Perfect Robust Implementation by Private Information Design*

Maxim Ivanov†

December 2023

Abstract

This paper studies the general principal-agent framework in which the principal aims to implement his first-best action that is monotone in the unknown state. The principal privately selects a signal structure of the agent whose payoff depends on the principal's action, the state, and the privately known type. The agent privately observes the generated signal and reports it to the principal, who takes action. 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. As to the economic application, we consider the bilateral trade model with non-quasilinear preferences and private multi-dimensional information of the buyer, and show that the seller can extract full surplus by privately designing the buyer's signal structures.

JEL classification: C72, D81, D82, D83

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

1 Introduction

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

*I am thankful to the Associate Editor and the anonymous referee for their insightful comments. I am also grateful to Kalyan Chatterjee, Mikhail Drugov, Seungjin Han, Peicong Hu, Vijay Krishna, Ming Li, Wooyoung Lim, Elliot Lipnowski, Vitor Farinha Luz, Tatiana Mayskaya, Alan Miller, Alessandro Pavan, Sergei Severinov, Joel Sobel, Chen Zhao, Charles Zheng, and the audiences at Concordia University, Western University, University of British Columbia, Canadian Economic Theory Conference 2022, Canadian Economic Association Meetings 2022, Stony Brook International Conference on Game Theory 2022, EEA-ESEM 2022 Conference, Econometric Society European Winter Meeting 2022, FES-ICEF-NES seminar, and CUHK-HKU-HKUST Joint Theory Seminar 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. This work was supported by the Canadian Social Sciences and Humanities Research Council (SSHRC) Insight Grant 435-2022-0137. 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.

information technology, which provides signals to players about an unknown variable and is a choice of the information designer, is privately known to the designer only. However, the signals generated by the technology become the private knowledge of the addressees. In this paper, we apply the private information design to the general principal-agent framework. As the main result, we show that the principal (he) can elicit perfect information from the agent (she) and implement his first-best outcome in a simple way by privately designing the agent's information structure. This result holds even if the principal's preferences are incomplete, the agent's preferences do not depend on this information, and/or the agent has separate private information about her preferences.

Before discussing the results, we start with a description of the environment. Its generality stems from substantially relaxing the standard assumptions about players' preferences and their information compared to standard models. First, our model uses only the principal's *ideal action*, which is a monotone function of the unknown *state*. In other words, our results hold for all principal's preferences with a given ideal action function. This type of robustness is economically important, because learning about the first-best choices of economic agents—for example, by observing their past behavior—is often much easier than obtaining information about non-preferred alternatives. Second, the agent's preferences are represented by a payoff function that is pseudo-concave in the principal's action. This class of functions is rather broad and includes monotone or single-peaked functions, which are used in many economic applications. Third, the agent's payoff depends on multi-dimensional information. Besides the state, it is also represented by the agent's *type*. The state affects the first-best decision of the principal and potentially the payoff of the agent. For example, in models of trade, the state can reflect the product quality, which determines the valuations of the object by both the seller (the principal) and the buyer (the agent). The agent's type reflects only the characteristics of the agent's payoff function, such as the measure of her risk-aversion. As a special case, the agent's preferences may be independent of the state and/or the type. Another key difference, which is also the third factor, is that the agent is a priori privately informed about her type, but uninformed about the state. The uncertainty about the state is a key motive in communication between the agent and the principal.

In short, our framework substantially extends the setup in the related paper by Ivanov and Sam (2022) who study an information design problem in Crawford and Sobel's (1982) cheap-talk framework, in three dimensions. First, they assume that the agent is ex-ante uninformed, while she is privately informed in our setup. Second, the agent in their model has a unique ideal action. Third, the agent's payoff is strictly supermodular in the action and the state. The last two assumptions are necessary in most cheap-talk models as they imply that the agent's ideal action is monotone in state. In contrast, neither of these assumptions about the agent's payoff—the existence of the ideal action, the dependence of this action on the state, or the strict supermodularity—is imposed in our setup. As a result, our framework covers a substantially broader class of economic applications than theirs. For example, our leading economic application—bilateral trade—does not fit into their model. Specifically, since the buyer's payoff is decreasing in her payment, she does not have an ideal action for any valuation of the product (i.e., the state). Furthermore, some payoff functions violate the supermodularity condition as well.[1] It is also worth noting that our model incorporates

[1]For example, if the buyer's payoff is a convex function of the difference between her product valuation

the worst-case scenario for the principal in terms of information extraction possibilities. For example, if the agent's preferences are monotone in the principal's action, then there is no meaningful communication given any publicly known signal structure of the agent.

We now turn to the information design component of the model. The information about the state is obtained by the agent whose information technology, called a *signal structure*, is privately selected by the principal.[2] Specifically, the principal randomizes between publicly known signal structures, where the realized signal structure is known only to the principal. Once the principal assigns this signal structure to the agent (without informing her about it), the structure generates a signal about the state. Upon privately observing the signal, the agent sends a report to the principal, who 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. 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 is that randomization between signal structures creates uncertainty for the agent about the impact of her message on the principal's action. The principal exploits this uncertainty by selecting signal functions with 'opposite monotonicities' in the state: a signal generated by one signal function is increasing in the state, while the signal generated by the other signal function is decreasing. This implies opposite reactions of posterior states (induced by each signal function) to the signal, and, hence, opposite reactions of the principal's decisions to the agent's message. As a result, any distortion of the signal in an attempt to increase the agent's payoff under one signal function is offset by the marginal losses under the other signal function. By properly selecting signal functions, the principal can sustain agent's truth-telling. 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 local separability condition on the agent's payoff. An implication of this condition is that the agent's marginal benefits from lying and, hence, the optimal signal structures become invariant to the agent's type.

Regarding potential applications of our results, it is important to note that the private information design framework is equivalent to communication between two players with identical preferences—the sender and the principal—via the strategic mediator (the agent) with conflicting preferences. In other words, the agent plays the role of the communication channel in which information distortions are determined endogenously by the agent's strategic motives. In this case, the privately known signal structure serves as the decoding key, which allows the principal to infer the sender's message from the agent's report.

As a more specific example, consider a large organization (a firm, a public institution, etc.) with the standard vertical organizational structure: the upper leadership, the middle management, and employees. The upper leadership wants to collect decision-relevant information from employees about some parameter, say, the human resource (HR)

and the payment, then it is submodular and, thus, violates the strict supermodularity condition.

[2]The higher opportunity costs of the principal compared to those of the agent is a common reason for assigning the task of collecting information to the agent rather than acquiring it directly.

effectiveness, by conducting an ad-hoc survey. The employees' responses, however, are collected, processed, and reported by the HR manager herself whose benefits—reputation, bonuses, and career opportunities—are increasing in this parameter. The HR manager thus has the incentive to misreport the acquired information.[3] In order to preclude such manipulations, the upper leadership can randomize between two surveys whose questions are known to the leadership but not to the HR manager.[4] 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 HR's *effectiveness* on a scale of 0 to 100, which corresponds to an increasing signal function. The other survey would ask to evaluate the *ineffectiveness*, with a score of 100 indicating a completely unproductive or mismanaged HR team. Because the HR 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.[5]

As the leading economic application, we consider the bilateral trade model with a buyer who has non-quasilinear preferences and private multidimensional information about her payoffs. As typically assumed in the literature (Bergemann and Pesendorfer, 2007; Li and Shi, 2019; and Bergemann et al., 2022), the seller determines: i) the terms of trade, i.e., the mechanism that enforces an allocation and payment; and ii) the buyer's signal structure.[6] In our model, the buyer's information is represented by two variables: the state and the type. The state reflects the intrinsic characteristics of the object, which determine the buyer's willingness to pay, that is, the highest payment, which makes her indifferent between trading and taking an outside option. The buyer's type reflects her preferences only, for instance, the degree of risk-aversion. The buyer is a priori uninformed about her state but informed about her type. Upon observing a private signal generated by the signal structure and learning her type, the buyer sends a message to the mechanism, which enforces the terms of trade. Similarly to the main framework, the outcome of the mechanism depends on the buyer's message and the signal structure privately known by the seller, i.e., the mechanism.

Our second main result is that the seller can extract the full surplus from trade upon eliciting the perfect information about all states in the *target* subinterval of states. These states reflect the fact that the seller might be interested in selling the object to a buyer only if her willingness to pay exceeds the seller's benefits from keeping the object. Full information and surplus extraction are feasible by employing private signal structures similar to those in the general framework. The only difference is that a randomization between signal functions

[3] 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.

[4] It is also assumed that the manager cannot access the questions in another way, for example, by spying on employees or cooperating with them.

[5] Two key factors differentiate this example from university processes, when students' feedback at the end of each semester is going directly to academic services by bypassing professors. First, the ad hoc character of the survey leaves the manager uncertain about the questions in it, while standard questions in the students' evaluation forms do not change over time and thus are publicly known. Second, involving external evaluators (similar to academic services at universities) for an ad-hoc survey would require substantial costs per survey.

[6] 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.

creates the buyer's uncertainty about the states in the target subinterval.[7] Specifically, the optimal private signal structure randomizes between two signal functions with the opposite monotonicities of signals in the state. Also, the seller sells the product to the buyer whenever the posterior state (which reflects the buyer's posterior willingness to pay) is in the target subinterval, at the price equal to the posterior state. Then, the opposite monotonicities of signal functions in the state imply the opposite monotonicities of the 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. Importantly, it does not depend on the absolute value of the buyer's willingness to pay. As a result, the seller can charge the buyer with the highest willingness to pay (which does not depend on the buyer's type) upon learning the state.

In this light, our paper extends the related models in the mechanism design literature in three dimensions. First, it does not assume that the players' preferences are quasilinear. It is a conceptual extension. If the parties' preferences are quasilinear in the state and the seller's valuation is constant, then he can extract the full surplus by informing the buyer whether the state is above or below the seller's valuation and setting the price at the higher posterior state (Saak, 2006). This construction, however, cannot be extended to setups with a more general form of players' preferences. For example, if the buyer is risk-averse, then hiding information about the state reduces her posterior willingness to pay. In addition, if the seller's valuation also depends on the state, he must also learn it before making a decision about selling the object. Otherwise, there is a chance of selling the object at the price below the seller's valuation. These sources of inefficiency do not allow the seller to extract the entire trade surplus.

