U} that are on the b^* -weak-Pareto frontier. Thus, neither Proposition 1 nor Remark 1 are useful to determine which of these points is a neo-optimum. Similar arguments as in our Spence example show that the 1st vertex is the unique neo-optimum.

We conclude that the 1st vertex, where the good is sold with probability 1 by both types, is the unique neo-optimum at all interior prior beliefs.

It is straightforward to show that the 1st vertex also is the ex-ante optimum. However, given any sufficiently large b^* , a trivial change of the setting will let the ex-ante optimum switch to the 2nd vertex while the neo-optimum remains at the 1st vertex: just assume that the payoff function of the type c_1 -seller is instead given by $\alpha(p - c_1q)$, where α is large; this will stretch Figure 5 horizontally by the factor α . The example is noteworthy because then the 2nd vertex is both a Kakutani PBE and yields a higher ex-ante expected payoff than the 1st vertex, yet the 1st vertex is the unique neo-optimum.

Appendix: omitted proofs

Proof of Lemma 3. Fix any $\epsilon > 0$. Denote by \overline{U} the (compact) convex hull of $\cup_{P \in \mathcal{P}, b' \in B} P(b')$. Define, for any $n \in \mathbb{N}$, any list of payoff vectors $(U_P)_{P \in \mathcal{P}}$ with $U_P \in \overline{U}$, and any $t \in T$, the choice probability

$$\Psi_n((U_P)_{P \in \mathcal{P}})_P(t) = \frac{n^{U_P(t)}}{\sum_{P' \in \mathcal{P}} n^{U_{P'}(t)}}.$$

Note that, for any n , the map

$$\Psi_n : \overline{U}^{|\mathcal{P}|} \longrightarrow \mathbb{R}^{|\mathcal{P}| \cdot |T|}, (U_P)_{P \in \mathcal{P}} \mapsto \Psi_n((U_P)_{P \in \mathcal{P}}) \quad \text{is continuous.}$$

Denote by \mathcal{C}_n the (compact) convex hull of $\Psi_n(\overline{U}^{|\mathcal{P}|})$. Note that $\sum_{P \in \mathcal{P}} c_P(t) = 1$ for all $t \in T$ and all $(c_P)_{P \in \mathcal{P}} \in \mathcal{C}_n$.

For any $(c_P)_{P \in \mathcal{P}} \in \mathcal{C}_n$, define $\hat{b}(c_P) \in B$ by

$$\hat{b}(c_P)(t) = \frac{b(t)c_P(t)}{\sum_{t' \in T} b(t')c_P(t')} \quad \text{for all } t \in T.$$

Define a correspondence

$$C_n : \bar{\mathcal{U}}^{|\mathcal{P}|} \times \mathcal{C}_n \longrightarrow \bar{\mathcal{U}}^{|\mathcal{P}|} \times \mathcal{C}_n$$

by

$$C_n((U_P)_{P \in \mathcal{P}}, (c_P)_{P \in \mathcal{P}}) = (P(\hat{b}(c_P)))_{P \in \mathcal{P}} \times \Psi_n((U_P)_{P \in \mathcal{P}}).$$

Because \hat{b} is continuous, the correspondence C_n is upper-hemicontinuous and convex-valued. Moreover, $\bar{\mathcal{U}}^{|\mathcal{P}|} \times \mathcal{C}_n$ is convex and compact. Hence, by Kakutani's Theorem, C_n has a fixed point

$$((U_{P,n})_{P \in \mathcal{P}}, (c_{P,n})_{P \in \mathcal{P}}).$$

Now choose a subsequence $n_l \rightarrow \infty$ such that, for all P ,

$$b_{P,n_l} \stackrel{\text{def}}{=} \hat{b}(c_{P,n_l}) \rightarrow b_P^* \text{ for some } b_P^* \in B,$$

and

$$U_{P,n_l} \rightarrow U_P^* \text{ for some } U_P^* \in \bar{\mathcal{U}}, \quad (11)$$

and, for all $t \in T$,

$$c_{P,n_l}(t) \rightarrow c_P^*(t) \text{ for some } c_P^*(t) \geq 0.$$

Hence,

$$s_{P,n_l} \stackrel{\text{def}}{=} \sum_{t' \in T} b_{P,n_l}(t')c_{P,n_l}(t') \rightarrow s_P^* \stackrel{\text{def}}{=} \sum_{t' \in T} b_P^*(t')c_P^*(t').$$

