Perfect Robust Implementation by Private Information Design*

Maxim Ivanov[†]

February 2022

Abstract

This paper studies the principal-agent framework in which the principal (e.g., the decision-maker or the seller) wants to implement his first-best action that is monotone in the unknown state. The principal privately selects a signal structure about the state of the agent (e.g., the sender or the buyer) whose preferences depend on the action and potentially on the state as well as on a privately known agent's type. The agent privately observes the signal generated by the signal structure and sends a message to the principal. We show that by randomizing between two perfectly informative signal structures, the principal can elicit perfect information from the agent about the state and implement his first-best action regardless of the agent's type. We provide the precise characterization of such signal structures. The key idea is that signal structures form posterior beliefs, which induce actions with opposite reactions to agent's messages. This sustains agent's truthtelling and allows the principal to implement his first-best action upon learning the state. As to the economic application, we consider the bilateral-trade model and show that the seller can extract the full surplus from the privately informed buyer with non-quasilinear preferences and multi-dimensional information.

JEL classification: C72, D81, D82, D83

Keywords: information design; Bayesian persuasion; mechanism design, surplus extraction.

1 Introduction

This paper studies the benefits of private information design as a novel implementation tool in economic environments. The term "private information design" refers to a situation in

*I am grateful to Kalyan Chatterjee, Seungjin Han, Vijay Krishna, and Elliot Lipnowski for helpful comments. A part of this work was written while I visited the University of Toronto. I especially thank Heski Bar-Isaac for hospitality and insightful discussions. Nikolai Penner provided invaluable help with copy-editing the manuscript. All errors are mine.

†Department of Economics, McMaster University, 1280 Main Street Hamilton, ON, Canada L8S 4M4. Phone: (905) 525-9140. Fax: (905) 521-8232. Email: mivanov@mcmaster.ca.

which the precision of the signal structure, i.e., the information technology, which generates players' information and is a choice of the information designer, is privately known to the designer only. At the same time, the information generated by the technology remains private knowledge of the addressee. Specifically, we apply the private information design to the general principal-agent framework. As the main result, we show that the principal can elicit perfect information from the agent and implement her first-best outcome in a simple way by privately designing the agent's signal structure.

Before discussing the results, we start with a brief introduction of the economic environment. The generality of our principal-agent framework stems from several factors. First, the agent's preferences are of the general form rather than, for instance, quasi-linear (as commonly assumed in the mechanism design), or quadratic (which is the standard assumption in models of delegation or communication). Second, the agent's information is multidimensional. Specifically, it is represented by the *state* and the *agent's type* (or simply the *type*). The state affects the payoffs of both players. It reflects, for example, the product quality, which determines the valuation of the object by the buyer and is proportional to the cost of the seller. The agent's type affects the payoff of the agent only. It broadly reflects the characteristics of the agent's payoff function, for example, the sensitivity of her preferences with respect to the state and payment, the magnitude of her risk-aversion, etc. As a special case, the agent's preferences may be independent of the state and/or the type. In our leading economic application—the bilateral-trade model—the key difference between the state and the type is that the agent's payoff is (weakly) monotone in the state, while the dependence on the type can be arbitrary. Another key difference, which is also the third factor, is that the agent is a priori perfectly and privately informed about her type, but uninformed about the state. The uncertainty about the state is a key motive in an interaction between the agent and the principal.

The role of the principal in our framework is dual. First, he wants to implement an action, which is a monotone function of the state. In turn, the action monotonically affects the payoff of the agent. Second, the principal selects a private signal structure of the agent. This implies that the principal randomizes between publicly known signal structures, where the realized signal structure is known to the principal only. Once the principal assigns this signal structure to the agent (without informing the agent about it), the structure generates a signal about the state. The signal is privately observed by the agent, who then sends a report about it to the principal. Finally, the principal takes an action based on the report and the realized signal structure.

The first main result of the paper establishes that the principal can elicit perfect information about all states and implement his first-best decision regardless of the agent's type. That is, perfect implementation is robust to the agent's privately known preferences under some local conditions on her preferences. Moreover, it can be achieved in a simple way by randomizing between two deterministic and perfectly informative signal structures (called signal functions). That is, each signal function maps the state into a single signal, and knowing this function allows one to perfectly infer the state from the agent's signal.

The key idea behind this result is that a randomization between signal structures can sustain agent's truthtelling due to the uncertainty about the impact of her message on the principal's action, which is also based on the realized signal structure. The principal can exploit this uncertainty by selecting signal functions with the 'opposite monotonicities' in the state. That is, a signal generated by one signal function is increasing in the state, while the signal generated by the other signal function is decreasing. As a result, the posterior states, which are induced by the signal under different signal functions, react to the agent's message in opposite ways. Because the principal's action is monotone in the state, and the agent's payoff is monotone in the principal's action, different reactions of posterior states induced by the agent's message create the trade-off for her. Specifically, any distortion of the signal by the agent in an attempt to increase her payoff under one signal function is offset by the marginal losses under the other signal function. By balancing these effects via properly selecting signal functions, the principal can sustain agent's truthful reporting. Finally, because the principal knows the realized signal function, he can infer the state from the agent's report and implement his first-best action.

Furthermore, this result is robust to the agent's privately known type under the separability condition on the agent's payoff. This condition requires the agent's marginal payoff with respect to the action at the principal's first-best action be factorizable into separate functions of the state and the type. Notably, this condition is local and holds, for example, if the payoff function depends on the difference between the principal's action and the state. As a result, it holds for a broad range of payoff functions. An implication of this condition is that the first-order conditions in the agent's problem of maximizing the posterior payoff and, hence, the optimal signal structures, become invariant to the agent's type.

Potential applications of our results can be illustrated by the following example.[1] Consider an organization (e.g., a university, a firm, etc.) with a standard vertical organizational structure: the upper leadership, the middle management, and regular employees. The upper leadership wants to collect the decision-relevant information from employees about some parameter, say, the perceived organizational effectiveness, by conducting a survey. The employees' responses, however, are aggregated and reported by the middle manager whose benefits (reputation, bonuses, career opportunities) are increasing in this parameter. The manager thus has the incentive to misreport the acquired information.[2] In order to preclude her manipulations, the upper leadership can randomize between two surveys whose questions are known to the leadership, but not to the manager.[3] The key idea is that the survey questions are formulated quantitatively, where the relationship between the reported number and the parameter—that is, an increasing or a decreasing signal function—is known only to the upper leadership. One survey, for example, might ask respondents to rate the department's *effectiveness* on a scale of 0 to 100, which corresponds to an increasing signal function. The other survey, however, would ask to evaluate the *ineffectiveness*, with a score of 100 indicating a completely unproductive and/or mismanaged department. Because the middle manager remains uncertain about the actual survey upon receiving employees' responses, then misreporting would distort the signal in the 'wrong direction' with a positive probability and thus stochastically penalize her.

As the leading economic application of our framework, we consider the generalized bilateral-trade model with non-quasilinear preferences and multidimensional information

[1]I am thankful to Heski Bar-Isaac for suggesting it.

[2]In a survey by the consulting firm McKinsey & Company (2007), 36% of top executives responded that managers hide, restrict, or misrepresent information at least "somewhat" frequently.

[3]It is also assumed that the manager cannot access the questions in another way, for example, by spying on employees.

of the buyer. Similar to Bergemann and Pesendorfer (2007) and Li and Shi (2019), the seller determines: i) the terms of trade, i.e., a mechanism, which enforces an allocation and payments on the basis of the buyer's report; and ii) the precision of the buyer's information, i.e., her private signal structure.[4] In our model, the buyer's information is represented by two variables: the state and the type. The state reflects the product quality, which determines the buyer's willingness to pay for the product and can also affect the seller's payoff.[5] The buyer's type affects her payoff only. The players' payoffs are increasing in the state, while the dependence of the buyer's payoff on her type is arbitrary. The buyer is a priori uninformed about the state and perfectly informed about her type. Upon observing a private signal generated by the signal structure and learning the type, the buyer sends a message to the mechanism, which enforces the terms of trade. Similarly to the main framework, the output of the mechanism depends on both the buyer's message and the privately known signal structure. This creates the uncertainty for the buyer regarding the impact of her message on the mechanism. The seller's preferences are quasilinear and depend on the state only. The goal of the seller is to maximize the surplus extracted from the buyer.

Our second main result shows that the seller can extract the full surplus from the buyer upon eliciting the perfect information about the state. This result is also robust to the buyer's knowledge of her type under local conditions on her payoff function and achieved by employing private signal structures similar to those in the implementation model. The only difference is that it they are applied to the target subset of states in which the buyer's willingness to pay (that is, the highest payment, which make her indifferent between trading and taking an outside option) exceeds the seller's payoff from keeping the object.[6] Specifically, the optimal private signal structure randomizes between two signal functions with the opposite reactions of signals to states. Also, the seller sells the product if and only if the posterior state belongs to the target subset and charges the buyer with her willingness to pay in each posterior state. Then, the opposite monotonicities of signal functions in the state imply the opposite monotonicities of buyer's payments in her message under these signal functions. That is, any marginal benefits from distorting the observed signal in an attempt to reduce the payment under one signal function are offset by the higher payment under the other signal function. This trade-off sustains the incentive-compatibility of the mechanism. Furthermore, it does not depend on the absolute value of the buyer's payments. As a result, the mechanism can extract the full surplus upon learning the state by charging the buyer with her willingness to pay, which does not depend on the buyer's type.

In this light, our paper extends the related models in the mechanism design literature in three dimensions. First, it does not assume that the buyer's preferences are quasilinear. It is a conceptual extension. If the parties' preferences are quasilinear in the state, and the seller's payoff from the product is constant, then he can extract the full surplus by informing the buyer whether her valuation of the object is above or below that of the seller and setting

[4] In practice, sellers often allow buyers to try or test the product before purchasing, provide a demo version of the product, or let buyers gather additional information about products in order to assess their quality.

[5] In general, the buyer's and seller's valuations can depend on the same information. For example, if there is another market in which buyers' valuations are correlated with the product quality, then this quality determines the seller's expectation about his opportunity cost and, hence, his valuation of the object.

[6] For low states, the precision of the buyer's information is not substantial. In fact, the seller can select identical perfectly informative signal functions for these states. Thus, the buyer perfectly infers the state from the signal.

the price at the higher posterior value (Saak, 2006). This construction, however, cannot be extended to setups in which the players' preferences take a more general form. For example, if the buyer is risk-averse, then hiding information about the product valuation reduces her willingness to pay and, as a consequence, does not allow the seller to extract the full surplus. In addition, if the seller's payoff also depends on the buyer's type, then he must also learn this information before making a decision about selling the object. Otherwise, there is a chance of selling the object whose value to the seller is above the price. This results in the allocative inefficiency, which does not allow the seller to extract all gains from trade.

Second, we show that the full surplus extraction is robust to the buyer's privately knowledge of her preferences. It is a hard and generally unsolvable problem even if the seller uses the private information design in the simplest model with quasilinear preferences (Krähmer, 2020). As we show, the problem can be circumvented under two conditions on the buyer's payoff function. First, the buyer's willingness to pay is solely determined by the state (i.e., the product quality). Second, the buyer's marginal payoff with respect to the payment at the payment equal to the buyer's willingness to pay can be factorized into separate functions of the state and the type. Notably, both conditions are local as they must hold only for the payment equal to the buyer's willingness to pay or equivalently, along the diagonal in the state-payment space. In the leading example, we show that these conditions hold for all payoff functions that depend on: i) the difference between the state and the payment; and ii) the type. This class of functions is very broad and includes many commons ones, for example, linear-quadratic and hyperbolic absolute risk aversion (HARA) functions. In turn, the latter one includes such functions as linear, quadratic, exponential (constant absolute risk-averse, isoelastic (constant relative risk-averse), and logarithmic. For these functions, the buyer's type represents the parameter, which determines the degree of risk aversion. Because this information is the buyer's private knowledge, the seller may be uncertain about the shape of the payoff function, for instance, linear or exponential. However, he is still able to extract the full surplus by using the private signal functions, which are robust to the buyer's private information about her type.

Third, in contrast to most of the literature on the full surplus extraction, the set of states in our model is continuous. It is also not a purely technical extension. The reason is that eliciting information about discrete states is a less difficult problem, since the set of states that the buyer can mimic and, thus, the set of the buyer's incentive-compatibility constraints, is substantially smaller. Importantly, the buyer cannot distort the state locally by mimicking nearby states. However, the local incentive-compatibility plays a crucial role in the mechanism design (Myerson, 1981). Extracting both perfect information and full surplus from a continuum of states is a more complicated problem as the buyer can mimic states arbitrarily close to the actual one.

