

Optimal Auction Design under Costly Learning

Kemal Ozbek*

December 3, 2025

Abstract

We study optimal auction design in an independent private values environment where bidders can endogenously—but at a cost—improve information about their own valuations. The optimal mechanism is two-stage: at stage-1 bidders register an information acquisition plan and pay a transfer; at stage-2 they bid, and allocation and payments are determined. We show that the revenue-optimal stage-2 rule is the Vickrey–Clarke–Groves (VCG) mechanism, while stage-1 transfers implement the optimal screening of types and absorb information rents consistent with incentive compatibility and participation. By committing to VCG ex post, the pre-auction information game becomes a potential game, so equilibrium information choices maximize expected welfare; the stage-1 fee schedule then transfers an optimal amount of payoff without conditioning on unverifiable cost scales. The design is robust to asymmetric primitives and accommodates a wide range of information technologies, providing a simple implementation that unifies efficiency and optimal revenue in environments with endogenous information acquisition.

1 Introduction

This paper studies the design of auctions for selling a single indivisible good under costly learning to a finite set of risk-neutral bidders with independent private values (IPV). Two benchmark objectives organize much of the modern theory. A welfare-maximizing mechanism seeks to allocate the good to the participant who values it most whenever trade is feasible, thereby implementing an efficient assignment (Vickrey [1961]; Clarke [1971]; Groves [1973]). A revenue-maximizing mechanism instead chooses

*Department of Economics, University of Southampton, University Road, Southampton, S017 1BJ, United Kingdom. Email: mkemalozbek@gmail.com

rules that maximize the seller’s expected gains subject to incentive compatibility and participation constraints (Myerson [1981]). Although these problems are posed in the same informational environment, they typically prescribe different allocation rules and, crucially, they need not be implemented by the same auction format—a distinction emphasized throughout the auction design literature (Myerson [1981]; Riley and Samuelson [1981]).

A central theme of the present paper is that another, often under-appreciated source of divergence between revenue and welfare may arise when information acquisition is endogenous. In many applications, bidders do not begin with perfectly precise knowledge of their own values; rather, they can invest in information—due diligence, inspections, expert reports—before bidding. The auction format that the seller commits to use after any learning has taken place shapes the marginal value of information for each participant. Formats that make a bidder’s allocation probability and payment more sensitive to her realized private information tend to induce more information acquisition; formats that dampen this sensitivity tend to induce less (Persico [2000]). Because equilibrium information choices shift the distribution of bids and the composition of winners, they influence both the revenue that can be extracted and the welfare that can be realized (Bergemann and Valimaki [2002]; Cremer, Spiegel, and Zheng [2009]).

In our setting, bidders can flexibly—but at a cost—improve what they know about their own valuations before bidding. We even allow them, in principle, to shift the mean of their belief at an additional cost; although this option is not exercised in equilibrium, keeping it available delivers a richer and more robust modeling environment. We study two-stage mechanisms: at stage-1, each bidder pays a registration fee and commits to an experiment (a signal technology); at stage-2, bidders submit bids in the auction. We work with a direct-revelation formulation in which a bidder reports at stage-1 her primitive type (a cost-scale parameter and a mean-value parameter) to determine the experiment she may run and the fee she pays, and at stage-2 reports her realized valuation to determine allocation and payment. Types are ex-ante random with a commonly known distribution, and equilibrium in experiment choices is Bayesian Nash. We first assume that registered experiments are verifiable by the seller and later relax this assumption, showing that the main results survive under weaker verification.

Our main result shows that, unlike in static models without learning, revenue maximization and welfare maximization coincide in the optimal two-stage design we characterize. The stage-2 allocation rule is VCG—delivering ex-post efficiency and truthful bidding—while stage-1 lets bidders choose any experiment compatible with their reported types after paying the corresponding fees. Stage-1 incentive-compatibility and individual-rationality

pin down a fee schedule that is independent of unverifiable cost-scale parameters yet robustly dependent on verifiable mean-value parameters via the registered experiments. The rationale for leaving experiment choice unconstrained is that, once stage-2 is VCG, the pre-auction information game is an exact potential game (Monderer and Shapley [1996]); therefore, any maximizer of the potential—here, expected welfare—coincides with a Bayesian Nash equilibrium of the bidders’ experiment-choice game.

The literature identifies two complementary channels on the divergence of revenue and welfare maximization objectives. In the first, bidders choose how much to learn before bidding. Here, the format’s payoff sensitivity to private signals determines the equilibrium investment in information; when expected payoffs respond sharply to incremental precision—because the probability of winning or the effective price is highly sensitive to the realized signal—participants optimally acquire more information (Persico [2000]). This matters for welfare because precision improves allocative matching; it matters for revenue because sharper selection can raise the expected virtual surplus that the seller can extract. In the second channel, the seller chooses what bidders learn by designing the information structure itself—deciding how informative pre-auction disclosures or signals will be—jointly with the auction rule. This transforms information release into a policy lever for revenue and can, in general, steer the allocation away from efficiency (Bergemann and Pesendorfer [2007]).¹

These forces qualify standard format comparisons. In the symmetric IPV benchmark without endogenous learning, a first-price auction and a second-price auction with a common reserve implement the same allocation rule and thus are revenue-equivalent. With endogenous information acquisition, however, the allocation rule becomes format-dependent, precisely because the stage-2 rule changes bidders’ learning incentives. Two formats that were revenue-equivalent absent learning can cease to be equivalent once learning is endogenous. From a welfare perspective, formats and transfer rules that internalize the social value of information—so that each bidder’s private incentive to learn matches the societal benefit of more precise allocation—are desirable; in private-value environments, VCG-type mechanisms can implement both efficient learning and efficient allocation under natural conditions (Bergemann and Valimaki [2002]). From a revenue perspective, the seller’s optimal policy typically co-determines reserves, disclosure, and, when feasible, multi-stage procedures that govern when and how information is produced; optimal reserve policies in this setting generally differ from the no-learning Myerson

¹Relatedly, classic analyses of interdependent or affiliated signals show how information release and format interact to shape both efficiency and revenue, even beyond IPV (Milgrom and Weber [1982]).

benchmark because they must account for the equilibrium response of information acquisition (Bergemann and Pesendorfer [2007]; Shi [2012]; Gershkov, Moldavanu, Strack, and Zhang [2021]).

In sum, even within the IPV paradigm, welfare- and revenue-maximizing designs can diverge for structural reasons—exclusion and asymmetry in the Myerson framework—and may diverge for strategic reasons once information acquisition is considered. Viewing the stage-2 auction format as a commitment to stage-1 learning incentives makes clear why the designer’s objective, the induced information environment, and the implementable allocation rule are tightly intertwined. Our results nevertheless show that, when registered experiments are verifiable, a single two-stage mechanism can align these forces and simultaneously achieve both welfare and revenue objectives. These findings have practical and theoretical implications. Practically, when pre-auction due diligence matters (e.g., spectrum, mineral rights, complex procurement), formats should be judged not only by static allocation rules but by the learning incentives they create; committing to VCG at stage-2, paired with a transparent, experiment-based registration fee, can deliver both high revenue and efficiency. Theoretically, we refine revenue equivalence: formats that are revenue-equivalent without learning need not remain so once learning is endogenous, but a suitable two-stage commitment can restore alignment by making the induced allocation rule format-invariant through efficient information acquisition.

The paper is organized as follows. Section 2 presents the framework, notation, and the incentive and participation constraints, and states the objective functions. Section 3 develops the main analysis, characterizing and comparing the revenue- and welfare-maximizing mechanisms. Section 4 extends the model to allow (costly) auditing of registered experiments and cost parameters and examines the robustness of the results. Section 5 provides a discussion of the related literature and highlights our contributions to the research area of optimal auction design under costly learning. Section 6 concludes.

2 Framework

We adapt the framework considered in Myerson [1981] to study the optimal auction design when buyers can flexibly learn.

Environment: There is one seller who has a single object to sell. Let $N = \{i_1, \dots, i_n\}$ denote the set of bidders with $n \geq 2$. We assume that the seller and the bidders are risk

neutral. Bidders have types $\theta_i = (r_i, s_i)$ drawn from the space $\Theta_i = [0, 1]^2$ equipped with the Borel sigma-algebra. We assume that bidder types θ_i are drawn i.i.d. according to the distribution $F_{rs} = F_r \times F_s$ where F_r and F_s are measurable c.d.f.'s with continuous density functions f_r and f_s over the space $[0, 1]$, respectively. Let $\Theta = \Theta_1 \times \dots \times \Theta_n$ denote the set of type profiles including every bidder and let Θ_{-i} denote the set of type profiles $\Theta_1 \times \dots \times \Theta_{i-1} \times \Theta_{i+1} \times \dots \times \Theta_n$ excluding bidder i . Let $\theta = (r, s)$ denote a generic profile of types $(r_1, s_1; \dots; r_n, s_n) \in \Theta$ and similarly, let $\theta_{-i} = (r_{-i}, s_{-i})$ denote the vector of realized types $(r_1, s_1; \dots; r_{i-1}, s_{i-1}; r_{i+1}, s_{i+1}; \dots; r_n, s_n) \in \Theta_{-i}$ excluding bidder i . We sometimes use the notation θ_{i1} to indicate r_i and θ_{i2} to indicate s_i when $\theta_i = (r_i, s_i)$. Let $F_{rs}^n(\hat{r}, \hat{s}) = (F_r \times F_s)^n(\hat{r}, \hat{s})$ denote the joint distribution $\prod_{i \in N} F_r(\hat{r}_i) \times F_s(\hat{s}_i)$ and let $F_{rs}^{n-1}(\hat{r}_{-i}, \hat{s}_{-i}) = (F_r \times F_s)^{n-1}(\hat{r}_{-i}, \hat{s}_{-i})$ denote the joint distribution $\prod_{j \neq i} F_r(\hat{r}_j) \times F_s(\hat{s}_j)$ excluding bidder $i \in N$.