Observe that, by the definition of the fixed point,

$$U_{P,n_l} \in P(b_{P,n_l}) \text{ for all } l.$$

Taking the limit $l \rightarrow \infty$ and using the upper-hemicontinuity of P , we obtain the conclusion

$$U_P^* \in P(b_P^*). \quad (12)$$

By definition of \hat{b} , for all P and t ,

$$b(t)c_{P,n_l}(t) = b_{P,n_l}(t)s_{P,n_l}.$$

Taking the limit $l \rightarrow \infty$, we obtain

$$b(t)c_P^*(t) = b_P^*(t)s_P^* \quad \text{for all } P \text{ and } t. \quad (13)$$

Observe that, for all $P \in \mathcal{P}$ and $t \in T$,

$$\text{if } c_P^*(t) > 0, \text{ then } U_P^*(t) \geq U_{P'}^*(t) \text{ for all } P' \in \mathcal{P}. \quad (14)$$

Indeed, if we had $U_P^*(t) < U_{P'}^*(t)$ for some P' , then

$$U_{P',n_l}(t) - U_{P,n_l}(t) > \epsilon \stackrel{\text{def}}{=} \frac{U_{P'}^*(t) - U_P^*(t)}{2} > 0$$

for all large l , implying—by definition of Ψ_{n_l} —that

$$\frac{c_{P',n_l}(t)}{c_{P,n_l}(t)} = (n_l)^{U_{P',n_l}(t) - U_{P,n_l}(t)} > (n_l)^\epsilon \rightarrow \infty,$$

contradicting the fact that $0 < c_P^*(t) = \lim_l c_{P,n_l}(t)$.

Next, $\sum_{P \in \mathcal{P}} c_P^*(t) = 1$ for all t by definition of \hat{b} . Using (13) with $t = t'$ and $t = t''$,

$$b_P^*(t')c_P^*(t'')b(t'') = c_P^*(t')b(t')b_P^*(t'') \quad \text{for all } t', t'' \in T.$$

The proof is completed by defining $U^*(t) = \max_{P \in \mathcal{P}} U_P^*(t)$ for all t .

Proof of Lemma 4. Note first that the feasibility set belongs to the set of realizable payoff vectors $[0, 3]^2$. Koessler and Skreta (2023) show that all four points mentioned in the lemma are b -feasible; let μ^1 denote a mechanism that yields the payoff vector $(0, 0)$, let μ^2 denote a mechanism that yields the payoff vector $(3, 0)$, let $\mu^{3,b}$ denote a mechanism that yields the payoff vector $(3, \frac{3}{2} \frac{b}{1-b})$, and let $\mu^{4,b}$ denote a mechanism that yields the payoff vector $(2, 4 \frac{b}{1-b})$. In the following, we fix $\mu^{4,b}$ as follows:

$$\mu^{4,b}(\underline{a}_0|G) = \mu^{4,b}(a_3|G) = \mu^{4,b}(a_3|I) = 0, \quad \mu^{4,b}(a_2|G) = 1,$$

$$\mu^{4,b}(\underline{a}_0|I) = \frac{1-3b}{1-b}, \quad \mu^{4,b}(a_2|I) = \frac{2b}{1-b}.$$