Second, we show that the full surplus extraction is robust to the buyer's private knowledge of her preferences. As noted by Krähmer (2020), it is a generally unsolvable problem if the seller employs the private information design in the model with quasilinear preferences, such that the buyer's state and type are independent. As we demonstrate, this problem can be circumvented by two conditions on the buyer's payoff function. First, the buyer's willingness to pay is solely determined by the state. As we elaborate below, this condition is not novel and is imposed, for instance, in auctions with non-linear buyers' preferences. In this case, the type parameterizes the buyer's payoff function and reflects, for instance, its measure of risk-aversion, which does not affect her willingness to pay. Second, the buyer's marginal payoff with respect to the payment, whenever it is equal to her willingness to pay, can be factored into separate functions of the state and the type. 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, which determines the shape of the payoff as a function of this difference. Under these conditions, the seller can extract the full surplus by using signal functions, which do not depend on the shape of the buyer's payoff function.

Third, in contrast to most of the literature on full surplus extraction, the set of states in our model is continuous. It is also not a purely technical extension. Eliciting information, which can take a finite number of values, is generally a less difficult problem since the

[7] For low states, the precision of the buyer's information is not substantial, since even the perfectly informed buyer prefers not to purchase the object.

set of the buyer's incentive-compatibility constraints is substantially smaller. Importantly, the buyer cannot distort her information locally by mimicking nearby values. At the same time, local incentive-compatibility plays a crucial role in mechanism design (Myerson, 1981). In addition, there are qualitative differences between spaces of finite and continuous distributions. Because of them, the results by Crémer and McLean (1988) and Krähmer (2020) about full surplus extraction from buyers with discrete valuations—which rely on eliciting posterior distributions from other correlated variables—are generically impossible for continuous distributions of states (Heifetz and Neeman, 2006).[8]

Literature

The first paper to introduce the idea of eliciting the agent's information by keeping her uncertain about how her message will influence the principal's action was Watson (1996). Specifically, he showed that the principal can elicit information from the perfectly informed agent and implement her first-best outcome if the agent is maximally uncertain about the relationship between the message and the action, whereas the relationship is known to the principal. The maximal uncertainty is modeled as the (approximately) uniform probability distribution over the set of all bijective mappings from reported states to the principal's ideal actions. In contrast, the information design in our work creates the endogenous uncertainty of the agent about the relationship between her message and the principal's action. Next, our construction is very simple and involves randomizing between only two bijective mappings. This randomization keeps the agent minimally uncertain about the relationship between her message and the principal's action rather than maximally uncertain, as in Watson's model. Finally, we show that the shapes of signal functions play a critical role in sustaining the agent's truth-telling.[9]

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 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 buyer's preferences in our model are of a general form. Second, while Krähmer's general model considers a privately informed buyer, full-surplus extraction is demonstrated for an a priori uninformed buyer only. We establish this result 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 truth-telling. 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

[8] However, almost full surplus extraction can be attained by partitioning the state space into a finite collection of subintervals (McAfee and Reny, 1992).

[9] Also, an important technical difference is the cardinality of the state space. Because it is finite in Watson's model, the set of all bijective mappings from reported states to ideal actions is also finite. In our setup, however, the state and action spaces are continuous. Applying Watson's idea to this setup is equivalent to the uniform randomization over the space of all bijective functions from the interval of reported states to the interval of ideal actions, which is a large and mathematically complicated space.

action enforces the buyer's incentive-compatibility.[10] 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 uncertainty for the buyer about her payments contingent on the realized signal structure. As a result, signal distortions create the 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.

As noted above, our framework is equivalent to communication between the sender and the principal with identical preferences via an agent with conflicting ones. In this regard, a recent work by Silva et al. (2023) studies perfect implementation in sender-principal communication via two agents. As they show, if the agents have state-independent preferences and cannot communicate with each other, the sender can convey the perfect information to the principal while sustaining the agents' truth-telling. The key insight is that the sender sends an encoded message to one agent and the decoding key to the other, whereas the encoding strategy involves redundant messages to prevent the agents from both learning the sender's information and misreporting her messages to the principal.

Our work is also related to the literature on surplus extraction. This topic drew significant attention due to a seminal work by Crémer and McLean (1988), who demonstrated the possibility of full surplus extraction based on eliciting beliefs about buyers' correlated valuations. Alternatively, recent literature on information design has demonstrated that it can be a powerful tool for the seller to extract the buyers' surplus. Lewis and Sappington (1994), Johnson and Myatt (2006), Bergemann and Pesendorfer (2007), Esö and Szentes (2007), Li and Shi (2019), and Ivanov (2021) show that the seller can benefit by shaping buyers' information in bilateral trade, auctions, or selling mechanisms. Larionov et al. (2021) and Zhu (2023) consider the mechanism/information design with multiple (at least four) agents. By combining Shannon's (1949) encryption technique with cross-checking agents' messages, they demonstrate that the principal can implement any state-contingent allocation as if he could directly observe the states. 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 full surplus extraction is feasible if the number of draws is large enough. Antsygina and Teteryatnikova (2023) achieve the same result in contests with private valuations of contestants whose signal structures about valuations are determined by the contest designer. These papers, however, do not consider private information design.

Finally, our construction of the private signal structure exploits the idea of stochastically penalizing the agent for distortions in her behavior. Besides information design, this idea has been used in many other economic applications. For example, buyers' truth-telling in the surplus-extracting mechanism by Crémer and McLean (1988) is sustained by payments that depend on the reports of all buyers. Because buyers' valuations are correlated,

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

distorting information by a buyer is more likely to induce profiles of reports that have low probabilities, which result in the buyer's significant payments. Similarly, the literature on communication via a mediator exploits the idea that a mediator can facilitate communication by stochastically penalizing the biased agent for reporting higher messages by recommending the principal to take unfavorable actions (Goltsman et al., 2009; Ivanov, 2010). A similar effect can be achieved by adding noise to the communication channel between the agent and the principal (Blume et al., 2007) or by the principal's imperfect ability to decode the agent's messages (Blume et al., 2019). Another application includes the principal-agent models with moral hazard. Ederer et al. (2018) study simple incentive schemes for an agent with privately known and asymmetric costs of performing different tasks. They show that optimal linear contracts include randomization between payments rules with opposite reactions to performance on different tasks, since these contracts induce more balanced efforts and eliminate the efficiency losses from the agent's private information.

The rest of the paper is organized as follows. Section 2 introduces the general framework, and Section 3 provides the main results for it. 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 players communicate about the ex-ante unknown *state* $\theta \in \mathbb{R}$, which is a random variable drawn from the state space $\Theta = \left[\underline{\theta}, \bar{\theta}\right]$ according to a continuous density $f(\theta)$, such that $f > 0$. (Hereafter, $u > 0$ for a function $u : X \to \mathbb{R}$ means $u(x) > 0$ for all $x \in X$.) For the subinterval of *target states* $\Theta_0 = [\theta_0, \theta_1] \subset \Theta$, the principal's goal is to elicit the perfect information about θ and implement his *first-best* (or *ideal*) *action* $y_P(\theta)$ from a closed interval $\mathbf{A} \subset \mathbb{R}$. We assume that $y_P(\theta)$ is continuous and strictly increasing in θ on Θ_0. For other states, knowing that $\theta < \theta_0$ or $\theta > \theta_1$ is sufficient to implement default actions y_0 or y_1, respectively. As a normalization, it is without loss of generality to put $y_P(\theta) = \theta$ for $\theta \in \Theta_0$.[11] Thus, Θ_0 is also a set of the principal's ideal actions induced by states $\theta \in \Theta_0$. We assume that $\Theta_0 \subset \mathbf{A}$, that is, all principal's ideal actions for target states are feasible.

Our specification of the principal's preferences has several important implications. First, it does not require specifying the principal's underlying preferences that are expressed, for example, by the payoff function $U(a, \theta)$. For our purposes, knowing only the maximizer $y_P(\theta)$ of $U(a, \theta)$ over a is sufficient. Second, due to such a broader specification of the principal's preferences, his problem is formulated as the *ex-post implementation problem* rather than the ex-ante optimization one. Formulating the principal's problem as the optimization one is impossible in our setup as it requires specifying the payoff function $U(a, \theta)$, which is not assumed in our model. Third, we show below that the principal can implement his ideal action $y_P(\theta) = \theta$ upon learning the state $\theta \in \Theta_0$ by the means of private information design. As a result, our solution to the implementation problem is a solution to the principal's optimization problem for any payoff function $U(a, \theta)$ with the ideal action $y_P(\theta)$. In other words, our results are robust to any modifications in the principal's payoff function as long as they do not affect the maximizer $y_P(\theta)$, or equivalently, knowing the principal's preferences

[11] Otherwise, if $y_P(\theta) \neq \theta$ for $\theta \in \Theta_0$, then the monotone transformation $z = y(\theta)$ results in $y_P(z) = z$.

over alternative outcomes (a, θ) different from the ideal ones is unnecessary.[12]

The agent's payoff function $V(a, \theta, \gamma)$ depends on the principal's action a, the state θ, and the privately known *buyer's type* (or simply *type*) γ. In general, V might not depend on θ and/or γ. The type γ is drawn from the type space $\mathbf{T} \subset \mathbb{R}$. We do not make any assumptions about $\mathbf{T}$, however, there can be bounds on it for some specific V in order to guarantee that γ has an economic meaning or V respects the conditions imposed below. Also, knowing the distribution of types over $\mathbf{T}$ is unnecessary. The variables θ and γ are independent.

