Theoretical Economics 15 (2020), 319–359 1555-7561/20200319

Information design and sequential screening with ex post participation constraint

Tibor Heumann
Department of Applied Economics, HEC Montreal

We study a principal–agent model. The parties are symmetrically informed at first; the principal then designs the process by which the agent learns his type and, concurrently, the screening mechanism. Because the agent can opt out of the mechanism ex post, it must leave him with nonnegative rents ex post. We characterize the profit-maximizing mechanism. In that optimal mechanism, learning proceeds in continuous time and, at each moment, the agent learns a lower bound on his type. For each type, there is one of two possible outcomes: the type is allocated the efficient quantity or is left with zero rents ex post.

Keywords. Information design, sequential screening, ex post participation constraint, sequential information disclosure, dynamic mechanism design.

JEL classification. D44, D82, D83.

1. Introduction

Motivation and model There is an extensive literature that studies principal–agent models in which the principal (she) sells a good to an agent (he) who is privately informed about his marginal valuation.[1] This paper examines situations in which the principal can control the process by which the agent learns his valuation of the good. By gradually disclosing information about the good, the principal can elicit information about the agent's preferences before he fully knows his own valuation. Although we allow for the principal to control the agent's learning process, we assume throughout that the agent can opt out of the mechanism ex post (i.e., *after* the agent has learned his valuation) if this leaves him with negative rents. We answer the following three questions. Does the principal derive any benefit from controlling the agent's learning process? Can the agent earn any informational rents when he initially has no private information and the principal can design his learning process? What is the optimal mechanism from the principal's perspective?

The model consists of a principal who supplies a good to an agent in return for transfers. Both principal and agent have quasi-linear preferences. The agent's valuation of

Tibor Heumann: tibor.heumann@hec.ca

For helpful suggestions and discussions, I thank Sümeyra Akin, Dirk Bergemann, Vitor Farinha, Daniel Krähmer, and Juuso Toikka. For thoughtful and detailed comments, I thank the three anonymous referees. All errors are my own.

[1]For seminal contributions to this literature, see Myerson (1981) and Baron and Myerson (1982). To streamline the presentation, we use feminine and masculine pronouns for (respectively) principal and agent.

Available at http://econtheory.org. https://doi.org/10.3982/TE2818

the good and the principal's cost of supplying the good are both determined by a one-dimensional parameter (henceforth, the agent's *type*). At the outset, the principal and the agent are symmetrically (un)informed about the agent's type. The agent learns his type by observing signals over time; after each observed signal, he sends a message to the principal. The principal designs the *mechanism*: the joint distribution of signals and types (which defines the agent's learning process) and the function that maps the messages sent by the agent into the payments received and the quantity of the good supplied by the principal. The principal's objective is profit maximization. Once the mechanism has ended, any residual uncertainty about the agent's type is resolved, and at this point the agent can either keep the allocation yielded by the mechanism or opt out of the mechanism. In other words, the agent can guarantee himself nonnegative rents ex post regardless of the messages he sent during the mechanism. We next discuss our key assumptions in detail. These assumptions are that (i) the agent can opt out of the mechanism ex post, (ii) the principal has full control over the agent's learning, and (iii) the principal and the agent are symmetrically informed about the agent's type before the mechanism starts.

In our model, the agent can opt out of the mechanism *after* he has learned his valuation (i.e., ex post). We interpret the agent's valuation as the expected utility he will derive from the good conditional on all available information about the good (that is, all information available to him and the principal). This assumption represents situations in which legal stipulations allow a buyer to renege on a contract after inspecting the good. These legal stipulations—usually present in consumer protection laws and other standard buyer–seller contracts—prevent the seller from *concealing* information about the good at the time the agent commits to buying the good.[2] For example, the European Union has legislated withdrawal rights that allow customers to return online purchases without penalty (cf. Krähmer and Strausz 2015). Another example consists of real estate contracts, which give the buyer time to do due diligence and allow him to void the contract (during this period) at no cost.[3] In both examples, the buyer can inspect the good and learn essentially everything there is to learn about the good before committing to purchasing the good.

Our second assumption—that the principal has full control over the agent's learning process—accounts for situations in which she can actually influence the agent's learning process. For example, by disclosing different characteristics of the good being sold and eliciting the agent's responses to these pieces of information, the principal can control how the agent learns about the good. Assuming that the principal has full control over the agent's learning process makes the problem tractable. This assumption is common in research addressing information design (Kamenica and Gentzkow 2011 or Bergemann and Pesendorfer 2007). In Section 6.2, we discuss learning processes

[2] Krähmer and Strausz (2015) discuss additional interpretations of the ex post participation constraint (e.g., bankruptcy laws and labor laws).

[3] For example, a 10-day inspection period is included in every Arizona Association of Realtors (AAR) standard residential contract. During those 10 days, the buyer is allowed to renege on the contract without penalty. This clause allows the buyer to get an estimated value of the house using all available information at the moment that he commits to buying it.

that have appeared in the literature and compare them with the learning process in our mechanism.

Third, the assumption that principal and agent are (initially) symmetrically informed simplifies the problem by allowing the principal to choose any learning process. More specifically, the principal's problem when the agent has private information before entering the mechanism is isomorphic to that in a model where again the agent has no private information before entering the mechanism yet the principal is *constrained* in the class of learning processes she can design.[4] Thus, our assumption that principal and agent are symmetrically informed at the outset plays a role similar to our assumption that the principal has full control over the agent's learning process.

Results The paper's first result is to establish a lower bound on the rents that the agent can attain in a mechanism (possibly, a suboptimal one). This lower bound is constructed by evaluating the agent's rents when he uses *simple* strategies in which he announces the same messages irrespective of the observed signals. If these messages result in a contract that yields positive rents when the agent's type is high enough, then the expected rents will be strictly positive. The agent can guarantee himself strictly positive expected rents because he can opt out of the mechanism ex post if his realized type turns out to be not high enough. Therefore, the possibility of informational rents is driven not by the presence of private information before the mechanism starts, but rather by the assumption that the agent can opt out of the mechanism ex post.

Our first theorem presents a "relaxed" problem that yields an upper bound on the principal's profits; later we show that this upper bound is achieved by the optimal mechanism (i.e., the upper bound is tight). In the relaxed problem, the principal maximizes the total surplus *less* the aforementioned lower bound on the agent's rents. Those rents are smaller in the relaxed problem than in a static mechanism, which leads us to conclude that (i) the principal does benefit from designing the agent's learning process, and (ii) the trade-off between maximizing the principal's surplus and minimizing the agent's informational rents is different than under a static mechanism. We emphasize that if the principal had no constraint on the ex post losses she can impose on the agent, then no trade-off would occur because the principal could maximize the total surplus and leave the agent with zero rents.

Our second theorem characterizes the profit-maximizing mechanism. The learning process is continuous in this mechanism: at each moment, the agent learns a lower bound on his type and is asked to report whether his type strictly exceeds the bound. The agent always learns his type, which happens when the lower bound in the learning process is equal to the agent's type. We refer to this learning process as *upward disclosure*.

[4]The isomorphism is constructed by considering a model in which the agent is initially uninformed but the principal is *constrained* to designing learning processes under which the first signal observed by the agent has the same distribution as the agent's private information in the original model. In the initial period, the agent has no information and, therefore, has nothing to announce to the principal. Because the participation constraint is satisfied ex post, it does not matter whether the agent sends the first message before or after he learns his initial private information.

Upward disclosure allows the principal to relax some of the incentive constraints that appear in the static mechanism design problem. Whenever the agent (mis)reports a type lower than his true type, it must happen without knowledge of his type's exact realization. The reason is that such misreporting would occur *before* the agent learns his type. So when misreporting downward, the agent is forced to report the same lie for all realizations of his type. Because the agent cannot tailor his reports when misreporting downward, the principal can extract more rents from the agent than under a static mechanism.

The *transfers* under the optimal mechanism are described in terms of a threshold type. The transfers for types above the threshold are constructed using the envelope theorem (as in the problem of designing a static mechanism). Types below the threshold are left with zero rents (in this case, the transfers are equal to the agent's utility). These transfers leave the agent with rents equal to the lower bound characterized by the paper's first result. It follows from our construction of this lower bound that the agent's rents under the optimal mechanism are the same as if his choice of announced messages were entirely independent of the observed signals. Under the optimal mechanism, then, the agent does not benefit from conditioning his reports on the signals he observes.

The *quantities* under the optimal mechanism are also described in terms of a threshold type. If the agent's realized type is above (resp., below) the threshold, then the quantity supplied is equal to (resp., less than) the surplus-maximizing quantity. These quantities enable the best trade-off between maximizing the principal's surplus and minimizing the agent's rents.

The paper proceeds as follows. Section 2 describes our model, and Section 3 presents the benchmark established by the optimal static mechanism. We give an upper bound on the principal's profits in Section 4 and characterize an optimal mechanism in Section 5. Finally, Section 6 details the model's main assumptions and discusses how our paper is related to the literature. Proofs are collected in the Appendix. Additional material is provided in the Supplemental Material, available in a supplementary file on the journal website, http://econtheory.org/supp/2818/supplement.pdf.

2. Model

2.1 *Payoffs*

We study a principal–agent model. The agent's rents and the principal's profits are determined by the quantity $q \in Q \triangleq [0, \overline{q}] \subset \mathbb{R}$ of a good supplied by the principal, the transfers $x \in \mathbb{R}$ from the agent to the principal, and the agent's type $\theta \in \Theta \triangleq [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}$. The agent has quasi-linear rent function

$$v(q, x, \theta) \triangleq u(q, \theta) - x.$$

By the agent's *utility* we refer to his rents net of transfers (i.e., $u(q, \theta)$); this utility is twice continuously differentiable, and

$$\forall (q \in Q, \theta \in \Theta), \quad \frac{\partial u(q, \theta)}{\partial \theta} \geq 0, \quad \text{and} \quad \frac{\partial^2 u(q, \theta)}{\partial q \partial \theta} > 0.$$

The principal has quasi-linear profits

$$\pi(q, \theta, x) \triangleq x - c(q, \theta).$$

We refer to her profits net of transfers (i.e., $c(q, \theta)$) as the *cost*, which we assume to be twice continuously differentiable also. The total surplus,

$$\mathcal{S}(q, \theta) \triangleq u(q, \theta) - c(q, \theta),$$

is strictly quasi-concave in q and has a unique maximum for all θ. The maximizer of this surplus,

$$q^s(\theta) \triangleq \arg\max_{q \in Q} \mathcal{S}(q, \theta),$$

is nondecreasing (i.e., $\partial q^s(\theta)/\partial\theta \geq 0$) and continuous. If $q = 0$ is implemented, then the agent's utility and the principal's cost are both zero:

$$\forall \theta \in \Theta, \quad c(0, \theta) = u(0, \theta) = 0.$$

The principal and the agent have a common prior over Θ. The measure over Θ is assumed to be absolutely continuous with respect to the Lebesgue measure and to have strictly positive density f (i.e., $f(\theta) > 0$ for all $\theta \in \Theta$).

The principal designs a mechanism that includes two parts: the learning process by which the agent learns his type and the mapping from the messages sent by the agent to the final allocation. The principal's objective is to maximize her own profit. In what follows, we formally describe that maximization problem.

2.2 *Mechanism design*

Time is indexed by $t \in T \subset \mathbb{R}$. At each t, the agent observes a signal $s_t \in S \subset \mathbb{R}$, where S denotes the set of signals. There is no discounting in our model, so time is only a subjective index used to keep track of the order in which signals are observed. The joint distribution of type and signals is denoted $\mu \in \Delta(\Theta \times S^T)$.[5] At each $t \in T$, the agent, after observing s_t, sends a message $z_t \in Z \subset \mathbb{R}$, where Z denotes the set of messages. For any report $z \in Z^T$, the outcome of the mechanism is determined by the functions[6]

$$q\colon Z^T \to Q \quad \text{and} \quad x\colon Z^T \to \mathbb{R}.$$

The principal chooses the functions (q, x) and the distribution $\mu \in \Delta(\Theta \times S^T)$ subject to the restriction that the marginal distribution of μ over Θ has density f. The sets T, S,

[5] We assume that μ is defined on the cylindrical σ-algebra. The cylindrical σ-algebra is the same as the Borel σ-algebra in finite-dimensional spaces. However, the cylindrical σ-algebra is the natural σ-algebra to consider in infinite-dimensional spaces. These technical considerations do not play an important role in our analysis.

[6] We assume that the functions are measurable with respect to the cylindrical σ-algebra.

and Z are primitives of the model; T is an interval of $\mathbb{R}$, and the sets S and Z each have at least two elements.[7] It follows that a mechanism is defined by

$$M \triangleq \big(\mu, (q, x)\big),$$

and $\mathcal{M}$ denotes the set of all possible mechanisms.

Finally, we note that the signal observed by the agent at t does not depend on the message he sent in previous periods. Adopting this convention serves only to reduce the amount of notation, since all the proofs go through unchanged in any event.

2.3 *The agent's strategies*

We now describe the agent's maximization problem under a mechanism. An agent's strategy specifies the message he sends at each $t \in T$ and the decision to opt out of the mechanism ex post. The set Σ of all valid mappings from signals to messages is defined as

$$\Sigma \triangleq \big\{\sigma : S^T \to Z^T \mid \sigma \text{ is adapted to the natural filtration associated to } \{s_t\}_{t\in T}\big\}.$$

When the signals disclosed by the mechanism are $s \in S^T$ and the agent uses strategy $\sigma \in \Sigma$, the messages he sends are $\sigma(s) \in Z^T$. The condition that σ is adapted to the natural filtration associated to $\{s_t\}_{t\in T}$ means that σ is a measurable function and the message sent by the agent at time t—denoted by σ_t—depends *only* on the history of signals $\{s_{t'}\}_{t'\leq t}$ (i.e., σ is non-anticipating). The set O of all opt-out strategies is defined as the set of all functions,

$$O \triangleq \big\{o : \Theta \times S^T \to \{0, 1\}\big\},$$

where $o(\theta, s) = 0$ (resp. $o(\theta, s) = 1$) means that the agent decides to opt out (resp. opt in) of the mechanism when his type is θ and the signals he observes are $s \in S^T$.[8] The agent's type is an argument in his decision to opt out of the mechanism because he makes this decision after learning his type.[9]

If the agent uses strategy $(\sigma, o) \in \Sigma \times O$, then his expected utility is equal to

$$\mathbb{E}^{\mu}_{(\theta,s)}\big[\big(u\big(q\big(\sigma(s)\big), \theta\big) - x\big(\sigma(s)\big)\big) \cdot o(\theta, s)\big],$$

where $\mathbb{E}^{\mu}_{(\theta,s)}[\cdot]$ denotes the expectation over $(\theta, s) \in \Theta \times S^T$ using measure μ.

2.4 *The principal's maximization problem*

The principal designs a mechanism to maximize profits, and her problem is expressed formally as

$$\Pi \triangleq \max_{M\in\mathcal{M}} \mathbb{E}^{\mu}_{(\theta,s)}\big[\big(x\big(\tilde{\sigma}(s)\big) - c\big(q\big(\tilde{\sigma}(s)\big), \theta\big)\big) \cdot \tilde{o}(\theta, s)\big] \tag{1a}$$

[7]The principal's profits in an optimal mechanism are *independent* of the choice of these sets.

[8]The opt-out strategy must also be measurable with respect to the cylindrical σ-algebra.

[9]Note that the signals the agent observed are a sufficient statistic for the messages he sent, so it is not necessary for the opt-out decision to explicitly condition on the sent messages.

$$\text{s.t. } (\tilde{\sigma}, \tilde{o}) \in \underset{(\sigma,o)\in\Sigma\times O}{\arg\max}\ \mathbb{E}^{\mu}_{(\theta,s)}\big[\big(u\big(q\big(\sigma(s)\big), \theta\big) - x\big(\sigma(s)\big)\big) \cdot o(\theta, s)\big]. \tag{1b}$$

The principal's profits are multiplied by $\tilde{o}(\theta, s)$ in (1a) because these are equal to $(x(\tilde{\sigma}(s)) - c(q(\tilde{\sigma}(s)), \theta))$ when the agent does not opt out of the mechanism and are zero otherwise. We say that a mechanism $M^* \in \mathcal{M}$ is optimal if it solves (1a) and (1b) (hence, referred to simply as (1)).