To further discuss the feasibility restrictions, we need to consider the obedience constraints. To this end, we must fix a payoff function for the agent that induces the agent's belief-dependent action as described in Koessler and Skreta (2023). Let

$$u_1(\underline{a}_0|G) = u_1(\underline{a}_0|I) = 0, \quad u_1(a_2|G) = 2, \quad u_1(a_2|I) = -1, \quad u_1(a_3|G) = 3, \quad u_1(a_3|I) = -3.$$

Define

$$\underline{g} = -2\frac{b}{1-b} \quad \text{and} \quad \bar{g} = \frac{5}{2}\frac{b}{1-b}.$$

We will show that, for all $g \in [\underline{g}, \bar{g}]$, the mechanism $\mu^{4,b}$ maximizes the weighted average of the principal-types' payoffs, where g denotes the weight on type t_G , while the payoff of type t_I has the weight 1. Formally, we show that (*) $\mu^{4,b}$ is a maximizer of the function

$$\phi_g(\mu) = g(2\mu(a_2|t_G) + 3\mu(a_3|t_G)) + 2\mu(a_2|t_I) + 3\mu(a_3|t_I)$$

subject to the b -feasibility constraints for the mechanism μ . This proves the lemma because

$$\phi_{\bar{g}}(\mu^{4,b}) = \phi_{\bar{g}}(\mu^{3,b}) \quad \text{and} \quad \phi_{\underline{g}}(\mu^{4,b}) = \phi_{\underline{g}}(\mu^1).$$

To show (*), we verify that the Karush-Kuhn-Tucker first-order conditions are satisfied at the point $\mu = \mu^{4,b}$ for all $g \in [\underline{g}, \bar{g}]$. Let $\lambda_{3,0}$ denote the Lagrangian multiplier for the obedience constraint that the agent when the recommended action is a_3 cannot gain from taking action \underline{a}_0 instead. Let $\lambda_{3,2}$, $\lambda_{2,0}$, $\lambda_{2,3}$, $\lambda_{0,2}$, and $\lambda_{0,3}$, denote the Lagrangian multipliers of the other obedience constraints, where in all variables the first lower index indicates the recommended action and the second indicates a non-recommended action. The KKT conditions require that

$$\lambda_{2,3} = \lambda_{0,2} = \lambda_{0,3} = 0$$

because the corresponding obedience conditions are not binding at $\mu^{4,b}$. The other three obedience constraints are binding at $\mu^{4,b}$, implying that the KKT conditions require the inequalities

$$\lambda_{3,0} \geq 0, \quad \lambda_{3,2} \geq 0, \quad \lambda_{2,0} \geq 0. \tag{15}$$

Let λ_G and λ_I denote the Lagrangian multipliers for the probability constraints

$$\begin{aligned}\mu(\underline{a}_0|t_G) + \mu(a_2|t_G) + \mu(a_3|t_G) &= 1, \\ \mu(\underline{a}_0|t_I) + \mu(a_2|t_I) + \mu(a_3|t_I) &= 1.\end{aligned}$$

We will not introduce Lagrangian multipliers for the probability-boundary constraints $0 \leq \mu(a|t) \leq 1$, but will write the KKT conditions as appropriate inequalities.

Differentiating the Lagrangian with respect to $\mu(\underline{a}_0|t_I)$ yields the first-order condition

$$\lambda_I = 0.$$

(Note that the condition is an equality because $0 < \mu^{4,b}(\underline{a}_0|t_I) < 1$, implying that the probability-boundary constraints are not binding.)