Literature

The most related paper from the technical aspect is a recent work by Ivanov and Sam (2022) who first suggested the idea of randomization over signal functions with the opposite monotonicities in the state to extract information from the sender in the cheap-talk model. Despite the similarity in private signal structures, the main difference between the two papers is that our environment is broader in several dimensions. The first dimension is the intensity of the conflict of interest with respect to the principal's action. It is not extreme in cheap-talk models. That is, even though players prefer different outcomes—the state-action pairs—the difference between their first-best outcomes is relatively small. Our setup, on the other hand, allows for any level of conflict of interest, including extreme cases. For example, in the bilateral-trade model, the buyer prefers to minimize her payment to the seller while the seller wants to maximize it regardless of characteristics of the model. Because of the extreme conflict of interest with respect to payment, eliciting meaningful information from the buyer becomes a substantially harder problem. For example, using the mechanism design alone does not allow the seller to extract the full surplus from the privately informed buyer (Myerson, 1981). The second crucial difference between the two papers is that standard cheap-talk models, including Ivanov and Sam (2022), assume the monotonicity of the agent's first-best action in the state. That is, both players—the agent and the principal—prefer higher actions if the state increases.[7] Our setup does not require this assumption. Furthermore, the agent's payoff can be independent of the state. Finally, Ivanov and Sam (2022) focus mostly on the case of the ex-ante uninformed sender, while the agent in our model can be privately informed about her type. As a result, our framework covers a substantially broader class of models than their.

In the context of the mechanism design, the most relevant paper to ours is Krähmer (2020) who first used the private information design to demonstrate the possibility of the full surplus extraction in the bilateral-trade model. There are several key distinctions between our and Krähmer's papers. First, he considers the quasilinear preferences of the buyer, whereas the buyers' preferences in our model are of a general form. Second, Krähmer demonstrates the full-surplus extraction result for an a priori uninformed buyer only, while we establish it even if the buyer is privately informed about her type. Third, Krähmer's and our constructions utilize conceptually different properties of private signal structures to extract the full information and surplus from the buyer. Signal structures in Krähmer (2020) are designed to monitor the buyer and detect his deviations from truthtelling. In particular, each signal structure is endowed with an individual signal set. This set is privately known to the seller, while the buyer privately observes the signal realization only. Thus, after receiving an 'incorrect' signal, the seller infers that the buyer lies and takes a penalizing action. A threat of this action enforces the buyer's incentive-compatibility.[8] In our model, the signal structures share a common signal space, which implies that the buyer's deviations are undetectable. However, a randomization between signal structures creates the uncertainty for the buyer about her payments contingent on the realized signal structure. As a result, signal distortions create a trade-off between her marginal benefits and losses. Fourth, a private signal structure in our model randomizes between two deterministic signal structures, whereas private signal structures in Krähmer (2021) are based on a randomization over a continuum of signal structures with individual signal spaces. Finally, our construction employs the continuity of the state space, whereas Krähmer's approach relies on its discreteness.

Our paper is also related to the literature on the buyer's surplus extraction by using the mechanism and/or information design. This topic drew significant attention due to a seminal work by Crémer and McLean (1988), who demonstrated the possibility of the full surplus extraction by using the mechanism design approach, which employs the correlation

[7] Formally, the payoff function of the agent must be strictly supermodular.

[8] Krähmer (2021) uses a similar idea in the cheap-talk context.

among buyers' values. On the other hand, recent developments in information design inspired by Kamenica and Gentzkow (2011) have demonstrated that it can also be a powerful tool for the seller to extract surplus from the buyer(s). Lewis and Sappington (1994), Johnson and Myatt (2006), Bergemann and Pesendorfer (2007), Esö and Szentes (2007), Ganuza and Penalva (2010), Li and Shi (2019), and Ivanov (2021) show that the seller(s) can benefit by designing or affecting the buyers' information about the product in standard bilateral trade or auction environments. Ivanov (2013) and Hwang et al. (2019) demonstrate the same effect in competitive markets with horizontally differentiated products. Bergemann and Wambach (2015) show that the seller can extract the full surplus by disclosing information to the buyer gradually over time. Zhu (2021) and Larionov et al. (2021) consider the implementation problem with multiple agents who can acquire additional information about their types. They demonstrate the possibility of implementing any social choice rule by using Shannon's (1949) encryption technique. Pastrian (2021) demonstrates the full surplus extraction in the reduced form framework of McAfee and Reny (1992) with a behavioral subset of buyer's types that are always truthful. Fu et al. (2021) consider a setup with a finite number of possible distributions of buyers' values, where the seller has access to a finite number of independent draws from the true distribution. They establish that the full surplus extraction is feasible if the number of draws is large enough. Neither of these papers, however, considers private signal structures.

The rest of the paper is organized as follows. Section 2 introduces the general implementation model. Section 3 provides the main result for this framework. Section 4 applies these results to the bilateral-trade model. Finally, Section 5 concludes the paper.

2 Model

We consider the framework with two players, an agent (she) and a principal (he). The principal's goal is to implement his *first-best* (or *ideal*) *action* $y(\theta) \in \mathbf{A} \subset \mathbb{R}$ from a closed and convex action set $\mathbf{A}$, where $y(\theta)$ depends on the ex-ante unknown *state* θ. The state θ is a random variable drawn from the *state space* $\Theta = \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbf{A}$ according to a continuous density $f(\theta)$, such that $f > 0$. (Hereafter, $u > 0$ for a function $u : X \rightarrow Y$ means $u(x) > 0$ for all $x \in X$.) We assume that $y(\theta)$ is continuous and strictly increasing in θ. As a normalization, it is without loss of generality to put $y(\theta) = \theta$.[9] Thus, Θ is also a set of the principal's ideal actions induced by states.

The agent's payoff is given by the function $V(a, \theta, \gamma)$, which depends on the principal's *action* a, the state θ, and the privately known *buyer's type* (or *type*) γ. As a special case, V can be independent of θ and/or γ. The type γ is drawn randomly from the type space $\mathbf{T}$ according to the cdf $H(\gamma)$. In general, we do not impose any restrictions on $\mathbf{T}$ and $H(\gamma)$, however, there can be bounds on $\mathbf{T}$ for some specific V in order to guarantee that the payoff function respects the conditions imposed below. The variables θ and γ are independent.

We assume that $V(a, \theta, \gamma)$ is continuously differentiable in (a, θ) and pseudo-concave in a for all $(a, \theta, \gamma) \in \Theta^2 \times \mathbf{T}$.[10] Next, consider the function $V'_a(a, \theta, \gamma)$, which represents the

[9] Otherwise, if $y(\theta) \neq \theta$, then the monotone transformation $z = y(\theta)$ results in $y(z) = z$.

[10] A differentiable function $\mathcal{V}(a)$ is *pseudo-concave* on a convex set $\mathbf{A} \subset \mathbb{R}$ if for every $(a, y) \in \mathbf{A}^2$, $\mathcal{V}(a) < \mathcal{V}(y)$ implies $\mathcal{V}'(a)(y - a) > 0$. If $\mathcal{V}'(y) = 0$ for $y \in \mathbf{A}$, then y is a maximizer of $\mathcal{V}$ (Proposition 2.4

agent's marginal payoff with respect to the principal's action. This function has another interpretation, which we employ through the paper. Because the principal wants to match the action to the state, $V_a'(a, \theta, \gamma)$ can be interpreted as the agent's marginal payoff from inducing the principal's (wrong) posterior belief that the state is a rather than θ. In other words, it represents the agent's marginal benefits or losses from manipulating the principal's action via his posterior belief. We impose the following separability condition on this function. Specifically, for a given subset of states $\Theta_0 \subset \Theta$, we require that $V_a'(\theta, \theta, \gamma)$ at $a = \theta$ can be expressed as

$$V_a'(\theta, \theta, \gamma) = g(\gamma)\zeta(\theta) \text{ for } (\theta, \gamma) \in \Theta_0 \times \mathbf{T}, \tag{1}$$

where $g(\gamma) > 0$ for all $\gamma \in \mathbf{T}$ and $\zeta(\theta) \neq 0$ for all $\theta \in \Theta_0$. That is, the agent's marginal payoff with respect to the action (or, equivalently, the induced posterior state a) at point $a = \theta$ can be factorized into $g(\gamma)$ and $\zeta(\theta)$, which are sole functions of the agent's type γ and state θ, respectively. For a given state θ, the condition (1) is local as the factorization is required at the single point $a = \theta$ only. Because $g > 0$, this implies that $V_a'(\theta, \theta, \gamma) \gtreqless 0$ if and only if $\zeta(\theta) \gtreqless 0$ for $\theta \in \Theta_0$. For concreteness, hereafter we assume that $\zeta(\theta) < 0$ for $\theta \in \Theta_0$. This results in $V_a'(\theta, \theta, \gamma) < 0$ for $(\theta, \gamma) \in \Theta_0 \times \mathbf{T}$. The case of $\zeta(\theta) > 0$ is symmetric.

Signal structures/functions. A *signal structure* $\xi(s|\theta)$ is a probability distribution over signals s conditional on the state θ. A *signal set* $\mathbf{S}_\xi \subset \mathbb{R}$ is the support of ξ. A signal structure ξ is called a *signal function* if it maps each state $\theta \in \Theta$ into a signal $s = \xi(\theta)$. In this case, the signal set $\mathbf{S}_\xi$ is the image of ξ. A signal function is *perfectly informative* if it is injective. Hereafter, we restrict the codomain $\mathbf{C}$ of each signal function ξ by its image $\mathbf{S}_\xi$. Hence, a perfectly informative $\xi : \Theta \to \mathbf{C}$ is bijective and thus has the inverse function (hereafter called the *inverse*) $\varphi = \xi^{-1} : \mathbf{C} \to \Theta_\varphi$, where Θ_φ is the image of φ. Similarly to signal functions, we restrict the codomain of φ by its image Θ_φ. Because of the restrictions on the codomains of ξ and φ, the existence of a function $\xi : \Theta \to \mathbf{C}$ (or $\varphi : \mathbf{C} \to \Theta$) also implies that the image of ξ is $\mathbf{C}$ (or Θ). Let $\mathcal{I}$ be the space of all signal structures. A *private signal structure* $\rho \in \Delta\mathcal{I}$ is a probability distribution over signal structures whose realization ξ is privately observed by the principal. Denote $\rho(\xi)$ the probability of drawing ξ by ρ, and $\mathcal{I}_\rho$ the support of ρ.

Timing. The game is played as follows. The agent is a priori uninformed about θ and perfectly informed about γ. That is, her information about θ is determined by the prior density $f(\theta)$. At the beginning of the game, the principal publicly selects a private signal structure $\rho \in \Delta\mathcal{I}$ and an action $y_\xi(m)$.[11] Then, the state θ and the signal structure $\xi \in \mathcal{I}_\rho$ are randomly drawn according to f and ρ, respectively, where ξ becomes the private information of the principal.[12] The agent then privately observes a signal s generated by ξ from θ and sends a message m from the message space $\mathbf{M}$ to the principal who takes an action $y_\xi(m)$. Hereafter, we assume that $\mathbf{M} = \mathbf{S} = \bigcup_{\xi \in \mathcal{I}_\rho} \mathbf{S}_\xi$, that is, the message space is large enough to convey all information about signals.

in Hadjisavvas et al., 2005).

[11]Formally, the principal's $y_\xi(m)$ action can also be based on the private signal structure ρ. Because neither of our results is driven by this dependence, we omit it for simplicity of notation.

[12]Since the probability $\rho(\xi)$ does not depend on θ, ξ and θ are independent random variables. Hence, knowing ξ does not provide any additional information about θ to the principal.

Two comments are necessary here. First, the principal can either commit to action $y_\xi(m)$ ex-ante, or it can be interim-optimal, that is, endogenously determined as a solution to the problem of maximizing the the principal's expected payoff conditional on the available information, ξ and m. In other words, the nature of the principal's action $y_\xi(m)$ is not essential. Second, because $y_\xi(m)$ depends on the signal structure ξ, this implies that ξ is verifiable ex-post. That is, the agent can infer ex-post all information that can be jointly derived from her signal s and the signal structure ξ. Obviously, this information is weakly more precise than the information that the principal can derive from ξ and m (with equality if the agent reports truthfully).