For each $\alpha \in [0, 1]$, let $z_\alpha = (a + \alpha(b - a))$ denote the mean value over the interval $[a, b]$ where $a > 0$ is the minimum and $b > a$ is the maximum level each individual is willing to pay for the object. Let $D = \Delta([a, b])$ denote the set of probability density functions over $[a, b]$ and assume that D is equipped with the weak topology. We call each distribution $f \in D$ an ‘‘experiment’’. For each $\alpha \in [0, 1]$, let $D(\alpha) = \{f \in \Delta([a, b]) : E_f(z) = \int_{[a, b]} z \cdot f(z) \cdot dz = z_\alpha\}$ be the set of experiments with center (i.e., mean value) $\mu(f)$ equal to z_α . Note that each $D(\alpha)$ is closed in D and we have $D = \cup_{\alpha \in [0, 1]} D(\alpha)$. Let $T = [a, b]^n$ and for each $i \in N$, let $T_i = [a, b]$ and $T_{-i} = [a, b]^{n-1}$.

For each type $\theta_i = (r_i, s_i) \in \Theta_i$, let $c_{\theta_i} : D \rightarrow \mathbb{R}_+$ denote an information processing cost function. We assume that each c_{θ_i} satisfies the following properties: (i) $c_{\theta_i}(\cdot)$ is increasing in θ_{i1} , (ii) $c_{\theta_i}(\delta_z) = 0$ if $z = z_{\theta_{i2}}$, (iii) $c_{\theta_i}(\cdot)$ is increasing in convex-order, (iv) $c_{\theta_i}(\cdot)$ is convex, and (v) $c_{\theta_i}(\cdot)$ is lower semi-continuous (l.s.c.). A feasible example of the abstract cost function c_{θ_i} can be given as $\hat{c}_{\theta_i}(f) = r_i \int k(|z_i - s_i|) f(z_i) dz_i$ for all $f \in D$ for some (strictly) increasing convex function $k : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ with $k(0) = 0$. As such, while r_i can be interpreted as a cost-scale, s_i can be interpreted as a mean-value parameter.²

Mechanism: The seller uses a two-stage direct-revelation mechanism with a reservation price to sell the good. For simplicity, assume that the object yields no value to the seller,

²The idea is that each experiment $f \in D$ determines the likelihood $f(z)$ of each class of posterior distributions with mean $z \in [a, b]$ where z_{s_i} denotes the mean of the initial belief of a bidder with type $\theta_i = (r_i, s_i)$. Bidders are allowed to consider experiments $f_i \in D$ with centers $\mu(f_i)$ such that $\mu(f_i) \neq z_{s_i} = z_{\theta_{i2}}$. Moreover, depending on the particular cost function, the cost of a degenerate experiment (i.e., δ_z for some $z \in [a, b]$) may or may not be zero. For instance, while $\hat{c}_{\theta_i}(f) \neq 0$ for any f with $\mu(f) \neq z_{\theta_{i2}}$, for the cost function $\hat{c}_{\theta_i}(f) = r_i \int k(|z_i - \mu(f)|) f(z_i) dz_i$, we have $\hat{c}_{\theta_i}(\delta_z) = 0$ for any $z \in [a, b]$.

and so the seller optimally sets the reservation price equal to 0. At stage-1, the seller offers to each bidder i a menu of contracts (σ_i, τ_i) such that $\sigma_i : \Theta_i \rightarrow D$ assigns an experiment $\sigma_i(\theta_i) \in D(\theta_{i2})$ and $\tau_i : \Theta_i \rightarrow \mathbb{R}$ charges a fee to each type $\theta_i = (r_i, s_i)$. Each bidder pays the fee $\tau_i(\theta_i)$ to conduct the experiment $\sigma_i(\theta_i)$ after revealing her type θ_i to the seller. The seller is assumed to be able to confirm that the experiment $\sigma_i(\theta_i)$ is conducted once a type $\theta_i \in \Theta_i$ is registered. Later we relax this assumption.

At stage-2, each bidder observes a value realization $t_i \in [a, b]$ within the support of her experiment $\sigma_i(\theta_i)$ and submits a bid $t'_i \in [a, b]$ where t'_i is not necessarily equal to t_i . The seller then allocates the object to the buyers according to the vector of probabilities $p(t', \sigma(\theta)) = (p_1(t', \sigma(\theta)), \dots, p_n(t', \sigma(\theta))) \geq 0$ such that $\sum_{i \in N} p_i(t', \sigma(\theta)) = 1$ together with $[x_1(t', \sigma(\theta)), \dots, x_n(t', \sigma(\theta))] \in \mathbb{R}^n$ as the vector of payments $x(t', \sigma(\theta))$ where $t' = (t'_1, \dots, t'_n) \in [a, b]^n$ is the vector of bids submitted by the buyers at stage-2 and $\sigma(\theta) = (\sigma_1(\theta_1), \dots, \sigma_n(\theta_n))$ is the vector of experiments registered by the buyers at stage-1. Let $(\sigma_{-i}, \tau_{-i}) = (\sigma_1, \tau_1; \dots, \sigma_{i-1}, \tau_{i-1}; \sigma_{i+1}, \tau_{i+1}; \dots; \sigma_n, \tau_n)$ denote the vector of menus of contracts given to the bidders excluding buyer i . For any given profile of types θ_{-i} , let $\sigma_{-i}(\theta_{-i}) = (\sigma_1(\theta_1), \dots, \sigma_{i-1}(\theta_{i-1}), \sigma_{i+1}(\theta_{i+1}), \dots, \sigma_n(\theta_n))$ denote the vector of experiments excluding bidder i . Let M denote the set of all two-stage mechanisms and let $m = (p, x, \sigma, \tau)$ denote a generic element of it.

Payoffs: For any vector of realized types other than bidder i , let $\sigma_{-i}[\theta_{-i}]$ denote the joint density function $\sigma_1(\theta_1) \times \dots \times \sigma_{i-1}(\theta_{i-1}) \times \sigma_{i+1}(\theta_{i+1}) \times \dots \times \sigma_n(\theta_n)$ and let its corresponding c.d.f. be denoted by $F_{\sigma_{-i}[\theta_{-i}]}$. Given a two-stage mechanism $m = (p, x, \sigma, \tau)$ in M , let

$$\begin{aligned} u_i^m(t'_i, t_i | f_i, \sigma_{-i}) &= E_{\theta_{-i}} [E_{\sigma_{-i}[\theta_{-i}]} [t_i p_i(t'_i, t_{-i}; f_i, \sigma_{-i}(\theta_{-i})) - x_i(t'_i, t_{-i}; f_i, \sigma_{-i}(\theta_{-i}))]] \\ &= \int_{\Theta_{-i}} \int_{T_{-i}} [t_i p_i(t'_i, t_{-i}; f_i, \sigma_{-i}(\theta_{-i})) \\ &\quad - x_i(t'_i, t_{-i}; f_i, \sigma_{-i}(\theta_{-i}))] dF_{\sigma_{-i}[\theta_{-i}]}(t_{-i}) d(F_r \times F_s)^{n-1}(\theta_{-i}) \end{aligned}$$

be the stage-2 expected payoff bidder i receives when she submits a bid $t'_i \in [a, b]$ while she observes a value $t_i \in [a, b]$ given that she registered the experiment f_i while all other bidders truthfully registered an experiment from their offered menus.

Whenever bidder i truthfully submits her bid, that is when $t'_i = t_i$, we simply write $u_i^m(t_i | f_i, \sigma_{-i})$. Let $u_i^m(f_i | \sigma_{-i}) = E_{f_i} [u_i^m(t_i | f_i, \sigma_{-i})]$ be the interim expected payoff bidder i obtains when she bids truthfully according to the distribution f_i . Given

$m = (p, x; \sigma, \tau) \in M$, let $v_i^m(\theta'_i|\theta_i) = u_i^m(\sigma_i(\theta'_i)|\sigma_{-i}) - c_{\theta_i}(\sigma_i(\theta'_i)) - \tau_i(\theta'_i)$ denote the net expected payoff of bidder i when other bidders truthfully follow the distributions $\sigma_{-i}(\theta_{-i})$ in their offered menus depending on their type realizations $\theta_{-i} = (r_{-i}, s_{-i})$ while bidder i with true type $\theta_i \in \Theta_i$ reveals $\theta'_i = (r'_i, s'_i)$ to register the experiment $\sigma_i(\theta'_i) \in D(\theta'_{i2})$ with the associated information processing cost $c_{\theta_i}(\sigma_i(\theta'_i))$ and transfer fee $\tau(\theta'_i)$. Whenever bidder i truthfully submits her type $\theta_i = (r_i, s_i)$, we write $v_i^m(\theta_i)$. Let

$$U_i(m) = E_{\theta_i}[v_i^m(\theta_i)] = \int_{\Theta_i} v_i^m(\theta_i) d(F_r \times F_s)(\theta_i)$$

denote the ex-ante net expected payoff bidder i receives when she plans to truthfully reveal her type at stage-1 and reveal her true value at stage-2 while all other bidders also truthfully submit their types and bids given the profile of assignment rules σ_{-i} .