Differentiating the Lagrangian with respect to $\mu(\underline{a}_0|t_G)$ yields the first-order condition

$$\lambda_G \leq 0.$$

Differentiating the Lagrangian with respect to $\mu(a_2|t_I)$ yields the first-order condition

$$2 + \lambda_{2,0}(1 - b)(-1) = 0,$$

where we have already used that $\lambda_I = 0$. Differentiating the Lagrangian with respect to $\mu(a_2|t_G)$ yields the first-order condition

$$r \cdot 2 + \lambda_G + \lambda_{2,0}b \cdot 2 \geq 0. \tag{16}$$

Differentiating the Lagrangian with respect to $\mu(a_3|t_I)$ yields the first-order condition

$$3 + \lambda_{3,0}(1 - b)(-3) + \lambda_{3,2}(1 - b)(-2) \leq 0.$$

Differentiating the Lagrangian with respect to $\mu(a_3|t_G)$ yields the first-order condition

$$r \cdot 3 + \lambda_G + \lambda_{3,0}b(-3) + \lambda_{3,2}b \cdot 1 \leq 0. \tag{17}$$

Here we can interpret (16) as an upper bound for $-\lambda_G$ and (17) as a lower bound (essentially we are applying the Fourier-Motzkin algorithm in the following). Thus we can remove λ_G from the first-order system of equations

and inequalities and conclude that a solution to the Karush-Kuhn-Tucker condition exists if and only if the following system has a solution:

$$\begin{aligned}
2 + \lambda_{2,0}(1-b)(-1) &= 0, \\
r \cdot 2 + \lambda_{2,0}b \cdot 2 &\geq 0, \\
3 + \lambda_{3,0}(1-b)(-3) + \lambda_{3,2}(1-b)(-2) &\leq 0, \\
r \cdot 2 + \lambda_{2,0}b \cdot 2 &\geq r \cdot 3 + \lambda_{3,0}b(-3) + \lambda_{3,2}b,
\end{aligned}$$

and (15). The first equation implies $\lambda_{2,0} = 2/(1-b) > 0$. Plugging this into the remaining conditions yields the simplified system

$$\begin{aligned}
r \cdot 2 + 4\frac{b}{1-b} &\geq 0, \\
\frac{b}{1-b}4 - r &\geq 3b\lambda_{3,0} + b\lambda_{3,2}, \\
3 &\leq 3(1-b)\lambda_{3,0} + 2(1-b)\lambda_{3,2}, \\
\lambda_{3,0} &\geq 0, \\
\lambda_{3,2} &\geq 0.
\end{aligned}$$

This can be rearranged as follows.

$$\begin{aligned}
r &\geq \underline{g}, \\
\lambda_{3,2} &\leq \frac{1}{1-b}4 - \frac{1}{b}r - 3\lambda_{3,0}, \\
\frac{3}{2}\frac{1}{1-b} - \frac{3}{2}\lambda_{3,0} &\leq \lambda_{3,2}, \\
\lambda_{3,0} &\geq 0, \\
\lambda_{3,2} &\geq 0.
\end{aligned}$$

Now we can remove $\lambda_{3,2}$ and simplify to

$$\begin{aligned}
r &\geq \underline{g}, \\
0 &\leq \frac{1}{1-b}4 - \frac{1}{b}r - 3\lambda_{3,0}, \\
\frac{3}{2}\frac{1}{1-b} - \frac{3}{2}\lambda_{3,0} &\leq \frac{1}{1-b}4 - \frac{1}{b}r - 3\lambda_{3,0}, \\
\lambda_{3,0} &\geq 0.
\end{aligned}$$

After rearranging terms,

$$\begin{aligned} r &\geq \underline{g}, \\ \lambda_{3,0} &\leq \frac{1}{1-b} \frac{4}{3} - \frac{1}{b} r \frac{1}{3}, \\ \lambda_{3,0} &\leq \frac{1}{1-b} \frac{5}{3} - \frac{1}{b} r \frac{2}{3}, \\ \lambda_{3,0} &\geq 0. \end{aligned}$$