Conditional on ρ and $y_\xi(m)$, the following subgame is the decision problem with a privately and imperfectly informed agent. A strategy of the agent $m(s,\gamma,\rho) \in \Delta\mathbf{S}$ specifies a (possibly random) message m given her information: the observed signal $s \in \mathbf{S}$, the type γ, and the private signal structure ρ. An *optimal strategy* $m^*(s,\gamma,\rho)$ is a maximizer of the agent's posterior payoff

$$EV(m|s,\gamma,\rho) = \int_{\mathcal{I}_\rho}\int_{\Theta} V(y_\xi(m),\theta,\gamma)\, dq(\theta|s,\rho)\, d\rho(\xi), \tag{2}$$

where $q(\theta|s,\rho) \in \Delta\Theta$ is the agent's *posterior belief*, which is a probability distribution over θ derived from s and ρ by using Bayes' rule.[13] We say that the state θ is *posterior* and *induced* by a signal s under a private signal structure ρ if θ is in the support of $q(.|s,\rho)$. In particular, if the support $\mathcal{I}_\rho$ of ρ contains only perfectly informative signal functions ξ, then $\theta = \varphi_\xi(s) = \xi^{-1}(s)$ represents the *posterior state* induced by a signal s under a signal function ξ, and thus the support of $q(\theta|s,\rho)$ is given by $\{\varphi_\xi(s) : \xi \in \mathcal{I}_\rho\}$.

Finally, the agent's *truthful* strategy is optimal under ρ if $m^*(s,\gamma,\rho) = s$ is in the set of maximizers of (2) for ρ and all $(s,\gamma) \in \mathbf{S}\times\mathbf{T}$. Equivalently,

$$EV(s|s,\gamma,\rho) = \max_{m\in\mathbf{S}} EV(m|s,\gamma,\rho) \text{ for all } (s,\gamma) \in \mathbf{S}\times\mathbf{T}. \tag{3}$$

3 Perfect information extraction and implementation

Before starting the general construction (hereafter, a *construction*) of private signal structures, which elicit the perfect information about the state from the agent and allow the principal to implement his ideal action, we provide an illustrative example. This example shows that perfect implementation can be achieved by randomizing between two perfectly informative signal functions. Notably, we consider the agent's preferences, which are independent on the state and the agent's type. (The second example below considers general preferences, which depend on both state and type.) The state-independent preferences of the agent make the principal's problem more difficult because two reasons. First, because the information about the state has no value to the agent, the principal cannot exploit the agent's incentive to acquire this information for his benefit.[14] Second, the agent is more willing to

[13]Since γ and θ are independent, $q(\theta|s,\rho)$ does not depend on γ.

[14]In the case of state-dependant agent's preferences, her incentives to acquire information can play a critical role for information extraction. For example, Ivanov (2015, 2016) shows that the principal can elicit perfect

share her information upon learning it if her preferences are closely aligned with those of the principal. Since the principal's ideal action depends on the state, while the agent's one does not, the players' preferences are substantially conflicting. Together, these factors suppress the agent's incentives to share her information truthfully.

3.1 Example A: state-independent agent's preferences

Suppose that the state is uniformly distributed on the unit interval, i.e., $f(\theta) = 1, \theta \in \Theta = [0, 1]$ and the agent's payoff function is of the form

$$V(a, \theta, \gamma) = V(a) = -a^b,$$

where $a \in \mathbf{A} = \mathbb{R}_+$ and $b \geq 1$ is a known parameter. Because V is strictly decreasing in a for $a \geq 0$, the agent's payoff is maximized at $a_0 = 0$.[15]

Now, consider the private signal structure ρ^o, which randomizes with equal probabilities between two perfectly informative signal functions

$$\xi_1(\theta) = \theta \text{ and } \xi_2(\theta) = \left(1 - \theta^{\frac{b+1}{2}}\right)^{\frac{2}{b+1}}. \tag{4}$$

Because the images of functions ξ_1 and ξ_2 are identical and equal to $\mathbf{S} = [0, 1]$, then any agent's deviation from truthtelling is undetectable by the principal. On the other hand, the agent cannot infer the realized signal function and, thus, the state θ upon observing the signal s. In particular, a signal s generates the agent's posterior beliefs

$$q(\theta|s, \rho^o) = \Pr\{\theta|s, \rho^o\} = \begin{cases} \frac{1}{1+|\varphi_2'(s)|} & \text{if } \theta = \varphi_1(s), \\ \frac{|\varphi_2'(s)|}{1+|\varphi_2'(s)|} & \text{if } \theta = \varphi_2(s), \text{ and} \\ 0 & \text{if } \theta \notin \{\varphi_1(s), \varphi_2(s)\}, \end{cases} \tag{5}$$

where

$$\varphi_1(s) = \xi_1^{-1}(s) = s \text{ and } \varphi_2(s) = \xi_2^{-1}(s) = \left(1 - s^{\frac{b+1}{2}}\right)^{\frac{2}{b+1}}$$

are the inverses of ξ_1 and ξ_2, respectively. Denote

$$q_i(s, \rho) = \Pr\{\theta = \varphi_i(s)|s, \rho\} = q(\varphi_i(s)|s, \rho) \tag{6}$$

the probability of the posterior state $\theta_i = \varphi_i(s)$.

Suppose that the principal believes that the agent is truthful. Then, sending a message $m \in \mathbf{S}$ induces the action

$$y_{\xi_i}(m) = \theta_i = \varphi_i(m), i = 1, 2$$

information from the agent and implement the ideal actions in the cheap-talk framework by exploiting these incentives in a dynamic way.

[15]Formally, $V_a'(0) = 0$ for $b > 1$, which violates the condition $V_a' < 0$. However, since $V(a)$ is strictly decreasing in a, neither of our results is affected by this technicality.

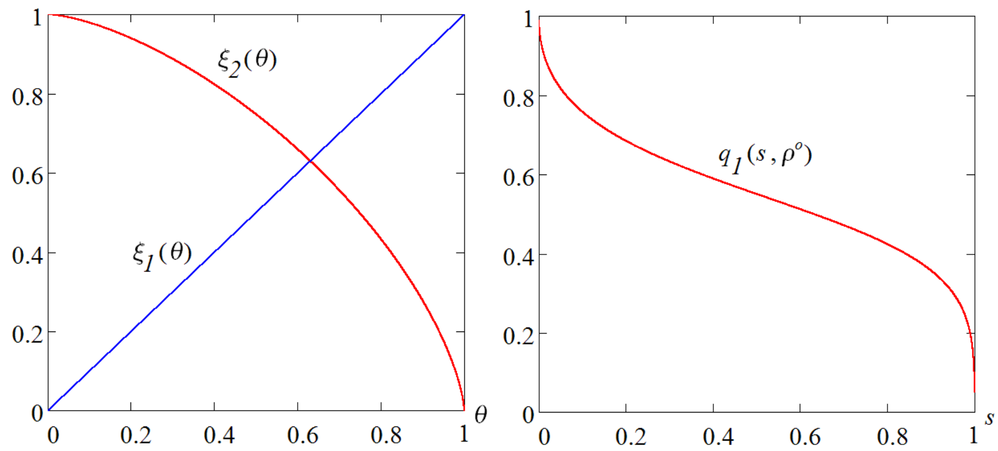

Figure 1: Signal functions $\xi_i(\theta), i = 1,2$ and the posterior probability $q_1(s,\rho^o)$ for $f(\theta) = 1$ and $V(a) = -a^2$.

under the signal function ξ_i. Therefore, the agent posterior payoff (2) is given by

$$\begin{aligned} EV(m|s,\rho^o) &= q_1(s,\rho^o)V(y_{\xi_1}(m)) + q_2(s,\rho^o)V(y_{\xi_2}(m)) \\ &= q_1(s,\rho^o)V(\varphi_1(m)) + q_2(s,\rho^o)V(\varphi_2(m)) \\ &= -q_1(s,\rho^o)(\varphi_1(m))^b - q_2(s,\rho^o)(\varphi_2(m))^b. \end{aligned}$$

As an illustration, consider the quadratic payoff function $V(a) = -a^2$, i.e., $b = 2$. Fig. 1 depicts the signal functions ξ_1, ξ_2, and the posterior probability $q_1(s,\rho^o)$ for this function. In this case,

$$\varphi_2(s) = \xi_2(s) = \left(1 - s^{3/2}\right)^{2/3},$$

$$q_1(s,\rho^o) = \frac{\left(1 - s^{3/2}\right)^{1/3}}{s^{1/2} + \left(1 - s^{3/2}\right)^{1/3}}, q_2(s,\rho^o) = \frac{s^{1/2}}{s^{1/2} + \left(1 - s^{3/2}\right)^{1/3}}, \text{ and}$$

$$EV(m|s,\rho^o) = -\frac{\left(1 - s^{3/2}\right)^{1/3} m^2 + s^{1/2}\left(1 - m^{3/2}\right)^{4/3}}{s^{1/2} + \left(1 - s^{3/2}\right)^{1/3}}.$$

By using basic calculations, it is easy to verify that $EV(m|s,\rho^o)$ is maximized at $m = s$ for all $s \in \mathbf{S}$, i.e., the agent reports truthfully.

Intuitively, this example demonstrates the keys factors of private signal structures, which sustain agent's truthtelling. Specifically, the induced posterior states $\theta_1 = \varphi_1(m)$ and $\theta_2 = \varphi_2(m)$ associated with signal structures ξ_1 and ξ_2, respectively, react oppositely to an agent's message m. Next, the principal's ideal action $y(\theta_i) = \theta_i$ is monotone in the induced posterior state $\theta_i, i = 1,2$. Finally, the agent's payoff $V(a)$ is monotone in the principal's action a. Together, these factors create the trade-off for the agent: any distortion of the signal in an attempt to marginally benefit from the receiver's action taken under one signal function are offset by the marginal losses caused by the action taken under the other signal function.

At the same time, the magnitude of the trade-off between the agent's marginal benefits and losses from distortions is driven by the shapes of signal functions, specifically, their inverses φ_1 and φ_2. Their effect on the trade-off is dual. First, they determine the marginal effects of an agent's message on the agent's posterior payoff via actions taken under different signal structures. Second, they reallocate the agent's posterior beliefs between the posterior states. We now explain in detail the relationship between the shapes of the inverses and their overall effect on the agent's incentives to report truthfully.

In order to explain the first effect, recall that the principal's action $a_i = y_{\xi_i}(m)$ taken under a signal structure ξ_i in response to the agent's message m is equal to the induced posterior state θ_i, which is given by the inverse function $\varphi_i(m)$. That is, $a_i = y_{\xi_i}(m) = \theta_i = \varphi_i(m)$. As a result, the shape of $\varphi_i(m)$ determines the marginal effect of an agent's message m on the principal's action a_i. Next, note that the agent's payoff function $V(a)$ is strictly concave in the principal's action a for $b > 1$. Hence, the agent's marginal benefits $|V'(a)|$ from a decrease in a are larger for high values of a. As a result, the overall effect of signal distortions on the agent's posterior payoff depends on the interaction between the inverses φ_1 and φ_2 and the marginal payoff $V'(a)$. Specifically, note that the first inverse φ_1 is linear in s. Therefore, the marginal effect of the signal s on the principal's action $a_1 = \varphi_1(s)$ is constant, $\varphi_1'(s) = 1$. On the other hand, because φ_2 is strictly concave, the absolute value of the marginal effect $|\varphi'(s)|$ on $a_2 = \varphi_2(s)$ is increasing in s. This implies that the 'counter-moving' action a_2 will increase at the faster rate in response to downward distortions if the signal s is high, whereas the rate of a decrease in the 'co-moving' action a_1 is constant. In other words, the marginal penalty from downward distortions caused by the unfavorable action a_2 relative to the benefits from the favorable action a_1 is increasing in an agent's signal s. Because the agent has the stronger incentives to decrease the action a_1 if it high and, thus, is associated with a high signal s, then understating such signals will result in the higher losses from the action a_2. The balance between these forces sustains agent's truthtelling.

Second, the shapes of the inverses reallocate the agent's posterior beliefs between posterior states $\theta_1 = \varphi_1(s)$ and $\theta_2 = \varphi_2(s)$. As follows from (5), the higher absolute value of the slope $|\varphi_2'|$ of the inverse φ_2 decreases the probability $q_1(s, \rho)$ of the posterior state $\theta_1 = \varphi_1(s)$ and increases the probability $q_2(s, \rho)$ of the posterior state $\theta_2 = \varphi_2(s)$. Because the value of $|\varphi_2'(s)|$ is increasing in s, the posterior probability of θ_2 and, hence, the probability of the penalizing action a_2 is increasing in s.