We make the following assumption about $V(a, \theta, \gamma)$.[13]

Condition 1 *$V(a, \theta, \gamma)$ is continuous in θ and differentiable and strictly pseudo-concave in a for all $(a, \theta, \gamma) \in \Theta^2 \times \mathbf{T}$.*

Intuitively, the pseudo-concavity is a generalized form of the concavity of differentiable functions for which stationary points are also global maximizers. As a special case, $V(a, \theta, \gamma)$ can be strictly monotone in a. Next, consider the function $V'_a(a, \theta, \gamma)$, which represents the 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, $V'_a(a, \theta, \gamma)$ is the marginal payoff of the agent's type γ in state θ from manipulating the principal's (wrong) belief that the state is a. That is, it represents the agent's marginal benefits or losses from manipulating the principal's action via the latter's posterior belief. We impose the following separability condition on this function:

$$V'_a(\theta, \theta, \gamma) = g(\gamma)\zeta(\theta) < 0 \text{ for } (\theta, \gamma) \in \Theta_0 \times \mathbf{T}. \tag{1}$$

For convenience, we put $g > 0$ and $\zeta < 0$. That is, the agent's marginal payoff with respect to action a at the principal's ideal point $a = \theta$ can be decomposed into the product of a positive function $g(\gamma)$ and a negative function $\zeta(\theta)$. Intuitively, condition (1) requires the agent's marginal payoff be negative and proportional to both the state and the agent's type at the principal's ideal action. (The case of $g(\gamma)\zeta(\theta) > 0$ is symmetric.) Importantly, for a given pair (θ, γ), this condition is local as the factorization is required at point $a = \theta$ only.

Finally, denote $y(\theta, \gamma)$ the maximizer of $V(a, \theta, \gamma)$ over $a \in \Theta_0$. By conditions on V, $y(\theta, \gamma)$ exists and is unique and continuous in θ. Also, by the strict pseudo-concavity of V in a and $V'_a(\theta, \theta, \gamma) < 0$, it follows that $y(\theta, \gamma) \leq \theta$ for all $(\theta, \gamma) \in \Theta_0 \times \mathbf{T}$ with the strict inequality if $\theta > \theta_0$, and $V'_a(a, \theta, \gamma) \lesseqgtr 0$ if and only if $a \gtreqless y(\theta, \gamma)$.

Information. A *signal structure* ξ determines a probability distribution $F_\xi(s|\theta)$ over signals s conditional on the state θ. For simplicity and with a minor abuse of notation, hereafter we use the term ξ for $F_\xi(s|\theta)$. 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

[12] For example, consider $\mathbf{A} = \mathbb{R}_+$ and the principal's payoff functions $U_1(a, \theta) = -(a - \theta)^2, U_2(a, \theta) = \theta \ln a - a$, and $U_3(a, \theta) = 1$ if $a = \theta$ and 0 otherwise. Because $y(\theta) = \theta$ is the unique maximizer for all these functions, the models with these principal's payoffs are equivalent.

[13] A differentiable function $\mathcal{U}(x)$ is *(strictly) pseudo-concave* on a convex set X if for every $(x, y) \in X^2, y \neq x$, $\mathcal{U}(x) < (\leq)\mathcal{U}(y)$ implies $\mathcal{U}'(x)(y - x) > 0$ (Hadjisavvas et al., 2005). If $\mathcal{U}'(x_0) = 0$ for $x_0 \in X$, then x_0 is a (unique) maximizer of $\mathcal{U}(x)$.

$\mathbf{S}_\xi$. 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 $\mathcal{I} = \{F(.|\theta) \,|\theta \in \Theta\}$ 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.[14] For our purposes, it is sufficient to restrict attention to ρ that randomizes over a finite number of signal structures. 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 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 $y_\xi(m)$.[15] Then, the state θ and the signal structure $\xi \in \mathcal{I}_\rho$ are randomly and independently drawn according to f and ρ, respectively, where ξ becomes the private information of the principal.[16] 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.[17]

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.[18] 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 $\{\varphi_\xi(s) : \xi \in \mathcal{I}_\rho\}$.

[14] Formally, $\mathcal{I}$ is a subspace of the space of distribution functions on $\mathbb{R}$ endowed with the Levy metric topology; and $\Delta\mathcal{I}$ is the space of probability measures defined on the Borel sets generated by that topology. I am thankful to the Associate Editor for providing these specifications.

[15] In general, 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.

[16] 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.

[17] In general, the principal can specify the richer message space $\mathbf{M} = \mathbf{S} \times \mathbf{T}$ and request the agent to report information not only about the state, but also about her type. However, our main results do not rely on the principal's knowledge of the agent's type and thus do not require the agent report it.

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

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,

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

Two comments are necessary here. First, since our model requires specifying the principal's ideal action $y(\theta)$ only, the nature of the principal's action $y_\xi(m)$ is unclear if he infers imperfect information about θ. However, it is not an issue as we focus on implementing $y(\theta)$ for all θ. Formally, we construct ρ, which randomizes between perfectly informative signals functions $\xi(\theta)$ only and sustains the truth-telling equilibrium. Hence, $y_\xi(s)|_{s=\xi(\theta)}=\theta$ for all $\xi\in\mathcal{I}_\rho$ and $\theta\in\Theta$. In other words, if a signal s is generated by any $\xi\in\mathcal{I}_\rho$ in state θ, and the agent reports s truthfully, then the principal infers θ and implements his ideal action $y(\theta)=\theta$. Second, the principal's ability to commit to $y_\xi(m)$ does not play a role. For example, he can commit to $y_\xi(m)$ ex-ante, or it can be interim-optimal, i.e., a solution to the problem of maximizing the principal's expected payoff conditional on ξ and m.

3 Perfect information extraction and implementation

Before starting the general 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. In this example, we consider the agent's preferences, which are strictly decreasing in the principal's action (i.e., pseudo-concave) and independent of the state and the agent's type. (The second example below considers general preferences, which depend on both the state and the type.) If the agent's signal structure were publicly known, then the above specification is the worst-case scenario for the principal, as he could not elicit any relevant information from the agent. This negative outcome is due to two factors. First, because the information about the state has no value to the agent, the principal cannot exploit the agent's incentives to acquire this information.[19] Second, since the agent's preferences are strictly monotone in action, irrespective of the information received, the agent will send only those messages that induce extreme feasible actions.[20] Together, these factors completely suppress the agent's incentives to convey any relevant information. In contrast, assigning a private signal structure to the agent allows the principal to extract perfectly precise information from her.

[19] In the case of state-dependant agent's preferences, the agent's incentives to acquire information can play a critical role for information extraction. For example, Ivanov (2015, 2016) shows that the principal can elicit perfect information from the agent and implement her ideal actions in the cheap-talk framework by exploiting these incentives in a dynamic way.

[20] Chakraborty and Harbaugh (2010) and Lipnowski and Ravid (2020) study cheap-talk communication with a perfectly informed agent whose preferences depend on the principal's action only. As they show, the agent can disclose relevant information if the action set is multi-dimensional or the agent's payoff function is not monotone in the action. In our example with single-dimensional actions and monotone preferences of the agent, informative communication is not feasible regardless of the agent's public signal structure.

3.1 Example A: action-only dependent 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 $a_0 = 0$.[21]

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 truth-telling 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[22]

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

where φ_k is the inverse of ξ_k:

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

Denote $q_k(s, \rho)$ the probability of the posterior state $\theta_k = \varphi_k(s)$:

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

If the principal believes that the agent is truthful, then a message $m \in \mathbf{S}$ induces the action

$$y_{\xi_k}(m) = \theta_k = \varphi_k(m), k = 1, 2$$

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

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

[21] Formally, $V_a'(0) = 0$ for $b > 1$, which violates the condition $V_a'(\theta) < 0$ for all $\theta \in \Theta$. However, neither of our results is affected by this technicality.

[22] Technically, the result follows from the Lebesgue differentiation theorem, which allows to evaluate the posterior belief $q(\theta|s, \rho^o)$ at posterior state $\theta \in \{\varphi_1(s), \varphi_2(s)\}$ as the limit of infinitesimal averages taken about the point.

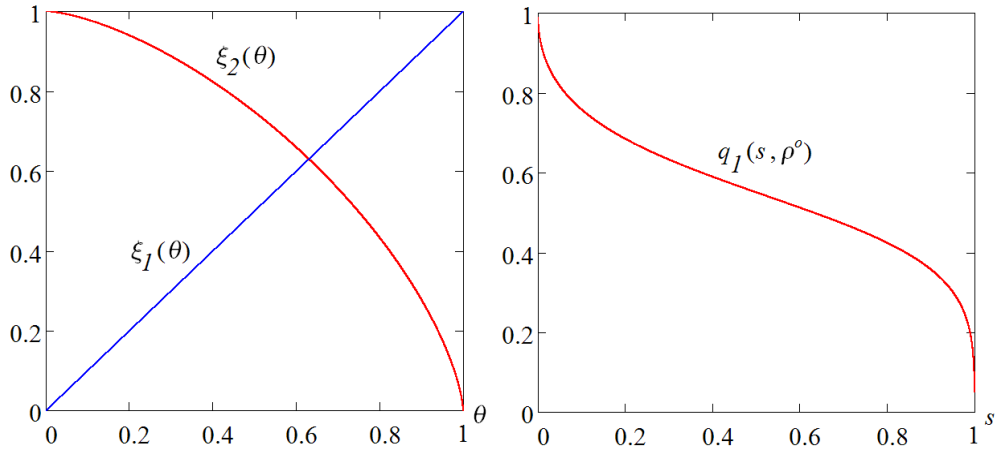

Figure 1: Signal functions $\xi_i(\theta), i = 1, 2$ and the posterior probability $q_1(s, \rho^o)$ for $f(\theta) = 1$ and $V(a) = -a^2$.

As an illustration, consider the quadratic payoff function $V(a) = -a^2$. 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 simple calculus, it is easy to verify that $EV(m|s, \rho^o)$ is maximized at unique $m = s$ for all $s \in \mathbf{S}$, i.e., the agent strictly benefits from reporting truthfully.