Thus, a solution exists if and only if

$$\begin{aligned} r &\geq \underline{g}, \\ 0 &\leq \frac{1}{1-b} \frac{4}{3} - \frac{1}{b} r \frac{1}{3}, \\ 0 &\leq \frac{1}{1-b} \frac{5}{3} - \frac{1}{b} r \frac{2}{3}, \end{aligned}$$

which is equivalent to $\underline{g} \leq r \leq \bar{g}$. □

References

- BALKENBORG, D., AND M. MAKRI (2015): “An undominated mechanism for a class of informed principal problems with common values,” *Journal of Economic Theory*, 157, 918–958.
- BALZER, B. (2017): “Collusion in auctions: an informed principal perspective,” *University of Technology, Sydney*.
- BERGEMANN, D., AND S. MORRIS (2019): “Information design: A unified perspective,” *Journal of Economic Literature*, 57(1), 44–95.
- CHO, I.-K., AND D. M. KREPS (1987): “Signaling games and stable equilibria,” *The Quarterly Journal of Economics*, 102(2), 179–221.
- DOSIS, A. (2022): “On the informed principal model with common values,” *The RAND Journal of Economics*, 53(4), 792–825.
- FARRELL, J. (1993): “Meaning and Credibility in Cheap-Talk Games,” *Games and Economic Behavior*, 5, 514–531.

- IZMALKOV, S., AND F. BALESTRIERI (2012): “The informed seller problem: The case of horizontal differentiation,” *mimeo*.
- KOCHERLAKOTA, N. R. (2006): “Advances in dynamic optimal taxation,” *Econometric Society Monographs*, 41, Vol I, ed. Richard Blundell, Chapter 7.
- KOESSLER, F., AND V. SKRETA (2016): “Informed seller with taste heterogeneity,” *Journal of Economic Theory*, 165, 456–471.
- (2019): “Selling with evidence,” *Theoretical Economics*, 14(2), 345–371.
- (2023): “Informed information design,” *Journal of Political Economy*, 131(11), 3186–3232.
- LAFFONT, J.-J. (2000): *Incentives and political economy*. OUP Oxford.
- LAFFONT, J.-J., AND J. TIROLE (1993): *A Theory of Incentives in Procurement and Regulation*. MIT Press.
- MAILATH, G. J. (1987): “Incentive compatibility in signaling games with a continuum of types,” *Econometrica: Journal of the Econometric Society*, pp. 1349–1365.
- MAILATH, G. J., M. OKUNO-FUJIWARA, AND A. POSTLEWAITE (1993): “Belief-Based Refinements in Signalling Games,” *Journal of Economic Theory*, 60(2), 241–276.
- 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.
- 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.
- (2025): “Neutral optimum in private-value settings,” *Unpublished manuscript*.
- NISHIMURA, T. (2022): “Informed principal problems in bilateral trading,” *Journal of Economic Theory*, 204, 105498.
- PEŖSKI, M. (2022): “Bargaining with mechanisms,” *American Economic Review*, 112(6), 2044–2082.
- RABIN, M. (1990): “Communication between rational agents,” *Journal of Economic Theory*, 51(1), 144–170.
- RENY, P. J. (2025): “Natural Language Equilibrium I: Off-Path Conventions,” *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2025-64).
- SKRETA, V. (2009): “On the informed seller problem: optimal information disclosure,” *Review of Economic Design*, 15(1), 1–36.
- SPENCE, A. M. (1973): “Job Market Signaling,” *The Quarterly Journal of Economics*, 87(3), 355–74.
- TAN, G. (1996): “Optimal Procurement Mechanisms for an Informed Buyer,” *Canadian Journal of Economics*, 29(3), 699–716.
- TRÖGER, T. (2025): “Optimal testing and social distancing of individuals with private health signals,” *Review of Economic Design*, pp. 1–41.
- 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.
- ZENG, H. (2025): “Persuasive Selection in Signaling Games,” *arXiv preprint arXiv:2511.00718*.
- ZHAO, X. (2023): “Auction design by an informed seller: A foundation of reserve price signalling,” *Canadian Journal of Economics/Revue canadienne d’économique*, 56(3), 1161–1190.