Constraints: The seller does not observe the true type $\theta_i = (r_i, s_i)$ of a bidder nor the realized value t_i . The seller, therefore, must provide the right incentives to each buyer to participate in the mechanism, submit truthfully their realized types and values, and conduct their intended experiments.

Given a two-stage auction mechanism $m = (p, x, \sigma, \tau)$, the constraints listed below constitute the two-stage individual rationality and incentive compatibility constraints for each buyer i for each $\theta_i \in \Theta_i$:

1. $u_i^m(t_i|\sigma_i(\theta_i), \sigma_{-i}) \geq 0$ for each $t_i \in [a, b]$ in the support of $\sigma_i(\theta_i)$ (*stage-2 IR*);
2. $u_i^m(t_i|\sigma_i(\theta_i), \sigma_{-i}) \geq u_i^m(t'_i, t_i|\sigma_i(\theta_i), \sigma_{-i})$ for each $t_i, t'_i \in [a, b]$ in the support of $\sigma_i(\theta_i)$ (*stage-2 IC*);
3. $v_i^m(\theta_i) \geq 0$ (*stage-1 IR*);
4. $v_i^m(\theta_i) \geq v_i^m(\theta'_i|\theta_i)$ for each $\theta'_i \in \Theta_i$ (*stage-1 IC*).

The first two constraints are the usual individual rationality and incentive compatibility constraints at the allocation stage. The third constraint is the pre-allocation individual rationality constraint implying that participation in the mechanism is voluntary. The fourth constraint is the pre-allocation incentive compatibility constraint incentivizing the bidders to tell the truth about their types (i.e., cost-scale and mean-value parameters).

We say that a two-stage mechanism $m = (p, x, \sigma, \tau)$ is feasible if it satisfies the above four constraints. Let \bar{M} denote the set of feasible two-stage mechanisms. There are two separate objectives that we focus on: revenue maximization and welfare maximization.

Revenue maximization: Given $m = (p, x, \sigma, \tau) \in M$, let

$$\begin{aligned} R(m) &= E_{\theta} [E_{\sigma(\theta)} \sum_{i \in N} [x_i(t, \sigma(\theta)) + \tau_i(\theta_i)]] \\ &= \int_{\Theta} [\int_T [\sum_{i \in N} [x_i(t, \sigma(\theta)) + \tau_i(\theta_i)]] dF_{\sigma(\theta)}(t)] d(F_r \times F_s)^n(\theta) \end{aligned}$$

denote the total expected payment received by the seller from the bidders if mechanism (p, x) is run at stage-2 and profiles of contracts (σ, τ) are used at stage-1, while the bidders truthfully reveal their types. Let $R(m^*) = \max_{m \in \bar{M}} R(m)$ denote the maximum level of expected revenue the seller can generate.

Welfare maximization: Let $W(m) = [\sum_{i \in N} U_i(m)] + R(m)$ denote the total expected welfare that can be achieved when mechanism $m = (p, x, \sigma, \tau) \in M$ is used. Let $W(m^*) = \max_{m \in \bar{M}} W(m)$ denote the maximum level of expected welfare that can be generated .

If a two-stage mechanism $m = (p, x, \sigma, \tau)$ maximizes the total surplus $\sum_{i \in N} t_i p_i(t, \sigma(\theta))$ for each profile of types $\theta \in \Theta$ and for each vector of values t within the support of $\sigma(\theta)$, we call it *ex-post efficient*; when the mechanism (p, x, σ, τ) maximizes the expected (total) net surplus

$$E_{\theta} [\sum_{i \in N} t_i p_i(t, \sigma(\theta))] - E_{\theta} [\sum_{i \in N} c_{\theta_i}(\sigma_i(\theta_i))],$$

we call it *ex-ante efficient*. In other words, a mechanism m is ex-ante efficient if it maximizes the total expected welfare (see also below).

3 Analysis

In this section, we study the optimal mechanisms that either maximize the expected revenue or the expected welfare. We start our analysis by first simplifying the objective functions within the class of feasible two-stage mechanisms. Suppose that $m = (p, x, \sigma, \tau)$ is a given feasible mechanism in \bar{M} . By definition, for each $\theta_i = (r_i, s_i)$ we have $v_i^m(\theta_i) = u_i^m(\sigma_i(\theta_i) | \sigma_{-i}) - c_{\theta_i}(\sigma_i(\theta_i)) - \tau_i(\theta_i)$ where

$$\begin{aligned} u_i^m(\sigma_i(\theta_i) | \sigma_{-i}) &= E_{\sigma_i(\theta_i)} [E_{\theta_{-i}} [E_{\sigma_{-i}[\theta_{-i}]} [t_i p_i(t'_i, t_{-i}; \sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})) \\ &\quad - x_i(t'_i, t_{-i}; \sigma_i(\theta_i), \sigma_{-i}(\theta_{-i}))]]]. \end{aligned}$$

Taking the expectation over Θ_i , we obtain

$$E_{\theta_i}[v_i^m(\theta_i)] = E_{\theta_i}[u_i^m(\sigma_i(\theta_i)|\sigma_{-i})] - E_{\theta_i}[c_{\theta_i}(\sigma_i(\theta_i))] - E_{\theta_i}[\tau_i(\theta_i)].$$

Summing over bidders, we derive that

$$E_{\theta}[\sum_{i \in N} v_i^m(\theta_i)] = E_{\theta}[\sum_{i \in N} t_i p_i(t, \sigma(\theta)) - x_i(t, \sigma(\theta))] - E_{\theta}[\sum_{i \in N} c_{\theta_i}(\sigma_i(\theta_i))] - E_{\theta}[\sum_{i \in N} \tau_i(\theta_i)].$$

Rearranging the terms, we first obtain that the expected revenue of the seller can be given as the difference between the expected net surplus (i.e., surplus net of cost of information) and the expected information rent:

$$\begin{aligned} R(m) &= E_{\theta}[\sum_{i \in N} (x_i(t, \sigma(\theta)) + \tau_i(\theta_i))] \\ &= E_{\theta}[\underbrace{\sum_{i \in N} t_i p_i(t, \sigma(\theta)) - \sum_{i \in N} c_{\theta_i}(\sigma_i(\theta_i))}_{\text{expected net surplus}}] - \underbrace{E_{\theta}[\sum_{i \in N} v_i^m(\theta_i)]}_{\text{expected information rent}}. \end{aligned}$$

We can further arrange the above equation to express the regulator's welfare function as:

$$\begin{aligned} W(m) &= R(m) + E_{\theta}[\sum_{i \in N} v_i^m(\theta_i)] \\ &= E_{\theta}[\underbrace{\sum_{i \in N} t_i p_i(t, \sigma(\theta)) - \sum_{i \in N} c_{\theta_i}(\sigma_i(\theta_i))}_{\text{expected welfare}}]. \end{aligned}$$

The two expressions above clearly manifest the tension between revenue and welfare maximization. As such, the seller's optimal mechanism may distort the level of welfare that can be maximally obtained, and similarly the efficient mechanism may cause the seller to leave additional rent to the bidders lowering the revenue. Naturally we come across similar efficiency-optimality conflict among many incentive compatible mechanisms. For instance, in the setting of Myerson [1981] where the distributions of values are fixed, the VCG mechanism (p^e, x^e) delivers an efficient outcome, whereas the Myerson mechanism (p^r, x^r) yields a revenue maximizing outcome. Since in our setting experiments are endogenously determined, a conflict between revenue maximization and welfare maximization, however, do not occur. In other words, the seller's revenue maximizing mechanism will also be delivering the maximum possible welfare. This means that the seller's optimal mechanism would not only be maximizing the expected net surplus, but also would be minimizing the expected information rent among the feasible two-stage mechanisms.

Efficient stage-2 rule: In the following, we will focus on two-stage mechanisms where at stage-2 the VCG mechanism, denoted by (p^e, x^e) , is used. The VCG mechanism awards the good to the highest reported value t_i (with a tie-break rule) and the winner pays the highest reported value by the other bidders. More formally, let $w(t) = \{i \in N : t_i \geq \max_{j \in N} t_j\}$ be the set of bidders who bid the highest value. Let

$$p_i^e(t) = \begin{cases} 1/|w(t)| & \text{if } i \in w(t), \\ 0 & \text{if } i \notin w(t), \end{cases}$$

be the winning probability and let

$$x_i^e(t) = \begin{cases} \max_{j \neq i} t_j & \text{if } i \text{ wins,} \\ 0 & \text{otherwise,} \end{cases}$$

be the payment for all $i \in N$ and $t \in T$. That is, a bidder wins only when she is among the highest bidders, and pays an amount only when she wins and the payment is equal to the highest bid among the opponents.