Given these observations, it is easy to notice the complementarity between the marginal effects of actions and their probabilities on the agent's incentives to report truthfully. Specifically, an increase in the slope $|\varphi_2'(s)|$ in response to the higher signal s results in both the higher marginal penalty from the counter-moving action a_2 and the higher probability $q_2(s, \rho)$ of inducing this action. In other words, the higher magnitude of one effect intensifies the second effect as well. Altogether, this implies that the agent's expected losses from downward distortions are increasing in her signal.

Finally, it is worth noting that the agent remains truthful regardless of the precision of her information (measured, for instance, by the variance of posterior states). In fact, if $\hat{s} = 2^{-2/3} \simeq 0.63$, then the agent is perfectly informed about the posterior state $\hat{\theta} = \hat{s}$. However, she still cannot use this information in her favor.

3.2 Optimal private signal structures

We start the general construction with the following lemma. It demonstrates how the sender's posterior beliefs are shaped by a private signal structure, which randomizes between two perfectly informative signal functions.

Lemma 1 *(Ivanov and Sam, 2022) Consider a private signal structure* ρ*, which randomizes between differentiable signal functions* $\xi_1 : \Theta \to \mathbf{S}$ *and* $\xi_2 : \Theta \to \mathbf{S}$ *with probabilities* $p_1 \in (0,1)$ *and* $p_2 = 1 - p_1$*, respectively, where* $\mathbf{S} = [s_0, s_1]$, $s_1 > s_0$*, and* $\xi_i' \neq 0$*. Denote* $\varphi_i = \xi_i^{-1}$ *the inverse of* ξ_i*. Then,*

$$q_i(s, \rho) = \frac{p_i f(\varphi_i(s)) |\varphi_i'(s)|}{p_1 f(\varphi_1(s)) |\varphi_1'(s)| + p_2 f(\varphi_2(s)) |\varphi_2'(s)|}. \tag{7}$$

Intuitively, the lemma highlights a key feature of signal functions, specifically, the possibility to induce the agent's posterior beliefs $q_i(s, \rho)$ about posterior states $\theta_1 = \varphi_1(s)$ and $\theta_2 = \varphi_2(\theta)$ anywhere between 0 and 1 by varying the ratio $\frac{|\varphi_2'(s)|}{|\varphi_1'(s)|}$ of the slopes of the inverses φ_i'. To see this feature, suppose $p_1 = p_2 = \frac{1}{2}$ and f is uniform, i.e., $f(\theta) = \frac{1}{\bar{\theta} - \underline{\theta}}$. It follows then that $q_1(s, \rho) = \frac{1}{1 + \frac{|\varphi_2'(s)|}{|\varphi_1'(s)|}}$ and $q_2(s, \rho) = 1 - q_1(s, \rho)$. By varying the ratio $\frac{|\varphi_2'(s)|}{|\varphi_1'(s)|}$ between 0 and ∞, the principal can induce $q_i, i = 1, 2$ anywhere between 0 and 1. An implication of this lemma is that the principal can reallocate the agent's posterior beliefs from less to more favorable posterior states, for example, to those in which the agent's and the principal's preferences are more aligned or the principal receives higher benefits.[16]

General construction. Consider the signal space $\mathbf{S} = [\underline{s}, \bar{s}]$ and a private signal structure ρ, which randomizes with equal probabilities between two perfectly informative signal functions ξ_1 and ξ_2 with the inverses $\varphi_i = \xi^{-1} : \mathbf{S} \to \Theta$ defined as follows. First, select a differentiable $\varphi_1 : \mathbf{S} \to \Theta$, such that $\varphi_1' > 0$. Thus, $\varphi_1(\underline{s}) = \underline{\theta}$ and $\varphi_1(\bar{s}) = \bar{\theta}$. The principal's problem is to derive $\varphi_2 : \mathbf{S} \to \Theta$, such that the private signal structure ρ sustains the agent's truthtelling and, thus, allows the principal to implement $y(\theta)$ upon inferring θ from m and ξ_i. Importantly, the signal sets, that is, the images of functions ξ_1 and ξ_2 are identical and equal to $\mathbf{S}$. First, this implies that the agent is unable to infer the realized ξ_i upon observing the signal s. Second, any agent's deviation from truth telling is undetectable by the principal.

Given the agent's truthful strategy $m^*(s, \gamma, \rho) = s$ for $(s, \gamma) \in \mathbf{S} \times \mathbf{T}$, the principal's best response to message m under the signal structure ξ_i is

$$y_{\xi_i}(m) = \varphi_i(m), i = 1, 2. \tag{8}$$

[16] Ivanov and Sam (2022) explain in detail how the slopes of the inverses $\varphi_i', i = 1, 2$ can be used to reallocate the posterior probabilities $q_i(s, \rho)$ for principal's benefits in the cheap-talk framework.

The agent's problem upon receiving a signal s is to maximize her posterior payoff (2) over messages $m \in \mathbf{S}$. Using (8), the posterior payoff can be expressed as

$$EV(m|s,\gamma,\rho) = \sum_{i=1}^{2} q_i(s,\rho) V(\varphi_i(m), \varphi_i(s), \gamma) \text{ for } (m,s,\gamma) \in \mathbf{S}^2 \times \mathbf{T}. \tag{9}$$

Then, the agent's marginal posterior payoff is given by

$$\begin{aligned} \frac{\partial}{\partial m} EV(m|s,\gamma,\rho) &= \sum_{i=1}^{2} q_i(s,\rho) V'_m(\varphi_i(m), \varphi_i(s), \gamma) \\ &= \sum_{i=1}^{2} q_i(s,\rho) V'_a(\varphi_i(m), \varphi_i(s), \gamma) \varphi'_i(m). \end{aligned} \tag{10}$$

The truthful strategy of the agent is optimal if and only if

$$EV(s|s,\gamma,\rho) = \max_{m\in\mathbf{S}} EV(m|s,\gamma,\rho) \text{ for all } (s,\gamma) \in \mathbf{S} \times \mathbf{T}. \tag{11}$$

By using (10), the first-order condition for the agent's maximization problem (11) is

$$\begin{aligned} \frac{\partial}{\partial m} EV(m|s,\gamma,\rho)|_{m=s} &= \sum_{i=1}^{2} q_i(s,\rho) V'_a(\varphi_i(s), \varphi_i(s), \gamma) \varphi'_i(s) \\ &= 0 \text{ for all } (s,\gamma) \in \mathbf{S} \times \mathbf{T}. \end{aligned} \tag{12}$$

Next, invoking condition (1) results in

$$V'_a(\varphi_i(s), \varphi_i(s), \gamma) = g(\gamma) \zeta(\varphi_i(s)),$$

where $g > 0$ and $\zeta > 0$. This means that (12) is independent of γ and thus can be written as

$$\frac{\partial}{\partial m} EV(m|s,\rho)|_{m=s} = \sum_{i=1}^{2} q_i(s,\rho) \zeta(\varphi_i(s)) \varphi'_i(s) = 0 \text{ for all } s \in \mathbf{S}.$$

By using Lemma 1, (12) can be written as a separable differential equation with respect to φ_2 for a given φ_1:

$$\varphi'_1(s) |\varphi'_1(s)| f(\varphi_1(s)) \zeta(\varphi_1(s)) + \varphi'_2(s) |\varphi'_2(s)| f(\varphi_2(s)) \zeta(\varphi_2(s)) = 0, \tag{13}$$

The equation (13) is identical to equation (9) in Ivanov and Sam (2022) for probabilities $p_1 = p_2 = \frac{1}{2}$. By applying their analysis, it follows that $\varphi'_2 < 0$ and (13) can be expressed as

$$\varphi'_1(s) h(\varphi_1(s)) = -\varphi'_2(s) h(\varphi_2(s)), \tag{14}$$

where

$$h(\theta) = \sqrt{-f(\theta)\zeta(\theta)}.$$

Note that $h > 0$ due to $f > 0$ and $\zeta < 0$.

By Lemma 2 in Ivanov and Sam (2022), the solution to (14) with the boundary condition $\varphi_2(\underline{s}) = \bar{\theta}$ is given by:

$$\varphi_2(s) = \Psi^{-1}\left(\Psi(\underline{\theta}) + \Psi(\bar{\theta}) - \Psi(\varphi_1(s))\right), \tag{15}$$

where $\Psi(x) = \int h(x)\,dx$ is the antiderivative of h.[17]

In general, a pair of inverses $\varphi_i : \mathbf{S} \to \Theta, i = 1, 2$ related by (15) is not necessarily a solution to the agent's maximization problem (11) as the second-order conditions might not hold. The following regularity condition addresses this issue.

Condition 1 *Given* $\Theta_0 \subset \Theta$, $\nu(a, \theta, \gamma) = \frac{V_a'(a,\theta,\gamma)}{h(a)}$ *is decreasing in* a *for all* $(a, \theta, \gamma) \in \Theta_0^2 \times T$.

Notably, this condition is imposed on the model primitives only, that is, the payoff function $V(a, \theta, \gamma)$ and the prior density $f(\theta)$. Therefore, the truthful and thus perfectly informative strategy can be optimal for various pairs $\{\varphi_1, \varphi_2\}$ parameterized by the inverse φ_1.

It is easier to explain this condition by noting that $\nu(a, \theta, \gamma)$ is the ratio of two functions, $V_a'(a, \theta, \gamma)$ and $h(\theta) = \sqrt{-f(\theta)\zeta(\theta)}$. The first function $V_a'(a, \theta, \gamma)$ is the marginal payoff with respect to the principal's action or equivalently, the induced posterior state. It is decreasing in a if $V(a, \theta, \gamma)$ is concave in a, and increasing if $V_a'(a, \theta, \gamma)$ is convex in a. The second function $h(\theta)$ reflects the marginal payoff $\zeta(\theta)$ weighted by the prior density $f(\theta)$. In this light, Condition 1 requires the function $V_a'(a, \theta, \gamma)$ be 'not very convex' in a, and the agent's marginal payoff $\zeta(\theta)$ weighted by the prior density $f(\theta)$ be 'relatively decreasing' in θ (since $V_a' < 0$ and $h > 0$ imply $\nu < 0$).[18]

Given these preliminaries, the following theorem establishes the main result of the paper. Consider the private signal structure ρ^*, which randomizes with equal probabilities between signal functions $\xi_1 = \varphi_1^{-1}$ and $\xi_2 = \varphi_2^{-1}$, such that the relationship between φ_1 and φ_2 is given by (15). Then ρ^* allows the principal to elicit the perfect information about θ from the agent and implement his ideal action $y(\theta)$ under the above regularity condition.

Theorem 1 *Suppose* V *satisfies (1) and* (f, V) *satisfy Condition 1 for* $\Theta_0 = \Theta$. *Consider differentiable* $\varphi_1 : \mathbf{S} \to \Theta, i = 1, 2$, *such that* $\varphi_1' > 0$ *and* φ_2 *is given by (15), and the private signal structure* ρ^* *that randomizes between* $\xi_1 = \varphi_1^{-1}$ *and* $\xi_2 = \varphi_2^{-1}$ *with equal probabilities. Then the agent's truthful strategy is optimal under* ρ^* *for all* $(\theta, \gamma) \in \Theta \times \mathbf{T}$.

This theorem generalizes Theorem 1 in Ivanov and Sam (2022) in three dimensions. First, in their model the agent is ex-ante uninformed, while our setup allows the agent to be privately informed about her type γ. Second, their model assumes the existence of the agent's ideal action $y_A(\theta)$ for each state θ. Third, they assume the strict supermodularity of the agent's preferences in (a, θ). This implies the dependence of the agent's payoff on state θ. Our setup does not require the existence of the agent's ideal action, the dependence of

[17] Note that $\varphi_1(\bar{s}) = \bar{\theta}$ implies $\varphi_2(\bar{s}) = \underline{\theta}$, which means that functions φ_1 and φ_2 have identical images $\mathbf{S}$.