Intuitively, this example demonstrates the keys factors of private signal structures, which sustain agent's truth-telling. 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 $y(\theta_k) = \theta_k$ is monotone in the induced posterior state $\theta_k, k = 1, 2$. 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 principal'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 $y_\xi(m)$ taken for a signal function ξ matches the induced posterior state $\theta_k = \varphi_k(m)$. As a result, the shape of $\varphi_k(m)$ determines the marginal effect of an agent's message m on the principal's action a_k. 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 inducing a lower action are larger for high actions. 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 increases 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 penalizing 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 incentive to decrease the action a_1 if it is high and, thus, is associated with a high signal s, then understating such signals will result in the higher losses from the penalizing action a_2.

Second, the shapes of the inverses reallocate the agent's posterior beliefs between the posterior states $\theta_1 = \varphi_1(s)$ and $\theta_2 = \varphi_2(s)$ from states in which the agent has the stronger incentives to lie toward those with weaker incentives. In our example, the buyer's payoff is steeper for high actions. Because the principal matches actions to states, the agent's benefits from distorting information are higher for high states. As a result, for high signals the slopes $|\varphi_1|$ and $|\varphi_2|$ of the inverses φ_1 and φ_2 assign a lower posterior probability to the higher posterior states. Because the agent places lower probabilities on these states, this decreases her posterior (that is, conditional on signal) benefits from manipulating the signal and thus her incentives to lie. The balance between these two forces sustains agent's truth-telling.

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. A steeper slope $|\varphi_2'(s)|$ in response to 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. Next, 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.

Finally, note that perfect information extraction does not rely on the strict concavity of the agent's preferences. Specifically, the main arguments above hold through if the agent' payoff is linear in a, that is, $V(a) = -a$. However, there are two key differences between the cases of the linear and strictly concave payoff functions. First, the agent with the linear payoff function is indifferent among all messages m for any signal $s \in \mathbf{S}$. At the same time, the example above demonstrates that the agent's incentive to report truthfully are not driven by this indifference. Second, (4) implies that the inverse φ_2 is linear for $V(a) = -a$. As a result, the posterior beliefs $q_k(s, \rho^o) = \frac{1}{2}, k = 1, 2$ given by (5) are constant. However, the linearity of the inverses does not allows us to see an interplay between their shapes, the posterior beliefs, and the agent's marginal benefits and losses from distorting her information.

3.2 Optimal private signal structures

We start the general construction with the following lemma. It demonstrates how the agent'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 $\xi_k' \neq 0$. Denote $\varphi_k = \xi_k^{-1}$ the inverse of ξ_k. Then,*

$$q_k(s,\rho) = \frac{p_k f(\varphi_k(s)) |\varphi_k'(s)|}{p_1 f(\varphi_1(s)) |\varphi_1'(s)| + p_2 f(\varphi_2(s)) |\varphi_2'(s)|}. \tag{7}$$

Intuitively, the lemma highlights the key feature of signal functions, specifically, the possibility to induce the agent's posterior beliefs $q_k(s,\rho)$ 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_k'. To see this feature, suppose $p_1 = p_2 = \frac{1}{2}$ and f is uniform on $[0,1]$, i.e., $f(\theta) = 1$. 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 any q_k between 0 and 1. As shown in the example above, the principal can use this ratio in order to reallocate the agent's posterior beliefs from the states with higher benefits from distorting the signal to the states with lower benefits and, as a result, reduce the agent's interim incentives to lie.

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 $\varphi_k = \xi^{-1} : \mathbf{S} \to \Theta$ 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 truth-telling and, thus, allows the principal to implement $y(\theta)$ upon inferring θ from m and ξ_k. Importantly, the signal sets, i.e., the images of 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_k from 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_k is

$$y_{\xi_k}(m) = \varphi_k(m), k = 1,2. \tag{8}$$

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

$$EV(m|s,\gamma,\rho) = \sum_{k=1}^{2} q_k(s,\rho) V(\varphi_k(m), \varphi_k(s), \gamma). \tag{9}$$

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

$$\frac{\partial}{\partial m}EV\left(m|s,\gamma,\rho\right) \ = \sum_{k=1}^{2} q_k\left(s,\rho\right) V_a'\left(\varphi_k\left(m\right),\varphi_k\left(s\right),\gamma\right)\varphi_k'\left(m\right), \tag{10}$$

where $V_a' = \frac{\partial V}{\partial a}$. The truthful strategy is optimal if

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

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

$$\frac{\partial}{\partial m}EV\left(m|s,\gamma,\rho\right)|_{m=s} \ = \sum_{k=1}^{2} q_k\left(s,\rho\right) V_a'\left(\varphi_k\left(s\right),\varphi_k\left(s\right),\gamma\right)\varphi_k'\left(s\right) = 0, (s,\gamma)\in\mathbf{S}\times\mathbf{T}. \tag{12}$$

Next, invoking (1) results in

$$V_a'\left(\varphi_k\left(s\right),\varphi_k\left(s\right),\gamma\right) = g\left(\gamma\right)\zeta\left(\varphi_k\left(s\right)\right),$$

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

$$\frac{\partial}{\partial m}EV\left(m|s,\rho\right)|_{m=s} = \sum_{k=1}^{2} q_k\left(s,\rho\right) g\left(\gamma\right)\zeta\left(\varphi_k\left(s\right)\right)\varphi_k'\left(s\right) = 0 \text{ for all } s\in\mathbf{S}.$$

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

$$\varphi_1'\left(s\right)|\varphi_1'\left(s\right)|\, f(\varphi_1\left(s\right))\zeta\left(\varphi_1\left(s\right)\right) + \varphi_2'\left(s\right)|\varphi_2'\left(s\right)|\, f(\varphi_2\left(s\right))\zeta\left(\varphi_2\left(s\right)\right) = 0, \tag{13}$$

Because $\varphi_1' = |\varphi_1'| > 0, f>0$, and $\zeta<0$, it immediately follows that $\varphi_2'<0$ and, hence, $|\varphi_2'\left(s\right)| = -\varphi_2'\left(s\right)$. Therefore, (13) can be expressed as

$$\varphi_1'\left(s\right)h(\varphi_1\left(s\right)) = -\varphi_2'\left(s\right)h(\varphi_2\left(s\right)), \tag{14}$$

where

$$h\left(\theta\right) = \sqrt{-f(\theta)\zeta\left(\theta\right)} > 0. \tag{15}$$

The solution to (14) with the boundary condition $\varphi_2\left(\underline{s}\right) = \bar{\theta}$ must satisfy

$$\Psi\left(\varphi_1\left(s\right)\right) + \Psi\left(\varphi_2\left(s\right)\right) = \Psi\left(\underline{\theta}\right) + \Psi\left(\bar{\theta}\right),$$

where

$$\Psi\left(x\right) = \int h\left(x\right)dx \tag{16}$$

is the antiderivative of h.[23] This gives the relationship between φ_1 and φ_2:

$$\varphi_2(s) = \Psi^{-1}\left(\Psi(\underline{\theta}) + \Psi\left(\bar{\theta}\right) - \Psi(\varphi_1(s))\right), \tag{17}$$

In general, a pair of inverses $\varphi_k : \mathbf{S} \to \Theta, k = 1, 2$ related by (17) 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 2 *Given $\Theta_0 \subset \Theta$, $\nu(a, \theta, \gamma) = \frac{V'_a(a,\theta,\gamma)}{h(a)}$ is decreasing in a for $a > y(\theta, \gamma)$ and $(\theta, \gamma) \in \Theta_0 \times \mathbf{T}$.*

Notably, this condition is imposed on the model primitives: the payoff function V and the prior density f. Therefore, agent's truth-telling can be optimal for various pairs $\{\varphi_1, \varphi_2\}$ parameterized by φ_1. Condition 2 can be explained by noting that $\nu(a, \theta, \gamma)$ is the ratio of two functions, $V'_a(a, \theta, \gamma)$ and $h(a) = \sqrt{f(a)|\zeta(a)|}$. 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 (increasing) in a if $V(a, \theta, \gamma)$ is concave (convex) in a. The second function $h(a)$ reflects the marginal benefits $|\zeta(a)|$ from distorting an action a weighted by the prior density $f(a)$. In this light, Condition 2 requires the function $V(a, \theta, \gamma)$ be 'not very convex' in a, and the agent's weighted marginal benefit $h(a)$ be 'relatively decreasing' in action a (since $V'_a < 0$ for $a > y(\theta, \gamma)$ and $h > 0$ imply $\nu < 0$).[24]

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 (17). Then ρ^* allows the principal to elicit the perfect information about the state from the agent and, hence, implement his ideal action if the above regularity condition holds.

Theorem 1 *Suppose V satisfies Condition 1 and (1), and (f, V) satisfy Condition 2 for $\Theta_0 = \Theta$. Consider a pair of functions (φ_1, φ_2) related by (17), where $\varphi_1 : \mathbf{S} \to \Theta$ is differentiable and $\varphi'_1 > 0$. Then the private signal structure ρ^* that randomizes between φ_1^{-1} and φ_2^{-1} with equal probabilities sustains agent's truth-telling for all $(\theta, \gamma) \in \Theta \times \mathbf{T}$.*

This result extends 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 $y_A(\theta) \in \mathbb{R}$ 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 her payoff function on the state, or the supermodularity. All these assumptions are replaced with the pseudo-concavity of the agent's payoff $V(a, \theta, \gamma)$ in the principal's action a and its local monotonicity at the principal's ideal action, $V'_a(\theta, \theta, \gamma) < 0$. Specifically, recall that the opposite monotonicities of inverses $\varphi_k(s), k = 1, 2$ in signal s and the monotonicity of the principal's action $y(\theta) = \theta$ in θ imply that the principal's actions $a_k = \varphi_k(m), k = 1, 2$ under different signal functions ξ_k react oppositely in response to the agent's message m.

[23] 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 $\mathbf{S}$.

[24] If $V'_a(\theta, \theta, \gamma) = g(\gamma)\zeta(\theta) > 0$, then $\nu > 0$. In this case $h(a)$ must be 'relatively increasing' in a.