It is well-known that for any fixed profile of distributions $f = (f_1, \dots, f_n)$, the VCG mechanism (p^e, x^e) maximizes the total surplus among the incentive compatible (IC) and individually rational (IR) allocation-payment rules (p, x) ; that is, for any stage-2 rule (p, x) satisfying the IC and IR constraints we have $\sum_{i \in N} t_i p_i^e(t, f) \geq \sum_{i \in N} t_i p_i(t, f)$. As such, for any given stage-1 contract σ , we have $E_\theta[\sum_{i \in N} t_i p_i^e(t, \sigma(\theta))] \geq E_\theta[\sum_{i \in N} t_i p_i(t, \sigma(\theta))]$ implying that

$$E_\theta[\sum_{i \in N} t_i p_i^e(t, \sigma(\theta)) - c_{\theta_i}(\sigma_i(\theta_i))] \geq E_\theta[\sum_{i \in N} t_i p_i(t, \sigma(\theta)) - c_{\theta_i}(\sigma_i(\theta_i))].$$

But this means that a two-stage mechanism which uses the efficient rule (p^e, x^e) at stage-2 pointwise dominates (in terms of the expected net surplus) a two-stage mechanism which uses another rule (p, x) at stage-2. In fact, the following result shows that for any given feasible two-stage mechanism $m = (p, x, \sigma, \tau) \in \bar{M}$, we can replace the stage-2 rule (p, x) with the efficient rule (p^e, x^e) while we use the same stage-1 assignments $\sigma' = \sigma$ but adjust the fees τ to some τ' such that the new mechanism is feasible and it weakly improves the revenue; that is, $(p^e, x^e, \sigma', \tau') \in \bar{M}$ and $R(p^e, x^e, \sigma', \tau') \geq R(p, x, \sigma, \tau)$.

Let \bar{M}^e denote the subset of the set of feasible mechanisms \bar{M} such that the stage-2 rule is fixed as the efficient second-price auction (i.e., the VCG mechanism).

Lemma 1. For any $m = (p, x, \sigma, \tau) \in \bar{M}$, there exists $m' = (p', x', \sigma', \tau') \in \bar{M}^e$ such that $R(m') \geq R(m)$.

Proof. Let $m = (p, x, \sigma, \tau) \in \bar{M}$ be a feasible mechanism. Let $m' = (p', x', \sigma', \tau') \in M$ be an alternative mechanism such that $(p', x') = (p^e, x^e)$ and $\sigma' = \sigma$, and for any bidder i , $\tau'_i(\theta_i) = \tau_i(\theta_i) + u_i^{m'}(\sigma_i(\theta_i)|\sigma_{-i}) - u_i^m(\sigma_i(\theta_i)|\sigma_{-i})$. We want to show that (i) $(p', x', \sigma', \tau') \in \bar{M}$ and (ii) $R(p', x', \sigma', \tau') \geq R(p, x, \sigma, \tau)$. As discussed before, constraints 1 and 2 hold when the VCG mechanism (p^e, x^e) is used at stage-2. Let $v_i^{m'}(\theta_i)$ be the net expected payoff of bidder i when she truthfully reveals her type $\theta_i = (r_i, s_i)$ under the mechanism m' . By definition, we have

$$\begin{aligned} v_i^{m'}(\theta_i) &= u_i^{m'}(\sigma'_i(\theta_i)|\sigma'_{-i}) - c_{\theta_i}(\sigma'_i(\theta_i)) - \tau'_i(\theta_i) \\ &= u_i^m(\sigma_i(\theta_i)|\sigma_{-i}) - c_{\theta_i}(\sigma_i(\theta_i)) - \tau_i(\theta_i) \\ &= v_i^m(\theta_i) \geq 0 \end{aligned}$$

showing that constraint 3 holds under m' . Now let $v_i^{m'}(\theta'_i|\theta_i)$ denote the net expected payoff of bidder i when she reveals her type as $\theta'_i = (r'_i, s'_i)$ instead of telling her true type $\theta_i = (r_i, s_i)$. By definition, we have

$$\begin{aligned} v_i^{m'}(\theta'_i|\theta_i) &= u_i^{m'}(\sigma'_i(\theta'_i)|\sigma'_{-i}) - c_{\theta_i}(\sigma'_i(\theta'_i)) - \tau'_i(\theta'_i) \\ &= u_i^m(\sigma_i(\theta'_i)|\sigma_{-i}) - c_{\theta_i}(\sigma_i(\theta'_i)) - \tau_i(\theta'_i) \\ &= v_i^m(\theta'_i|\theta_i) \leq v_i^m(\theta_i) = v_i^{m'}(\theta_i) \end{aligned}$$

showing that constraint 4 holds under m' . Thus, $m' \in \bar{M}$. Moreover we have,

$$\begin{aligned} R(m) &= E_\theta[\sum_{i \in N} t_i p_i(t, \sigma(\theta))] - E_\theta[\sum_{i \in N} c_{\theta_i}(\sigma_i(\theta_i))] - E_\theta[\sum_{i \in N} v_i^m(\theta_i)] \\ &= E_\theta[\sum_{i \in N} t_i p_i(t, \sigma'(\theta))] - E_\theta[\sum_{i \in N} c_{\theta_i}(\sigma'_i(\theta_i))] - E_\theta[\sum_{i \in N} v_i^{m'}(\theta_i)] \\ &\leq E_\theta[\sum_{i \in N} t_i p_i^e(t, \sigma'(\theta))] - E_\theta[\sum_{i \in N} c_{\theta_i}(\sigma'_i(\theta_i))] - E_\theta[\sum_{i \in N} v_i^{m'}(\theta_i)] \\ &= R(m'). \end{aligned}$$

where the inequality follows from the fact that the VCG rule (p^e, x^e) maximizes the surplus for any fixed profile of distributions. \square

Lemma 1 shows that in searching for an optimal mechanism maximizing the expected

revenue, one can focus on the set of mechanisms which employ the efficient rule at stage-2. In fact, the proof of this result shows that from the given mechanism m an alternative mechanism m' can be constructed such that each bidder's information rent will remain the same; that is, $v_i^{m'}(\theta_i) = v_i^m(\theta_i)$ for all $i \in N$ and $\theta_i \in \Theta_i$. As such, we also have

$$\begin{aligned} W(m) &= R(m) + E_\theta[\sum_{i \in N} v_i^m(\theta_i)] \\ &\leq R(m') + E_\theta[\sum_{i \in N} v_i^{m'}(\theta_i)] = W(m') \end{aligned}$$

showing that to maximize the expected welfare, one can focus on the efficient mechanisms in \bar{M}^e . Thus we have the following corollary.

Corollary 1. *For any $m = (p, x, \sigma, \tau) \in \bar{M}$, there exists $m' = (p', x', \sigma', \tau') \in \bar{M}^e$ such that $W(m') \geq W(m)$.*

Corollary 1 implies that shows any feasible ex-ante efficient two-stage mechanism must use the VCG rule at stage-1, and, thus, by Lemma 1 any revenue maximizing two-stage mechanism must be ex-ante efficient. There are related studies in the literature which show, similar to the above corollary, that the VCG rule maximizes the welfare whenever bidder participation or preference determination is costly. For instance, in an IPV auction setting where sending messages to the seller (i.e., participation in the auction) is costly, Stegeman [1996] shows that the second-price auction has an equilibrium that is ex-ante efficient. In another study with an IPV auction setting where bidders can acquire (covert) information before participating in the auction, Bergemann and Valimaki [2002] show that the VCG rule guarantees both ex-ante and ex-post efficiency. When we have an efficient mechanism $(p, x, \sigma, \tau) \in \bar{M}^e$ and we want to focus on the assignment rule σ , we also write $W^e(\sigma)$ to denote the expected welfare under the mechanism (p, x, σ, τ) before fees are applied.

Potential game: Given that the optimal stage-2 rule should be the efficient rule (p^e, x^e) , we are now interested in understanding the optimal stage-1 assignments σ and schedules τ which determine an experiment $\sigma_i(\theta_i)$ and a fee $\tau_i(\theta_i)$ to each declared type $\theta_i \in \Theta_i$ by each bidder i . Notice that stage-1 fees do not depend on the experiments registered by the bidders. As such, the only remaining constraint for bidder i 's choice of experiment $\sigma_i(\theta_i)$ is that its center $\mu(\sigma_i(\theta_i))$ must meet the declared mean value $z_{\theta_{i2}}$; that is, $\sigma_i(\theta_i) \in D(\theta_{i2})$. Beyond this restriction the bidders are (and should be) free to choose the best experiment they can obtain based on their types. The reason for this

fact is that we have a Bayesian exact potential game where each bidder's switch of an experiment leads to the same amount of change in the expected welfare and the expected payoff of the bidder. To see this, let $m, m' \in \bar{M}^e$ be two given efficient mechanisms with assignment and fee rules $(\sigma, \tau), (\sigma', \tau')$, respectively such that $\tau = \tau'$ and $\sigma_{-i} = \sigma'_{-i}$, but $\sigma_i \neq \sigma'_i$ for some bidder $i \in N$. Note that under the efficient rule (p^e, x^e) , for any given profile of bids $t \in T$, payoff of bidder i is

$$u_i^e(t) = t_i p_i^e(t) + \sum_{j \neq i} t_j p_j^e - \max_{j \neq i} t_j = \sum_{k \in N} t_k p_k^e - \max_{j \neq i} t_j.$$