[18] If $V_a' > 0$, then $\nu > 0$. In this case, Condition 1 implies that the prior density $f(\theta)$ must be 'relatively increasing' in θ.

her payoff function on the state, or the supermodularity. All these assumptions are replaced with the strict monotonicity of the agent's payoff in the principal's action. Specifically, recall that the opposite monotonicities of inverses $\varphi_i(s), i = 1, 2$ in signal s and the monotonicity of the principal's action $y(\theta) = \theta$ in θ imply that the principal's actions $a_i = \varphi_i(m), i = 1, 2$ under different signal functions ξ_i react oppositely in response to the agent's message m. Then, the monotonicity of the agent's payoff V in the principal's action a implies that the opposite reactions of $\varphi_i(m), i = 1, 2$ to m are mapped in the opposite marginal payoffs to the agent. That is, agent's misreporting in an attempt to obtain extra gains under one signal function are offset by the extra losses under the other signal function. These marginal effects are balanced by the relationship (15) between the inverses $\varphi_i(s) = 1, 2$ in order to sustain agent's truthtelling. Notably, the logic above does not depend on the concavity of the agent's preferences in a. As a result, theorem holds for non-concave payoff functions V as long as they satisfy Condition 1 and (1).

At the same time, the proofs of the two theorems share common features. In both theorems, the critical part is to establish the optimality of the agent's truthtelling strategy under the private signal structure ρ. The main technical tension comes from three facts. First, the agent's posterior payoff EV is a convex combination of pseudo-concave functions $V(a, \theta, \gamma)$, which is generally not pseudo-concave. Second, each payoff $V(\varphi_i(m), \varphi_i(s), \gamma)$ is a composite function of V and φ_i. As a result, the pseudo-concavity of the composite function is violated if the pseudo-concavity of V in a is dominated by the convexity of $\varphi_i(s)$. Third, φ_1 and φ_2 are functionally dependent by (15). Therefore, the posterior payoff is pseudo-concave if each composite function $V(\varphi_i(m), \varphi_i(s), \gamma), i = 1, 2$ is pseudo-concave in m, and a convex combination of these functions is also pseudo-concave.

Condition 1 resolves all three issues. Specifically, the necessary and sufficient condition for the pseudo-concavity of $EV(m|s, \gamma, \rho)$ is the pseudo-monotonicity of the marginal posterior payoff $\frac{\partial}{\partial m} EV(m|s, \gamma, \rho)$ (Hadjisavvas et al., 2005).[19] To show that this function is pseudo-monotone, we use the results by Quah and Strulovici (2012) who establish conditions for the pseudo-monotonicity of a convex combination of pseudo-monotone functions. We apply these conditions to composite functions $V(\varphi_i(m), \varphi_i(s), \gamma), i = 1, 2$, and use the functional relationship (14) between φ_1 and φ_2. This completes the proof of theorem.

4 Application: bilateral trade

As the leading economic application of the results above, we consider the bilateral trade model with generalized buyer's preferences. Specifically, her preferences are non-quasilinear and generally depend on two random variables, the state and the buyer's type. The buyer is privately informed about her type γ, which affects her payoff only. The seller determines the buyer's information about the state θ, which can be interpreted as the product quality. As the main result, we demonstrate that the seller can extract the perfect information about the state and the full buyer's surplus by using a private signal structure that randomizes between two perfectly informative signal functions.

[19] A function $\phi(x)$ is *pseudo-monotone* on a convex set $\mathbf{A} \subset \mathbb{R}$ if for every $(x, y) \in \mathbf{A}^2$, $\phi(x)(y - x) \leq 0$ implies $\phi(y)(y - x) \leq 0$. Equivalently, $\phi(x) \leq 0$ implies $\phi(y) \leq 0$ for all $y > x$.

4.1 Setup

A buyer (she) and a seller (he) are involved in trading a single indivisible object. The buyer's net utility from obtaining the object and making a payment t to the seller is determined by the payoff function $V(t,\theta,\gamma)$. Similarly to the general setup above, θ denotes the state, which represents intrinsic characteristics of the object, for example, its quality. It can also affect the seller's payoff in the case of keeping the object. Also, γ denotes the buyer's type, which represents some properties of her function. This variable has some antecedents in the mechanism design literature. For instance, in Dworczak et al. (2021), a privately known variable reflects the marginal value for money of agents in a market. In our model, the meaning of γ is broader. As shown below, it can reflect the marginal (dis-)utility with respect to the state θ and payment t similarly to Dworczak et al. (2021). In addition, it can determine the concavity of the buyer's payoff function (i.e., the magnitude of the risk-aversion) or other characteristics.

The state θ is a random variable drawn according to a continuous density $f(\theta) > 0$ from the state space $\Theta = \left[\underline{\theta}, \bar{\theta}\right] \subset \mathbb{R}_+$, where $\bar{\theta} > \underline{\theta}$. The type γ is drawn randomly from the type space $\mathbf{T}$ according to the cdf $H(\gamma)$. The variables θ and γ are independent.

Seller's preferences. The seller's preferences are quasilinear, and the seller's payoff from keeping the object in state θ is $U(\theta)$, which is continuous in θ. The seller's goal is to extract the maximum surplus from the buyer. As a result, he wants to sell the product if the buyer's willingness to pay, which is also represented by θ, exceeds his payoff, i.e., $\theta \geq U(\theta)$ at the highest price that the buyer is willing to accept. Denote the subset of these 'target' states $\Theta_0 = [\theta \in \Theta | \theta \geq U(\theta)]$. We assume that Θ_0 is a subinterval of high types, i.e., $\Theta_0 = \left[\theta_0, \bar{\theta}\right]$, where $\theta_0 \in \Theta$. It is the case under mild conditions on $U(\theta)$.[20]

Buyer's preferences. The buyer's payoff function $V(t,\theta,\gamma)$ is continuously differentiable in (t,θ), pseudo-concave in t, and $V'_\theta(t,\theta,\gamma) > 0 > V'_t(t,\theta,\gamma)$ for all $(t,\theta,\gamma) \in \Theta^2 \times \mathbf{T}$. As a normalization, we assume that not obtaining the object is equivalent to obtaining the object of quality $\theta = 0$ and $V(0,0,\gamma) = 0$. Equivalently, 0 is the value of the buyer's outside option, which she receives in the case of not obtaining the object and not making a payment. Also, $t = \theta$ represents the maximum payment, which makes the buyer indifferent between paying it for the object and taking the outside option. In other words, θ represents the buyer's willingness to pay. We assume that this value does not depend on γ. For example, if V depends on the difference $\theta - t$ between the object's quality θ and payment t, then γ does not affect the maximum acceptable payment $t = \theta$ (as we show below). At the same time, the buyer's payoff for $t \neq \theta$ can depend on γ. That is, for a given state $\theta \in \Theta$, the payment $t = \theta$ is a solution to the equation

$$V(t,\theta,\gamma) = 0 \text{ for all } \gamma \in \mathbf{T}. \tag{16}$$

We assume that this equation has a solution for all $\theta \in \Theta$. Furthermore, since $V'_t(t,\theta,\gamma) < 0$, then $t = \theta$ is the unique solution to (16). This implies

$$V(\theta,\theta,\gamma) \equiv 0 \text{ for all } (\theta,\gamma) \in \Theta \times \mathbf{T}. \tag{17}$$

[20]For example, $\Theta_0 = \left[\theta_0, \bar{\theta}\right]$ for $\theta_0 \in \Theta$ if $U(\theta)$ is differentiable, $U(\underline{\theta}) \geq \underline{\theta}, U\left(\bar{\theta}\right) \leq \bar{\theta}$, and $U'(\theta) \leq 1$ for all $\theta \in \Theta$.

Finally, we assume that the separability condition (1) holds for the subset of target states Θ_0, i.e., the buyer's marginal payoff with respect to the payment at point $t = \theta$ can be expressed as

$$V_t'(t, \theta, \gamma)\,|_{t=\theta} = g(\gamma)\,\zeta(\theta) \text{ for } (\theta, \gamma) \in \Theta_0 \times \mathbf{T}, \tag{18}$$

where $g(\gamma) > 0$ for all $\gamma \in \mathbf{T}$. Because $V_t' < 0$, it follows that $\zeta(\theta) < 0$ for all $\theta \in \Theta_0$.

Trade. The terms of trade are enforced by a trading mechanism (hereafter, a *mechanism*) $\mathcal{M}$ defined as follows. A mechanism $\mathcal{M} = (\mathbf{M}, Q_\xi(m), t_\xi(m))$ consists of a message space $\mathbf{M}$, an allocation rule $Q_\xi(m) \in [0, 1]$, and a transfer rule $t_\xi(m) \geq 0$. Here, Q and t are the buyer's probability of obtaining the object and her payment to the seller, respectively. Importantly, Q and t depend on both the buyer's message m and the realized structure ξ privately known to the seller (i.e., the mechanism).[21] This implies that ξ is verifiable ex-post. That is, the buyer can infer ex-post all information that can be jointly derived from her signal s and ξ. Obviously, this information is weakly more precise than the information that that seller can derive from ξ and m (with the equality if the buyer reports truthfully).

Timing. The game is played as follows. The buyer is a priori perfectly informed about γ and uninformed about θ. At the beginning of the game, the seller publicly selects a private signal structure $\rho \in \Delta\mathcal{I}$ and a mechanism $\mathcal{M} = (\mathbf{M}, Q_\xi(m), t_\xi(m))$. Then, the state θ and the signal structure $\xi \in \mathcal{I}_\rho$ are randomly drawn according to f and ρ, respectively, where ξ becomes the private information of the seller.[22] The buyer then privately observes a signal s generated by ξ from θ and decides whether to participate in trade or take the outside option. In the former case, the buyer sends a message $m \in \mathbf{M}$ to the mechanism $\mathcal{M}$. Finally, the terms of trade are enforced by the mechanism.

Because ρ is publicly observable, the following subgame is a standard selling mechanism with a privately and imperfectly informed buyer. Thus, we can invoke the Revelation Principle and restrict attention to direct interim incentive-compatible mechanisms, that is, such that $\mathbf{M} = \mathbf{S} = \bigcup_{\xi \in \mathcal{I}_\rho} \mathbf{S}_\xi$ and the buyer is truthful for all signals generated by the private signal structure ρ. Direct mechanisms are denoted $\mathcal{M} = (Q_\xi(s), t_\xi(s))$ hereafter. We also require mechanisms be interim individually-rational. This implies that the buyer does not receive a negative posterior payoff from trade upon receiving any signal (since the value of her outside option is normalized to 0). Thus, hereafter we consider only those mechanisms, which are interim incentive-compatible (IC) and interim individually-rational (IR), i.e., those which satisfy the following conditions:

$$EV_B(s|s, \gamma, \rho) = \max_{m \in \mathbf{S}} EV_B(m|s, \gamma, \rho) \text{ for all } (s, \gamma) \in \mathbf{S} \times \mathbf{T}, \text{ (IC)} \tag{19}$$

$$EV_B(s|s, \gamma, \rho) \geq 0 \text{ for all } (s, \gamma) \in \mathbf{S} \times \mathbf{T}, \text{ (IR)} \tag{20}$$

[21] In general, Q and t may also depend on the private signal structure ρ. Since our results do not rely on this dependence, we omit it in notation.

[22] Since the probability $\rho(\xi)$ does not depend on θ, ξ and θ are independent random variables. Hence, knowing ξ does not provide any additional information about θ to the seller.

where

$$EV_B(m|s,\gamma,\rho) = \int\limits_{\mathcal{I}_\rho}\int\limits_{\Theta} Q_\xi(m) V(t_\xi(m),\theta,\gamma) + (1 - Q_\xi(m)) V(t_\xi(m), 0, \gamma)\, dq(\theta|s,\rho)\, d\rho(\xi).$$

Given this framework, we start with an example, which provides the key insights into the general construction of private signal structures and mechanisms that extract the full information and the surplus from the buyer.

4.2 Example B: non-quasilinear buyer's preferences

Suppose that the prior density of states is uniform on the unit interval, that is, $f(\theta) = 1, \theta \in \Theta = [0,1]$. The type $\gamma \in \mathbf{T}$ is drawn randomly according to the cdf $H(\gamma)$. The variables θ and γ are independent. The seller's preferences are quasilinear with the payoff from keeping the object $U(\theta)$, which intersects θ once from above at $\theta_0 \in \Theta$. Thus, the subset of target states is $\Theta_0 = [\theta \in \Theta | \theta \geq U(\theta)] = [\theta_0, 1]$.