2.5 *Discussion on the interpretation*

Before we proceed, we interpret our model as a single-unit, indivisible good model. For this, we focus on the parametrization

$$u(q, \theta) = \theta \cdot q \quad \text{and} \quad c(q, \theta) = c_0,$$

where $q \in [0, 1]$ and $c_0 \geq 0$. We interpret the quantity q as the probability that the agent is assigned the good and interpret the transfer x as a payment equal to x/q (resp., zero) whenever the lottery is such that the agent receives (resp., does not receive) the good.[10]

The agent's type θ is interpreted as the agent's expected valuation of the good at the moment he commits to buying the good. In the examples discussed in the Introduction, the interpretation is as follows. In an online purchase (in the European Union), the agent's type represents the buyer's expected valuation of the good after inspecting it but before the withdrawal period ends. In a real estate transaction, the agent's type represents the buyer's expected valuation of the house after inspecting it but before the inspection period ends. In both instances, by the time the withdrawal period ends, the buyer can get an estimate of the utility he would derive from the good based on all available information. Note that the agent's valuation could change after buying the good (e.g., the utility derived from a house may change due to unexpected flooding). However, by the time the agent needs to decide whether (or not) to opt out of the mechanism, his expected valuation of the good conditions on all available information.

From the principal's perspective, the ex post participation constraint is as if she cannot *conceal* information from the agent at the time he agrees to buy the good. The agent's decision to opt out of the mechanism conditions on θ and she does not have any additional information that would change the agent's expected valuation after he already learned θ. Thus, it is as if the principal is obligated to disclose all payoff-relevant information to the agent by the time he makes the final decision to buy the good.

More broadly, $u(q, \theta)$ is interpreted as the agent's expected utility at the moment he commits to buying q units of the good. In the Supplemental Material, we extend our model by allowing the agent's valuation to be affected by additional shocks (i.e., in addition to θ) after the mechanism ends. We show that in this extension, the principal's profit-maximization problem remains unchanged.

[10] In most of the mechanism design literature, an allocation $\{q, x\} \in [0, 1] \times \mathbb{R}$ can be implemented by any lottery that sells the object with total probability q and that requires the agent to pay x in expectation. In models with an ex post participation constraint, the agent pays the principal only when he gets the object.

3. Benchmark: Optimal static mechanism

We first study the optimal static mechanism. In other words, we assume that the agent knows $\theta \in \Theta$. The principal's problem is then[11]

$$\max_{\{q(\theta),x(\theta)\}_{\theta\in\Theta}\in(Q\times\mathbb{R})^{\Theta}} \quad \mathbb{E}_{\theta}^{f}\left[x(\theta)-c\big(q(\theta),\theta\big)\right] \tag{2a}$$

$$\text{subject to} \quad \theta \in \arg\max_{\theta'\in\Theta} u\big(q(\theta'),\theta\big)-x(\theta') \tag{2b}$$

$$\forall\theta\in\Theta,\ u\big(q(\theta),\theta\big)-x(\theta)\geq 0. \tag{2c}$$

This maximization problem is solved by using standard techniques. To simplify the analysis, we assume that

$$\mathcal{S}(q,\theta)-\frac{\big(1-F(\theta)\big)}{f(\theta)}\frac{\partial u(q,\theta)}{\partial\theta}$$

has a unique maximum and that the cross-derivative with respect to q and θ is positive. These two assumptions are not used in any result of this paper other than Lemma 1. We represent the optimal static mechanism by the allocation policy $\{q^c(\theta),x^c(\theta)\}_{\theta\in\Theta}$.

Lemma 1 (Optimal static mechanism). *The optimal static mechanism* $\{q^c(\theta),x^c(\theta)\}_{\theta\in\Theta}$ *is given by*

$$\begin{aligned} q^c(\theta) &= \arg\max_{q\in Q}\mathcal{S}(q,\theta)-\frac{\big(1-F(\theta)\big)}{f(\theta)}\frac{\partial u(q,\theta)}{\partial\theta} \\ x^c(\theta) &= u\big(q^c(\theta),\theta\big)-\int_{\underline{\theta}}^{\theta}\frac{\partial u\big(q^c(s),s\big)}{\partial\theta}\,ds. \end{aligned} \tag{3}$$

Lemma 1 characterizes the optimal static mechanism. It will prove useful that, under this mechanism, an agent's informational rents can be written as

$$U^c(\theta)\triangleq u\big(q^c(\theta),\theta\big)-x^c(\theta)=\int_{\underline{\theta}}^{\theta}\frac{\partial u\big(q^c(s),s\big)}{\partial\theta}\,ds. \tag{4}$$

This expression is the standard construction (using the envelope theorem) of the agent's rents.

4. Upper bound on profits

In this section, we give an upper bound on the principal's profits Π, after which we show that the optimal mechanism achieves the same profits as this upper bound. We start

[11] In (2), the agent is guaranteed nonnegative rents only if he reports truthfully. In other words, if the agent reports a type different than his true type, then he may earn negative rents. So the participation constraint in (2) is more lax than that which we assume in Section 2. However, the analysis of this benchmark remains unchanged if we strengthen the participation constraint by allowing the agent to opt out of the mechanism after he has reported his type. We present the benchmark model as in (2) because this is the standard formulation in the literature.

by constructing a lower bound on the agent's rents and then maximize the total surplus minus that lower bound, which yields an upper bound on the principal's profits. The analysis here demonstrates that the agent can guarantee himself positive rents even in the absence of private information before the mechanism starts. For this reason, the principal faces a trade-off between maximizing total surplus and minimizing the agent's rents.

4.1 *Lower bound on the agent's rents*

We establish a lower bound on the agent's rents by considering a class of simple strategies. Under any mechanism $M \in \mathcal{M}$, there exists a strategy that corresponds to reporting the same messages $z \in Z^T$ regardless of the observed signals. If the agent optimally opts out of the mechanism ex post, then reporting the same messages $z \in Z^T$ yields expected rents

$$\mathbb{E}_{\theta}^{f}\big[\big\{u\big(q(z),\theta\big)-x(z)\big\}^{+}\big]. \tag{5}$$

Here $\{\cdot\}^+$ is the "positive part" function, which accounts for the ex post participation constraint; the agent opts out of the mechanism whenever messages $z \in Z^T$ yield negative rents. The expectation (5) is taken over random variable θ using measure f because the agent is (by construction) reporting the same messages irrespective of the particular signals observed. As a result, the expected rents depend only on the realization of θ.[12]

Of course, reporting the same messages regardless of the observed signals must yield weakly lower rents than an optimal strategy. It follows that, for any mechanism $M \in \mathcal{M}$ and any optimal strategy $(\tilde{\sigma},\tilde{o}) \in \Sigma \times O$,

$$\mathbb{E}_{(\theta,s)}^{\mu}\big[\big(u\big(q\big(\tilde{\sigma}(s)\big),\theta\big)-x\big(\tilde{\sigma}(s)\big)\big)\tilde{o}(\theta,s)\big] \geq \sup_{z\in Z^T} \mathbb{E}_{\theta}^{f}\big[\big\{u\big(q(z),\theta\big)-x(z)\big\}^{+}\big]. \tag{6}$$

Note that the previous lower bound is computed using only the set of quantities and transfers implemented by the mechanism; that is, without reference to the learning process. In other words, we calculate the right-hand side of (6) without referencing μ. Next we describe a bound whose calculation references neither μ nor the transfers $\{x(\tilde{\sigma}(s))\}_{s\in S^T}$.

Before we proceed with our analysis, we simplify the notation by defining

$$\Omega \triangleq \Theta \times S^T.$$

We use $\theta(\omega)$ (resp. $s(\omega)$) to denote the projection of $\omega \in \Omega$ onto $\theta \in \Theta$ (resp. $s \in S^T$) and write more simply $\sigma(\omega) = \sigma(s(\omega))$. The allocation implemented in mechanism (q,x) when the agent sends messages according to $\sigma \in \Sigma$ and he opts out of the mechanism when the allocation yields negative rents, is denoted by

$$\big(q_\sigma(\omega),x_\sigma(\omega)\big) \triangleq \begin{cases} \big(q\big(\sigma(\omega)\big),x\big(\sigma(\omega)\big)\big) & \text{if } u\big(q\big(\sigma(\omega)\big),\theta(\omega)\big)-x\big(\sigma(\omega)\big)\geq 0 \\ (0,0) & \text{otherwise.}\end{cases} \tag{7}$$

[12]The density of the marginal distribution of μ over θ is always equal to f.

The definition of $(q_\sigma(\omega), x_\sigma(\omega))$ takes the value $(0, 0)$ whenever the allocation yields negative rents for the agent because it incorporates that, in this case, $o(\omega) = 0$. In other words, the definition of $\{(q_\sigma(\omega), x_\sigma(\omega))\}_{\omega\in\Omega}$ already reflects the agent's decision to opt out of the mechanism ex post whenever it is optimal for him to do so.

We now fix an optimal strategy $(\tilde{\sigma}, \tilde{o}) \in \Sigma \times O$ and then consider a strategy that reports the exact same messages $\tilde{\sigma}(\omega') \in Z^T$ (i.e., regardless of the observed signals) for some $\omega' \in \Omega$. This strategy yields expected rents

$$\mathbb{E}^f_\theta\big[\big\{u\big(q_{\tilde{\sigma}}\big(\omega'\big), \theta\big) - x_{\tilde{\sigma}}\big(\omega'\big)\big\}^+\big].$$

Of course, an agent who reports "as if" the state were always $\omega' \in \Omega$ captures rents that must be weakly lower than those under the optimal strategy $\tilde{\sigma}$, from which it follows that

$$\mathbb{E}^\mu_{(\theta,s)}\big[\big(u\big(q\big(\tilde{\sigma}(s)\big), \theta\big) - x\big(\tilde{\sigma}(s)\big)\big)\tilde{o}(\theta, s)\big] \geq \sup_{\omega'\in\Omega} \mathbb{E}^f_\theta\big[\big\{u\big(q_{\tilde{\sigma}}\big(\omega'\big), \theta\big) - x_{\tilde{\sigma}}\big(\omega'\big)\big\}^+\big]. \tag{8}$$

In other words, the expected rents under the optimal strategy are weakly greater than reporting as if the state were always $\omega' \in \Omega$—a statement that holds for all $\omega' \in \Omega$. Inequality (8) is the analogue of (6) but considers the messages sent under strategy $\tilde{\sigma}$ for some state ω'.

We can now use the ex post participation constraint to write an inequality that makes no reference to the transfers $\{x_{\tilde{\sigma}}(\omega)\}_{\omega\in\Omega}$. The definition of $(q_\sigma(\omega), x_\sigma(\omega))$ (see (7)) implies that $x_{\tilde{\sigma}}(\omega) \leq u(q_{\tilde{\sigma}}(\omega), \theta(\omega))$ and so, for all $\omega \in \Omega$, we have

$$\mathbb{E}^f_\theta\big[\big\{u\big(q_{\tilde{\sigma}}(\omega), \theta\big) - x_{\tilde{\sigma}}(\omega)\big\}^+\big] \geq \mathbb{E}^f_\theta\big[\big\{u\big(q_{\tilde{\sigma}}(\omega), \theta\big) - u\big(q_{\tilde{\sigma}}(\omega), \theta(\omega)\big)\big\}^+\big]. \tag{9}$$

We use this inequality to bound the agent's rents, as formalized in our next lemma.

Lemma 2 (Bound on the agent's rents). *For any mechanism $M \in \mathcal{M}$ and an optimal strategy $(\tilde{\sigma}, \tilde{o}) \in \Sigma \times O$, the agent's rents under mechanism M satisfy*

$$\begin{aligned}&\mathbb{E}^\mu_{(\theta,s)}\big[\big(u\big(q\big(\tilde{\sigma}(s)\big), \theta\big) - x\big(\tilde{\sigma}(s)\big)\big)\tilde{o}(\theta, s)\big]\\ &\quad\geq \sup_{\omega\in\Omega} \mathbb{E}^f_\theta\big[\big\{u\big(q_{\tilde{\sigma}}(\omega), \theta\big) - u\big(q_{\tilde{\sigma}}(\omega), \theta(\omega)\big)\big\}^+\big].\end{aligned} \tag{10}$$

This lemma establishes a lower bound on the agent's rents. Of particular interest is that this lower bound is computed using only the set of quantities implemented by the mechanism, that is, without reference to either the learning process or the transfers associated with each quantity. In other words, the right-hand side of (10) can be computed using only $\{q_{\tilde{\sigma}}(\omega)\}_{\omega\in\Omega}$ (i.e., neither μ nor $\{x_{\tilde{\sigma}}(\omega)\}_{\omega\in\Omega}$ is required).

Lemma 2 shows that there is a limit to how much rent the principal can extract from the agent, a limit that depends solely on the set of quantities implemented in the mechanism. If the mechanism provides type θ' with quantity $q(\theta')$, then the agent is guaranteed an expected rent of no less than

$$\mathbb{E}^f_\theta\big[\big\{u\big(q\big(\theta'\big), \theta\big) - u\big(q\big(\theta'\big), \theta'\big)\big\}^+\big]. \tag{11}$$

This conclusion shows that the ex post participation constraint allows the agent to extract rents even when, initially, he and the principal were symmetrically informed. We emphasize that this lower bound hinges on the participation constraint being satisfied ex post: the agent's rents would be zero if there were no limit on the losses that the agent can incur ex post (see Section 6.1 for a discussion).

4.2 *Upper bound on the principal's profits*

Lemma 2 gives a lower bound on the agent's rents that depends only on the quantities implemented by the optimal mechanism. In any mechanism, the principal's profits are equal to the total surplus minus the agent's rents. Hence, we can use the lower bound on the agent's rents to compute an upper bound on the principal's profits.

THEOREM 1 (Upper bound on the profits). *The principal's profits satisfy*

$$\Pi \leq \sup_{I\in\mathbb{R}} \sup_{\{q(\theta)\}_{\theta\in\Theta}\in\mathbb{R}^{\Theta}} \{\mathbb{E}_{\theta}^{f}[\mathcal{S}(q(\theta),\theta)] - I\} \tag{12a}$$

$$s.t.\ I \geq \sup_{\theta'\in\Theta} \mathbb{E}_{\theta}^{f}[\{u(q(\theta'),\theta) - u(q(\theta'),\theta')\}^{+}]. \tag{12b}$$

This theorem provides an upper bound on the principal's profits. The upper bound can be found by solving a maximization problem over a function $\{q(\theta)\}_{\theta\in\Theta} \in \mathbb{R}^{\Theta}$ and a scalar $I \in \mathbb{R}$, but with no reference to the transfers or the learning process. This two-step maximization problem can be viewed as first maximizing over the rents left to the agent (which corresponds to maximizing over $I \in \mathbb{R}$) and then maximizing over the quantities implemented subject to the constraint that the lower bound on the agent's rents is less than I. Theorem 1 is proved by (i) writing the principal's profits as the total surplus minus the agent's rents and (ii) replacing the agent's rents with the lower bound characterized in Lemma 2, a procedure that yields an upper bound on the principal's profits.

Theorem 1 illustrates that the principal faces a trade-off between maximizing the total surplus and minimizing the rents left to the agent. Increasing the quantity allocated to type θ enables the principal to increase the total surplus, but it also increases the agent's rents (which must be no less than the term given in (11)). The trade-off between total surplus and the agent's rents follows from the ex post participation constraint, not from the presence of private information before the agent enters the mechanism. Hence, the trade-off differs from that which arises in the optimal *static* mechanism.