Therefore, the interim payoff of bidder i when she bids t_i is

$$\begin{aligned} u_i^e(t_i | \sigma_i(\theta_i), \sigma_{-i}) &= E_{\theta_{-i}} [E_{\sigma_{-i}[\theta_{-i}]} [t_i p_i^e(t | \sigma_i(\theta_i), \sigma_{-i}) + \sum_{j \neq i} t_j p_j^e(t | \sigma_i(\theta_i), \sigma_{-i}) - \max_{j \neq i} t_j]] \\ &= E_{\theta_{-i}} [E_{\sigma_{-i}[\theta_{-i}]} [\sum_{k \in N} t_k p_k^e(t | \sigma_i(\theta_i), \sigma_{-i}) - \max_{j \neq i} t_j]], \end{aligned}$$

given that all other bidders truthfully bid according to their assigned experiments in $\sigma_{-i}(\theta_{-i})$ for each realization of profile of types θ_{-i} . Therefore, the interim payoff of bidder i when she follows her experiment $\sigma_i(\theta_i)$ is equal to

$$\begin{aligned} u_i^e(\sigma_i(\theta_i) | \sigma_{-i}) &= E_{\sigma_i(\theta_i)} [E_{\theta_{-i}} [E_{\sigma_{-i}[\theta_{-i}]} [\sum_{k \in N} t_k p_k^e(t | \sigma_i(\theta_i), \sigma_{-i}(\theta_{-i}))]] \\ &\quad - E_{\theta_{-i}} [E_{\sigma_{-i}[\theta_{-i}]} [\max_{j \neq i} t_j]]. \end{aligned}$$

But then when there is a change of assignment for bidder i from σ_i to σ'_i , for each type realization $\theta_i \in \Theta_i$ we have the payoff difference as

$$\begin{aligned} u_i^e(\sigma'_i(\theta_i) | \sigma_{-i}) - u_i^e(\sigma_i(\theta_i) | \sigma_{-i}) &= E_{\sigma'_i(\theta_i)} [E_{\theta_{-i}} [E_{\sigma_{-i}[\theta_{-i}]} [\sum_{k \in N} t_k p_k^e(t | \sigma'_i(\theta_i), \sigma_{-i}(\theta_{-i}))]] \\ &\quad - E_{\sigma_i(\theta_i)} [E_{\theta_{-i}} [E_{\sigma_{-i}[\theta_{-i}]} [\sum_{k \in N} t_k p_k^e(t | \sigma_i(\theta_i), \sigma_{-i}(\theta_{-i}))]]]. \end{aligned}$$

Taking an expectation over Θ_i , we derive that the expected payoff difference for bidder i by switching from σ_i to σ'_i is equal to

$$E_{\theta_i} [u_i^e(\sigma'_i(\theta_i) | \sigma_{-i})] - E_{\theta_i} [u_i^e(\sigma_i(\theta_i) | \sigma_{-i})] = E_{\theta} [\sum_{k \in N} t_k p_k^e(t, \sigma'(\theta))] - E_{\theta} [\sum_{k \in N} t_k p_k^e(t, \sigma(\theta))]. \quad (3.1)$$

Let $\Pi_i^e(\sigma_i, \sigma_{-i}) = E_{\theta_i}[u_i^e(\sigma_i(\theta_i)|\sigma_{-i})] - E_{\theta_i}[c_{\theta_i}(\sigma_i(\theta_i))]$ denote the expected net interim payoff of bidder i before the fees are applied. Whenever the fees are set in such a way that the stage-1 IC and IR constraints hold, then the bidder will be maximizing the objective function $\Pi^e(\sigma_i, \sigma_{-i})$ by choosing an experiment $\sigma_i(\theta_i)$ for each realization of her type $\theta_i \in \Theta_i$. Notice that if we subtract $E_{\theta_i}[c_{\theta_i}(\sigma'_i(\theta_i))] - E_{\theta_i}[c_{\theta_i}(\sigma_i(\theta_i))]$ from the left hand side in the equation (3.1), then we obtain $\Pi_i^e(\sigma'_i, \sigma_{-i}) - \Pi_i^e(\sigma_i, \sigma_{-i})$.

By definition, we have

$$W^e(\sigma_i, \sigma_{-i}) = E_{\theta}[\sum_{k \in N} t_k p_k^e(t, \sigma(\theta))] - E_{\theta}[\sum_{k \in N} c_{\theta_k}(\sigma_k(\theta_k))]$$

as the expected welfare. If we subtract $E_{\theta}[\sum_{k \in N} c_{\theta_k}(\sigma'_k(\theta_k))] - E_{\theta}[\sum_{k \in N} c_{\theta_k}(\sigma_k(\theta_k))]$ from the right hand side of the equation (3.1), we obtain $W^e(\sigma'_i, \sigma_{-i}) - W^e(\sigma_i, \sigma_{-i})$. But also note that since $\sigma_{-i} = \sigma'_{-i}$, when switching from σ_i to σ'_i the expected cost difference for bidder i and the total expected cost difference are the same, and so we have

$$E_{\theta_i}[c_{\theta_i}(\sigma'_i(\theta_i))] - E_{\theta_i}[c_{\theta_i}(\sigma_i(\theta_i))] = E_{\theta}[\sum_{k \in N} c_{\theta_k}(\sigma'_k(\theta_k))] - E_{\theta}[\sum_{k \in N} c_{\theta_k}(\sigma_k(\theta_k))].$$

This means, however, that the following equality holds

$$\Pi_i^e(\sigma'_i, \sigma_{-i}) - \Pi_i^e(\sigma_i, \sigma_{-i}) = W^e(\sigma'_i, \sigma_{-i}) - W^e(\sigma_i, \sigma_{-i})$$

which shows that we have a Bayesian exact potential game with the potential function W^e ; that is, the marginal change in the expected welfare is the same as the marginal change in the payoff of bidder i when she switches from σ_i to σ'_i . As such, the game bidders are playing at stage-1 is a Bayesian exact potential game, and so any maximizer σ^* of the expected welfare function $W^e(\sigma)$ must be a Bayesian Nash equilibrium of the game where each bidder freely chooses an experiment $f_i \in D(s_i)$ whenever her type is some $\theta_i = (r_i, s_i)$.³

Optimal assignments: We now show that the potential game we described above has a solution σ^* . For any given type θ_i , the set $D(\theta_{i2})$ is a non-empty, convex, closed subset of $\Delta([a, b])$. Recall that the set $\Delta([a, b])$ is weakly compact, and so $D(\theta_{i2})$ is also compact. A pure (measurable) strategy for bidder i is a map $\sigma_i : \Theta_i \rightarrow \Delta([a, b])$ such

³In a related study, Bergemann and Pesendorfer [2007] consider an IPV auction setting and show that when the seller is in full control of the information precision of each bidder, then the optimal information structures are partitional and the partitions are asymmetric among the bidders.

that $\sigma_i(\theta_i) \in D(\theta_{i2})$ for almost all $\theta_i \in \Theta_i$ with respect to the measure $F_r \times F_s$. Let A_i be the space of such maps. We equip A_i with the pointwise weak convergence topology since $A_i \subset \prod_{\theta_i} \Delta([a, b])$, and therefore it inherits the product of weak topologies. Moreover, A_i is closed and since $\prod_{\theta_i} \Delta([a, b])$ is compact by Tychonoff's theorem, A_i is compact. Now let $A = A_1 \times \dots \times A_n$ be the joint strategy space with the product of the product-weak topologies. A is compact since it is the product of compact spaces.

Given $\sigma = (\sigma_1, \dots, \sigma_n)$, we have the Bayesian exact potential as

$$W^e(\sigma) = E_\theta \left[\sum_{k \in N} t_k p_k^e(t, \sigma(\theta)) \right] - E_\theta \left[\sum_{k \in N} c_{\theta_k}(\sigma_k(\theta_k)) \right].$$

We want to show that $W^e(\sigma)$ is upper semi-continuous (u.s.c). Clearly, the first term above is continuous since it is a linear function of σ . For each θ_i , the cost function $c_{\theta_i}(\cdot)$ is l.s.c. on $D(\theta_{i2})$. Thus the function $c_{\theta_i}(\sigma_i(\theta_i))$ is l.s.c. pointwise. Taking an expectation over Θ_i yields that $E_{\theta_i}[c_{\theta_i}(\sigma_i(\theta_i))]$ is l.s.c., and finally summing over the bidders gives us the fact that $E_\theta[\sum_{k \in N} c_{\theta_k}(\sigma_k(\theta_k))]$ is a l.s.c. functional of f_σ . Thus, we derive that $W^e(\sigma)$ is an u.s.c functional of σ . Therefore, by the Weierstrass theorem there exists at least one maximizer σ^* such that $W^e(\sigma^*) = \max_{\sigma \in A} W^e(\sigma)$.

Furthermore, we can argue that there exists a symmetric solution such that $\sigma_i^* = \sigma_j^*$ for all $i, j \in N$. Now let π be a permutation of the set N and define by $\sigma_\pi^* = (\sigma_{\pi(1)}^*, \dots, \sigma_{\pi(n)}^*)$ the permuted profile. Since every bidder's type realization is i.i.d, both the first and second terms in the welfare functional are symmetric in the bidders, we have $W^e(\sigma^*) = W^e(\sigma_\pi^*)$ for all π . Thus, every permutation σ_π^* of the maximizer σ^* is also a maximizer. Let $\bar{\sigma} = \frac{1}{n!} \sum_{\pi} \sigma_\pi^*$ be a symmetrized strategy profile. Since A is convex, we have $\bar{\sigma} \in A$. Recall that the first term of the welfare functional is linear in σ , while the second term is convex in σ since $c_{\theta_i}(\cdot)$ is a convex function for each i and θ_i . But then the welfare functional $W^e(\sigma)$ is concave in σ . Therefore, by Jensen's inequality we have

$$W^e(\bar{\sigma}) = W^e\left(\frac{1}{n!} \sum_{\pi} \sigma_\pi^*\right) \geq \frac{1}{n!} \sum_{\pi} W^e(\sigma_\pi^*) = W^e(\sigma^*)$$

showing that the symmetrized strategy profile $\bar{\sigma}$ is a maximizer of W^e . Is the solution unique? With our current set of assumptions on the primitives, this is not necessarily the case. However, if we assume that the cost function c_{θ_i} is strictly convex instead of only convex, then the potential function becomes strictly concave. In that case, there exists a unique solution. Moreover, due to the permutation argument given above, the unique solution must be symmetric.