The buyer's payoff from consuming the product and making a payment t is given by the function

$$V(t,\theta,\gamma) = v(\theta - t, \gamma), \tag{21}$$

where $v(x,\gamma)$ is strictly increasing, concave, and continuously differentiable in x. Thus, $V(t,\theta,\gamma)$ is concave in t. Also, we put $v(0,\gamma) = 0$ for all $\gamma \in \mathbf{T}$. This property and (21) imply that the buyer's willingness to pay is θ for all $(\theta,\gamma) \in \Theta \times \mathbf{T}$. That is, (21) satisfies condition (17). Furthermore, (21) implies

$$V_t'(t,\theta,\gamma)|_{t=\theta} = -v_x'(0,\gamma),$$

that is, condition (18) holds, where $g(\gamma) = v_x'(0,\gamma) > 0$ and $\zeta(\theta) = -1$.

Next, we verify that Condition 1 also holds. First, note that $h(\theta) = \sqrt{-f(\theta)\zeta(\theta)} = 1$. Second, because $v(x,\gamma)$ is concave in x, then $V_t'(t,\theta,\gamma) = -v_x'(\theta - t,\gamma)$ is decreasing in t. As a result, the function $\nu(t,\theta,\gamma) = \frac{-v_x'(\theta-t,\gamma)}{h(t)}$ is decreasing in t.

An example of function (21) is the linear-quadratic function $v(x,\gamma) = \gamma x - x^2$, where $\gamma > 2$. In this case, the buyer's type γ determines the weight of the linear component in the payoff and, hence, the marginal payoff with respect to the difference $\theta - t$ at $t = \theta$. Another example is the hyperbolic absolute risk aversion (HARA) payoff function

$$v(x,\gamma) = \frac{1-\gamma}{\gamma}\left(\frac{\alpha}{1-\gamma}x + \beta\right)^\gamma - \frac{1-\gamma}{\gamma}\beta^\gamma,$$

where $\alpha > 0$ and $\frac{\alpha}{1-\gamma}x + \beta > 0$.[23] Depending on values of γ, α, and β, this form encompasses many standard payoff functions, such as linear, exponential (constant absolute risk aversion), power (and, hence, constant relative risk aversion), and logarithmic.[24] For this payoff

[23] The HARA function is defined as $v(x,\gamma) = \frac{1-\gamma}{\gamma}\left(\frac{\alpha}{1-\gamma}x + \beta\right)^\gamma$. Adding the constant term $-\frac{1-\gamma}{\gamma}\beta^\gamma$ is a normalization, which guarantees that $v(0,\gamma) = 0$ for all γ.

[24] See Ingersoll (1987).

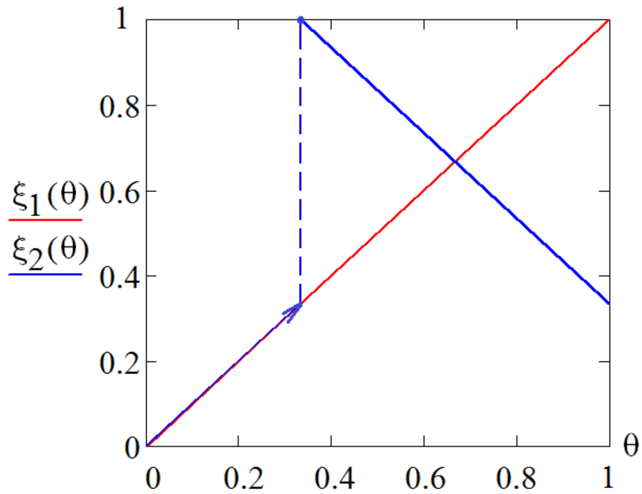

Figure 2: Signal functions $\xi_i(\theta), i = 1, 2$ for $\theta_0 = \frac{1}{3}$.

function, the buyer's type γ determines the degree of risk aversion. Because the value of γ is the buyer's private information, the seller can be uncertain about the specific shape of the payoff function, for example, power or exponential.

Now, consider the private signal structure ρ^c, which randomizes with equal probabilities between two perfectly informative signal functions

$$\xi_1(\theta) = \theta, \text{ and}$$
$$\xi_2(\theta) = \begin{cases} \theta & \text{if } \theta < \theta_0, \\ 1 + \theta_0 - \theta & \text{if } \theta \geq \theta_0. \end{cases} \tag{22}$$

Fig. 2 depicts signal functions $\xi_i(\theta), i = 1, 2$. The signal sets, that is, the images of ξ_1 and ξ_2 are identical and equal to $\mathbf{S} = [0, 1]$. Thus, any buyer's deviation from truthtelling is undetectable by the seller. Also, the buyer perfectly infers $\theta = s$ upon observing a signal $s < \theta_0$, but is uncertain about θ upon observing $s \geq \theta_0$. In the latter case, the posterior probability $q(\theta|s, \rho^c)$ of state θ induced by signal s under a private signal structure ρ^c is the binary distribution, which places probabilities $\frac{1}{2}$ on the posterior states

$$\theta_1 = \varphi_1(s) = s, \text{ and}$$
$$\theta_2 = \varphi_2(s) = \begin{cases} s & \text{if } s < \theta_0, \\ 1 + \theta_0 - s & \text{if } s \geq \theta_0. \end{cases}$$

where $\varphi_i = \xi_i^{-1}, i = 1, 2$.

Also, consider the direct mechanism $\mathcal{M}^c = \left(Q^c_{\xi_i}(s), t^c_{\xi_i}(s)\right), i = 1, 2$ with the message set $\mathbf{M} = \mathbf{S} = [0,1]$, such that

$$Q^c_{\xi_i}(s) = Q^c(s) = \begin{cases} 0 & \text{if } s < \theta_0, \\ 1 & \text{if } s \geq \theta_0, \end{cases}$$

$$t^c_{\xi_i}(s) = \begin{cases} 0 & \text{if } s < \theta_0, \\ \varphi_i(s) & \text{if } s \geq \theta_0. \end{cases}$$

Given the pair $(\rho^c, \mathcal{M}^c)$, the buyer's posterior payoff is

$$\begin{aligned} &EV_B(m|s,\gamma,\rho^c) = 0 \text{ if } m < \theta_0, s \in \mathbf{S}, \text{ and} \\ &EV_B(m|s,\gamma,\rho^c) = q(\theta_1|s,\rho^c)\, V\left(t^c_{\xi_1}(m), \theta_1, \gamma\right) + q(\theta_2|s,\rho^c)\, V\left(t^c_{\xi_2}(m), \theta_2, \gamma\right) \\ &\qquad = \frac{1}{2} v(\varphi_1(s) - \varphi_1(m), \gamma) + \frac{1}{2} v(\varphi_2(s) - \varphi_2(m), \gamma) \\ &\qquad = \begin{cases} \frac{1}{2} v(s-m,\gamma) + \frac{1}{2} v(m-s,\gamma) & \text{if } m \geq \theta_0, s \geq \theta_0, \\ \frac{1}{2} v(s-m,\gamma) + \frac{1}{2} v(s-(1+\theta_0-m),\gamma) & \text{if } m \geq \theta_0, s < \theta_0. \end{cases} \end{aligned}$$

It is straightforward to show that $EV_B(m|s,\gamma,\rho^c)$ is maximized at $m = s$ for all $(s,\gamma) \in \mathbf{S} \times \mathbf{T}$, and $EV_B(s|s,\gamma,\rho^c) = 0$ for all $s \in \mathbf{S}$. Hence, the interim incentive-compatibility constraints (19) and the interim individual-rationality constraints (20) hold. Furthermore, the buyer's ex-post payoff for $\theta_i = \varphi_i(s) \geq \theta_0$ is

$$\begin{aligned} V\left(\varphi_i(s), t^c_{\xi_i}(s), \gamma\right) &= v\left(\varphi_i(s) - t^c_{\xi_i}(s), \gamma\right) \\ &= v(\varphi_i(s) - \varphi_i(s), \gamma) = 0 \text{ if } s \geq s_0, \gamma \in \mathbf{T}, i = 1, 2. \end{aligned}$$

This implies that the seller extracts the full surplus in each state $\theta \geq \theta_0$ for any type $\gamma \in \mathbf{T}$ upon inferring θ from m and ξ_i.

Intuitively, the possibility for the seller to extract the full information and surplus from the buyer without violating her interim incentive-compatibility and individual-rationality constraints is driven by a combination of three factors. First, the object is sold to the buyer if and only if the mechanism infers that the state is above the cutoff θ_0. That is, the object is allocated to the buyer if and only if her ex-post acceptable payment exceeds the seller's ex-post benefits from keeping the object. Second, the incentive-compatibility in these state is sustained by the opposite reactions of the buyer's payments under different signal functions. That is, any buyer's deviation in an attempt to reduce the payment under one signal function is offset by the larger payment under the other signal function. A proper selection of φ_1 and φ_2 eliminates the buyer's marginal benefits from both local distortions (that is, when $s \geq \theta_0$ and $m \geq \theta_0$) and global ones (that is, when $s < \theta_0$ and $m \geq \theta_0$) and thus sustains buyer's truthtelling.[25] As a result, the mechanism perfectly infers the posterior state θ_i from m and the realized signal function ξ_i. Third, the above effect does not depend on the absolute values of buyers' payments. Thus, the seller can charge the buyer with the maximum payment, which precludes her from selecting the outside option. Because the value of this payment does not depend on the buyer's type γ, the mechanism extracts the full surplus from the buyer in all target states.

[25] Verifying that $\mathcal{M}$ is interim incentive-compatible for other combinations of s and m is trivial.

Two comments are worth mentioning. First, similarly to Example A, the full surplus extraction is feasible regardless of the precision of the buyer's information. In particular, if $s = \frac{1+\theta_0}{2}$, then the buyer is perfectly informed about the posterior state $\hat{\theta} = \hat{s} = \frac{1+\theta_0}{2}$. However, she is still unable to receive a positive payoff by using this information.

Second, the analysis above does not require the strict concavity of $V(t, \theta, \gamma)$ in t. Hence, it is equally applicable to the buyer's payoff function, which is linear in θ and t for all γ:[26]

$$V(t, \theta, \gamma) = v(\theta - t, \gamma) = \alpha(\gamma)(\theta - t).$$

Because the buyer is risk-neutral in this case, her interim payoff EV_B is unaffected by lotteries over payments under different signal functions, which are induced by her message. In other words, the buyer is indifferent between all messages upon receiving a signal $s \geq \theta_0$:

$$EV_B(m|s, \gamma, \rho^c) = \frac{1}{2}v(s - m, \gamma) + \frac{1}{2}v(m - s, \gamma) = 0 \text{ for } m \in \mathbf{S}, s \geq \theta_0.$$

4.3 Full surplus extraction

In this subsection we establish the possibility of the full information and surplus extraction for states above an arbitrary cutoff $\theta_0 \in \Theta$ in the general case. Consider the private signal structure ρ^*, which randomizes with equal probabilities between signal functions $\xi_1 = \varphi_1^{-1} : \Theta \to \mathbf{S}$ and $\xi_2 = \varphi_2^{-1} : \Theta \to \mathbf{S}$, where $\mathbf{S} = [\underline{s}, \bar{s}], \varphi_1' > 0$, and

$$\varphi_2(s) = \begin{cases} \varphi_1(s) & \text{if } s < s_0, \\ \Psi^{-1}\left(\Psi(\theta_0) + \Psi(\bar{\theta}) - \Psi(\varphi_1(s))\right) & \text{if } s \geq s_0, \end{cases} \tag{23}$$

where $s_0 = \xi_1(\theta_0)$, or equivalently, $\varphi_1(s_0) = \theta_0$. Thus, upon receiving a signal $s < s_0$ the buyer perfectly infers the state $\theta = \varphi_1(s)$. For $s \geq s_0$, the buyer's posterior belief is a binary distribution over $\{\varphi_1(s), \varphi_2(s)\}$, where $\varphi_2(s)$ satisfies the differential equation (14) with the boundary condition $\varphi_2(s_0) = \bar{\theta}$.

Next, consider a mechanism $\mathcal{M}^{\rho^*} = \left(Q_{\xi_i}^{\rho^*}(s), t_{\xi_i}^{\rho^*}(s)\right)$, such that

$$Q_{\xi_i}^{\rho^*}(s) = Q^{\rho^*}(s) = \begin{cases} 0 & \text{if } s < s_0, \\ 1 & \text{if } s \geq s_0, \end{cases} \tag{24}$$

$$t_{\xi_i}^{\rho^*}(s) = \begin{cases} 0 & \text{if } s < s_0, \\ \varphi_i(s) & \text{if } s \geq s_0. \end{cases} \tag{25}$$

The theorem below establishes that the pair $\left(\rho^*, \mathcal{M}^{\rho^*}\right)$ extracts the full information and surplus from the buyer for states $\theta \geq \theta_0$ under Condition 1.