Then, the monotonicity of the agent's payoff V in the principal's action at $a = \theta$ implies that the opposite reactions of $\varphi_k(m), k = 1, 2$ 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 (17) between the inverses $\varphi_k(s) = 1, 2$ in order to sustain agent's truth-telling. Notably, the logic above does not depend on the concavity of the agent's preferences in a. As a result, the theorem may hold for arbitrarily convex payoff functions V as long as (V, f) satisfy Condition 2.[25]

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 truth-telling strategy under the private signal structure ρ. Once the agent's posterior payoff function (9) is single-peaked in m and achieves its maximum at $m = s$, this prevents the agent's local distortions of her signal (i.e., small lies) as well as global distortions (large lies). In order to guarantee the single-peakedness of the posterior payoff, we employ the pseudo-concavity of V. The tension with pseudo-concavity, however, comes from three factors. First, while each function $V(\varphi_k(m), \varphi_k(s), \gamma), k = 1, 2$ in (9) is strictly monotone and, hence, pseudo-concave in m, the agent's posterior payoff EV is a convex combination of these functions, which is generally not pseudo-concave.[26] Second, a function $V(\varphi_k(m), \varphi_k(s), \gamma), k = 1, 2$ is a composite function of V and φ_k. As a result, the pseudo-concavity of this function is violated if the (local) concavity of V in a is dominated by the convexity of $\varphi_k(m)$. Third, φ_1 and φ_2 are functionally dependent by (17). Therefore, the posterior payoff EV is pseudo-concave in m if: i) each composite function $V(\varphi_k(m), \varphi_k(s), \gamma), k = 1, 2$ is pseudo-concave, and ii) a convex combination of these functions is also pseudo-concave.

Condition 2 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).[27] 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 $V(\varphi_k(m), \varphi_k(s), \gamma), k = 1, 2$, and use the functional relationship (14) between φ_1 and φ_2. This completes the proof of the theorem.

[25] Consider $V(a) = e^{-ba}$. It is decreasing in a, and the parameter b is the Arrow–Pratt measure of absolute risk aversion $R(a) = -\frac{V_a''}{V_a'} = b$. According to this measure, V becomes more convex as b increases and eventually converges to the extreme case of convexity: $V(0) = 0$ and $V(\theta) = -1$ for $\theta > 0$. Then $V_a' = -be^{-ab}$, and $\zeta(\theta) = -V_a'(a)|_{a=\theta} = be^{-b\theta} > 0$. Also, consider the truncated exponential distribution on $[0, 1]$ with the density $f(\theta) = Ce^{c\theta}$, where $C = \frac{c}{1-e^{-c}}$. Then $h(a) = \sqrt{\zeta(a) f(a)} = \sqrt{Cb} e^{-\frac{b+c}{2}a}$, and $\nu(a) = \frac{V_a'(a,\theta,\gamma)}{h(a)} = -\sqrt{\frac{b}{C}} e^{\frac{c-b}{2}a}$ is decreasing in a for any b and $c > b$.

[26] For example, e^m and e^{-m} are strictly pseudo-concave in m, but $\frac{e^m + e^{-m}}{2}$ is not.

[27] 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 Application: surplus extraction

As the leading economic application, we consider the problem of full surplus extraction from a buyer with non-quasilinear preferences and multidimensional information. Specifically, the buyer's payoff function depends on the state θ and the type γ. The state θ reflects the willingness of the buyer to pay for the object and, thus, is of interest to the seller as well, while the type γ affects the buyer's payoff only. The buyer is privately informed about γ and is uninformed about θ. Our main result establishes that the seller can extract the perfect information about the state from the buyer and, as a consequence, her full surplus, by privately designing the buyer's signal structure that randomizes between two perfectly informative signal functions.

4.1 Bilateral trade: setup

A single buyer (she) and a seller (he) are involved in trading a single indivisible object. The buyer's utility from obtaining the object and making a payment t to the seller is determined by the payoff function $V(t, \theta, \gamma)$. The state θ determines the buyer's willingness to pay, i.e., $t = \theta$ is the highest price that the buyer is willing to pay in state θ. Similarly to the main framework, γ is the buyer's type. 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. For example, it can reflect the marginal 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 of its shape.

The state θ is a random variable distributed according to a continuous density $f(\theta) > 0$ from the state space $\Theta = \left[0, \bar{\theta}\right]$. The type γ is drawn from the type space $\mathbf{T} \subset \mathbb{R}$. The variables θ and γ are independent. The statistical independence between the 'quality' characteristics of the object and some intrinsic buyer's preferences also has antecedents in the mechanism design literature.[28]

Seller's preferences. As in the main framework, we do not define the seller's utility function $U(t, \theta)$. Instead, we assume that the seller aims to extract the full surplus θ from the buyer in *target states* $\Theta_0 = \left[\theta_0, \bar{\theta}\right]$, where $\theta_0 \in \Theta$. Intuitively, Θ_0 is the subset of states in which the buyer's willingness to pay θ exceeds the seller's utility from keeping the object.[29]

Buyer's preferences. We impose the following assumptions about the buyer's payoff.

Condition 3 *$V(t, \theta, \gamma)$ is strictly increasing in θ, and differentiable and strictly pseudo-concave in t for all $(t, \theta, \gamma) \in \Theta^2 \times \mathbf{T}$.*

[28] According to Esö and Szentes (2007): "for a specific example, suppose that the object for sale is a car, and assume that the buyer knows its make, model, age, and mileage, but not its colour, which the seller can reveal. It seems reasonable to assume that a buyer's initial willingness to pay for the car and his colour preference are statistically independent".

[29] As a special case, suppose that the seller's preferences are quasilinear, $U(t, \theta) = u(\theta) + t$, where $u(\theta)$ is the seller's payoff from keeping the object in state θ, such that $u(\theta) \leq \theta$ if and only if $\theta \geq \theta_0$ for $\theta_0 \in \Theta$. As a result, the seller's goal is to extract the buyer's surplus θ if and only if $\theta \in \Theta_0 = \left[\theta_0, \bar{\theta}\right]$.

The value of the buyer's outside option is 0 (a normalization), which she receives in the case of not obtaining the object and not making a payment. As noted above, θ represents the buyer's willingness to pay, i.e., the payment, which makes her indifferent between making this payment for the object and taking the outside option. That is, V must satisfy the condition

$$V(\theta, \theta, \gamma) \equiv 0 \text{ for all } (\theta, \gamma) \in \Theta_0 \times \mathbf{T}. \tag{18}$$

Condition (18) implies that the buyer's willingness to pay does not depend on γ. This condition is not novel and imposed, for instance, in the literature on auctions with buyers' non-linear preferences (see Section 4.1 in Krishna, 2009). Specifically, this literature considers buyers with payoff functions $V(t, \theta) = v(\theta - t)$, where $v(.)$ is in the class of functions $\mathbf{V} = \{\mathbf{C}^1 | v(0) = 0, v' > 0\}$.[30] Thus, if V is an element of a subset $\mathbf{T} \subset \mathbf{V}$ parameterized by γ, that is, $V(t, \theta, \gamma) = v(\theta - t, \gamma)$, then it must satisfy (18). For such payoff functions, while γ does not affect the buyer's maximum acceptable payment $t = \theta$, it affects her payoff for $t \neq \theta$. For example, the buyer's willingness to pay for a car can be determined entirely by its quality characteristics such as the seating capacity or horsepower. However, the pleasure of driving the car is also affected by its color or the shape of the buyer's payoff function.

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

$$V_t'(\theta, \theta, \gamma) = g(\gamma)\zeta(\theta) < 0 \text{ for } (\theta, \gamma) \in \left(\theta_0, \bar{\theta}\right) \times \mathbf{T}, \tag{19}$$

where $g > 0$ and $\zeta < 0$.

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 mechanism.

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 mechanism.[31] 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

[30] For risk-averse buyers, the concavity condition $v'' < 0$ is additionally imposed.

[31] 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.

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:

$$EV(s|s,\gamma,\rho) = \max_{m\in\mathbf{S}} EV(m|s,\gamma,\rho) \text{ for all } (s,\gamma)\in\mathbf{S}\times\mathbf{T}, \text{ (IC)} \tag{20}$$

$$EV(s|s,\gamma,\rho) \geq 0 \text{ for all } (s,\gamma)\in\mathbf{S}\times\mathbf{T}, \text{ (IR)} \tag{21}$$

Here, $EV(m|s,\gamma,\rho)$ is the buyer's posterior payoff

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

where $V(t,0,\gamma)$ is the buyer's value in the case of paying t for the worthless object, i.e., the one for which the buyer's willingness to pay is $\theta = 0$. It is equivalent to not obtaining the object while still paying t.

Given this framework, we start with an simple example, which provides the key insights into the general construction of private signal structures and mechanisms that extract full information and 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, $f(\theta) = 1, \theta \in \Theta = [0,1]$, and the type γ is drawn from $\mathbf{T} \subset \mathbb{R}$. The variables θ and γ are independent. The subset of seller's target states is $\Theta_0 = [\theta_0, 1]$.

Following the literature on auctions with risk-averse buyers, the buyer's payoff from consuming the product and making a payment t is given by the function

$$V(t,\theta,\gamma) = v(\theta - t,\gamma), \tag{22}$$

where $v(x,\gamma)$ is strictly increasing, differentiable, and concave in x, and $v(0,\gamma) = 0$ for all $\gamma \in \mathbf{T}$. The last property implies that the buyer's willingness to pay is θ for all $(\theta,\gamma)\in\Theta\times\mathbf{T}$. That is, (22) satisfies condition (18). Furthermore, (22) implies

$$V'_t(t,\theta,\gamma)|_{t=\theta} = -v'_x(0,\gamma) < 0,$$

that is, condition (19) holds, where $g(\gamma) = v'_x(0,\gamma) > 0$ and $\zeta(\theta) = -1$.

Next, we verify that Condition 2 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 as well.