The optimal mechanism we construct will implement quantities that solve (12a) and (12b) (henceforth, referred to simply as (12)). Therefore, the properties of the quantities that solve this maximization problem are also satisfied in this optimal mechanism. We revisit the solution to (12)—and give its properties—after first describing the optimal mechanism. The only property of (12) that is used in the construction of this mechanism is that a nondecreasing solution exists (as stated formally in Lemma 4).

5. An optimal mechanism

In this section, we describe a mechanism that achieves the upper bound characterized by Theorem 1 and, therefore, that maximizes profits. Our description consists of three steps. First we formalize the learning process, then we calculate the optimal transfers for some fixed quantities, and finally we characterize the quantities implemented under the optimal mechanism. The second and third steps echo the logic of our analysis in Section 4: after calculating the transfers that minimize the agent's rents for any quantities (for this, we use Lemma 2), we show that the quantities solving (12) also maximize the principal's profits (for this, we use Theorem 1).

5.1 *The learning process*

The learning process of the optimal mechanism is defined as follows.

Definition 1 (Upward disclosure). Upward disclosure is the learning process in which $T=\Theta$ and[13]

$$s_t(\theta)=\begin{cases}1 & \text{if } t=\theta\\ 0 & \text{if } t\neq\theta.\end{cases}$$

Under upward disclosure, the agent learns, at each $t\in T$, whether (or not) his type is $\theta=t$. If the agent's type is $\theta<t$, then the agent learned his type prior to t; if the agent's type is $\theta>t$, then at time t, the agent knows only that his type is strictly higher than t. In other words, at each $t\in T$, the agent either learns his type or learns a lower bound on his type.

5.2 *Incentive-compatible mechanisms*

A strategy is truth-telling if the agent reports the signal he observes. A mechanism is incentive compatible if it is optimal for the agent to report the signals he observes truthfully. In this section, we restrict our attention to incentive-compatible mechanisms.

An incentive-compatible mechanism is defined by a quantity and a transfer for each type $\{q(\theta),x(\theta)\}_{\theta\in\Theta}$. At each $t\in T$, the agent sends a message $z\in Z=\{0,1\}$.[14] If the agent sends message $z=1$ at $t\in T$ and *only* at t, then the principal will implement allocation $\{q(t),x(t)\}$ (provided that it yields the agent nonnegative rents). If the agent sends message $z=1$ either at more than one $t\in T$ or at no $t\in T$, then the principal will implement $q=x=0$. A mechanism $\{q(\theta),x(\theta)\}_{\theta\in\Theta}$ is incentive compatible if and only if

$$\mathbb{E}^f_\theta\big[u\big(q(\theta),\theta\big)-x(\theta)\big]\geq\sup_{(\sigma,o)\in\Sigma\times O}\mathbb{E}^f_{(\theta,s)}\big[\big(u\big(q\big(\sigma(s)\big),\theta\big)-x\big(\sigma(s)\big)\big)o(s,\theta)\big].$$

[13] In setting up the model, we require only that T be an interval; however, we can assume without loss of generality that $T=\Theta$ because it is always possible to relabel T.

[14] In the description of the model, we assume only that Z has at least two elements. However, we can assume $Z=\{0,1\}$ by relabeling the elements and ignoring any extras.

We shall find the incentive-compatible optimal mechanism. Any incentive-compatible mechanism will also give the agent incentives to opt in the mechanism ex post, and so we focus on mechanisms that satisfy $u(q(\theta), \theta) \geq x(\theta)$.

5.3 *Incentive-compatibility constraints*

When the learning process proceeds via upward disclosure, the agent knows either his type or a lower bound on his type. If the agent wants to misreport a type higher than his true type, then he can do so without uncertainty; the time to report the lie is *after* the agent learns his type. Yet, if the agent wants to misreport being a type lower than his true type, then he must do so without knowing the exact realization of his true type because the time to report that lie is *before* the agent learned his type. Our next lemma characterizes the set of constraints that must be satisfied for a mechanism to be incentive compatible.

Lemma 3 (Incentive-compatibility constraints under upward disclosure). *When the learning proceeds by upward disclosure, an allocation* $\{(q(\theta), x(\theta))\}_{\theta\in\Theta}$ *is incentive compatible if and only if the following statements hold:*

$$\forall \theta'' < \theta' \in \Theta, \quad u\big(q\big(\theta'\big), \theta''\big) - x\big(\theta'\big) \leq u\big(q\big(\theta''\big), \theta''\big) - x\big(\theta''\big) \tag{13}$$

$$\forall \theta' \in \Theta, \quad \mathbb{E}_\theta^f\big[u\big(q\big(\theta'\big), \theta\big) - x\big(\theta'\big) | \theta \geq \theta'\big] \leq \mathbb{E}_\theta^f\big[u\big(q(\theta), \theta\big) - x(\theta) | \theta \geq \theta'\big]. \tag{14}$$

Lemma 1 characterizes the incentive-compatibility (IC) constraints that must be satisfied when learning proceeds by upward disclosure. On the one hand, the *upward* IC constraints of (13) are the same as in the static mechanism design problem. This follows because, if an agent of type θ'' wants to report θ' and if $\theta' > \theta''$, then he can do so without uncertainty (as described previously). On the other hand, the principal need not satisfy all the *downward* IC constraints in the design of a static mechanism, for which these constraints are

$$\forall \theta > \theta', \quad u\big(q\big(\theta'\big), \theta\big) - x\big(\theta'\big) \leq u\big(q(\theta), \theta\big) - x(\theta). \tag{15}$$

Constraint (14) can be derived by taking expectations of the previous inequalities with respect to θ conditional on $\theta > \theta'$. It is clear that satisfying (15) is more difficult than satisfying (14). The implication is that there are mechanisms under which (14) is satisfied but (15) is not.

At this point, two questions arise. (i) Why should the principal focus on relaxing the downward IC constraints when she designs the learning process? (ii) How does upward disclosure relax these particular constraints?

To answer the first question, it is useful to consider a mechanism design problem in which the principal must satisfy only the individual rationality constraints. Then the quantities $\{q(\theta)\}_{\theta\in\Theta}$ would be implemented via transfers:

$$x(\theta) = u\big(q(\theta), \theta\big).$$

In other words, the principal would leave the agent with zero rents. Even though all the upward IC constraints (i.e., (13)) are satisfied, the downward IC constraints are violated. Thus, it is only the downward IC constraints that prevent the principal from extracting all the agent's rents. In general, if the principal tries to implement some quantities $\{\tilde{q}(\theta)\}_{\theta\in\Theta}$ using transfers that increase more rapidly than prescribed by the envelope theorem (see (3)), then the downward IC constraints will be violated. So in a static mechanism, what limits the principal's reduction of the agent's rents is the existence of these downward IC constraints.

As for the second question, upward disclosure relaxes downward IC constraints by confounding the local with the global IC constraints that appear in a static mechanism. An agent who is given only a lower bound on his type cannot gauge precisely who to mimic, which confounds local with global IC constraints. In a static incentive-compatible mechanism, the local IC constraints are binding; if the transfers are perturbed by even a small amount, then the agent misreports a type in the vicinity of his true type. However, in a static incentive-compatible mechanism, the global IC constraints are satisfied "with slack": the agent has strict incentives *not* to misreport a type that is far from his true type. If the principal does not allow the agent to tailor his misreporting to his true type, then she can relax the local IC constraints by using the slack in the global IC constraints.

5.4 *Optimal transfers*

We now fix some nondecreasing quantities $\{\tilde{q}(\theta)\}_{\theta\in\Theta}$—possibly different from the corresponding quantities of the optimal mechanism—and find transfers that implement $\{\tilde{q}(\theta)\}_{\theta\in\Theta}$ and so leave the agent with rents equal to

$$\max_{\theta'\in\Theta}\mathbb{E}_{\theta}^{f}\big[\big\{u\big(\tilde{q}\big(\theta'\big),\theta\big)\big)-u\big(\tilde{q}\big(\theta'\big),\theta'\big)\big)\big\}^{+}\big]. \tag{16}$$

These rents are equal to the lower bound described in Lemma 2, which means that these transfers minimize the rents left to the agent across all transfers that implement $\{\tilde{q}(\theta)\}_{\theta\in\Theta}$. The intuition behind the construction is as follows. A mechanism cannot be incentive compatible unless the agent's expected informational rents from reporting truthfully in that mechanism exceed the expected rents from misreporting his type too early. As long as the agent is capturing informational rents, the principal's interests are best served by providing all the informational rents later in the mechanism. This approach maintains, throughout the mechanism, a high level of expected informational rents from engaging with it truthfully. In other words, providing informational rents to a high type relaxes the IC constraint of all lower types because those rents keeps the agent's expected informational rents high. Of course, positive rents cannot be allocated only to the highest type because this would violate the upward IC constraints.

We define a class of transfers in which "low" types are left with zero rents and the transfers of "high" types are constructed (as in a static mechanism) according to the envelope theorem:

$$\tilde{x}(\theta) \triangleq \begin{cases} u\big(\tilde{q}(\theta), \theta\big) & \text{if } \theta \le \tilde{\theta} \\ u\big(\tilde{q}(\theta), \theta\big) - \displaystyle\int_{\tilde{\theta}}^{\theta} \frac{\partial u\big(\tilde{q}(s), s\big)}{\partial \theta}\, ds & \text{if } \theta > \tilde{\theta}. \end{cases} \tag{17a}$$

Under transfers (17a), if the agent's type is below $\tilde{\theta}$, then the transfers are equal to the agent's utility. If the agent's type is above $\tilde{\theta}$, then the transfers are constructed using the envelope theorem (as in (3)), but starting from type $\tilde{\theta}$ instead of from type $\underline{\theta}$.

The transfers (17a) can be defined a priori for any threshold $\tilde{\theta}$. However, we will fix the threshold such that the agent's rents under the truth-telling strategy are equal to (16); this approach corresponds to finding a $\tilde{\theta}$ that satisfies

$$\mathbb{E}_{\theta}^{f}\left[\left\{\int_{\tilde{\theta}}^{\theta} \frac{\partial u\big(\tilde{q}(s), s\big)}{\partial \theta}\, ds\right\}^{+}\right] = \max_{\theta' \in \Theta} \mathbb{E}_{\theta}^{f}\big[\big\{u\big(\tilde{q}(\theta'), \theta\big) - u\big(\tilde{q}(\theta'), \theta'\big)\big\}^{+}\big]. \tag{17b}$$

We remark that there exists a unique $\tilde{\theta}$, which appears only in the integral's limit, that solves (17b) (this claim is proved in the Supplemental Material). The left-hand side of (17b) is the agent's expected rents under truth-telling, which we compute much as in (4), while the right-hand side is the lower bound characterized in Lemma 2. The following proposition states that under the transfers defined by (17a) and (17b) (henceforth, referred to simply as (17)), the mechanism is incentive compatible.

Proposition 1 (Minimal rents). *Let $\{\tilde{q}(\theta)\}_{\theta\in\Theta}$ be nondecreasing and let $\{\tilde{x}(\theta)\}_{\theta\in\Theta}$ be defined as in (17). Then $\{(\tilde{q}(\theta), \tilde{x}(\theta))\}_{\theta\in\Theta}$ is incentive compatible under upward disclosure and*

$$\mathbb{E}_{\theta}^{f}\big[u\big(\tilde{q}(\theta), \theta\big) - \tilde{x}(\theta)\big] = \max_{\theta' \in \Theta} \mathbb{E}_{\theta}^{f}\big[\big\{u\big(\tilde{q}(\theta'), \theta\big) - u\big(\tilde{q}(\theta'), \theta'\big)\big\}^{+}\big]. \tag{18}$$

This proposition shows that any nondecreasing quantities $\{\tilde{q}(\theta)\}_{\theta\in\Theta}$ can be implemented by leaving the agent with the minimum rents characterized in Lemma 2. Because the threshold $\tilde{\theta}$ in (17a) is constructed according to (17b), it is hardly surprising that the mechanism—conditional on being truth-telling—leaves the agent with the minimum rents characterized in Lemma 2. We now offer some intuition concerning just why the quantities and transfers are incentive compatible under an upward disclosure learning process.

It is easy to check that any mechanism under which transfers are constructed according to (17) does not satisfy the downward IC constraints (15) required of a static incentive-compatible mechanism. In fact, every type $\theta < \tilde{\theta}$ captures more rents if he misreports as type $\theta' < \theta$. Under an upward disclosure learning process, however, the applicable downward IC constraints (i.e., (14)) *are* satisfied. Note that an agent may be left with zero rents ex post, but the mechanism is still incentive compatible because the learning process is upward disclosure. So even though an agent may regret (ex post) not misreporting a type lower than his true type, at the moment of reporting that lie, the agent could know only the lower bound on his true type. Hence the agent reports truthfully because he expects thereby to receive high informational rents.

Finally, we explain why the transfer's threshold constructed in (17b) guarantees that the downward IC constraints (14) are satisfied for some fixed type $\theta' < \tilde{\theta}$. On the one hand, the agent's expected rents if he always reports type θ' (i.e., regardless of the signals he observes) are equal to the probability of his type being higher than θ' *multiplied by* his expected rents conditional on his type actually being higher than θ':

$$\mathbb{E}_\theta^f\big[\{u(q(\theta'),\theta)-u(q(\theta'),\theta')\}^+\big]=(1-F(\theta'))\mathbb{E}_\theta^f\big[u(q(\theta'),\theta)-u(q(\theta'),\theta')|\theta\geq\theta'\big]. \quad (19)$$

On the other hand, the agent's expected rents if he reports truthfully are equal to the probability that his type is higher than θ' *multiplied by* his expected rents conditional on his type actually being higher than θ':[15]

$$\mathbb{E}_\theta^f\left[\left\{\int_{\tilde{\theta}}^{\theta}\frac{\partial u(\tilde{q}(s),s)}{\partial\theta}\,ds\right\}^+\right]=(1-F(\theta'))\mathbb{E}_\theta^f\left[\left\{\int_{\tilde{\theta}}^{\theta}\frac{\partial u(\tilde{q}(s),s)}{\partial\theta}\,ds\right\}^+\middle|\theta>\theta'\right]. \quad (20)$$

It is easy to see that (17b) implies (19) is smaller than (20) for all $\theta' < \tilde{\theta}$. If we now simplify the $(1 - F(\theta'))$ term on the right-hand side of each equation, then (17b) implies also that the expected informational rents conditional on being an agent of type higher than θ' *exceed* the expected rents of misreporting type θ'. It follows that the downward IC constraint for θ' is satisfied; that is, (14) is satisfied for $\theta' < \tilde{\theta}$.

5.5 *Optimal mechanism*

We now construct the optimal mechanism, under which the principal's profits will be as given in (12). Here we show that a nondecreasing solution to (12), when combined with transfers constructed as in Proposition 1, constitute (under upward disclosure) an optimal mechanism. In Section 5.6, we show that a nondecreasing solution to (12) always exists.

Theorem 2 (Optimal mechanism). *Let $\{q^*(\theta)\}_{\theta\in\Theta}$ be a nondecreasing solution to (12) and let $\{x^*(\theta)\}_{\theta\in\Theta}$ be constructed as in (17). Then $\{(q^*(\theta), x^*(\theta))\}_{\theta\in\Theta}$, as implemented via upward disclosure, is an optimal mechanism and yields profits equal to (12).*

Theorem 2 characterizes an optimal mechanism. As explained previously, there is a natural sense in which upward disclosure relaxes the relevant IC constraints; the transfers constructed as in (17) leverage these relaxed constraints to minimize the agent's rents. The intuition underlying the trade-offs faced by the principal are analogous to those arising under the optimal static mechanism: providing the agent with higher quantities of the good increases not only the social surplus, but also the agent's rents. Yet here the agent's rents differ from those under a static mechanism, which ultimately leads to different quantities and different transfers in this optimal mechanism. The following sections explore the optimal mechanism in more detail.

[15] This equality follows because the agent earns positive rents only if his type is higher than θ'.