Theorem 2 *Suppose V satisfies (17)–(18) and (f, V) satisfy Condition 1 for Θ_0. Consider the private signal structure ρ^* that randomizes between $\xi_1 = \varphi_1^{-1}$ and $\xi_2 = \varphi_2^{-1}$ with equal probabilities, where $\varphi_1 : \mathbf{S} \to \Theta$ is differentiable, $\varphi_1' > 0$, and φ_2 is given by (23). Then ρ^* and the mechanism $\mathcal{M}^{\rho^*} = \left(Q_{\xi_i}^{\rho^*}(s), t_{\xi_i}^{\rho^*}(s)\right)$ extract the full surplus for $(\theta, \gamma) \in \Theta_0 \times \mathbf{T}$.*

[26] In general, $v(\theta - t, \gamma) = \alpha(\gamma)(\theta - t) + \beta(\gamma)$. However, the condition $v(0, \gamma) = 0$ for all $\gamma \in \mathbf{T}$ results in $\beta(\gamma) = 0$.

The proof of theorem consists of two parts. The first part demonstrates that the mechanism $\mathcal{M}^{\rho^*}$ is interim individually-rational under the private signal structure ρ^*. Specifically, for signals $s < s_0$, the buyer receives the outside option with value 0. For $s \geq s_0$, the buyer pays θ_i in each posterior state $\theta_i = \varphi_i(s), i = 1, 2$, which is equal to her willingness to pay. That is, the mechanism extracts the buyer's full surplus upon learning the state θ from $m = s$ and ξ_i.

The main part of the proof is to establish the interim incentive-compatibility of the mechanism $\mathcal{M}^{\rho^*}$, which is done in a few steps depending on the values of a signal s and a message m. For $s \geq s_0$ and $m \geq s_0$, the interim incentive-compatibility is an implication of Theorem 1. First, the incentive-compatibility for these values of s and m is equivalent to the optimality of the truthful strategy in the implementation model with the state space Θ_0, the prior density $f_0(\theta) = f(\theta|\theta \in \Theta_0)$, the signal set $\mathbf{S}_0$, and the private signal structure ρ_0 that randomizes between $\xi_1(\theta)$ and $\xi_2(\theta)$ with the domains restricted to Θ_0. Second, the inverses φ_1 and φ_2 satisfy the first-order condition (14) with the boundary condition $\varphi_1(s_0) = \theta_0$ in the equivalent implementation model. Third, because V and f satisfy (21) and Condition 1 for Θ_0, then applying Theorem 1 to the equivalent implementation model means that the agent's truthful strategy is optimal. This in turn results in the interim incentive-compatibility of $\mathcal{M}^{\rho^*}$ for $(s, m) \in \mathbf{S}_0^2$. Next, for $m < s_0$ the incentive-compatibility holds as the buyer receives her outside option of value 0 for all $s \in \mathbf{S}$, which is identical to her payoff from truthful reporting. This is because truthful reporting provides the buyer with the outside option for $s < s_0$. For $s \geq s_0$, truthful reporting results in the full surplus extraction, so the buyer receives 0 as well. The final step is to show that the buyer with a signal $s < \theta_0$ cannot benefit from deviating to $m \geq s_0$. This step is based on the monotonicity of the buyer's payoff $V(t, \theta, \gamma)$ in θ and the fact that the buyer with signal $s \geq s_0$ does not receive a positive surplus. This completes the proof of the theorem.

Two other remarks are necessary. First, conditions (1) and (17) are essential for the full surplus extraction. Without additional assumptions about the impact of buyer's private information on her preferences, the seller cannot extract the full surplus by using the private information design even in the simplest model with quasi-linear preferences.[27] Importantly, these conditions are local. This is because for a given state θ, they must hold only at the ' full surplus extraction' point $t = \theta$, i.e., for the payment equal to the buyer's willingness to pay. Equivalently, they must holds only along the diagonal (θ, θ) in the (t, θ) space.

Second, the construction can be equally applied to the setup with a single product and many buyers if their states and types are distributed independently. In this case, the seller can utilize the pair of the optimal signal structure ρ_j^* and the mechanism $\mathcal{M}_j^{\rho_j^*}$ to elicit the perfect information about state θ_j from each buyer j. If the highest willingness to pay among buyers is $\theta_k = \max_j \theta_j \geq \theta_0$, the seller then sells the product to buyer k for price $t = \theta_k$.[28]

[27] See Remark 8 in Krähmer (2020).

[28] Even though each buyer can influence the probability of getting the object by manipulating her signal, she cannot benefit from it as her interim (and ex-post payoff) is at most 0 in either case.

5 Conclusion and discussion

This paper adds to the literature on the agency problem by showing how the principal can use private information design in a simple way to implement her ideal action for a target subset of states or the entire state space. The result holds even if the agent's preferences are non-quasilinear, non-convex, depend on the privately known component, and are independent of the state.

We conclude the paper by suggesting possible avenues for future research. First, the proposed construction of private signal structures can be potentially used in other economic environments. These may include models in which players' ideal actions are non-monotone to the unknown information or the buyer's payoff is non-monotone in the principal's action. Intuitively, truthtelling of the agent is driven by opposite monotonicities of the payoffs in her message for different posterior states. In general, each of these payoffs is a composition of three functions: i) the payoff as a function of the principal's action; ii) the principal's ideal action as a function of the posterior state; and iii) the induced posterior state as a function of the agent's message, which is the inverse of the signal function.[29] In our paper, we assume the strict monotonicity of the first two functions. If one or both of these functions are non-monotone, then the monotonicity of the composite function can be potentially restored by selecting a non-monotone (but bijective and, hence, perfectly informative) inverse.

Another avenue for future research is to extend the setup to multidimensional state and action spaces. If the agent's payoff function is additively separable, then our construction can be easily applied coordinatewise.[30] However, the question of whether our construction can be extended to multidimensional spaces in the case of payoff functions of the general form remains open.

Appendix

Proof of Theorem 1 Consider functions $\varphi_i : \mathbf{S} \to \Theta, i = 1, 2$, such that φ_1 is differentiable, $\varphi_1' > 0$, and φ_2 is given by (15). By construction, the pair $\{\varphi_1, \varphi_2\}$ satisfies the first-order condition (12). Because $\varphi_2 : \mathbf{S} \to \Theta$ is such that $\varphi_2' < 0$, then it is bijective. Hence, the functions $\xi_i = \varphi_i^{-1} : \Theta \to \mathbf{S}, i = 1, 2$ exist and are perfectly informative signal functions. Consider the private signal structure ρ^*, which randomizes between ξ_1 and ξ_2 with equal probabilities.

Next, the truthful strategy is optimal for the agent if $EV(m|s, \gamma, \rho^*)$ given by (9) is pseudo-concave in m for all $(s, \gamma) \in \Theta \times \mathbf{T}$. To establish the pseudo-concavity of

[29] The agent's posterior payoff is $EV(m|s, \gamma, \rho) = \sum_{i=1}^{2} q_i(s, \rho) V(y(\varphi_i(m)), \varphi_i(s), \gamma)$. Hence, the payoff conditional on the posterior state $\theta_i = \varphi_i(s)$ is given by $V(y(\varphi_i(m)), \varphi_i(s), \gamma)$.

[30] Consider, for instance, $V\left(\vec{a}, \vec{\theta}\right) = -\sum_{i=1}^{2} (a_i - \theta_i - b_i)^2$, where $\vec{a} = (a_1, a_2), \vec{\theta} = (\theta_1, \theta_2)$, and $\vec{\theta}$ is uniformly distributed on $[0, 1]^2$. Then the private signal structure, which randomizes between signal functions $\xi_1\left(\vec{\theta}\right) = \vec{\theta}$ and $\xi_2\left(\vec{\theta}\right) = (1, 1) - \vec{\theta}$ with equal probabilities, sustains agent's truthtelling.

$EV(m|s,\gamma,\rho^*)$ in m, it is sufficient to show that the function

$$\phi(m|s,\gamma,\rho^*) = \frac{\partial}{\partial m} EV(m|s,\gamma,\rho^*)$$

is pseudo-monotone in m on $\mathbf{S}$ for all $(s,\gamma) \in \mathbf{S} \times \mathbf{T}$ (Proposition 2.5, Hadjisavvas et al. 2005). A function $\phi(m|s,\gamma,\rho^*)$ is pseudo-monotone in m on an interval $\mathbf{S} \subset \mathbb{R}$ if $\phi(m_1|s,\gamma,\rho^*)(m_2 - m_1) \leq 0$ implies $\phi(m_2|s,\gamma,\rho^*)(m_2 - m_1) \leq 0$ for all $(m_1, m_2) \in \mathbf{S}^2$.

To guarantee the pseudo-monotonicity of $\phi(m|s,\gamma,\rho^*)$, we use the aggregation result by Quah and Strulovici (Proposition 1, 2012). It says that a linear combination $\alpha_1 \mathcal{V}_1(m) + \alpha_2 \mathcal{V}_2(m)$ of two pseudo-monotone functions $\mathcal{V}_1(m)$ and $\mathcal{V}_2(m)$ is pseudo-monotone for all $\alpha_i \geq 0, i = 1, 2$ if and only if: (i) $-\frac{\mathcal{V}_1(m)}{\mathcal{V}_2(m)}$ is decreasing in m for all m such that $\mathcal{V}_1(m) > 0$ and $\mathcal{V}_2(m) < 0$; and (ii) $-\frac{\mathcal{V}_2(m)}{\mathcal{V}_1(m)}$ is decreasing in m for all m such that $\mathcal{V}_1(m) < 0$ and $\mathcal{V}_2(m) > 0$.[31]

Fix $(s,\gamma) \in \mathbf{S} \times \mathbf{T}$. It follows from (12) that

$$\phi(m|s,\gamma,\rho^*) = \sum_{i=1}^{2} q_i(s,\rho^*) V_a'(\varphi_i(m), \varphi_i(s), \gamma) \varphi_i'(m) = \sum_{i=1}^{2} q_i(s,\rho^*) V_i(m,s,\gamma),$$

where

$$\mathcal{V}_i(m,s,\gamma) = V_a'(\varphi_i(m), \varphi_i(s), \gamma) \varphi_i'(m), i = 1, 2.$$

Because $V_a' < 0, \varphi_1' > 0$, and $\varphi_2' < 0$, we have

$$\mathcal{V}_1(m,s,\gamma) < 0 \text{ and } \mathcal{V}_2(m,s,\gamma) > 0 \text{ for all } (m,s,\gamma) \in \mathbf{S}^2 \times \mathbf{T}.$$

Thus, $\phi(m|s,\gamma,\rho^*)$ is pseudo-monotone in m for $q_i(s,\rho^*) \geq 0, i = 1, 2$ if and only if

$$-\frac{\mathcal{V}_2(m,s,\gamma)}{\mathcal{V}_1(m,s,\gamma)} = -\frac{V_a'(\varphi_2(m), \varphi_2(s), \gamma) \varphi_2'(m)}{V_a'(\varphi_1(m), \varphi_1(s), \gamma) \varphi_1'(m)}$$

is decreasing in m. By using (14), we get

$$\begin{aligned} -\frac{\mathcal{V}_2(m,s,\gamma)}{\mathcal{V}_1(m,s,\gamma)} &= \frac{V_a'(\varphi_2(m), \varphi_2(s), \gamma)}{V_a'(\varphi_1(m), \varphi_1(s), \gamma)} \frac{\sqrt{-f(\varphi_1(m))\zeta(\varphi_1(m))}}{\sqrt{-f(\varphi_2(m))\zeta(\varphi_2(m))}} \\ &= \frac{\nu(\varphi_2(m), \varphi_2(s), \gamma)}{\nu(\varphi_1(m), \varphi_1(s), \gamma)}, \end{aligned}$$

where

$$\nu(a,\theta,\gamma) = \frac{V_a'(a,\theta,\gamma)}{\sqrt{-f(a)\zeta(a)}} = \frac{V_a'(a,\theta,\gamma)}{h(a)}.$$

Because $V_a' < 0$ and $h > 0$, it follows that $\nu(a,\theta,\gamma) < 0$ for all $(a,\theta,\gamma) \in \Theta^2 \times \mathbf{T}$. Also, $\varphi_i \in \Theta, i = 1, 2$. Then, $\varphi_1' > 0$ and Condition 1 imply that $\nu(\varphi_1(m), \varphi_1(s), \gamma)$ is decreasing

[31] Quah and Strulovici (2012) use the term *single crossing* ϕ, which is equivalent to a pseudo-monotone $-\phi$. Formally, a single crossing function can intersect the $x-$axis at a single interval from below, whereas a pseudo-monotone function can intersect the $x-$axis at a single interval from above.

in m for all $(s,\gamma) \in \mathbf{S}\times\mathbf{T}$. Similarly, $\varphi_2' < 0$ and Condition 1 imply that $\nu(\varphi_2(m), \varphi_2(s), \gamma)$ is increasing in m for all $(s,\gamma) \in \mathbf{S}\times\mathbf{T}$. By combining these arguments, it follows that $-\frac{V_1(m,s,\gamma)}{V_2(m,s,\gamma)}$ is decreasing in m for all $(s,\gamma) \in \mathbf{S}\times\mathbf{T}$. ■

Proof of Theorem 2 Consider functions $\varphi_i : \mathbf{S} \to \Theta, i = 1,2$, such that φ_1 is differentiable, $\varphi_1' > 0$, and φ_2 is given by (23). Because $\varphi_2 : \mathbf{S} \to \Theta$ is piecewise continuous and strictly monotone for $s < \theta_0$ and $s \geq \theta_0$, and $\varphi_2(s) < \theta_0 \leq \varphi_2(z)$ for all $s < \theta_0 \leq z$, then φ_2 is bijective. Hence, the functions $\xi_i = \varphi_i^{-1} : \Theta \to \mathbf{S}, i = 1,2$ exist and are perfectly informative signal functions. Next, consider the private signal structure ρ^*, which randomizes with equal probabilities between ξ_1 and ξ_2, and the mechanism $\mathcal{M}^{\rho^*}$ with the allocation and payment rules given by (24) and (25), respectively.