An example of function (22) 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 relative 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

$$v(x, \gamma) = \begin{cases} \frac{1-\gamma}{\gamma}\left(\frac{\alpha}{1-\gamma}x + \beta\right)^{\gamma} - \frac{1-\gamma}{\gamma}\beta^{\gamma} & \text{if } \gamma \neq 0, \\ \ln\left(1 + \frac{\alpha}{\beta}x\right) & \text{if } \gamma = 0, \end{cases}$$

where $\alpha > 0$ and $\frac{\alpha}{1-\gamma}x + \beta > 0$.[32] Depending on values of γ, α, and β, this form encompasses many standard payoff functions, such as linear, exponential (constant absolute risk aversion), power (constant relative risk aversion), and logarithmic.[33] For this payoff function, the value of γ determines the degree of buyer's risk aversion and thus substantially affect her utility and incentives in different mechanisms without the full surplus extraction. Because γ is the buyer's private information, the seller can be uncertain about the specific shape of the payoff function in our model. For example, he might not know whether it takes the power, exponential, or linear form. As we demonstrate below, the structure of the private signal structure and the mechanism are invariant to this information. At the same time, these tools jointly allow the seller to extract perfect information and full surplus from the buyer.

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

$$\begin{aligned} &\xi_1(\theta) = \theta, \text{ and} \\ &\xi_2(\theta) = \begin{cases} \theta & \text{if } \theta < \theta_0, \\ 1 + \theta_0 - \theta & \text{if } \theta \geq \theta_0. \end{cases} \end{aligned} \tag{23}$$

Fig. 2 depicts signal functions $\xi_k(\theta), k = 1, 2$. 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 truth-telling 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 $q(\theta|s, \rho^*)$ of state θ induced by signal s under a private signal structure ρ^* is the binary distribution, which places probabilities $\frac{1}{2}$ on the posterior states

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

where $\varphi_k = \xi_k^{-1}, k = 1, 2$.

Also, consider the direct mechanism $\mathcal{M}^* = \left(Q_\xi^*(s), t_\xi^*(s)\right)$ with the message set $\mathbf{M} =$

[32] 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 does not affect the shape of the function, but guarantees that condition (18) holds.

[33] See Ingersoll (1987).

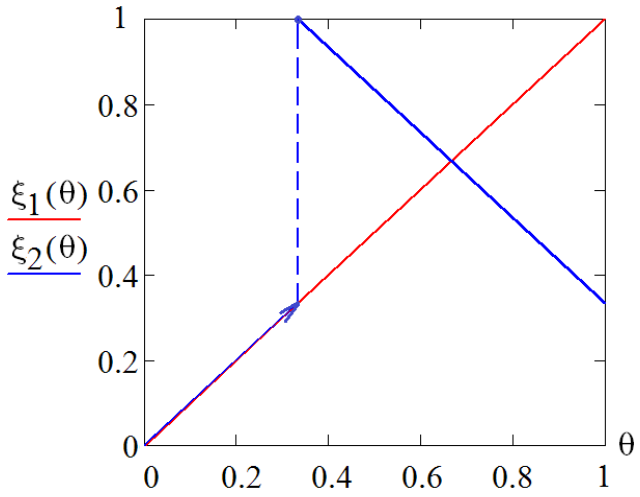

Figure 2: Signal functions $\xi_i(\theta), i = 1, 2$ for $\theta_0 = \frac{1}{3}$.

$\mathbf{S} = [0, 1]$, such that

$$Q_\xi^*(s) = Q^*(s) = \begin{cases} 1 & \text{if } s \geq \theta_0, \\ 0 & \text{if } s < \theta_0, \end{cases}$$
$$t_\xi^*(s) = \begin{cases} \varphi(s) = \xi^{-1}(s) & \text{if } s \geq \theta_0, \\ 0 & \text{if } s < \theta_0. \end{cases}$$

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

$$\begin{aligned} EV(m|s, \gamma, \rho^*) &= 0 \text{ if } m < \theta_0, s \in \mathbf{S}, \text{ and} \\ EV(m|s, \gamma, \rho^*) &= q_1(s, \rho^*) V\left(t_{\xi_1}^*(m), \varphi_1(s), \gamma\right) + q_2(s, \rho^*) V\left(t_{\xi_2}^*(m), \varphi_2(s), \gamma\right) \\ &= \frac{1}{2} v(\varphi_1(s) - \varphi_1(m), \gamma) + \frac{1}{2} v(\varphi_2(s) - \varphi_2(m), \gamma) \\ &= \begin{cases} \frac{1}{2} v(s - m, \gamma) + \frac{1}{2} v(m - s, \gamma) & \text{if } m \geq \theta_0, s \geq \theta_0, \\ \frac{1}{2} v(s - m, \gamma) + \frac{1}{2} v(s - (1 + \theta_0 - m), \gamma) & \text{if } m \geq \theta_0, s < \theta_0. \end{cases} \end{aligned}$$

It is straightforward to show that $EV(m|s, \gamma, \rho^*)$ is maximized at $m = s$ for all $(s, \gamma) \in \mathbf{S} \times \mathbf{T}$, and $EV(s|s, \gamma, \rho^*) = 0$ for all $s \in \mathbf{S}$. Hence, the interim incentive-compatibility constraints (20) and the interim individual-rationality constraints (21) hold. Furthermore, the buyer's ex-post payoff for $\theta_k = \varphi_k(s) \geq \theta_0$ is

$$\begin{aligned} V\left(\xi_k(s), t_{\xi_k}^*(s), \gamma\right) &= v\left(\varphi_k(s) - t_{\xi_k}^*(s), \gamma\right) \\ &= v(\varphi_k(s) - \varphi_k(s), \gamma) = 0 \text{ if } s \geq \theta_0, \gamma \in \mathbf{T}, k = 1, 2. \end{aligned}$$

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_k.

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 highest acceptable payment exceeds the seller's ex-post benefits from keeping the object. Second, the incentive-compatibility in these states 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 truth-telling.[34] As a result, the mechanism perfectly infers the posterior state θ_k from m and the realized signal function ξ_k. Third, the above effect does not depend on the absolute values of buyer's 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.

It is worth mentioning that 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:[35]

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

Because the buyer is risk-neutral in this case, her interim payoff EV 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 $s \geq \theta_0$:

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

4.3 Information and mechanism design

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 $\xi_2 = \varphi_2^{-1} : \Theta \to \mathbf{S}$, such that $\mathbf{S} = [\underline{s}, \bar{s}]$, $\varphi_1' > 0$, and

$$\varphi_2(s) = \begin{cases} \varphi_1(s) & \text{if } s < s_0, \\ \Psi^{-1}\left(\Psi(\theta_0) + \Psi\left(\bar{\theta}\right) - \Psi(\varphi_1(s))\right) & \text{if } s \geq s_0, \end{cases} \tag{24}$$

where $s_0 = \xi_1(\theta_0)$, or equivalently, $\varphi_1(s_0) = \theta_0$, and $\Psi(x)$ is given by (15) and (16). Thus, upon receiving a signal $s < s_0$ the buyer perfectly infers the state $\theta = \varphi_1(s) < \theta_0$. 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 $\varphi_2(s_0) = \bar{\theta}$.

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

[35] In general, $v(\theta - t, \gamma) = \alpha(\gamma)(\theta - t) + \beta(\gamma)$. However, condition (18) implies $\beta(\gamma) = 0$.

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

$$Q_\xi^*(s) = Q^*(s) = \begin{cases} 1 & \text{if } s \geq s_0, \\ 0 & \text{if } s < s_0, \end{cases} \tag{25}$$

$$t_\xi^*(s) = \begin{cases} \varphi(s) = \xi^{-1}(s) & \text{if } s \geq s_0, \\ 0 \text{ if } s < s_0. \end{cases} \tag{26}$$

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

$$EV(m|s,\gamma,\rho^*) = \begin{cases} \sum_{k=1}^{2} q_k(s,\rho^*) V(\varphi_k(m), \varphi_k(s), \gamma) & \text{if } m \geq s_0, \\ 0 \text{ if } m < s_0, \end{cases} \tag{27}$$

where the top line is identical to (9), and the bottom line holds due to $V(0,0,\gamma) = 0$.

The theorem below establishes that a combination of ρ^* and $\mathcal{M}^*$ extracts the full information and surplus from the buyer for states $\theta \geq \theta_0$ under Condition 2.

Theorem 2 *Suppose V satisfies Condition 3, (18)–(19), and (f, V) satisfy Condition 2 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 (24). Then ρ^* and the mechanism $\mathcal{M}^* = \left(Q_\xi^*(s), t_\xi^*(s)\right)$ extract the full surplus for $(\theta, \gamma) \in \Theta_0 \times \mathbf{T}$.*

The proof of theorem consists of two parts. The first part demonstrates that the mechanism $\mathcal{M}^*$ 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_k in each posterior state $\theta_k = \varphi_k(s), k = 1, 2$, 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_k.

The main part of the proof is to establish the interim incentive-compatibility of the mechanism $\mathcal{M}^*$, 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 original principal-agent model with the state space Θ_0, the prior density $f_0(\theta) = f(\theta|\theta \in \Theta_0)$, the signal set $\mathbf{S}_0 = [s_0, \bar{s}]$, 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 (22) and Condition 2 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 $\mathcal{M}^*$ 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.

Importantly, conditions (18) and (19) 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 players' preferences and discrete states.[36] However, 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.

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 his 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-concave, depend on the privately known component, and are independent of the state.