Optimal fees: We now analyze how the optimal fees τ^* can be obtained. Recall that we fixed (p^e, x^e) as stage-2 rule and without loss of generality the optimal assignment profile σ^* at stage-1 can be taken symmetric such that $\sigma_i^* = \sigma_j^*$ for all bidders i, j . For each truly reported type $\theta_i = (r_i, s_i)$, let $\Phi_i(r_i, s_i) = \max_{f_i \in D(s_i)} (u_i^e(f_i | \sigma_{-i}^*) - c_{r_i, s_i}(f_i))$ denote the highest stage-2 utility obtained before the fees applied. If instead the type is misreported as $\theta'_i = (r'_i, s'_i)$, then her best (before fees are applied) utility is $\Phi_i(r'_i, s'_i | r_i, s_i) = \max_{f_i \in D(s'_i)} (u_i^e(f_i | \sigma_{-i}^*) - c_{r_i, s_i}(f_i))$.

Note that the misreported type $\theta'_i = (r'_i, s'_i)$ affects pre-fee payoff only through s'_i ; misreported r'_i has no effect. Fix some s_i and let $\theta_i = (r_i, s_i)$ be the true type and (r'_i, s_i) be the reported type. Let $\tau_i : \Theta_i \rightarrow \mathbb{R}$ be a feasible fee scheme for each bidder i . As such, τ_i (together with stage 2 rule (p^e, x^e) and stage 1 experiment scheme σ^*) must satisfy stage-1 IR and IC constraints. The bidder i 's net payoff under the fee schedule τ_i is $\Phi_i(r_i, s_i) - \tau_i(r'_i, s_i)$. The stage-2 IC condition in r then implies that

$$\Phi_i(r_i, s_i) - \tau_i(r_i, s_i) \geq \Phi_i(r_i, s_i) - \tau_i(r'_i, s_i)$$

and so $\tau_i(r_i, s_i) \leq \tau_i(r'_i, s_i)$. But then swapping the roles of r_i and r'_i , we obtain that $\tau_i(r_i, s_i) = \tau_i(r'_i, s_i)$ for all s_i ; that is, the optimal fee should be independent of the reported parameter r_i . Therefore, let $\tau_i(s_i) : [0, 1] \rightarrow \mathbb{R}$ denote in short the feasible fee for each bidder i and reported s_i .

The stage-1 IR constraint requires that $\Phi_i(r_i, s_i) - \tau_i(s_i) \geq 0$ for each r_i and s_i . Notice that $\Phi_i(\bar{r}_i, s_i) \leq \Phi_i(r_i, s_i)$ for all s_i where $\bar{r}_i = 1$. Thus, the fee schedule must satisfy

$$\tau_i(s_i) \leq \Phi_i(\bar{r}_i, s_i) \tag{3.2}$$

for each $s_i \in [0, 1]$. Now consider the stage-1 IC constraint in s_i . We have

$$\Phi_i(r_i, s_i) - \tau_i(s_i) \geq \Phi_i(r_i, s'_i | r_i, s_i) - \tau_i(s'_i)$$

for every r_i , s_i , and s'_i . Let $H_i(s_i, s'_i) = \inf_{r_i \in [0, 1]} (\Phi_i(r_i, s_i) - \Phi_i(r_i, s'_i | r_i, s_i))$ denote the lowest pre-fee utility gap among the true types (r_i, s_i) who report (r_i, s'_i) . Due to the above stage-1 IC constraint, we must have

$$\tau_i(s_i) \leq \tau_i(s'_i) + H_i(s_i, s'_i) \tag{3.3}$$

for each $s_i, s'_i \in [0, 1]$. It is clear that the above two conditions (equations 3.2 and 3.3)

characterize the set of all feasible stage-1 fees (see, e.g., Rochet [1987], Theorem 1).

Now we want to find out which feasible fee can maximize the expected revenue for the seller. Let $\Omega = \cup_{k \in \mathbb{N}} [0, 1]^k$ denote the set of finite chains (e.g., $\omega = (\omega_1, \dots, \omega_k)$ for some $k \in \mathbb{N}$) generated from the set $[0, 1]$. Let $\Omega(s)$ denote the set of finite chains which end with $s \in [0, 1]$; that is, if $\omega \in \Omega(s)$ with $\omega \in [0, 1]^k$ for some $k \geq 0$, then $\omega_k = s$. Let $l(\omega)$ denote the length of a given chain $\omega \in \Omega$; that is, $l(\omega) = k \in \mathbb{N}$ if $\omega \in [0, 1]^k$. Define the chain-closure fee for each bidder i as follows:

$$\tau_i^{cc}(s_i) = \inf_{\omega \in \Omega(s_i)} \left(\Phi_i(\bar{r}_i, \omega_1) + \sum_{k=2}^{l(\omega)} H_i(\omega_k, \omega_{k-1}) \right)$$

for any $s_i \in [0, 1]$. We now show that the chain-closure fee is feasible; that is, it satisfies stage-1 IR and IC constraints. Since $H_i(\omega_k, \omega_{k-1}) \geq 0$ for any $\omega \in \Omega$, we have $\inf_{\omega \in \Omega} \left(\Phi_i(\bar{r}_i, \omega_1) + \sum_{k=2}^{l(\omega)} H_i(\omega_k, \omega_{k-1}) \right) \leq \Phi_i(\bar{r}_i, \omega_1)$, and so $\tau^{cc}(s_i) \leq \Phi_i(\bar{r}_i, s_i)$ for any given $s_i \in [0, 1]$ showing that stage-1 IR holds. Now fix some $s_i, s'_i \in [0, 1]$. Since $\tau^{cc}(s'_i)$ is the infimum, for any $\epsilon > 0$ we can find a chain $\omega \in \Omega(s'_i)$ such that $\Phi_i(\bar{r}_i, \omega_1) + \sum_{k=2}^{l(\omega)} H_i(\omega_k, \omega_{k-1}) \leq \tau^{cc}(s'_i) + \epsilon$. But then if we add $H_i(s_i, s'_i)$ to both hand side of the inequality we derive

$$\Phi_i(\bar{r}_i, \omega_1) + \sum_{k=2}^{l(\omega)} H_i(\omega_k, \omega_{k-1}) + H_i(s_i, s'_i) \leq \tau^{cc}(s'_i) + H_i(s_i, s'_i) + \epsilon.$$

By definition, the left hand side of the above inequality is bigger than $\tau^{cc}(s_i)$, and so we have $\tau_i^{cc}(s_i) \leq \tau_i^{cc}(s'_i) + H_i(s_i, s'_i) + \epsilon$. Since ϵ was arbitrary, we can let it go to 0, and obtain in the limit that $\tau_i^{cc}(s_i) \leq \tau_i^{cc}(s'_i) + H_i(s_i, s'_i)$ showing that stage-1 IC condition holds. Thus, the chain-closure fee τ_i^{cc} is feasible.

Let $\omega \in \Omega(s_i)$ be a given finite chain for some $s_i \in [0, 1]$. Since for any feasible fee τ_i stage-1 IR and IC conditions hold, we have $\tau_i(\omega_1) \leq \Phi_i(\bar{r}_i, \omega_1)$ and $\tau_i(\omega_k) \leq \tau_i(\omega_{k-1}) + H_i(\omega_k, \omega_{k-1})$ for any $k = 2, \dots, l(\omega)$. By iterating and summing up the terms, we obtain that $\tau_i(s_i) = \tau_i(\omega_k) \leq \Phi_i(\bar{r}_i, \omega_1) + \sum_{k=2}^{l(\omega)} H_i(\omega_k, \omega_{k-1})$ showing that $\tau_i(s_i)$ is a lower-bound for the set $T(s_i) = \{ \Phi_i(\bar{r}_i, \omega_1) + \sum_{k=2}^{l(\omega)} H_i(\omega_k, \omega_{k-1}) : \omega \in \Omega(s_i) \}$. Since, by definition, $\tau_i^{cc}(s_i)$ is the infimum of the set $T(s_i)$, we conclude that $\tau_i(s_i) \leq \tau_i^{cc}(s_i)$ and so the chain-closure fee yields the maximum amount of fee for each s_i given that stage-2 allocation rule and stage-1 experiment rule are fixed as (p^e, x^e) and $\bar{\sigma}$, respectively.

Optimal two-stage mechanism: To sum up, the optimal two-stage mechanism (p, x, σ, τ) in \bar{M} is obtained by letting stage-2 rule to be efficient rule (p^e, x^e) , while letting stage-1 assignment rule σ to be a symmetric maximizer $\bar{\sigma}$ of the potential function $W^e(\sigma)$, and letting stage-1 fees τ to be determined by the chain-closure fees τ^{cc} .