First, the interim individual-rationality constraints (20) hold and are binding for all $s \in \mathbf{S}$. If $s < s_0$, then $Q_{\xi_i}^{\rho^*}(s) = 0$ and $t_{\xi_i}^{\rho^*}(s) = 0, i = 1,2$. That is, the buyer receives the outside option, and her posterior payoff $EV_B(s|s,\gamma,\rho) = V(0,0,\gamma) = 0$ for all $s < s_0$ and $\gamma \in \mathbf{T}$. If $s \in \mathbf{S}_0 = [s_0, \bar{s}]$, then $Q_{\xi_i}^{\rho^*}(s) = 1$ and $t_{\xi_i}^{\rho^*}(s) = \varphi_i(s), i = 1,2$. This results in

$$EV_B(s|s,\gamma,\rho^*) = \sum_{i=1}^{2} q_i(s,\rho^*) V(\varphi_i(s), \varphi_i(s), \gamma) = 0 \text{ for all } (s,\gamma) \in \mathbf{S}_0 \times \mathbf{T}. \tag{26}$$

where the second equality holds because (17) implies $V(\varphi_i(s), \varphi_i(s), \gamma) = 0$ for $(s,\gamma) \in \Theta\times\mathbf{T}, i = 1,2$.

Second, we prove the interim incentive-compatibility of the pair $\left(\rho^*, \mathcal{M}^{\rho^*}\right)$ by considering three cases depending on the values of $(m,s) \in \mathbf{S}^2$.

(i) $(s,m) \in \mathbf{S}_0^2$. Because $Q_{\xi_i}^{\rho^*}(m) = 1$ and $t_{\xi_i}^{\rho^*}(m) = \varphi_i(m), i = 1,2$ for $m \in \mathbf{S}_0$, we have

$$EV_B(m|s,\gamma,\rho^*) = \sum_{i=1}^{2} q_i(s,\rho^*) V(\varphi_i(m), \varphi_i(s), \gamma) \text{ for } (m,s,\gamma) \in \mathbf{S}_0^2 \times \mathbf{T}. \tag{27}$$

Next, note that states $\theta \in \Theta_0$ generate signals $s_i = \varphi_i^{-1}(\theta) \in \mathbf{S}_0, i = 1,2$ under ρ^*. Hence, the agent's posterior belief induced by a signal $s \in \mathbf{S}_0$ is the binary distribution over states $\theta_i = \varphi_i(s) \in \Theta_0, i = 1,2$. By using (7), the posterior probability of θ_i is

$$\begin{aligned} q_i(s,\rho^*) &= \frac{f(\varphi_i(s))\varphi_i'(s)}{f(\varphi_1(s))\varphi_1'(s) - f(\varphi_2(s))\varphi_2'(s)} \\ &= \frac{f_0(\varphi_i(s))\varphi_i'(s)}{f_0(\varphi_1(s))\varphi_1'(s) - f_0(\varphi_2(s))\varphi_2'(s)} = q_i\left(s,\rho^0\right), i = 1,2. \end{aligned}$$

Here, $f_0(\theta) = f(\theta|\theta \in \Theta_0) = \frac{f(\theta)}{1-F(\theta_0)}$ is the prior density of θ conditional on $\theta \in \Theta_0, F(\theta)$ is the cdf of θ, and ρ^0 is the private signal function that randomizes with equal probabilities between ξ_1^0 and ξ_2^0, where $\xi_i^0 = \xi_i : \Theta_0 \to \mathbf{S}_0$ is a signal function ξ_i with the domain restricted by Θ_0 and, thus, the image $\mathbf{S}_0$.

Now, consider the implementation model with the signal set $\mathbf{S}_0$, the prior density $f_0(\theta)$, and the private signal structure ρ^0 (called the *equivalent implementation model* hereafter). By combining the arguments above and comparing (27) with (9), it follows that the interim incentive-compatibility condition (19) for $(s,\gamma)\in\mathbf{S}_0\times\mathbf{T}$ on the restricted message space $\mathbf{S}_0$ is identical to the optimality condition (11) for the agent's truthful strategy in the equivalent implementation model. Next, $\varphi_2(\theta)$ given by (23) satisfies the first-order condition (14) with the boundary condition $\varphi_1(s_0)=\theta_0$ in this model. Also, conditions (21) and 1 hold for Θ_0. Then, by applying Theorem 1 to the equivalent implementation model, it follows that the agent's truthful strategy is optimal. This means that the incentive-compatibility constraints in the bilateral-trade model hold for $(s,m)\in\mathbf{S}_0^2$.

(ii) $s\in\mathbf{S}, m<s_0$. Then $Q_{\xi_i}^{\rho^*}(m)=0, t_{\xi_i}^{\rho^*}(m)=0, i=1,2$, and

$$EV_B(m|s,\gamma,\rho^*)=V(0,0,\gamma)=0=EV_B(s|s,\gamma,\rho^*),$$

where $EV_B(s|s,\gamma,\rho^*)=0$ for $(s,\gamma)\in\mathbf{S}\times\mathbf{T}$ follows from the interim individual-rationality of $\mathcal{M}^{\rho^*}$.

(iii) $s<s_0, m\geq s_0$. Since $s<s_0$, then $\varphi_i(s)<\theta_0=\varphi_1(s_0), i=1,2$. Also, $m\geq s_0$ implies $Q_{\xi_i}^{\rho^*}(m)=1$ and $t_{\xi_i}^{\rho^*}(m)=\varphi_i(s), i=1,2$. Then, we have

$$\begin{aligned}EV_B(m|s,\gamma,\rho^*)&=\sum_{i=1}^{2}q_i(s,\rho^*)V(\varphi_i(m),\varphi_i(s),\gamma)<\sum_{i=1}^{2}q_i(s,\rho^*)V(\varphi_i(m),\varphi_1(s_0),\gamma)\\&\leq\sum_{i=1}^{2}q_i(s,\rho^*)V(\varphi_i(m),\varphi_i(s_0),\gamma)\leq\sum_{i=1}^{2}q_i(s,\rho^*)V(\varphi_i(s_0),\varphi_i(s_0),\gamma)\\&=EV_B(s_0|s_0,\gamma,\rho^*)=0,\end{aligned}$$

where the first inequality holds since $\varphi_i(s)<\varphi_1(s_0), i=1,2$ and $V'_\theta(t,\theta,\gamma)>0$ imply $V(\varphi_i(m),\varphi_i(s),\gamma)<V(\varphi_i(m),\varphi_1(s_0),\gamma), i=1,2$, the second inequality holds due to $\varphi_2(s_0)=\bar{\theta}\geq\theta_0=\varphi_1(s_0)$ and $V'_\theta(t,\theta,\gamma)>0$, and the last one holds since $m=s_0$ maximizes $EV_B(m|s_0,\gamma,\rho)$ over $m\geq s_0$. ■

References

Bereanu, B., 1972. Quasi-convexity, strictly quasi-convexity and pseudo-convexity of composite objective functions. Revue française d'automatique, informatique, recherche opérationnelle. Mathématique 6, 15–26

Bergemann, D. and M. Pesendorfer, 2007. Information structures in optimal auctions. Journal of Economic Theory 137, 580–609

Bergemann, D. and A. Wambach, 2015. Sequential information disclosure in auctions. Journal of Economic Theory 159, 1074–1095

Crawford, V., and J. Sobel, 1982. Strategic information transmission. Econometrica 50, 1431–1451

Crémer, J., and R.P. McLean, 1988. Full extraction of the surplus in Bayesian and dominant strategy auctions. Econometrica 56, 1247–1257

Dworczak, P., Kominers, S.D., and M. Akbarpour, 2021. Redistribution through markets. Econometrica 89, 1665–1698

Esö, P., and B. Szentes, 2007. Optimal information disclosure in auctions. Review of

Economic Studies 74, 705–731

Fu, H., Haghpanah, N., Hartline, J., and R. Kleinberg, 2021. The full surplus extraction from samples. Journal of Economic Theory 193, 105230

Ganuza, J.J., and J. Penalva, 2010. Signal orderings based on dispersion and the supply of private information in auctions. Econometrica 78, 1007–1030

Hadjisavvas, N., Komlòsi, S., and S. Schaible, 2005. Handbook of generalized convexity and generalized monotonicity. Springer, Boston

Hwang, I., Kim, T., and R. Boleslavsky, 2019. Competitive advertising and pricing. Working paper

Ingersoll, J., 1987. Theory of financial decision making. Totowa, NJ: Rowman and Littlefield Publishers

Ivanov, M. and A. Sam, 2022. Cheap talk with private signal structures, Games and Economic Behavior 132, 288–304

Ivanov, M., 2021. Optimal monotone signals in Bayesian persuasion mechanisms. Economic Theory 72, 955–1000

Ivanov, M., 2016. Dynamic learning and strategic communication. International Journal of Game Theory 45, 627–653

Ivanov, M., 2015. Dynamic information revelation in cheap talk. The B.E. Journal of Theoretical Economics 15, 251–275

Ivanov, M., 2013. Information revelation in competitive markets. Economic Theory 52, 337–365

Johnson, J., and D. Myatt, 2006. On the simple economics of advertising, marketing, and product design. American Economic Review 93, 756–784

Kamenica, E. and M. Gentzkow, 2011. Bayesian persuasion. American Economic Review 101, 2590–2615

Krähmer, D., 2020. Information disclosure and the full surplus extraction in mechanism design. Journal of Economic Theory 187, 105020

Krähmer, D., 2021. Information design and strategic communication. American Economic Review: Insights 3, 51–66

Larionov, D., Pham, H., and T. Yamashita, 2021. First best implementation with costly information acquisition, Working paper

Lewis, T. and D. Sappington (1994). Supplying information to facilitate price discrimination. International Economic Review 35, 309–327

Li, H., and X. Shi, 2019. Discriminatory information disclosure. American Economic Review 107, 3363–3385

McAfee, P. and P. Reny, 1992. Correlated information and mechanism design. Econometrica 60, 395–421

McKinsey & Company. How companies spend their money: a McKinsey global survey, McKinsey Quarterly, June 2007.

Myerson, R., 1981. Optimal auction design. Mathematics of Operations Research 6, 58–73

Pastrian, N., 2021. Surplus extraction with behavioral types, Working paper

Quah, J. and B. Strulovici, 2012. Aggregating the single crossing property. Econometrica 80, 2333–2348

Saak, A. 2006. The optimal private information in single unit monopoly. Economics Letters 91, 267–272

Shannon, C., 1949. Communication theory of secrecy systems, Bell System Technical Journal 28, 656–715

Zhu, S., 2021. Private disclosure with multiple agents, Working paper.