We conclude the paper by discussing the limits of our framework for robust full surplus extraction and suggesting possible avenues for future research. Regarding the former question, surplus extraction is predicated on several key elements. The first main element is the dependence of the private signal structure on the buyer's preferences and, hence, on the principal's knowledge of them. Specifically, the buyer's marginal payoff with respect to payment at her willingness to pay must be decomposable into separate functions of her state and type. This is necessary to neutralize the effect of the buyer's type on her willingness to report the truth about the state whenever her surplus is fully extracted. In addition, the buyer's willingness to pay must be independent of her taste preferences, which are reflected in the type. This precludes the principal's incentives to learn the buyer's type. The second key element is the flexibility of the principal over the buyer's information about the state. In particular, the principal must have access to signal functions with opposite monotonicities. Finally, the state and type must be independent, that is, any information about the state does not update the buyer's beliefs about her type. The reason is that the signal functions finely balance the buyer's marginal benefits and losses for each posterior belief induced by a signal. Thus, any distortion in posterior beliefs about the state caused by the buyer's type will distort the buyer's incentives to report her signal truthfully. Furthermore, since the set of signals is continuous, the buyer is always able to distort her information locally. At the same time, the main idea behind our construction and the arguments do not depend on the cardinality of the state space and thus can be equally applied to discrete distributions of states. The main difference in this case is that the buyer's posterior beliefs must be updated by using the standard Bayes formula rather than Lemma 1 based on the Lebesgue differentiation theorem.

Regarding potential avenues for future research, 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

[36] See Remark 8 in Krähmer (2020).

buyer's payoff is non-monotone in the principal's action. Intuitively, truth-telling 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.[37] 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.[38] 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 $\varphi_k : \mathbf{S} \to \Theta, k = 1, 2$, such that φ_1 is differentiable, $\varphi_1' > 0$, and φ_2 is given by (17). 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 $\xi_k = \varphi_k^{-1} : \Theta \to \mathbf{S}, k = 1, 2$ 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 $EV(m|s, \gamma, \rho^*)$ 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).[39]

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

[37] 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 $V(y(\varphi_i(m)), \varphi_i(s), \gamma)$.

[38] 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.

[39] A function $\phi(x)$ is *(strictly) pseudo-monotone* on a convex set X if for every $(x, y) \in X^2, y \neq x$, $\phi(x)(y - x) \leq 0$ implies $\phi(y)(y - x) \leq (<) 0$. Equivalently, $\phi(x) \leq 0$ implies $\phi(y) \leq (<) 0$ for all $y > x$; and $\phi(x) \geq 0$ implies $\phi(y) \geq (>) 0$ for all $y < x$.

$\alpha_k \geq 0, k = 1, 2$ if and only if: (i) $-\frac{\mathcal{V}_2(m)}{\mathcal{V}_1(m)}$ is decreasing in m for all m such that $\mathcal{V}_1(m) < 0$ and $\mathcal{V}_2(m) > 0$; and (ii) $-\frac{\mathcal{V}_1(m)}{\mathcal{V}_2(m)}$ is decreasing in m for all m such that $\mathcal{V}_1(m) > 0$ and $\mathcal{V}_2(m) < 0$.[40]

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

$$\phi(m|s, \gamma, \rho^*) = \sum_{k=1}^{2} q_k(s, \rho^*) V_a'(\varphi_k(m), \varphi_k(s), \gamma) \varphi_k'(m) = \sum_{k=1}^{2} q_k(s, \rho^*) \mathcal{V}_k(m, s, \gamma),$$

where

$$\mathcal{V}_k(m, s, \gamma) = V_a'(\varphi_k(m), \varphi_k(s), \gamma) \varphi_k'(m), k = 1, 2.$$

Since $V(a, \theta, \gamma)$ is strictly pseudo-concave in a, then $V_a'(a, \theta, \gamma)$ is strictly pseudo-monotone in a by Proposition 2.5 in Hadjisavvas et al. (2005). Because $\varphi_1' > 0 > \varphi_2'$, it follows that $\mathcal{V}_k(m, s, \gamma), k = 1, 2$ is strictly pseudo-monotone in m. Specifically, let $z > m$ and $\mathcal{V}_1(m, s, \gamma) \leq 0$. Then $\varphi_1' > 0$ implies $\varphi_1(z) > \varphi_1(m)$ and $\mathcal{V}_1(m, s, \gamma) \leq 0$ if and only if (denoted $\Leftrightarrow$ hereafter) $V_a'(\varphi_1(m), \varphi_1(s), \gamma) \leq 0$. This inequality, $\varphi_1(z) > \varphi_1(m)$, and the strict pseudo-monotonicity of $V_a'(a, \theta, \gamma)$ imply $V_a'(\varphi_1(z), \varphi_1(s), \gamma) < 0$, which results in $\mathcal{V}_1(z, s, \gamma) < 0$. Similarly, let $z > m$ and $\mathcal{V}_2(m, s, \gamma) \leq 0$. Then $\varphi_2' < 0$ implies $\varphi_2(z) < \varphi_2(m)$ and $\mathcal{V}_2(m, s, \gamma) \leq 0 \Leftrightarrow V_a'(\varphi_2(m), \varphi_2(s), \gamma) \geq 0$. This inequality, $\varphi_2(z) < \varphi_2(m)$, and the strict pseudo-monotonicity of $V_a'(a, \theta, \gamma)$ imply $V_a'(\varphi_2(z), \varphi_2(s), \gamma) > 0$, which results in $\mathcal{V}_2(z, s, \gamma) < 0$.

Next, define the subsets

$$\mathbf{M}^{<>} = \left\{m, s, \gamma \in \mathbf{S}^2 \times \mathbf{T} | \mathcal{V}_1(m, s, \gamma) < 0 \text{ and } \mathcal{V}_2(m, s, \gamma) > 0\right\}, \text{ and}$$
$$\mathbf{M}^{><} = \left\{m, s, \gamma \in \mathbf{S}^2 \times \mathbf{T} | \mathcal{V}_1(m, s, \gamma) > 0 \text{ and } \mathcal{V}_2(m, s, \gamma) < 0\right\}.$$

Consider $\mathbf{M}^{<>}$. Because $V_a'(a, \theta, \gamma) < 0$ if and only if $a > y(\theta, \gamma)$ and $\varphi_1' > 0$, then

$$\mathcal{V}_1(m, s, \gamma) < 0 \Leftrightarrow V_a'(\varphi_1(m), \varphi_1(s), \gamma) < 0 \Leftrightarrow \varphi_1(m) > y(\varphi_1(s), \gamma). \tag{28}$$

For $x \in [\underline{s}, s)$, consider the function

$$\varphi_1(x) - y(\varphi_1(s), \gamma),$$

which is continuous in x. Because $y(\varphi_1(s), \gamma) \geq \underline{\theta} = \varphi_1(\underline{s})$, and $y(\varphi_1(s), \gamma) < \varphi_1(s)$, then

$$\varphi_1(\underline{s}) - y(\varphi_1(s), \gamma) \leq 0 < \varphi_1(s) - y(\varphi_1(s), \gamma).$$

Because φ_1 is strictly increasing, there is the unique $x_1(s, \gamma) \in [\underline{s}, s)$, such that

$$\varphi_1(x_1(s, \gamma)) = y(\varphi_1(s), \gamma).$$

Since $x_1(s, \gamma) < s$, then $\Delta_1(s, \gamma) = s - x_1(s, \gamma) > 0$. Because φ_1 is strictly increasing, then

[40] 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.

(28) is equivalent to

$$m > x_1(s,\gamma) = s - \Delta_1(s,\gamma).$$

Similarly, because $\varphi_2' < 0$, then

$$\mathcal{V}_2(m,s,\gamma) > 0 \Leftrightarrow V_a'(\varphi_2(m),\varphi_2(s),\gamma) < 0 \Leftrightarrow \varphi_2(m) > y(\varphi_2(s),\gamma). \tag{29}$$

For $x \in (s,\bar{s}]$, consider the function

$$\varphi_2(x) - y(\varphi_2(s),\gamma),$$

which is continuos in x. Because $y(\varphi_2(s),\gamma) \geq \underline{\theta} = \varphi_2(\bar{s})$ and $y(\varphi_2(s),\gamma) < \varphi_2(s)$, then

$$\varphi_2(\bar{s}) - y(\varphi_2(s),\gamma) \leq 0 < \varphi_2(s) - y(\varphi_2(s),\gamma).$$

Because φ_2 is strictly decreasing, there is the unique $x_2(s,\gamma) \in (s,\bar{s}]$, such that

$$\varphi_2(x_2(s,\gamma)) = y(\varphi_2(s),\gamma).$$

Since $x_2(s,\gamma) > s$, then $\Delta_2(s,\gamma) = x_2(s,\gamma) - s > 0$. Because φ_2 is strictly decreasing, then (29) is equivalent to

$$m < x_2(s,\gamma) = s + \Delta_2(s,\gamma).$$

As a result, $\mathbf{M}^{<>}$ can be expressed as

$$\mathbf{M}^{<>} = \left\{m,s,\gamma \in \mathbf{S}^2 \times \mathbf{T} | s - \Delta_1(s,\gamma) < m < s + \Delta_2(s,\gamma)\right\}.$$

Now, consider $\mathbf{M}^{><}$. By the same arguments as above, we have

$$\mathbf{M}^{><} = \left\{m,s,\gamma \in \mathbf{S}^2 \times \mathbf{T} | s + \Delta_2(s,\gamma) < m < s - \Delta_1(s,\gamma)\right\}.$$

Because $\Delta_k(s,\gamma) > 0, k = 1,2$, the inequalities $s + \Delta_2(s,\gamma) < m < s - \Delta_1(s,\gamma)$ cannot hold jointly, which implies $\mathbf{M}^{><} = \varnothing$.

Therefore, $\phi(m|s,\gamma,\rho^*)$ is pseudo-monotone in m for $q_k(s,\rho^*) \geq 0, k = 1,2$ 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 for all $(m,s,\gamma) \in \mathbf{M}^{<>}$. By using (14), we get

$$-\frac{\varphi_2'(s)}{\varphi_1'(s)} = \frac{h(\varphi_1(s))}{h(\varphi_2(s))},$$

and

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

where

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

For any $(m,s,\gamma) \in \mathbf{M}^{<>}$, (28) implies $V_a'(\varphi_1(m),\varphi_1(s),\gamma) < 0$ and, hence, $\nu(\varphi_1(m),\varphi_1(s),\gamma) < 0$. Also, (28) means that $a_1 = \varphi_1(m) > y(\varphi_1(s),\gamma) = y(\theta_1,\gamma)$. By Condition 2 and $\varphi_1' > 0$, it follows that $\nu(\varphi_1(m),\varphi_1(s),\gamma)$ is decreasing in m for all $(m,s,\gamma) \in \mathbf{M}^{<>}$.