Theorem 1. *The two-stage mechanism $m^* = (p^e, x^e, \bar{\sigma}, \tau^{cc})$ maximizes both the expected revenue $R(m)$ and the expected welfare $W(m)$ among the feasible mechanisms in \bar{M} .*

The proof of Theorem 1 follows from above discussions (including Lemma 1 and Corollary 1), therefore it is omitted. We see that $(p^e, x^e, \bar{\sigma}, \tau^{cc})$ is the optimal two-stage mechanism where at stage-2 the efficient outcome is achieved, while at stage-1 the seller lets the bidders experiment as they wish, but charges them a fee according to the chain-closure scheme and generates the maximum level of expected revenue, while leaving a minimum level of expected information rent to the bidders. From the welfare maximization perspective this outcome is optimal since the efficient rule already ensures that the maximum level of welfare is achieved.

4 Extensions

We now consider some possible extensions of the mechanism design problem we have considered above. Firstly, we relax the assumption that the seller can verify for free the experiment used f_i or its center $\mu(f_i)$ whenever the bidder i reports s_i . We instead assume that the seller pays a cost for auditing each bidder's choice of experiment. We identify the effects of this change on the equilibrium outcomes. In an extreme case, the seller cannot verify any statistics about the experiment (no auditing); in that situation, we observe that the optimal transfer is a flat fee. Secondly, we allow for the possibility that the principal can also verify the cost parameter $r \in [0, 1]$ of each bidder with costly auditing. We identify the effects of this change on the optimal fee schedule.

Auditing bidder experiments: Suppose that the seller cannot verify for free the experiment undertaken by a bidder. Instead, the seller must pay a cost for auditing the experiments. Let $k_e : [0, 1] \rightarrow \mathbb{R}_+$ denote the cost of auditing such that when the seller incurs the cost $k_e(q) \geq 0$, then with probability $q \in [0, 1]$ the seller can perfectly observe the conducted experiment, and with probability $1 - q$ he does not observe the experiment at all. In case a misreport by bidder i is detected, then bidder i is punished by at most an amount of $P_i \leq 0$.

As before, the optimal two-stage mechanism with auditing will be used $(p^e, x^e, \bar{\sigma}, \tau^{cc})$ as part of the revenue maximization process. Moreover, the seller needs to determine an auditing rule where for each $s_i \in [0, 1]$ reported by bidder i , the seller's success probability of catching a non-compliance will be $q(s_i) \in [0, 1]$. For each reported type $\theta_i = (r_i, s_i)$, let

$$\Psi_i(r_i, s_i) = \max_{f_i \in D \setminus D(s_i)} (u_i^e(f_i | \sigma_{-i}^*) - c_{r_i, s_i}(f_i))$$

denote the best non-compliant pre-fee utility of bidder i . If bidder i reports $\theta_i = (r_i, s_i)$ truthfully and complies, she obtains a payoff of $\Phi_i(r_i, s_i) - \tau_i^{cc}(s_i)$. If, on the other hand, she deviates the best she can expect to receive will be $(1 - q(s_i)) \Psi_i(r_i, s_i) + q(s_i) P_i - \tau_i^{cc}(s_i)$. Therefore, truth-telling is enforced if $\Phi_i(r_i, s_i) \geq (1 - q(s_i)) \Psi_i(r_i, s_i) + q(s_i) P_i$ for all r_i .

Rearranging the above terms, we have the lower-bound condition on the deterrence probability $q(s_i) \geq \frac{\Psi_i(r_i, s_i) - \Phi_i(r_i, s_i)}{\Psi_i(r_i, s_i) - P_i} \in [0, 1]$. Thus, the minimal auditing probability when s_i is reported (regardless of r_i) is

$$q_{P_i}^*(s_i) = \sup_{r_i \in [0, 1]} \left[\frac{\Psi_i(r_i, s_i) - \Phi_i(r_i, s_i)}{\Psi_i(r_i, s_i) - P_i} \right]^+$$

where $[\cdot]^+$ refers to the $\max\{0, \cdot\}$ operator. Note that the higher $|P_i|$ is, the lower $q_{P_i}^*(s_i)$ is for each s_i , and so the lower auditing cost $k_e(q_{P_i}^*(s_i))$ will be. Suppose that the auditing costs are low enough that it is optimal for the seller to audit. In that case, if $P = (P_1, \dots, P_n)$ is the profile of maximum punishments that can be imposed on bidders, then in equilibrium the maximum expected revenue under auditing experiments will be $E_\theta[\sum_{i \in N} (x_i^e(t, \sigma^*(\theta)) + \tau_i^{cc}(\theta_{i2}))] - E_\theta[\sum_{i \in N} k_e(q_{P_i}^*(\theta_{i2}))]$.

Auditing bidder costs: Suppose that the seller can verify for free the experiments. Moreover, suppose that the seller can audit bidder costs. Specifically, let $k_r : [0, 1] \rightarrow \mathbb{R}_+$ denote the cost of auditing such that when the seller incurs the cost $k_r(q) \geq 0$, then with probability $q \in [0, 1]$ the seller can perfectly observe bidder i 's cost parameter r_i , and with probability $1 - q$ he does not observe the cost parameter. In case a misreport is detected, each bidder i is punished by at most the amount $P_i \leq 0$.

Let $H_i(s_i, s'_i | r_i) = \Phi_i(r_i, s_i) - \Phi_i(r_i, s'_i | r_i, s_i)$ denote the pre-fee utility gap when a true type (r_i, s_i) reports (r_i, s'_i) . Define the chain-closure fee for each bidder i as follows:

$$\tau_i^{cc}(r_i, s_i) = \inf_{\omega \in \Omega(s_i)} \left(\Phi_i(r_i, \omega_1) + \sum_{k=2}^{l(\omega)} H_i(\omega_k, \omega_{k-1}) \right)$$

for any $\theta_i = (r_i, s_i) \in [0, 1]^2$. For each fixed r_i , the fee scheme $\tau_i^{cc}(r_i, \cdot)$ satisfies stage-1 IR and IC constraints, and maximizes the amount of fee that can be collected among the feasible schemes. Suppose that bidder i with true type (r_i, s_i) contemplates reporting (r'_i, s_i) . Then, the auditing deterrence requires

$$\Phi_i(r_i, s_i) - \tau_i^{cc}(r_i, s_i) \geq (1 - q(r'_i))[\Phi_i(r_i, s_i) - \tau_i^{cc}(r'_i, s_i)] + q(r'_i)P_i.$$

Rearranging the terms gives us the inequality that

$$q(r'_i) \geq \frac{\tau_i^{cc}(r_i, s_i) - \tau_i^{cc}(r'_i, s_i)}{\Phi_i(r_i, s_i) - \tau_i^{cc}(r'_i, s_i) - P_i}.$$

If $\tau_i^{cc}(r'_i, s_i) \geq \tau_i^{cc}(r_i, s_i)$, then the inequality trivially holds; the fees already establish the deterrence, and so there is no need for costly auditing. If $\tau_i^{cc}(r'_i, s_i) < \tau_i^{cc}(r_i, s_i)$, then the auditing probability should be at least equal to the right hand side. Choosing the minimum detection probability deters all possible misreports for every s :

$$q_{P_i}^*(r'_i) = \sup_{r_i, s_i: \tau_i^{cc}(r'_i, s_i) \leq \tau_i^{cc}(r_i, s_i)} \left[\frac{\tau_i^{cc}(r_i, s_i) - \tau_i^{cc}(r'_i, s_i)}{\Phi_i(r_i, s_i) - \tau_i^{cc}(r'_i, s_i) - P_i} \right]^+.$$

By stage-1 IR and $P_i \leq 0$, we have $q_{P_i}^*(r'_i) \in [0, 1]$. Once again the higher $|P_i|$ is, the lower $q_{P_i}^*(r_i)$ is for each r_i , and so the lower auditing cost $k_r(q_{P_i}^*(r_i))$ will be. Suppose that the auditing costs are low enough that it is optimal for the seller to audit. In that case, if $P = (P_1, \dots, P_n)$ is the profile of maximum punishments that can be imposed on bidders, then in equilibrium the maximum expected revenue under auditing will be $E_\theta[\sum_{i \in N} (x_i^e(t, \sigma^*(\theta)) + \tau_i^{cc}(\theta_i))] - E_\theta[\sum_{i \in N} k_r(q_{P_i}^*(\theta_{i1}))]$.

No auditing: Suppose that the seller cannot audit bidder experiments or costs. In this case, the seller cannot enforce stage-1 IC constraint in the cost-scale parameter r or mean-value parameter s . In other words, any r or s dependent stage-1 fee is not incentive compatible, and so stage-1 fees should be flat in both r and s . Since we still have a potential game, the optimal assignments of experiments can be found by considering the maximizer of the potential function in D^n :

$$\max_{\sigma \in D^n} W^e(\sigma) = \max_{\sigma \in D^n} \left(E_\theta \left[\sum_{k \in N} t_k p_k^e(t, \sigma(\theta)) \right] - E_\theta \left[\sum_{k \in N} c_{\theta_k}(\sigma_k(\theta_k)) \right] \right).$$

Then the total expected payment at stage-2 will be $x^{na} = E_{\theta}[\sum_{i \in N}(x_i^e(t, \sigma^{na}(\theta)))]$ where $\sigma^{na} \in D^n$ is the maximizer of the potential function when there is no audit, and so the bidders can choose any experiment without any center restriction. In this case, the highest flat fee that the seller can enforce τ_i^{na} should satisfy stage-1 IR constraint. As such, τ_i^{na} should satisfy $\tau_i^{na} = \text{ess inf}_{(r_i, s_i)} \Phi_i^{na}(r_i, s_i)$ where $\Phi_i^{na}(r_i, s_i) = \max_{f_i \in D} (u_i^e(f_i | \sigma_{-i}^{na}) - c_{r_i, s_i}(f_i))$. Thus, the total expected revenue for the seller will be $x^{na} + \sum_{i \in N} \tau_i^{na}$.