5.6 *Properties of the optimal mechanism*

According to Theorem 2, there exists an optimal mechanism under which the quantities implemented are a solution to (12). These quantities are easy to find because, for a fixed I, constraint (12b) is a pointwise constraint. In other words, if I is fixed, then one can see whether the quantities $\{q(\theta)\}_{\theta\in\Theta}$ satisfy (12b) by checking each individual value of $q(\theta)$ independently. For a fixed I, then, it is optimal to set $q(\theta)$ equal to the surplus-maximizing quantity $q^s(\theta)$ whenever doing so does not violate (12b); otherwise, it is optimal to find the maximum $q(\theta)$ such that (12b) is satisfied.

To characterize the solution to (12), we first define

$$\hat{q}_I(\theta') \triangleq \max\{q \in Q : \mathbb{E}^f_\theta[\{u(q,\theta) - u(q,\theta')\}^+] \le I\}. \tag{21}$$

That is to say, $\hat{q}_I(\theta')$ is the maximum quantity that can be allocated to type θ' such that (12b) is not violated. For a fixed I, the optimal quantities are given by

$$q_I(\theta) \triangleq \min\{q^s(\theta), \hat{q}_I(\theta)\}. \tag{22}$$

In the next lemma, we use $q_I(\theta)$ to find a simple characterization of (12).

Lemma 4 (Equivalent formulation of maximization). *The maximization problem (12) can be written as*

$$\max_{I\in\mathbb{R}}\{\mathbb{E}^f_\theta[\mathcal{S}(q_I(\theta),\theta)] - I\}. \tag{23}$$

Moreover, for all $I > 0$, (a) the quantity $\{q_I(\theta)\}_{\theta\in\Theta}$ is continuous and nondecreasing, and (b) there exists a threshold $\hat{\theta}_I < \bar{\theta}$ such that $q_I(\theta) = q^s(\theta)$ for all $\theta > \hat{\theta}_I$.

This lemma offers a compact characterization of the optimal solution to the problem (12). Because $q_I(\theta)$ is nondecreasing, Lemma 4 establishes that a nondecreasing solution to this problem exists and that it can be used to construct (per Theorem 2) an optimal mechanism. The solution to (12) has the property that "high-enough" types are always allocated the surplus-maximizing quantity. At the same time, in the optimal mechanism, "low-enough" types are left with zero informational rents. In fact, the optimal mechanism has the property that every type is either assigned the surplus-maximizing quantity or is left with zero rents ex post.

Proposition 2 (Property of the optimal mechanism). *Let $\{q^*(\theta)\}_{\theta\in\Theta}$ be a nondecreasing solution to (12) and let $\{x^*(\theta)\}_{\theta\in\Theta}$ be given by (17). Then*

$$\forall\theta\in\Theta, \quad q^*(\theta) = q^s(\theta) \quad \textit{or} \quad x^*(\theta) = u(q^*(\theta),\theta). \tag{24}$$

Proposition 2 states that every type is either allocated his surplus-maximizing quantity or is left with zero rents ex post. This statement clearly illustrates the basic intuitions behind the optimal mechanism. High types are given rents, so there is no need to distort their allocation; in contrast, allocation to low types is distorted and they are left with zero rents. Observe that an agent's allocation is never greater than the efficient quantity. The reason for this is that an agent's informational rents are increasing in the quantity allocated to him, so giving the agent a quantity higher than the surplus-maximizing quantity decreases the total surplus and also increases the agent's rents.

5.7 *Uniqueness of the optimal mechanism*

We now discuss the uniqueness of the optimal mechanism. Our results in this section rely on the assumptions (i.e., in addition to those given in Section 2)

$$\frac{\partial^2 S(q,\theta)}{\partial q^2} \leq 0 \quad \text{and} \quad \frac{\partial^3 u(q,\theta)}{\partial q^2 \partial \theta} \geq 0. \tag{25}$$

These assumptions guarantee that (12) has a unique solution.

Lemma 5 (Uniqueness of solution to (12)). *If (25) is satisfied, then there exists a unique set of quantities* $\{q^*(\theta)\}_{\theta\in\Theta}$*, up to a set of measure* 0*, that solve (12).*

This lemma guarantees that the solution to (12) is unique. Because Lemma 4 guarantees the existence of a nondecreasing solution to (12), we henceforth assume that $\{q^*(\theta)\}_{\theta\in\Theta}$ is a nondecreasing solution—in essence, the unique solution—to (12).

Our next proposition states that in every optimal mechanism, the quantities implemented are the same; also, in every optimal mechanism, types that are allocated a socially inefficient quantity are left with zero rents.

Proposition 3 (Uniqueness of the optimal mechanism). *If (25) is satisfied, then, in every optimal mechanism and associated optimal strategy* $(M^*, (\sigma^*, o^*))$ *for* μ*-a.e.* $(\theta, s) \in \Theta \times S$*, the implemented quantities and transfers satisfy*

$$q\big(\sigma^*(s)\big)\cdot o^*(s,\theta) = q^*(\theta) \tag{26}$$

$$\text{if } q^*(\theta) < q^s(\theta), \quad x\big(\sigma^*(s)\big)\cdot o^*(s,\theta) = u\big(q^*(\theta),\theta\big). \tag{27}$$

Here $q^*(\theta)$ *is the unique solution to the maximization problem (12).*

According to Proposition 3, the quantities implemented by any optimal mechanism, which can be found by solving (12), are essentially unique. It is immediate that every optimal mechanism induces the same surplus and the same expected rents for the agent. In addition, transfers are uniquely determined for types that are allocated an inefficient quantity. Note that a direct corollary of this proposition is that Proposition 2 holds across all optimal mechanisms.

Proposition 3 makes no reference to the transfers of types that are allocated a socially efficient quantity. Transfers associated with types that are assigned the efficient quantity are not, in general, uniquely determined across all optimal mechanisms (see Supplemental Material B for a counterexample). Yet in a model with linear utility and zero marginal costs (i.e., when $u(q, \theta) = q \cdot \theta$ and $c(q, \theta) = 0$), the transfers in every optimal mechanism are uniquely determined.[16]

Like transfers, information disclosure is not uniquely defined across all optimal mechanisms (a counterexample is given in Supplemental Material B). Although the information disclosure is not unique, it must satisfy some constraints. Consider two types $\theta > \theta'$ such that both are allocated a socially inefficient quantity (i.e., $q^*(\theta) < q^s(\theta)$ and $q^*(\theta') < q^s(\theta')$). Proposition 3 shows that both of these types receive zero rents ex post in every optimal mechanism. However, an agent of type θ can usually earn positive rents by (falsely) reporting that he is of type θ'. Therefore, an agent who learns he is of type θ must not be allowed to report being type θ'; it follows that to satisfy the IC constraints, type θ' must learn and report his type *before* type θ. So even though the learning process is not unique, there is a natural sense in which it must have a similar structure to that of upward disclosure.

Social surplus and agent's rents An immediate implication of Proposition 3 is that every optimal mechanism induces the same social surplus and agent's rents. It is thus natural to compare these quantities with those induced by the optimal *static* mechanism. It is possible to find examples of the surplus under the optimal mechanism being larger (or smaller) than under the optimal static mechanism. Some of these examples are discussed in Section 5.8. In those examples, the agent's rents in the optimal mechanism end up being smaller than his rents in the optimal static mechanism. We were not able to prove, or to disprove via counterexample, that the agent's rents are always less in the optimal mechanism than in the optimal static mechanism.

5.8 *Examples*

This section gives two examples that illustrate the optimal mechanism. In the first example, we also explain how we find the optimal mechanism.

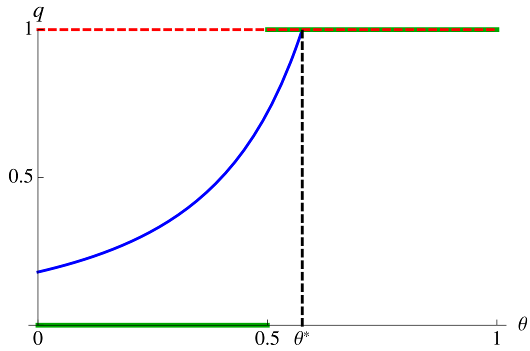
5.8.1 *Indivisible good model* We study the example

$$u(q, \theta) = \theta \cdot q \quad \text{and} \quad c(q, \theta) = 0.$$

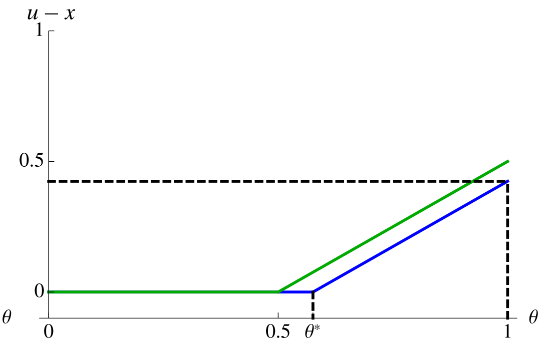
The space of types and quantities is $\Theta = Q = [0, 1]$, and the prior distribution over types is uniform (i.e., $f = 1$).

In Figure 1(a), we plot the quantities under the optimal mechanism ($\{q^*(\theta)\}_{\theta\in\Theta}$) and under the optimal static mechanism ($\{\hat{q}(\theta)\}_{\theta\in\Theta}$) as well as the surplus-maximizing quantities ($\{q^s(\theta)\}_{\theta\in\Theta}$). There are types under which $\{q^*(\theta)\}_{\theta\in\Theta}$ assigns the object to the agent with probability less than 1 even though the optimal static mechanism would assign the object to the agent with probability 1. However, there are also types that are

[16]The intuition here is that, in this model, the efficient allocation is $q^s(\theta) = 1$ for all $\theta \in \Theta$. In addition, incentive compatibility dictates that there exists a unique transfer associated with the quantity $q = 1$ and that this transfer must equal the utility of the *lowest* type that is assigned an efficient quantity. The reason is that the transfers associated with all quantities $q(\theta) < 1$ are designed to leave the agent with zero rents ex post and so, by continuity, the transfer associated with $q = 1$ must likewise leave zero rents for the lowest type to be assigned $q = 1$.



(a) Quantities



(b) Agent's Rents

FIGURE 1. Indivisible good model. In blue is plotted the optimal mechanism ($\{q^*(\theta), x^*(\theta)\}_{\theta\in\Theta}$), in green, the optimal static mechanism ($\{q^s(\theta), x^s(\theta)\}_{\theta\in\Theta}$), and in red, the surplus-maximizing quantity ($q^s(\theta)$). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

assigned the object with probability 0 under the optimal static mechanism yet are assigned the object with positive probability under the optimal dynamic mechanism. We can check numerically that the dynamic mechanism generates a higher expected surplus.

A remarkable property of the optimal dynamic mechanism is that there is always a positive probability of the object being sold to the agent. A sharp intuition can be gained by analyzing the agent's decision about whether to misreport, at the beginning of the mechanism, that his type is $\theta = 0$. The agent can obtain the object free of charge ($x^*(0) = 0$) with probability $q^*(0) \approx 0.2$ if he immediately reports his type to be $\theta = 0$. Nevertheless, the agent reports truthfully that his type is not $\theta = 0$ because his expected rents under truthful reporting are (weakly) greater than the rents from misreporting his type (in fact, the agent is indifferent between these options). It is clear that allowing the agent to win the object with small probability but at an extremely low price when his type is very low does not necessarily increase the agent's rents. Formally, this situation is captured in the pointwise nature of the constraint in the relaxed problem (i.e., in (12b)). In other words, since the agent is provided with some rents I in expectation, low types can be assigned the object with positive probability *without* having to increase those expected rents I.

In Figure 1(b), we plot the rents left to the agent under the optimal mechanism ($\{\theta \cdot q^*(\theta) - x^*(\theta)\}_{\theta\in\Theta}$) and under the optimal static mechanism ($\{\theta \cdot q^s(\theta) - x^s(\theta)\}_{\theta\in\Theta}$). The graphs illustrate the properties of the optimal mechanism stated in Proposition 2, which include that the agent earns strictly positive rents only when he receives the object with probability 1 (this is the efficient allocation). We can make the following sharp comparison: for every type, the agent's rents are (weakly) lower under the optimal dynamic mechanism than under the optimal static mechanism.

In this example, the optimal mechanism implements a higher social surplus than does the static mechanism. That property does not hold in general, however. For instance, if the distribution of the types is uniform in $[1, 2]$, then the optimal static mechanism implements the efficient allocation for all types, whereas the optimal mechanism will not (this claim can be verified numerically). Hence it is clear that, in general, one cannot know ex ante whether the optimal mechanism will induce a higher or a lower surplus than the optimal static mechanism. (Along these lines, recall that we were unable to prove or disprove that the optimal mechanism *always* yields lower rents to the agent than the optimal static mechanism.)

Construction of the optimal mechanism We use this example to illustrate how to find the optimal mechanism. If the utility function of the agent is linear (i.e., if $u(q, \theta) = \theta \cdot q$), then $\hat{q}_I(\theta)$ (as defined in (21)) can be computed in closed form as

$$\hat{q}_I(\theta) = \frac{I}{(1 - F(\theta))\mathbb{E}^f_{\theta'}[\theta' - \theta | \theta' > \theta]} = \frac{I}{(1-\theta)\dfrac{1-\theta}{2}}. \tag{28}$$

Here the second equality is implied by the assumption that the types are uniformly distributed in $[0, 1]$. Now we can use (28) to write (22) explicitly:

$$q_I(\theta) = \begin{cases} 1 & \text{if } \theta > 1 - \sqrt{2 \cdot I} \\ \dfrac{2I}{(1-\theta)^2} & \text{if } \theta \leq 1 - \sqrt{2 \cdot I}. \end{cases}$$

Note that $I \leq 1/2$ in this example because the expected valuation of the agent is $1/2$; hence the agent's expected rents never exceed $1/2$. The quantities in the optimal mechanism are found by solving (23), which corresponds to

$$\max_{I \leq 1/2} \mathbb{E}^f_\theta[\theta \cdot q_I(\theta)] - I.$$

Given the optimal I^* that solves the previous maximization problem, we can use (24) and compute the transfers

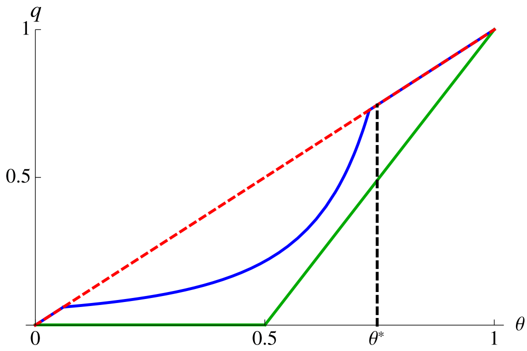
$$x^*(\theta) = \begin{cases} 1 - \sqrt{2 \cdot I^*} & \text{if } \theta > 1 - \sqrt{2 \cdot I^*} \\ q^*(\theta) \cdot \theta & \text{if } \theta \leq 1 - \sqrt{2 \cdot I^*}. \end{cases}$$

This completes the description of the optimal mechanism.

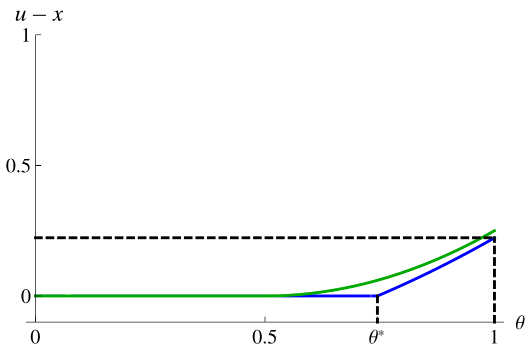
5.8.2 *Quadratic cost* Next we study the example

$$u(q, \theta) = \theta \cdot q \quad \text{and} \quad c(q, \theta) = q^2/2.$$

As before, the space of types and quantities is $\Theta = Q = [0, 1]$ and the prior distribution over types is uniform ($f = 1$). The optimal screening policy is constructed analogously to the construction of the indivisible good example, but here we find the solution numerically.