Similarly, (29) implies $V_a'(\varphi_2(m),\varphi_2(s),\gamma) < 0$ and, hence, $\nu(\varphi_2(m),\varphi_2(s),\gamma) < 0$. Also, (29) means that $a_2 = \varphi_2(m) > y(\varphi_2(s),\gamma) = y(\theta_2,\gamma)$. By Condition 2 and $\varphi_2' > 0$, it follows that $\nu(\varphi_2(m),\varphi_2(s),\gamma)$ is increasing in m for all $(m,s,\gamma) \in \mathbf{M}^{<>}$. By combining these arguments, it follows that $-\frac{\mathcal{V}_2(m,s,\gamma)}{\mathcal{V}_1(m,s,\gamma)} = \frac{\nu(\varphi_2(m),\varphi_2(s),\gamma)}{\nu(\varphi_1(m),\varphi_1(s),\gamma)}$ is decreasing in m for all $(m,s,\gamma) \in \mathbf{M}^{<>}$. ■

Proof of Theorem 2 Consider functions $\varphi_k : \mathbf{S} \to \Theta, k = 1,2$, such that φ_1 is differentiable, $\varphi_1' > 0$, and φ_2 is given by (24). Because $\varphi_2 : \mathbf{S} \to \Theta$ is piecewise continuous and strictly monotone for $s < s_0$ and $s \geq s_0$, 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 $\xi_k = \varphi_k^{-1} : \Theta \to \mathbf{S}, k = 1,2$ 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 $\mathcal{M}^*$ with the allocation and payment rules given by (25) and (26), respectively.

First, it follows from (27) that the interim individual-rationality constraints (21) are binding for all (s,γ):

$$EV(s|s,\gamma,\rho^*) = \left\{ \begin{array}{c} \sum_{k=1}^{2} q_k(s,\rho^*) V(\varphi_k(s),\varphi_k(s),\gamma) \text{ if } s \geq s_0, \\ 0 \text{ if } s < s_0 \end{array} \right. \equiv 0 \qquad (30)$$

for all $(s,\gamma) \in \mathbf{S} \times \mathbf{T}$. Here, the first equality holds because (18) and $\varphi_k(s) \geq \theta_0$ for $s \geq s, k = 1,2$ imply that $V(\varphi_k(s),\varphi_k(s),\gamma) = 0$ for $(s,\gamma) \in \mathbf{S}_0 \times \mathbf{T}, k = 1,2$, where $\mathbf{S}_0 = [s_0, \bar{s}]$.

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

(i) $(s,m) \in \mathbf{S}_0^2$. By (27), the buyer's posterior payoff is

$$EV(m|s,\gamma,\rho^*) = \sum_{k=1}^{2} q_k(s,\rho^*) V(\varphi_k(m),\varphi_k(s),\gamma)$$

Note that states $\theta \in \Theta_0$ generate signals $s_k = \varphi_k^{-1}(\theta) \in \mathbf{S}_0, k = 1,2$ under ρ^*. Hence, the

agent's posterior belief induced by a signal $s \in \mathbf{S}_0$ is the binary distribution over states $\theta_k = \varphi_k(s) \in \Theta_0, k = 1, 2$. By using (7), the posterior probability of θ_k is

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

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 $\xi_k^0 = \xi_k : \Theta_0 \to \mathbf{S}_0$ is a signal function ξ_k with the domain restricted by Θ_0 and, thus, the image $\mathbf{S}_0$.

Now, consider the principal-agent model with the signal set $\mathbf{S}_0$, the prior density $f_0(\theta)$, and the private signal structure ρ^0. By combining the arguments above and comparing (27) with (9), it follows that the interim incentive-compatibility condition (20) 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 principal-agent model. Next, $\varphi_2(\theta)$ given by (24) satisfies the first-order condition (14) with the boundary condition $\varphi_1(s_0) = \theta_0$ in this model. Also, conditions (22) and 2 hold for Θ_0. Then, by applying Theorem 1 to the principal-agent 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 $Q_{\xi_k}^*(m) = 0, t_{\xi_k}^*(m) = 0, k = 1, 2$, and $V(0, 0, \gamma) = 0$ imply

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

where the last equality follows from (30).

(iii) $s < s_0, m \geq s_0$. Since $s < s_0$, then $\varphi_k(s) < \theta_0 = \varphi_1(s_0), k = 1, 2$. Also, $m \geq s_0$ implies $Q_{\xi_k}^*(m) = 1$ and $t_{\xi_k}^*(m) = \varphi_k(s), k = 1, 2$. This results in

$$\begin{aligned} EV(m|s, \gamma, \rho^*) &= \sum_{k=1}^{2} q_k(s, \rho^*) V(\varphi_k(m), \varphi_k(s), \gamma) < \sum_{k=1}^{2} q_k(s, \rho^*) V(\varphi_k(m), \varphi_1(s_0), \gamma) \\ &< \sum_{k=1}^{2} q_k(s, \rho^*) V(\varphi_k(m), \varphi_k(s_0), \gamma) \leq \sum_{k=1}^{2} q_k(s, \rho^*) V(\varphi_k(s_0), \varphi_k(s_0), \gamma) \\ &= EV(s_0|s_0, \gamma, \rho^*) = EV(s|s, \gamma, \rho^*) = 0, \end{aligned}$$

where the first inequality holds since $\varphi_k(s) < \varphi_1(s_0) = \theta_0, k = 1, 2$ and the strict monotonicity $V(t, \theta, \gamma)$ in θ imply $V(\varphi_k(m), \varphi_k(s), \gamma) < V(\varphi_k(m), \varphi_1(s_0), \gamma), k = 1, 2$, the second inequality holds due to $\varphi_2(s_0) = \bar{\theta} > \theta_0 = \varphi_1(s_0)$ and the strict monotonicity of $V(t, \theta, \gamma)$ in θ, and the last one holds since $m = s_0$ maximizes $EV(m|s_0, \gamma, \rho)$ over $m \geq s_0$. ■

References

Antsygina, A., Teteryatnikova, M.: Optimal information disclosure in contests with stochastic prize valuations. Econ. Theory 75, 743–780 (2023)

Bergemann, D., Pesendorfer, M.: Information structures in optimal auctions. J. Econ. Theory 137, 580–609 (2007)

Bergemann, D., Morris, S., Heumann, T.: Screening with persuasion. Working paper (2023)

Blume, A., Board, O., Kawamura, K.: Noisy talk. Theor. Econ. 2, 395–440 (2007)

Blume, A., Lai, E., Lim, M.: Eliciting private information with noise: the case of randomized response. Games Econom. Behav. 113, 356–380 (2019)

Chakraborty, A., Harbaugh, R.: Persuasion by cheap talk. Am. Econ. Rev. 100, 2361–2382 (2010)

Crawford, V., Sobel, J.: Strategic information transmission. Econometrica 50, 1431–1451 (1982)

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

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

Ederer, F., Holden, R., Meyer, M.: Gaming and strategic opacity in incentive provision. Rand J. Econ. 49, 819–854 (2018)

Esö, P., Szentes, B.: Optimal information disclosure in auctions. Rev. Econ. Stud. 74, 705–731 (2007)

Fu, H., Haghpanah, N., Hartline, J., Kleinberg, R.: The full surplus extraction from samples. J. Econ. Theory 193, 105230 (2021)

Goltsman, M., Hörner, J., Pavlov, G., Squintani, F.: Mediation, arbitration and negotiation. J. Econ. Theory 144, 1397–1420 (2009)

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

Heifetz, A., Neeman, Z.: On the generic (im)possibility of full surplus extraction in mechanism design. Econometrica 74, 213–233 (2006)

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

Ivanov, M., Sam, A.: Cheap talk with private signal structures. Games Econom. Behav. 132, 288–304 (2022)

Ivanov, M.: Optimal monotone signals in Bayesian persuasion mechanisms. Econ. Theory 72, 955–1000 (2021)

Ivanov, M.: Dynamic learning and strategic communication. Int. J. Game Theory 45, 627–653 (2016)

Ivanov, M.: Dynamic information revelation in cheap talk. The B.E. J. Theor. Econ. 15, 251–275 (2015)

Ivanov, M.: Communication via a strategic mediator. J. Econ. Theory 145, 869–884 (2010)

Johnson, J., Myatt, D.: On the simple economics of advertising, marketing, and product design. Am. Econ. Rev. 93, 756–784 (2006)

Krähmer, D.: Information disclosure and the full surplus extraction in mechanism design. J. Econ. Theory 187, 105020 (2020)

Krähmer, D.: Information design and strategic communication. Am. Econ. Rev: Insights 3, 51–66 (2021)

Krishna, V.: Auction Theory, 2nd edition. Elsevier Academic Press, Boston (2009)

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

Lewis, T. Sappington, D.: Supplying information to facilitate price discrimination. Int. Econ. Rev. 35, 309–327 (1994)

Li, H., Shi, X.: Discriminatory information disclosure. Am. Econ. Rev. 107, 3363–85 (2017)

Lipnowski, E., Ravid, D: Cheap talk with transparent motives. Econometrica 88, 1631–1660 (2020)

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

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

Myerson, R.: Optimal auction design. Math. Oper. Res. 6, 58–73 (1981)

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

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

Saak, A.: The optimal private information in single unit monopoly. Econ. Lett. 91, 267–272 (2006)

Shannon, C.: Communication theory of secrecy systems. Bell Syst. Tech. J. 28, 656–715 (1949)

Silva, F., Figueroa, N., Jara, R. Communication through biased intermediators. Working paper (2023)

Watson, J.: Information transmission when the informed party is confused. Games Econom. Behav. 12, 143–161 (1996)

Zhu, S.: Private disclosure with multiple agents. J. Econ. Theory 212, 105705 (2023)