5 Discussion

Our two-stage mechanism—VCG at stage-2 combined with a stage-1 experiment and fee schedules (where fees are constructed via the shortest-path (i.e., the chain-closure) solution to the s -IC/IR envelope)—differs from classic optimal auction models in two intertwined ways: (i) it preserves efficiency ex-post (VCG) while extracting rents ex-ante through a fee that internalizes information choices, and (ii) it treats the stage-1 information-acquisition game as a Bayesian exact potential game, so the equilibrium experiment profile maximizes expected welfare. Because bidders are ex-ante symmetric, the stage-1 experiment assignment rule and the fee schedule can be taken symmetric across bidders without loss, which greatly simplifies both characterization and implementation.

Relative to Myerson [1981] and Milgrom and Weber [1982], who characterize revenue-optimal allocation/pricing with fixed information, our two-stage mechanism keeps the allocation rule efficient (VCG) rather than ironing virtual values. The rationale is that with endogenous signals the informational externalities run through stage-2 payoffs; VCG maximizes the potential (expected surplus minus information costs), so any distortion of stage-2 would reduce the potential and, by the standard revenue decomposition, reduce revenue. In contrast, Stegeman [1996] and Cremer, Spiegel, and Zheng [2009] introduce participation/inspection fees and show how entry or inspection can be screened; our fee, however, is not a uniform entry or inspection fee but the shortest-path envelope over pairwise s -IC constraints at a given r , which achieves maximal rent extraction subject to truthful centers and stage-1 IR.

Within the auction design with endogenous information literature, mechanisms typically either (a) restrict or recommend information structures, or (b) embed precision choices into the objective. Our approach is closest in spirit to Bergemann and Valimaki [2002]’s efficiency with costly information and to Bergemann and Pesendorfer [2007]’s focus on information structures, but it leans on two design commitments: (1) stage-2 remain VCG precisely because the stage-1 signal game is a Bayesian exact potential game,

so welfare-maximization at stage-2 aligns with equilibrium information choices; and (2) rent extraction occurs entirely via a stage-1 fee, pinned down by the chain-closure construction that aggregates all local s -deviation bounds into a global maximum-revenue, s -IC-and-IR-feasible tariff. Compared to Shi [2012]) and Gershkov, Moldavanu, Strack, and Zhang [2021], which allow allocation distortions to influence information incentives, our two-stage mechanism isolates information provision from allocation by keeping VCG at stage-2 and using the fee to implement the efficient information profile while minimizing information rents.

There are other related work which treat information as a design primitive but differ on who controls it, what is optimized, and whether allocation is distorted. Mensch [2022] studies a monopoly screening problem with limited attention: buyers choose attention and the seller uses menus (attributes/prices) to separate types, accepting distortions to extract surplus. Eso and Szentes [2007] analyze competitive auctions where the seller designs public disclosure/handicaps to raise revenue, often tilting competition and sacrificing efficiency when profitable. By contrast, our IPV auction keeps allocation efficient by committing to VCG at stage-2 and prices bidder-chosen information ex ante via a stage-1 experiment/fee menu (built from an s -IC/IR shortest-path envelope). This makes the pre-auction learning game an exact potential game, implements efficient information acquisition, and lets the seller achieve optimal revenue without allocation distortions.

Finally, Persico [2000] studies signal precision choice at a cost in first- and second-price auctions within a model of affiliated values; Pernoud and Gleyze [2025] take the second-price allocation rule as given and characterize how competition shapes how much and what type of information bidders seek (e.g., learning about own values or rival values). By contrast, our objective is to find an optimal mechanism which maximizes the revenue and/or welfare. Our two-stage mechanism designs the information stage (and fees) so that competition remains effective and revenue-optimal under IPV with endogenous learning. The symmetry of the environment implies a symmetric fee/experiment mapping across bidders, and the potential-game structure implies that the induced information profile is socially optimal among feasible signals. In short, our two-stage mechanism sits at the intersection of efficient allocation (VCG), optimal information provision (through the potential), and maximal rent extraction (via chain-closure), offering a clean benchmark against which the classic revenue-maximizing distortions or inspection-fee designs can be contrasted.

6 Conclusion

This paper revisits single-object auctions under independent private values when bidders can flexibly—but at a cost—improve what they know about their own valuations before bidding. We contrasted the welfare benchmark, which implements efficient allocation in VCG/second-price environments, with the revenue benchmark characterized by virtual values and reserves. We then asked how endogenous information acquisition reshapes this classic dichotomy. Our main result shows that, once the seller commits to a two-stage design in which experiments are registered up front and the stage-2 allocation rule is VCG, the incentives to learn line up with efficiency, and the seller can implement a transfer scheme that is incentive compatible and individually rational while also attaining the Myerson revenue objective. Intuitively, fixing VCG at stage-2 turns the pre-auction information game into an exact potential game, so equilibrium learning maximizes expected welfare; the stage-1 fee schedule then extracts the value generated by better selection without needing to condition on unverifiable cost scales, relying instead on verifiable features of the registered experiments. This delivers a unification: in our environment, the divergence between revenue and welfare that arises from exclusion, asymmetries, and reserves in static models can be neutralized when learning is properly internalized at stage-1.

These findings carry practical and theoretical implications. Practically, they suggest that when pre-auction diligence is salient—as in spectrum, mineral rights, or complex procurement—designers should evaluate formats not only by their static allocation rules but by the learning incentives they embed; a simple commitment to VCG in stage-2 coupled with a transparent, experiment-based registration fee can simultaneously deliver high revenue and efficiency. Theoretically, our results refine the scope of revenue equivalence: formats that are revenue-equivalent in the no-learning benchmark need not remain so once learning is endogenous, but an appropriate two-stage commitment can restore alignment by making the induced allocation rule format-invariant through efficient information acquisition. Several extensions merit further work: robustness to richer frictions in verifying experiments; budget or liquidity constraints; risk aversion; correlated or affiliated signals beyond IPV; dynamic entry and disclosure; and optimal auditing technologies. Each of these can interact with learning incentives and reserves in ways that either reinforce or erode the alignment we document, offering a roadmap for future research at the intersection of information design and auction theory.

References

- BERGEMANN, D. AND M. PESENDORFER (2007): “Information structures in optimal auctions,” *Journal of Economic Theory*, 137, 580–609.
- BERGEMANN, D. AND J. VALIMAKI (2002): “Information acquisition and efficient mechanism design,” *Econometrica*, 70, 1007–1033.
- CLARKE, E. H. (1971): “Multipart pricing of public goods,” *Public Choice*, 11, 17–33.
- CREMER, J., Y. SPIEGEL, AND C. ZHENG (2009): “Auctions with costly information acquisition,” *Economic Theory*, 38, 41–72.
- ESO, P. AND B. SZENTES (2007): “Optimal information disclosure in auctions and the handciap auction,” *Review of Economic Studies*, 74, 705–731.
- GERSHKOV, A., B. MOLDAVANU, P. STRACK, AND M. ZHANG (2021): “A theory of auctions with endogenous valuations,” *Journal of Political Economy*, 129, 1011–1051.
- GROVES, T. (1973): “Incentives in teams,” *Econometrica*, 41, 617–631.
- MENSCH, J. (2022): “Screening Inattentive Buyers,” *American Economic Review*, 112, 1949–1984.
- MILGROM, P. R. AND R. J. WEBER (1982): “A theory of auctions and competitive bidding,” *Econometrica*, 50, 1089–1122.
- MONDERER, D. AND L. S. SHAPLEY (1996): “Potential games,” *Games and Economic Behavior*, 14, 124–143.
- MYERSON, R. B. (1981): “Optimal auction design,” *Mathematics of Operations Research*, 6, 58–73.
- PERNOUD, A. AND S. GLEYZE (2025): “How competition shapes information in second-price auctions,” Working Paper.
- PERSICO, N. (2000): “Information acquisition in auctions,” *Econometrica*, 68, 135–148.
- RILEY, J. G. AND W. F. SAMUELSON (1981): “Optimal auctions,” *American Economic Review*, 71, 381–392.
- ROCHET, J.-C. (1987): “A necessary and sufficient condition for rationalizability in a quasi-linear context,” *Journal of Mathematical Economics*, 16, 191–200.

SHI, X. (2012): “Optimal auctions with information acquisition,” *Games and Economic Behavior*, 74, 227–243.

STEGEMAN, M. (1996): “Participation costs and efficient auctions,” *Journal of Economic Theory*, 71, 228–259.

VICKREY, W. (1961): “Counterspeculation, auctions, and competitive sealed tenders,” *Journal of Finance*, 16, 8–37.