(a) Quantities



(b) Agent's Rents

FIGURE 2. Quadratic cost model. In blue is plotted the optimal mechanism $(\{q^*(\theta), x^*(\theta)\}_{\theta\in\Theta})$, in green, the optimal static mechanism $(\{q^s(\theta), x^s(\theta)\}_{\theta\in\Theta})$, and in red, the surplus-maximizing quantity $(q^s(\theta))$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

As in Figure 1(a), in Figure 2(a) we plot the quantities under the optimal mechanism and under the optimal static mechanism as well as the surplus-maximizing quantities, and just as in Figure 1(b), in Figure 2(b) we plot the rents left to the agent under the optimal mechanism and under the optimal static mechanism. We can see that sufficiently high types receive the surplus-maximizing quantity while some intermediate types receive less than the surplus-maximizing quantity. The optimal quantities are equal to the surplus-maximizing quantity even for some types below the threshold; for these types, the IC constraint does not bind. We can also see that $\{q^*(\theta)\}_{\theta\in\Theta}$ creates more surplus than the optimal static mechanism for all $\theta \in \Theta$. Finally, the optimal allocation $\{q^*(\theta), x^*(\theta)\}_{\theta\in\Theta}$ leaves the agent with less informational rents.

6. Discussion and literature review

In this section, we discuss the results in this paper and relate them to the literature. Table 1 tabulates some classic papers, grouping them in terms of their assumptions on the timing of the participation constraint and on the principal's ability to control the agent's learning process. Our paper is the first to study optimal sequential screening when the principal designs the learning process and the participation constraint must be satisfied ex post.

6.1 *Role of the participation constraint*

One of the key assumptions of our model is that the principal must satisfy the agent's participation constraint ex post. We now explain how different assumptions about the participation constraint shape the agent's ability to gain informational rents and the principal's trade-offs when she designs a profit-maximizing mechanism.

Timing of participation	Learning process is exogenous	Principal can control learning process
Ex ante	Harris and Raviv (1978)	
Interim	Courty and Hao (2000)	Esö and Szentes (2007)
Posterior		Bergemann and Wambach (2015)
Ex post	Krähmer and Strausz (2015)	*This paper*

Table 1. Literature review.

Ex ante participation constraint An ex ante participation constraint is evaluated before the agent learns any private information. If the participation constraint must be satisfied ex ante, then the optimal mechanism implements the surplus-maximizing quantity and the agent pays a transfer equal to his ex ante expected utility. In that case, the mechanism is efficient, but the agent is left with zero rents. This problem was studied by Harris and Raviv (1978), among others.[17]

Interim participation constraint The interim participation constraint is evaluated *before* the agent enters the mechanism but *after* he observes some private information. Because the agent in our model has no private information before entering the mechanism, relaxing the ex post participation constraint to an interim constraint is equivalent to relaxing the former to an ex ante constraint.[18] Courty and Hao (2000) study the optimal mechanism under an interim participation constraint and when the principal cannot control the agent's learning process.

Esö and Szentes (2007) characterize the optimal mechanism when satisfying an interim participation constraint and when the principal *can* design the information disclosure. They show that, in the optimal mechanism, the principal can optimize "as if" she could observe all the additional information that the agent observes. The implications are that (i) the principal does not benefit from controlling the information disclosure under an interim participation constraint and (ii) the agent's informational rents result only from the private information he observes before entering the mechanism. Neither of these properties holds when the participation constraint is satisfied ex post.

More recently, Li and Shi (2017) show that the principal can do better than the mechanism proposed by Esö and Szentes (2007) if she uses discriminatory disclosure policies. This shows that by studying discriminatory disclosure policies, the principal can reduce the agent's informational rents. However, those rents are still solely attributable to the agent's private information before entering the mechanism.

[17]The nature of the solution is similar to that of Gershkov (2002) in a multi-agent context and also to the optimal regulation of a natural monopoly, which suggests the ex ante sale of all future rents (see Demsetz 1968 and Loeb and Magat 1979).

[18]Under an interim participation constraint, the agent (usually) observes some private information before entering the mechanism; otherwise, this constraint is indistinguishable from the ex ante participation constraint.

Posterior participation constraint A posterior participation constraint is evaluated at the end of the mechanism but before the agent learns his valuation.[19] If the principal were to maximize her profits subject only to posterior participation constraint, then her maximization problem would be as in (1), but the agent's decision to opt out of the mechanism would be determined by a function $o : S^T \to \{0, 1\}$. In other words, the agent's decision to opt out of the mechanism does not condition explicitly on the agent's type.[20] If we relax the participation constraint to posterior and focus on a model with linear utility and zero cost (i.e., with $u(q, \theta) = q \cdot \theta$ and $c(q, \theta) = 0$), then the optimal mechanism discloses no information and sells the good to the agent at his (ex ante) expected valuation.[21]

The conceptual difference between ex post and posterior participation constraints is that under the latter, the principal is allowed to conceal information from the agent at the time he must decide whether (or not) to buy the good.[22] A posterior participation constraint does not adequately capture consumer protection laws that allow a buyer to inspect the good before committing to buying it. These laws prevent the seller from concealing any substantive amount of information about the good before the buyer commits to purchase it, and so a posterior participation constraint is too lax to model these situations appropriately.

Bergemann and Wambach (2015) study the optimal mechanism when there are multiple buyers who have private information about their own valuation before entering the mechanism, the principal can control the buyers' learning process, and she must satisfy the posterior participation constraint. They show that, provided the principal can control the information disclosure, the allocation implemented by Esö and Szentes (2007) can also be implemented when the participation constraint is strengthened to a posterior participation constraint. Thus, there is no loss in strengthening the participation constraint when the principal can control the learning process. Since Bergemann and Wambach (2015) implement the same allocation as Esö and Szentes (2007), it clearly follows that the source of informational rents, as well as the trade-off between minimizing those rents and maximizing total surplus, is the same under a posterior participation constraint as under an interim constraint.

[19] The notion of posterior implementation was first proposed by Green and Laffont (1987).

[20] The participation constraint studied by Bergemann and Wambach (2015) is slightly weaker than that which we present here: they require that the agent's expected utility is nonnegative after every history if the agent reports truthfully. In other words, they require the agent to get nonnegative expected rents under his optimal strategy, but he is not guaranteed nonnegative rents if he deviates. We give a slightly stronger version of the participation constraints because it is simpler to make the connection with the notation in our paper, and the optimal mechanism found by Bergemann and Wambach (2015) can be implemented under both versions of the posterior participation constraint.

[21] Bergemann and Wambach (2015) study only an indivisible good model; hence, there is no clear identification of the optimal mechanism under a posterior participation constraint with other utility and cost functions (i.e., when $u(q, \theta) \neq q \cdot \theta$ or a nonconstant $c(q, \theta)$). If the cost is not zero, then the principal discloses to the agent only whether his valuation is above the cost. As before, the principal can extract the full surplus.

[22] Bergemann and Wambach (2015, p. 1077) explain it as follows: "These notions of posterior constraints were first introduced by Green and Laffont (1987) to reflect the possibility that the mechanism may reveal some, but not necessarily all, payoff-relevant information to the agents."

The analysis of our model differs in important ways from the analysis in Bergemann and Wambach (2015). Under an ex post participation constraint, the optimal mechanism does not implement quantities or transfers that equal some benchmark in the literature; hence we must construct a "relaxed" problem and then show that the optimal mechanism achieves the same solution. Our approach has the further benefit of allowing for more general preferences than those that are accommodated by much of the literature. In particular, we allow the agent to have a general utility function and allow the principal to have a general cost function. While the analysis in our paper differs from that in Bergemann and Wambach (2015), the learning process in the optimal mechanism in our paper is very similar to that which appears in their paper. In the following section, we explain the similarities between our optimal mechanism and theirs, while relating them also to other papers in the literature.

Ex post participation constraint An ex post participation constraint is evaluated using the agent's type θ. This constraint amounts to a strengthening of the posterior participation constraint because the ex post constraint is evaluated using the agent's realized type. It can also be viewed as a constraint that forces the principal to disclose all available information to the agent *before* he decides whether (or not) to opt out of the mechanism. This strengthening generates fundamental differences between the ex post constraint and all weaker forms of the participation constraint.

We show that under the optimal mechanism with an ex post participation constraint, the agent can gain informational rents and the quantities that the principal allocates to the agent are distorted to reduce those rents. As previously explained, if we relax the participation constraint to posterior and focus on a model with linear utility and constant marginal cost, then the allocation implemented would be efficient and the agent would gain zero rents. Thus, we conclude that the informational rents attainable by the agent under an ex post participation constraint would not be attainable if the participation constraint was weakened to a posterior constraint (or any further weakening). These rents generate new trade-offs for the principal as she balances the maximization of surplus and the reduction of informational rents.

In our model, the principal must leave the agent with nonnegative rents ex post. Yet one can imagine cases where the agent could incur limited but nonzero losses ex post. Imposing limited liability conditions that allow for some (but limited) losses would allow for the study of models that smoothly connect the ex post participation constraint with the ex ante participation constraint. Doing so would help us understand how strength of the participation constraint affects the optimal mechanisms in sequential screening and the agent's ability to earn informational rents. Sappington (1983) studies a model in which the principal and the agent are initially symmetrically informed and the principal must satisfy a limited liability restriction. The class of limited liability constraints studied therein is more general than those allowed in our paper; however, the agent learns his type immediately after entering the mechanism (i.e., the principal has *no* control over the learning process). We believe that studying a model in which the principal can control the learning process—and in which the agent can incur limited but nonzero losses ex post—would be a fruitful avenue for future work to explore.

Finally, there is an alternative version of the ex post participation constraint in which the agent is guaranteed nonnegative rents if he reports truthfully, but he may get negative rents if he deviates from the truth-telling strategy. This alternative formulation of the ex post participation constraint would be more suitable to study problems in which the principal wants to guarantee the agent nonnegative rents ex post, but the agent is not allowed to opt out of the mechanism ex post. We give a discussion of this alternative version of the ex post participation constraint in the Supplemental Material.

6.2 *Learning process*

Under the optimal mechanism (characterized in Theorem 1), the learning process is upward disclosure; this process is similar to other disclosure policies that have appeared in the literature. The optimal mechanism of Bergemann and Wambach (2015) features a learning process similar to upward disclosure; the only difference is that their learning process stops before each agent learns his type. Thus, the learning process does not disclose all available information to the agents. This difference allows the principal to leave the agent with ex post losses.[23]

Ata and Dana (2015) study the profit-maximizing mechanism in a model in which an agent has private information about the distribution of his valuations at the outset and the participation constraint must be satisfied interim. The learning proceeds in continuous time (much as in our model), where agents with different interim expected valuations learn at different times. They specify conditions under which the principal can implement the surplus-maximizing quantity and extract all of the agent's rents, while also characterizing her second-best strategy whenever the first-best is not possible (see also Akan et al. 2015).

The reason why the learning process in our optimal mechanism is similar to those that have appeared in the literature is that these learning processes relax the downward incentive constraints. A similar intuition underlies the optimal mechanism of our model. Our paper helps to explain how sequential screening relaxes the downward incentive constraints.

6.3 *Initial private information*

Krähmer and Strausz (2015) study a model of sequential screening in which the agent has private information before entering the mechanism, the principal has no control over the agent's learning process, and the participation constraint must be satisfied ex post. They show that, for a wide range of (exogenous) learning processes, sequential screening does not yield higher profits than the optimal static mechanism. They interpret this result to mean that imposing an ex post participation constraint protects the agent because it increases his informational rents.

[23] In Bergemann and Wambach (2015), the learning process is as in upward disclosure, but it sometimes stops before the agent learns his type. Because the learning process does not always disclose all the information to the agent, the principal can leave the agent with losses ex post and still satisfy the posterior participation constraint. Bergemann and Wambach (2015) describe a model that identifies particular agent preferences and object characteristics that allow the principal to implement upward disclosure as a learning process.

Our results are consistent with their interpretation of the ex post participation constraint: in our model, the agent earns positive informational rents only if there is no weakening in the timing of the participation constraint (in which case his informational rents would be zero). Although our results support the notion that the ex post participation constraint increases the agent's rents, we also show that this constraint does not preclude the possibility of profiting from sequential screening (as also demonstrated in a multi-unit model by Krähmer and Strausz 2016). In sum, the rents that an agent can guarantee himself under an ex post participation constraint are greater than zero but less than his informational rents under a static mechanism.

The agent in our model learns his type in continuous time, which maximizes the principal's capacity to screen the agent. Yet even with a two-stage information disclosure (as in Krähmer and Strausz 2015), the principal can do better than in the optimal static mechanism.[24] Bergemann et al. (2018) analyze two-stage disclosure policies with ex post participation constraints when the principal cannot control the agent's learning process. Our paper (indirectly) gives an upper bound on the principal's profits under any mechanism in which the participation constraint must be satisfied ex post.

Appendix: Proofs

The proof of Lemma 1 is well known in the literature (see, for example, Baron and Myerson 1982).

Proof of Lemma 2. Let $(\tilde{\sigma}, \tilde{o}) \in \Sigma \times O$ be an optimal strategy in mechanism M. Then

$$\forall \sigma \in \Sigma, \quad \mathbb{E}_{\omega}^{\mu}\big[u\big(q_{\tilde{\sigma}}(\omega), \theta(\omega)\big) - x_{\tilde{\sigma}}(\omega)\big] \geq \mathbb{E}_{\omega}^{\mu}\big[u\big(q_{\sigma}(\omega), \theta(\omega)\big) - x_{\sigma}(\omega)\big].$$

For all $\omega' \in \Omega$, there exists $\sigma' \in \Sigma$ such that

$$\big(q_{\sigma'}(\omega), x_{\sigma'}(\omega)\big) = \begin{cases} \big(q_{\tilde{\sigma}}(\omega'), x_{\tilde{\sigma}}(\omega')\big) & \text{if } u\big(q_{\tilde{\sigma}}(\omega'), \theta\big) - x_{\tilde{\sigma}}(\omega') \geq 0 \\ (0, 0) & \text{otherwise.} \end{cases}$$

That is, there exists a strategy $\sigma' \in \Sigma$ that yields the same outcomes as strategy $\tilde{\sigma}$ when the agent reports as if the state is ω' regardless of the observed signals.[25] The expected rents under strategy $\sigma' \in \Sigma$ are

$$\mathbb{E}_{\omega}^{\mu}\big[u\big(q_{\sigma'}(\omega), \theta(\omega)\big) - x_{\sigma'}(\omega)\big] = \mathbb{E}_{\theta}^{f}\big[\big\{u\big(q_{\tilde{\sigma}}(\omega'), \theta\big) - x_{\tilde{\sigma}}(\omega')\big\}^{+}\big].$$

Thus, for all $\omega' \in \Omega$,

$$\mathbb{E}_{\omega}^{\mu}\big[u\big(q_{\tilde{\sigma}}(\omega), \theta(\omega)\big) - x_{\tilde{\sigma}}(\omega)\big] \geq \mathbb{E}_{\theta}^{f}\big[\big\{u\big(q_{\tilde{\sigma}}(\omega'), \theta\big) - x_{\tilde{\sigma}}(\omega')\big\}^{+}\big]. \tag{29}$$

The result follows from (29) and (9). □

[24] In Supplemental Material B, we give an example that uses the intuition of upward disclosure but in a two-stage mechanism. The same example is cited by Krähmer and Strausz (2015), who provide a more detailed discussion.

[25] It is sufficient to define $\sigma'_t(\{s_{t'}\}_{t' \leq t}) = \tilde{\sigma}_t(\omega')$.

Proof of Theorem 1. Let $M \in \mathcal{M}$ be any mechanism and let $(\tilde{\sigma}, \tilde{o}) \in \Sigma \times O$ be an optimal strategy in this mechanism

$$(\tilde{\sigma}, \tilde{o}) \in \underset{(\sigma', o') \in \Sigma \times O}{\arg\max}\ \mathbb{E}^{\mu}_{(\theta,s)}\big[\big(u\big(q\big(\sigma'(s)\big), \theta\big) - x\big(\sigma'(s)\big)\big)o'(\theta, s)\big].$$

The principal's profits in this mechanism are equal to the total surplus minus the agent's rents:

$$\begin{aligned}\pi^M &\triangleq \mathbb{E}^{\mu}_{(\theta,s)}\big[\big(x\big(\tilde{\sigma}(s)\big) - c\big(q\big(\tilde{\sigma}(s)\big), \theta\big)\big)\tilde{o}(\theta, s)\big] \\ &= \mathbb{E}^{\mu}_{(\theta,s)}\big[\mathcal{S}\big(q\big(\tilde{\sigma}(s)\big), \theta\big)\tilde{o}(\theta, s) - \big(u\big(q\big(\tilde{\sigma}(s)\big), \theta\big) - x\big(\tilde{\sigma}(s)\big)\big)\tilde{o}(\theta, s)\big].\end{aligned}$$

Here π^M is the principal's profits in mechanism M, possibly different than the optimal mechanism. Using the lower bound on the agent's rents (see Lemma 2),

$$\begin{aligned}\pi^M \le{}& \mathbb{E}^{\mu}_{(\theta,s)}\big[\mathcal{S}\big(q\big(\tilde{\sigma}(s)\big), \theta\big)\tilde{o}(\theta, s)\big] \\ &- \sup_{\omega' \in \Omega} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q_{\tilde{\sigma}}(\omega'), \theta\big) - u\big(q_{\tilde{\sigma}}(\omega'), \theta(\omega')\big)\big\}^{+}\big].\end{aligned}$$

Because $q(\tilde{\sigma}(\omega)) \neq q_{\tilde{\sigma}}(\omega)$ only when $\tilde{o}(\theta, s) = 0$, the previous inequality can be written as

$$\begin{aligned}\pi^M \le{}& \mathbb{E}^{\mu}_{\omega}\big[\mathcal{S}\big(q_{\tilde{\sigma}}(\omega), \theta(\omega)\big)\tilde{o}(\omega)\big] \\ &- \sup_{\omega' \in \Omega} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q_{\tilde{\sigma}}(\omega, \theta)\big) - u\big(q_{\tilde{\sigma}}(\omega'), \theta(\omega')\big)\big\}^{+}\big].\end{aligned}$$

The right-hand side of this inequality is computed using quantities $\{q_{\tilde{\sigma}}(\omega)\}_{\omega \in \Omega}$. By maximizing over all possible quantities $\{q(\omega)\}_{\omega \in \Omega}$, we get the upper bound

$$\begin{aligned}\pi^M \le \max_{\{q(\omega)\}_{\omega \in \Omega} \in Q^{\Omega}} \Big\{&\mathbb{E}^{\mu}_{\omega}\big[\mathcal{S}\big(q(\omega), \theta(\omega)\big)\big] \\ &- \sup_{\omega' \in \Omega} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q(\omega'), \theta\big) - u\big(q(\omega'), \theta(\omega')\big)\big\}^{+}\big]\Big\}.\end{aligned}$$

Since the above inequality holds for all mechanisms $M \in \mathcal{M}$, we have that

$$\Pi \le \max_{\{q(\omega)\}_{\omega \in \Omega}} \Big\{\mathbb{E}^{\mu}_{\omega}\big[\mathcal{S}\big(q(\omega), \theta(\omega)\big)\big] - \sup_{\omega' \in \Omega} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q(\omega'), \theta\big) - u\big(q(\omega'), \theta(\omega')\big)\big\}^{+}\big]\Big\}. \tag{30}$$

The maximization on the right-hand side of (30) is over $\{q(\omega)\}_{\omega \in \Omega}$. We now prove that this can be reduced to a maximization over $\{q(\theta)\}_{\theta \in \Theta}$.

Consider any set of quantities $\{q(\omega)\}_{\omega \in \Omega}$, and construct $\{\tilde{q}(\theta)\}_{\theta \in \Theta}$ as

$$\begin{aligned}\tilde{q}(\theta) = \arg\sup_{q \in \mathbb{R}} \mathcal{S}(q, \theta) \\ \text{subject to: } \exists \omega' \in \Omega, \theta = \theta\big(\omega'\big) \text{ and } q \le q\big(\omega'\big).\end{aligned}$$

For each $\theta \in \Theta$, $\tilde{q}(\theta)$ is constructed as the quantity that yields the highest surplus subject to the constraint that it must be lower than some quantity implemented when the agent's type is θ (i.e., lower than some $q(\omega')$ with $\theta = \theta(\omega')$). We make two observations:

(i) Quantities $\{\tilde{q}(\theta)\}_{\theta\in\Theta}$ yield a weakly larger expected surplus than $\{q(\omega)\}_{\omega\in\Omega}$:

$$\mathbb{E}^{\mu}_{\omega}\big[\mathcal{S}\big(q(\omega), \theta(\omega)\big)\big] \leq \mathbb{E}^{f}_{\theta}\big[\mathcal{S}\big(\tilde{q}(\theta), \theta\big)\big]. \tag{31}$$

This is because for each $\theta \in \Theta$, $\tilde{q}(\theta)$ is constructed by considering the quantity that yields the highest surplus across all $\omega' \in \Omega$ such that $\theta = \theta(\omega')$.

(ii) The lower bound on the agent's rents constructed in Lemma 2 is weakly lower when computed with quantities $\{\tilde{q}(\theta)\}_{\theta\in\Theta}$ than when computed with quantities $\{q(\omega)\}_{\omega\in\Omega}$:

$$\begin{aligned}&\sup_{\omega'\in\Omega} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q(\omega'), \theta\big) - u\big(q(\omega'), \theta(\omega')\big)\big\}^{+}\big]\\ &\quad\geq \sup_{\theta'\in\Theta} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(\tilde{q}(\theta'), \theta\big) - u\big(\tilde{q}(\theta'), \theta'\big)\big\}^{+}\big].\end{aligned} \tag{32}$$

To prove (32), note that

$$\mathbb{E}^{f}_{\theta}\big[\big\{u(q, \theta) - u\big(q, \theta'\big)\big\}^{+}\big]$$

is an increasing function of q (we show this separately in the Supplemental Material). Thus, for all $\theta' \in \Theta$,

$$\mathbb{E}^{f}_{\theta}\big[\big\{u\big(\tilde{q}(\theta'), \theta\big) - u\big(\tilde{q}(\theta'), \theta'\big)\big\}^{+}\big] \leq \sup_{\omega'\in\Omega} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q(\omega'), \theta\big) - u\big(q(\omega'), \theta'\big)\big\}^{+}\big]$$

subject to $\theta' = \theta\big(\omega'\big)$.

This is because $\tilde{q}(\theta')$ is less than or equal to the supremum of $\{q(\omega)\}_{\omega\in\Omega:\theta(\omega)=\theta'}$. By taking the supremum of the previous equation with respect to $\theta' \in \Theta$, we prove (32).

From (31) and (32), it follows that

$$\begin{aligned}&\mathbb{E}^{\mu}_{\omega}\big[\mathcal{S}\big(q(\omega), \theta(\omega)\big)\big] - \sup_{\omega'\in\Omega} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q(\omega'), \theta\big) - u\big(q(\omega'), \theta(\omega')\big)\big\}^{+}\big]\\ &\quad\leq \mathbb{E}^{f}_{\theta}\big[\mathcal{S}\big(\tilde{q}(\theta), \theta\big)\big] - \sup_{\theta'\in\Theta} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(\tilde{q}(\theta'), \theta\big) - u\big(\tilde{q}(\theta'), \theta'\big)\big\}^{+}\big].\end{aligned}$$

Thus, for all quantities $\{q(\omega)\}_{\omega\in\Omega}$,

$$\begin{aligned}&\mathbb{E}^{\mu}_{\omega}\big[\mathcal{S}\big(q(\omega), \theta(\omega)\big)\big] - \max_{\omega'\in\Omega} \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q(\omega'), \theta\big) - u\big(q(\omega'), \theta(\omega')\big)\big\}^{+}\big]\\ &\quad\leq \max_{\{\tilde{q}(\theta)\}_{\theta\in\Theta}\in\mathbb{R}^{\Theta}} \Big\{\mathbb{E}^{f}_{\theta}\big[\mathcal{S}\big(\tilde{q}(\theta), \theta\big)\big] - \max_{\theta'\in\Theta}\mathbb{E}^{f}_{\theta}\big[u\big(\tilde{q}(\theta'), \theta\big) - u\big(\tilde{q}(\theta'), \theta'\big)\big]\Big\}.\end{aligned} \tag{33}$$

The result follows from (30) and (33). $\square$

Proof of Lemma 3. We provide the proof in three steps. We first reformulate the IC constraint, then show the sufficiency of the conditions, and, finally, show the necessity of the conditions.

Step 1: Reformulation of IC constraints. In a mechanism, the set of strategies Σ is the set of functions $\{\sigma_t\}_{t\in T}$ of the form $\sigma_t : \times_{t'\leq t} S \to \{0,1\}$. That is, at each period t, the agent needs to report whether his current signal is 0 or 1; his report may condition on the history of all signals. However, the only relevant decision for the agent is when to report that the process $\{s_t\}_{t\in T}$ changed from 0 to 1. If the agent reports message $z=1$ more than once, then he will get a payoff of 0, which he can guarantee ex post anyway. Therefore, the set of strategies for the agent is a family of functions $\sigma : T \to T$ such that, if the process s_t changes from 0 to 1 at t, then the agent reports $z=1$ at time $\sigma(t)$. However, the agent's reports must be measurable with respect to the information he has at the time. So we define the set of functions $\sigma : T \to T$ that constitute a valid strategy:

$$\tilde{\Sigma} \triangleq \{\sigma : T \to T : \text{ if } \sigma(t) < t \text{ then for all } t' > \sigma(t), \sigma(t') = \sigma(t)\}. \tag{34}$$

The set of available strategies for the agent is equal to $\tilde{\Sigma}$. Recall that under upward disclosure, $\Theta = T$ and so the same condition can be reinterpreted in terms of the agent's type. If type $\theta \in \Theta$ reports being a type $\theta' < \theta$, then he must do so without knowing the exact realization of his type. Thus, for all types $\theta'' > \theta'$, the agent must also report θ'.

For any strategy $\sigma \in \tilde{\Sigma}$, let $\hat{\theta}_\sigma \in \Theta$ be such that

$$\exists \theta > \hat{\theta}_\sigma, \quad \sigma(\theta) = \hat{\theta}_\sigma.$$

That is, type θ reports being $\hat{\theta}_\sigma$ and $\theta > \hat{\theta}_\sigma$. However, this implies that $\forall \theta > \hat{\theta}_\sigma$, $\sigma(\theta) = \hat{\theta}_\sigma$. Thus, we can write the expected utility that strategy σ yields

$$\mathbb{E}^f_\theta\big[\{u(q(\sigma(\theta)), \theta) - x(\sigma(\theta))\}^+\big] = \mathbb{E}^f_\theta\big[\{u(q(\sigma(\theta)), \theta) - x(\sigma(\theta))\}^+ | \theta \leq \hat{\theta}_\sigma\big] \tag{35}$$

$$+ \mathbb{E}^f_\theta\big[u(q(\hat{\theta}_\sigma), \theta) - x(\hat{\theta}_\sigma) | \theta > \hat{\theta}_\sigma\big]. \tag{36}$$

The positive part function accounts for the ex post participation constraint; the agent opts out of the mechanism whenever σ yields a negative rent. We do not need to add the positive part function in the second term (i.e., in (36)) because the ex post participation constraint guarantees that $u(q(\hat{\theta}_\sigma), \hat{\theta}_\sigma) - x(\hat{\theta}_\sigma) \geq 0$ and so $u(q(\hat{\theta}_\sigma), \theta) - x(\hat{\theta}_\sigma) \geq 0$ for all $\theta > \hat{\theta}_\sigma$ (this is implied by $\partial u(q, \theta)/\partial \theta \geq 0$). Finally, note that in (35), $\sigma(\theta) \geq \theta$ for all $\theta \leq \hat{\theta}_\sigma$ (this is implied by the measurability condition (34)).

We now denote by $\sigma^i : \Theta \to \Theta$ the set of functions such that $\sigma^i(\theta) > \theta$. A mechanism $\{q(\theta), x(\theta)\}_{\theta\in\Theta}$ is incentive compatible if

$$\mathbb{E}^f_\theta\big[u(q(\theta), \theta) - x(\theta)\big] \geq \sup_{\sigma\in\tilde{\Sigma}} \mathbb{E}^f_\theta\big[\{u(q(\sigma(\theta)), \theta) - x(\sigma(\theta))\}^+\big]. \tag{37}$$

We can rewrite the right-hand side of (37) as

$$\sup_{\substack{\{\sigma^i : T \to T\} \\ \hat{\theta}_\sigma \in \Theta}} \mathbb{E}^f_\theta\big[u(q(\sigma^i(\theta)), \theta) - x(\sigma^i(\theta)) | \theta \leq \hat{\theta}_\sigma\big] + \mathbb{E}^f_\theta\big[u(q(\hat{\theta}_\sigma), \theta) - x(\hat{\theta}_\sigma) | \theta > \hat{\theta}_\sigma\big]. \tag{38}$$

In other words, the maximization over all strategies $\tilde{\Sigma}$ can be written in two parts: (i) as the optimal cutoff type $\hat{\theta}_\sigma \in \Theta$, such that all types $\theta > \hat{\theta}_\sigma$ report $\sigma(\theta) = \hat{\theta}_\sigma$, and (ii) as the optimal reports for types $\theta \leq \hat{\theta}_\sigma$, which must satisfy that $\theta \leq \sigma(\theta)$. Observe that neither term in (38) has the positive part function. As previously explained, the positive part function in the second term is redundant because $u(q(\hat{\theta}_\sigma), \hat{\theta}_\sigma) - x(\hat{\theta}_\sigma) \geq 0$ and $\partial u(q, \theta)/\partial\theta \geq 0$ implies that $u(q(\hat{\theta}_\sigma), \theta) - x(\hat{\theta}_\sigma) \geq 0$ for all $\theta > \hat{\theta}_\sigma$. In the first term, it is also redundant because, if for some σ^i and some θ', $u(q(\sigma^i(\theta')), \theta') - x(\sigma^i(\theta')) < 0$, then the agent would opt out of the mechanism ex post. So in this case, we can redefined the strategy so that $\sigma^i(\theta') = \theta'$, which would obviously give weakly larger rents. Thus, it is never optimal for the agent to use a strategy in which $\sigma^i(\theta') > \theta'$ and $u(q(\sigma^i(\theta')), \theta') - x(\sigma^i(\theta')) < 0$.

We conclude this part of the proof by writing (37) as

$$\begin{aligned}\mathbb{E}_\theta^f\big[u\big(q(\theta), \theta\big) - x(\theta)\big] \geq \sup_{\substack{\{\sigma^i: T \to T\\ \hat{\theta}_\sigma \in \Theta\}}} & \mathbb{E}_\theta^f\big[u\big(q\big(\sigma^i(\theta)\big), \theta\big) - x\big(\sigma^i(\theta)\big) | \theta \leq \hat{\theta}_\sigma\big] \\ & + \mathbb{E}_\theta^f\big[u\big(q(\hat{\theta}_\sigma), \theta\big) - x(\hat{\theta}_\sigma) | \theta > \hat{\theta}_\sigma\big]. \end{aligned} \tag{39}$$

Step 2: Equations (13) and (14) are necessary for incentive compatibility. Inequality (39) implies that

$$\mathbb{E}_\theta^f\big[u\big(q(\theta), \theta\big) - x(\theta)\big] \geq \sup_{\{\sigma^i: T \to T\}} \mathbb{E}_\theta^f\big[u\big(q\big(\sigma^i(\theta)\big), \theta\big) - x\big(\sigma^i(\theta)\big)\big].$$

By definition, $\sigma^i(\theta) \geq \theta$ for all $\theta \in \Theta$ and so the previous inequality implies that (13) is satisfied. Similarly, if (39) is satisfied, then

$$\begin{aligned}&\mathbb{E}_\theta^f\big[u\big(q(\theta), \theta\big) - x(\theta)\big] \\ &\quad \geq \sup_{\{\hat{\theta}_\sigma \in \Theta\}} \mathbb{E}_\theta^f\big[u\big(q(\theta), \theta\big) - x(\theta) | \theta \leq \hat{\theta}_\sigma\big] + \mathbb{E}_\theta^f\big[u\big(q(\hat{\theta}_\sigma), \theta\big) - x(\hat{\theta}_\sigma) | \theta > \hat{\theta}_\sigma\big].\end{aligned}$$

The previous inequality can be written as

$$0 \geq \sup_{\{\hat{\theta}_\sigma \in \Theta\}} \mathbb{E}_\theta^f\big[u\big(q(\hat{\theta}_\sigma), \theta\big) - x(\hat{\theta}_\sigma) - \big(u\big(q(\theta), \theta\big) - x(\theta)\big) | \theta > \hat{\theta}_\sigma\big].$$

The previous inequality implies that (14) is satisfied (this can be seen directly by relabeling $\hat{\theta}_\sigma$ with θ').

Step 3: Equations (13) and (14) are sufficient for incentive compatibility. We write (39) as

$$\begin{aligned}0 \geq \sup_{\substack{\{\sigma^i: T \to T\\ \hat{\theta}_\sigma \in \Theta\}}} & \mathbb{E}_\theta^f\big[u\big(q\big(\sigma^i(\theta)\big), \theta\big) - x\big(\sigma^i(\theta)\big) - \big(u\big(q(\theta), \theta\big) - x(\theta)\big) | \theta \leq \hat{\theta}_\sigma\big] \\ & + \mathbb{E}_\theta^f\big[u\big(q(\hat{\theta}_\sigma), \theta\big) - x(\hat{\theta}_\sigma) - \big(u\big(q(\theta), \theta\big), x(\theta)\big) | \theta > \hat{\theta}_\sigma\big].\end{aligned}$$

If (13) and (14) are satisfied, then both terms are going to be nonpositive, so (39) will be satisfied. Therefore, (13) and (14) imply incentive compatibility. This concludes the proof. □

Proof That (17b) Has a Solution. We prove that there exists a $\tilde{\theta}$ that solves (17b). For this, note that the right-hand side of (17b) does not depend on $\tilde{\theta}$. Also note that the left-hand side is continuous in $\tilde{\theta}$. Because $\frac{\partial^2 u(q,\theta)}{\partial\theta\partial q} > 0$ and $\frac{\partial u(q,\theta)}{\partial\theta} \geq 0$, we have that $\frac{\partial u(q,\theta)}{\partial\theta} > 0$ for all $q > 0$. Thus, the left-hand side is additionally strictly decreasing in $\tilde{\theta}$. This is because $\frac{\partial u(q,\theta)}{\partial\theta}$ is strictly positive and, hence, the term inside the integral is positive. Thus, there is at most one $\tilde{\theta}$ that solves for (17b).

We now show that a solution exists. We show this by proving that in the limit $\tilde{\theta} \to \bar{\theta}$, the left-hand side of (17b) is smaller than the right-hand side of (17b), while in the limit $\tilde{\theta} \to \underline{\theta}$, the inequality is reversed. Because the left-hand side of (17b) is continuous in $\tilde{\theta}$, the intermediate value theorem guarantees that there exists $\tilde{\theta}$ such that the inequality is satisfied with equality.

In the limit $\tilde{\theta} \to \bar{\theta}$, we have that the left-hand side of (17b) converges to 0 and, hence, the right-hand side of (17b) is larger than the left-hand side of (17b). We now show that in the limit $\tilde{\theta} \to \underline{\theta}$, the right-hand side of (17b) is smaller than the left-hand side of (17b). For this, we note that

$$\max_{\theta'\in\Theta} \mathbb{E}_\theta^f\big[\big\{u\big(q(\theta'),\theta\big) - u\big(q(\theta'),\theta'\big)\big\}^+\big]$$

$$= \max_{\theta'\in\Theta}\big(1 - F(\theta')\big)\mathbb{E}_\theta^f\left[\int_{\theta'}^{\theta} \frac{u\big(q(\theta'),s\big)}{\partial\theta}\,ds\middle|\theta > \theta'\right] \tag{40}$$

$$\leq \max_{\theta'\in\Theta}\big(1 - F(\theta')\big)\mathbb{E}_\theta^f\left[\int_{\theta'}^{\theta} \frac{\partial u\big(q(s),s\big)}{\partial\theta}\,ds\middle|\theta > \theta'\right] \tag{41}$$

$$= \big(1 - F(\underline{\theta})\big)\mathbb{E}_\theta^f\left[\int_{\underline{\theta}}^{\theta} \frac{\partial u\big(q(s),s\big)}{\partial\theta}\,ds\middle|\theta > \underline{\theta}\right]. \tag{42}$$

The explanation of the steps is as follows. Equation(40) uses that $\partial u(q,\theta)/\partial\theta \geq 0$ and rewrites the difference as an integral. Equation (41) uses that $\partial^2 u(q,\theta)/\partial\theta\partial q \geq 0$ and that $q(\theta)$ is nondecreasing. Equation (42) follows from the fact that the term inside the expectation is positive and, hence, the whole term on the right-hand side of (41) is decreasing in θ'. Hence, the maximum is achieved at $\theta' = \underline{\theta}$. However, (42) is equal to the left-hand side of (17b) evaluated at $\tilde{\theta} = \underline{\theta}$. Hence, the left-hand side of (17b) evaluated at $\tilde{\theta} = \underline{\theta}$ is greater than the right-hand side of (17b). Thus, there is always a $\tilde{\theta}$ that solves (17b). □

Proof of Proposition 1. To prove incentive compatibility, we need to prove that if the transfers are defined as in (17a) and (17b), then the IC constraints (13) and (14) are satisfied.

Step 1: Constraint (13). We first note that for $\theta \in [\tilde{\theta}, \bar{\theta}]$, the transfers are essentially the same as those dictated by a static incentive-compatible mechanism (shifted only by a constant) and, thus,

$$\forall \theta'', \theta' \in [\tilde{\theta}, \bar{\theta}], \quad u(q(\theta''), \theta'') - x(\theta'') \geq u(q(\theta'), \theta'') - x(\theta'). \tag{43}$$

Hence, for all $\theta'', \theta' \in [\tilde{\theta}, \bar{\theta}]$, (13) is satisfied. If $\theta'' < \theta' \leq \tilde{\theta}$, then

$$u(q(\theta'), \theta'') - x(\theta') = u(q(\theta'), \theta'') - u(q(\theta'), \theta') \leq u(q(\theta'), \theta') - x(\theta') = 0,$$

where the inequality is implied by $\partial u(q, \theta)/\partial \theta \geq 0$. If $\theta'' < \tilde{\theta} \leq \theta'$, then

$$u(q(\theta'), \theta'') - x(\theta') \leq u(q(\theta'), \tilde{\theta}) - x(\theta') \leq u(q(\tilde{\theta}), \tilde{\theta}) - x(\tilde{\theta}) = 0.$$

In the first inequality, we use the fact that $\partial u(q, \theta)/\partial \theta \geq 0$. In the second inequality, we use that for all $\theta'', \theta' \in [\tilde{\theta}, \bar{\theta}]$, constraint (13) is satisfied. Thus (13) is satisfied for all $\theta'', \theta' \in \Theta$.

Step 2: Constraint (14). For all $\theta' \geq \tilde{\theta}$, we can take expectation of (43) with respect to $\theta'' \geq \theta'$. We get

$$\mathbb{E}_\theta^f[u(q(\theta), \theta) - x(\theta) | \theta \geq \theta'] \geq \mathbb{E}_\theta^f[u(q(\theta'), \theta) - x(\theta') | \theta \geq \theta'].$$

Thus, (14) is satisfied for $\theta' \geq \tilde{\theta}$. For all $\theta' < \tilde{\theta}$,

$$\begin{aligned}
\mathbb{E}_\theta^f[u(q(\theta'), \theta) - x(\theta') | \theta \geq \theta'] &= \mathbb{E}_\theta^f[u(q(\theta'), \theta) - u(q(\theta'), \theta') | \theta \geq \theta'] && (44)\\
&= \frac{1}{1 - F(\theta')} \mathbb{E}_\theta^f[\{u(q(\theta'), \theta) - u(q(\theta'), \theta')\}^+] && (45)\\
&\leq \frac{1}{1 - F(\theta')} \max_{\theta'' \in \Theta} \mathbb{E}_\theta^f[\{u(q(\theta''), \theta) - u(q(\theta''), \theta'')\}^+] && (46)\\
&= \frac{1 - F(\tilde{\theta})}{1 - F(\theta')} \mathbb{E}_\theta^f\left[\int_{\tilde{\theta}}^{\theta} \frac{\partial u(q(s), (s))}{\partial \theta} ds \middle| \theta \geq \tilde{\theta}\right] && (47)\\
&= \frac{1 - F(\tilde{\theta})}{1 - F(\theta')} \mathbb{E}_\theta^f[u(q(\theta), \theta) - x(\theta) | \theta \geq \tilde{\theta}] && (48)\\
&= \frac{1}{1 - F(\theta')} \mathbb{E}_\theta^f[u(q(\theta), \theta) - x(\theta)] && (49)\\
&= \mathbb{E}_\theta^f[u(q(\theta), \theta) - x(\theta) | \theta > \theta']. && (50)
\end{aligned}$$

The explanation of the previous steps are as follows. Equation (44) uses that for $\theta' < \tilde{\theta}$, $x(\theta') = u(q(\theta'), \theta')$. Equation (45) is because $\partial u(q, \theta)/\partial \theta \geq 0$ and, hence, $u(q(\theta'), \theta) < u(q(\theta'), \theta')$ for $\theta < \theta'$. By taking the positive part and normalizing, we get (45). Equation (46) is by the definition of the maximum. Equation (47) follows from the definition of $\tilde{\theta}$ (see (17b)). Equation (48) follows from the fact that the transfers are constructed using the envelope theorem as in a static incentive-compatible mechanism; (48) is analogous to (4). Equations (49) and (50) are implied by $u(q(\theta), \theta) - x(\theta) = 0 \; \forall \theta \leq \tilde{\theta}$. Thus, inequality (14) is satisfied for all $\theta' \in \Theta$.

Finally note that (18) is obtained by replacing $\tilde{x}(\theta)$ with (17a) and then using (17b). This proves the result. □

Proof of Theorem 2. Lemma 4 (proved later in the Supplemental Material) states that there exists a nondecreasing solution to (12). Proposition 1 shows that the solution to (12) can be implemented with transfers as in (17a) and (17b). The principal's profits are

$$\pi = \mathbb{E}_\theta^f\big[x^*(\theta) - c\big(q^*(\theta), \theta\big)\big] = \mathbb{E}_\theta^f\big[\mathcal{S}\big(q^*(\theta), \theta\big) - \big(u\big(q^*(\theta), \theta\big) - x^*(\theta)\big)\big].$$

Using Proposition 1,

$$\mathbb{E}_\theta^f\big[u\big(q^*(\theta), \theta\big) - x^*(\theta)\big] = \max_{\theta' \in \Theta} \mathbb{E}_\theta^f\big[\big\{u\big(q^*(\theta'), \theta\big) - u\big(q^*(\theta'), \theta'\big)\big\}^+\big].$$

Hence,

$$\pi = \mathbb{E}_\theta^f\big[\mathcal{S}\big(q^*(\theta), \theta\big)\big] - \max_{\theta' \in \Theta} \mathbb{E}_\theta^f\big[u\big(q^*(\theta'), \theta\big) - u\big(q^*(\theta'), \theta'\big)\big].$$

However, $\{q^*(\theta)\}_{\theta\in\Theta}$ is a nondecreasing solution to (12). Hence,

$$\pi = \max_{\{q(\theta)\}_{\theta\in\Theta} \in \mathbb{R}^\Theta} \Big\{\mathbb{E}_\theta^f\big[\mathcal{S}\big(q(\theta), \theta\big)\big] - \max_{\theta' \in \Theta} \mathbb{E}_\theta^f\big[\big\{u\big(q(\theta'), \theta\big) - u\big(q(\theta'), \theta'\big)\big\}^+\big]\Big\}.$$

It follows from Theorem 1 that this is an optimal mechanism. □

Proof of Lemma 4. Let $\{\tilde{q}_I(\theta)\}_{\theta\in\Theta}$ be a solution to the maximization (12) for a fixed I:

$$\begin{aligned} \big\{\tilde{q}_I(\theta)\big\}_{\theta\in\Theta} \in & \operatorname*{arg\,max}_{\{q(\theta)\}_{\theta\in\Theta} \in \mathbb{R}^\Theta} \big\{\mathbb{E}_\theta^f\big[\mathcal{S}\big(q(\theta), \theta\big)\big] - I\big\} \\ & \text{subject to } I \geq \max_{\theta' \in \Theta} \mathbb{E}_\theta^f\big[\big\{u\big(q(\theta'), \theta\big) - u\big(q(\theta'), \theta'\big)\big\}^+\big]. \end{aligned} \tag{51}$$

It follows from the definition of $\hat{q}_I(\theta)$ (see (21)) that $\tilde{q}_I(\theta) \leq \hat{q}_I(\theta)$; otherwise (51) would be violated. Because $\mathcal{S}(q, \theta)$ is strictly quasi-concave if $\hat{q}_I(\theta) < q^s(\theta)$, then

$$\forall q < \hat{q}_I(\theta), \quad \mathcal{S}(q, \theta) < \mathcal{S}\big(\hat{q}_I(\theta), \theta\big).$$

Alternatively, since $\mathbb{E}_\theta^f[\{u(q, \theta) - u(q, \theta')\}^+]$ is increasing in q (we prove this separately in the Supplemental Material), if $q^s(\theta) < \hat{q}_I(\theta)$, then implementing $q^s(\theta)$ would not violate (51). Thus, for almost every $\theta \in \Theta$, $\tilde{q}_I(\theta)$ must be equal to $q^s(\theta)$ whenever possible or equal to $\hat{q}_I(\theta)$. Thus, $\tilde{q}(\theta)$ is given by

$$\tilde{q}_I(\theta) = \min\big\{q^s(\theta), \hat{q}(\theta)\big\}.$$

However, this is the same as the definition of $q_I(\theta)$ in (22). In other words, $q_I(\theta)$ solves the maximization (12) for a fixed I. This proves that (12) is equivalent to (23).

Properties of $q_I(\theta)$. We now show that for all $I > 0$, (a) the quantity $\{q_I(\theta)\}_{\theta\in\Theta}$ is continuous and nondecreasing, and (b) there exists a threshold $\hat{\theta}_I < \bar{\theta}$ such that $q_I(\theta) = q^s(\theta)$ for all $\theta > \hat{\theta}_I$. We now observe that $\{\hat{q}_I(\theta)\}_{\theta\in\Theta}$ is continuous and nondecreasing (we prove this later). Therefore, $\{q_I(\theta)\}_{\theta\in\Theta}$ is a minimum of two continuous nondecreasing functions, so it is also continuous and nondecreasing (remember that

$\{q^s(\theta)\}_{\theta\in\Theta}$ is also nondecreasing and continuous). Additionally, because $\partial u(q,\theta)/\partial\theta \geq 0$, we can write (21) as

$$\hat{q}_I(\theta) = \max\{q \in Q : (1-F(\theta))\mathbb{E}^f_{\theta'}[u(q,\theta') - u(q,\theta)|\theta' > \theta] \leq I\}.$$

Because all functions are continuous, and because Q and Θ have bounded support, we have that

$$\forall q \in Q, \quad \lim_{\theta\to\bar{\theta}} (1-F(\theta))\mathbb{E}^f_{\theta'}[u(q,\theta') - u(q,\theta)|\theta' > \theta] = 0.$$

Thus, if $I > 0$, then $\hat{q}(\theta) = \bar{q}$ for all θ high enough. Thus, $q_I(\theta)$ is equal to $q^s(\theta)$ for θ high enough. This proves the properties of $q_I(\theta)$.

Proof that $\hat{q}_I(\theta)$ is continuous and nondecreasing. First note that $u(0,\theta) = 0$ for all θ and, thu $\hat{q}_I(\theta) > 0$ for all $I > 0$. Because $\partial u(q,\theta)/\partial\theta \geq 0$ and $\partial^2 u(q,\theta)/\partial\theta\partial q > 0$, we in fact have that $\partial u(q,\theta)/\partial\theta > 0$ for all $q > 0$.

As before, we can write (21) as

$$\hat{q}_I(\theta) = \max\{q \in Q : (1-F(\theta))\mathbb{E}^f_{\theta'}[u(q,\theta') - u(q,\theta)|\theta' > \theta] \leq I\}.$$

Writing the difference as an integral, we get

$$\hat{q}_I(\theta) = \max\left\{q \in Q : (1-F(\theta))\mathbb{E}^f_{\theta'}\left[\int_\theta^{\theta'} \frac{\partial u(q,s)}{\partial\theta}\, ds \middle| \theta' > \theta\right] \leq I\right\}.$$

Because $\partial^2 u(q,\theta)/\partial\theta\partial q > 0$, the term inside the integral is strictly increasing in q. Thus, for all $\theta \in \Theta$ such that $\hat{q}_I(\theta) < \bar{q}$, we have that

$$I = (1-F(\theta'))\mathbb{E}^f_\theta[u(\hat{q}_I(\theta'),\theta) - u(\hat{q}_I(\theta'),\theta')|\theta \geq \theta']. \tag{52}$$

The previous equation defines $\hat{q}_I(\theta)$. The implicit function theorem guarantees that $\hat{q}_I(\theta)$ is differentiable for all θ such that $\hat{q}_I(\theta) < \bar{q}$. We now take the derivative of (52) with respect to θ':

$$\begin{aligned} 0 = {} & (1-F(\theta'))\mathbb{E}^f_\theta\left[-\frac{\partial u(\hat{q}_I(\theta'),\theta')}{\partial\theta}\middle|\theta \geq \theta'\right] \\ & + (1-F(\theta'))\frac{\partial\hat{q}_I(\theta')}{\partial\theta'}\mathbb{E}^f_\theta\left[\frac{\partial u(\hat{q}_I(\theta'),\theta)}{\partial q} - \frac{\partial u(\hat{q}_I(\theta'),\theta')}{\partial q}\middle|\theta \geq \theta'\right]. \end{aligned}$$

Rearranging terms, we write

$$\frac{\partial\hat{q}_I(\theta')}{\partial\theta'} = \frac{\mathbb{E}^f_\theta\left[\frac{\partial u(\hat{q}_I(\theta'),\theta')}{\partial\theta}\middle|\theta \geq \theta'\right]}{\mathbb{E}^f_\theta\left[\int_{\theta'}^{\theta}\frac{\partial^2 u(\hat{q}_I(\theta'),s)}{\partial q\partial\theta}\, ds\middle|\theta \geq \theta'\right]}.$$

Because $\partial u(q,\theta)/\partial\theta \geq 0$ and $\partial^2 u(q,\theta)/\partial q\partial\theta > 0$, the numerator and denominator are positive, which implies that $\{\hat{q}_I(\theta)\}_{\theta\in\Theta}$ is nondecreasing for all values of θ such that

$\hat{q}_I(\theta) < \bar{q}$. Moreover, since it is differentiable, it is obviously also continuous for all θ such that $\hat{q}_I(\theta) < \bar{q}$.

Since $\hat{q}_I(\theta)$ is nondecreasing and continuous when the value is less than $\bar{q}$, it follows that there exists a type θ'' such that $\hat{q}_I(\theta) = \bar{q}$ if and only if $\theta \geq \theta''$. It follows that $\hat{q}_I(\theta)$ is nondecreasing for all values of θ and it is also continuous at θ''. Therefore, $\hat{q}_I(\theta)$ is continuous and nondecreasing. ◁

The proof of Proposition 2 is a corollary of Proposition 3 (proved later in the Supplemental Material). □

Proof of Lemma 5. Remember that we can write (12) as

$$\max_{I \in \mathbb{R}} \mathbb{E}_\theta^f\big[\mathcal{S}\big(q_I(\theta), \theta\big)\big] - I. \tag{53}$$

We define

$$\bar{I} \triangleq \max_{\theta' \in \Theta} \mathbb{E}_\theta^f\big[\big\{u\big(q^s(\theta'), \theta\big) - u\big(q^s(\theta'), \theta'\big)\big\}^+\big].$$

In other words, $\bar{I}$ is the minimum rent such that $q_{\bar{I}}(\theta) = q^s(\theta)$. It is clear that the maximizer of (53) is never larger than $\bar{I}$. It is also clear that if $I < \bar{I}$, then there exists a positive measure of types $\theta \in \Theta$ such that $q_I(\theta) < q^s(\theta)$ (i.e., such that $q_I(\theta) = \hat{q}_I(\theta)$).

We now show that for all $I \in [0, \bar{I}]$, the objective function is strictly concave as a function of I (i.e., $\mathbb{E}_\theta^f[\mathcal{S}(q_I(\theta), \theta)] - I$ is strictly concave). We define $\{\theta_1^-, \theta_1^+, \dots, \theta_k^-, \theta_k^+,\}$ as the cutoffs such that $q_I(\theta) = \hat{q}_I(\theta)$, i.e.,

$$q_I(\theta) = \begin{cases} \hat{q}_I(\theta) & \text{if } \theta \in \big[\theta_j^-, \theta_j^+\big] \text{ for some } j \in \{1, \dots, k\} \\ q^s(\theta) & \text{otherwise.} \end{cases}$$

We write $\mathbb{E}_\theta^f[\mathcal{S}(q_I(\theta), \theta)] - I$ in terms of these thresholds:

$$\mathbb{E}_\theta^f\big[\mathcal{S}\big(q_I(\theta), \theta\big)\big] - I = \sum_{j=1}^{k} \int_{\theta_j^-}^{\theta_j^+} \mathcal{S}\big(\hat{q}_I(\theta), \theta\big) f(\theta)\, d\theta + \sum_{j=0}^{k} \int_{\theta_j^+}^{\theta_{j+1}^-} \mathcal{S}\big(q^s(\theta), \theta\big) f(\theta)\, d\theta - I. \tag{54}$$

Here we adopted the convention that $\theta_0^+ = \underline{\theta}$ and $\theta_{k+1}^- = \bar{\theta}$ (these two terms appear in the second summation). The derivative of $\mathbb{E}_\theta^f[\mathcal{S}(q_I(\theta), \theta)] - I$ with respect to I is given by

$$\frac{\partial \mathbb{E}_\theta^f\big[\mathcal{S}\big(q_I(\theta), \theta\big)\big] - I}{\partial I} = \sum_{j \in k} \int_{\theta_j^-}^{\theta_j^+} \frac{\partial \mathcal{S}\big(q_I(\theta), \theta\big)}{\partial q} \frac{\partial \hat{q}_I(\theta)}{\partial I} f(\theta)\, d\theta - 1.$$

We remind the reader that $q_I(\theta)$ is continuous, and so the derivative of (54) with respect to the thresholds $\{\theta_1^-, \theta_1^+, \dots, \theta_k^-, \theta_k^+,\}$ is zero; also $\partial q^s(\theta)/\partial I = 0$ (i.e., the surplus-maximizing quantity does not depend on I) and so the derivative of $\mathcal{S}(q^s(\theta), \theta)$ with respect to I is equal to 0. If we take the second derivative of $\mathbb{E}_\theta^f[\mathcal{S}(q_I(\theta), \theta)] - I$ with respect to I, we get

$$
\begin{aligned}
&\frac{\partial^2 \mathbb{E}_\theta^f\big[\mathcal{S}\big(q_I(\theta),\theta\big)\big]-I}{\partial I^2} \\
&\quad=\sum_{j\in k}\int_{\theta_j^-}^{\theta_j^+}\left(\frac{\partial^2\mathcal{S}\big(q_I(\theta),\theta\big)}{\partial q^2}\left(\frac{\partial \hat{q}_I(\theta)}{\partial I}\right)^2+\frac{\partial\mathcal{S}\big(q_I(\theta),\theta\big)}{\partial q}\frac{\partial^2\hat{q}_I(\theta)}{\partial I^2}\right)f(\theta)\,d\theta \qquad (55)\\
&\qquad-\sum_{j\in k}\frac{\partial\theta_j^-}{\partial I}\frac{\partial\mathcal{S}\big(q_I(\theta_j^-),\theta_j^-\big)}{\partial q}\frac{\partial\hat{q}_I(\theta_j^-)}{\partial I}f(\theta_j^-) \qquad (56)\\
&\qquad+\sum_{j\in k}\frac{\partial\theta_j^+}{\partial I}\frac{\partial\mathcal{S}\big(q_I(\theta_j^+),\theta_j^+\big)}{\partial q}\frac{\partial\hat{q}_I(\theta_j^+)}{\partial I}f(\theta_j^+). \qquad (57)
\end{aligned}
$$

Remember that $\{\theta_1^-,\theta_1^+,\dots,\theta_k^-,\theta_k^+,\}$ are the cutoffs in which $q_I(\theta)$ is equal to $\hat{q}_I(\theta)$. We remind the reader that if $\hat{q}_I(\theta)<\bar{q}$, then $\hat{q}_I(\theta)$ is strictly increasing in I. Thus, whenever $q_I(\theta)=\hat{q}_I(\theta)$, we have that $\hat{q}_I(\theta)$ is strictly increasing in I. Thus, the intervals $\{\theta_1^-,\theta_1^+,\dots,\theta_k^-,\theta_k^+,\}$ get smaller as I increases (i.e., the zones where $q_I(\theta)=q^s(\theta)$ get larger). Therefore,

$$\frac{\partial\theta_j^-}{\partial I}\geq 0\quad\text{and}\quad\frac{\partial\theta_j^+}{\partial I}\leq 0.$$

Moreover, the previous inequalities are strict whenever $\theta_1^-\neq\underline{\theta}$ and $\theta_j^+\neq\bar{\theta}$. In other words, the intervals get strictly smaller as long as the cutoffs are not equal to the limits of the type space. This is because $q_I(\theta)=\min\{\hat{q}_I(\theta),q^s(\theta)\}$ and both functions inside the minimum are continuous, so if $\hat{q}_I(\theta)$ is strictly increasing, then the zones where $q_I(\theta)=q^s(\theta)$ must also increase. Since we showed in Lemma 4 that $q_I(\theta)=q^s(\theta)$ for θ high enough, we have that $\theta_j^+\neq\bar{\theta}$ always and, hence,

$$\frac{\partial\theta_j^+}{\partial I}<0.$$

We also have that

$$\forall\theta\in\big[\theta_j^-,\theta_j^+\big],\qquad\frac{\partial\mathcal{S}\big(q_I(\theta),\theta\big)}{\partial q}\frac{\partial\hat{q}_I(\theta)}{\partial I}f(\theta)>0.$$

The previous inequality follows from the fact that $\mathcal{S}(q,\theta)$ is strictly quasi-concave and, hence, it is strictly increasing for $q<q^s(\theta)$, and $\hat{q}_I(\theta)$ is strictly increasing in I. Thus (56) is weakly negative (with strict inequality whenever $\theta_1^-\neq\underline{\theta}$) and (57) is strictly negative. If we show that (55) is weakly negative, then we will conclude that

$$\frac{\partial^2\mathbb{E}_\theta^f\big[\mathcal{S}\big(q_I(\theta),\theta\big)\big]-I}{\partial I^2}<0.$$

We now show that (55) is weakly negative. We first check that $\hat{q}_I(\theta)$ is a concave function of I. For this, we remind the reader that whenever $\hat{q}_I(\theta)<\bar{q}$, then $\hat{q}_I(\theta)$ is implicitly defined as in (52). Taking the second derivative of (52) with respect to I, we have that for all θ such that $\hat{q}_I(\theta) < \bar{q}$,

$$\frac{\partial^2 \hat{q}_I(\theta')}{\partial I^2} = -\frac{\mathbb{E}_\theta^f\Big[\Big(\frac{\partial^2 u(\hat{q}_I(\theta'), \theta)}{\partial q^2} - \frac{\partial^2 u(\hat{q}_I(\theta'), \theta')}{\partial q^2}\Big)\Big(\frac{\partial \hat{q}_I(\theta')}{\partial I}\Big)^2 \Big| \theta \geq \theta'\Big]}{\mathbb{E}_\theta^f\Big[\Big(\frac{\partial u(\hat{q}_I(\theta'), \theta)}{\partial q} - \frac{\partial u(\hat{q}_I(\theta'), \theta')}{\partial q}\Big) \Big| \theta \geq \theta'\Big]}.$$

Because $\partial^2 u(q, \theta)/\partial q \partial \theta > 0$ when $q > 0$, then

$$\forall \theta > \theta', \quad \Big(\frac{\partial u(\hat{q}_I(\theta'), \theta)}{\partial q} - \frac{\partial u(\hat{q}_I(\theta'), \theta')}{\partial q}\Big) > 0$$

and, thus, the denominator of the fraction is strictly positive. Because $\partial^3 u/\partial q \partial q \partial \theta \geq 0$, then

$$\forall \theta > \theta', \quad \frac{\partial^2 u(\hat{q}_I(\theta'), \theta)}{\partial q^2} - \frac{\partial^2 u(\hat{q}_I(\theta'), \theta')}{\partial q^2} \geq 0$$

and, thus, the numerator of the fraction is weakly positive. Thus, the fraction is positive, which multiplied by -1 gives a negative number. Hence, $\hat{q}_I(\theta)$ is a concave function of I. Looking at the term inside the integral in (55), we have that $\mathcal{S}(q, \theta)$ is weakly concave, and we showed that $q_I(\theta)$ is also concave and that $\mathcal{S}(q, \theta)$ is strictly increasing in q for $q \leq q^s(\theta)$;[26] thus, all the terms inside the integral in (55) are positive. Hence, the term (55) is nonnegative, which implies that $\partial^2(\mathbb{E}_\theta^f[\mathcal{S}(q_I(\theta), \theta)] - I)/\partial I^2 < 0$. This concludes the proof. □

Proof of Proposition 3. Let $M^* = (q, x)$ and (σ^*, o^*) be an optimal mechanism and an optimal strategy in this mechanism. We define

$$\ddot{q}(\omega) \triangleq q(\sigma^*(\omega)) \cdot o^*(\omega).$$

In other words, $\ddot{q}(\omega)$ is equal to the quantity implemented in mechanism M^*. Using the same arguments as in the proof of Theorem 1, we get that

$$\begin{aligned}\Pi &= \mathbb{E}_\omega^\mu\big[x(\sigma^*(\omega)) - c(q^*(\sigma(\omega)), \theta(\omega)\big] \\ &\leq \mathbb{E}_\omega^\mu\big[\mathcal{S}(\ddot{q}(\omega), \theta(\omega)\big] - \max_{\omega' \in \Omega} \mathbb{E}_\omega^\mu\big[\big\{u(\ddot{q}(\omega'), \theta(\omega)) - u(\ddot{q}(\omega'), \theta(\omega'))\big\}^+\big].\end{aligned}$$

This expression is derived by writing the principal's profits as the total surplus minus the lower bound on the agent's rents characterized in Lemma 2. We now construct $\{\tilde{q}(\theta)\}_{\theta \in \Theta}$ as in the proof of Theorem 1, that is, we construct it as

$$\tilde{q}(\theta) = \arg\sup_{q \in \mathbb{R}} \mathcal{S}(q, \theta)$$

$$\text{subject to } \exists \omega' \in \Omega, \theta = \theta(\omega') \text{ and } q \leq \ddot{q}(\omega').$$

[26] The strict quasi-concavity of $\mathcal{S}(q, \theta)$ implies that $\mathcal{S}(q, \theta)$ strictly increases in q for $q < q^s(\theta)$.

Using the same arguments as in the proof of Theorem 1, we get that

$$\mathbb{E}^{\mu}_{\omega}\big[\mathcal{S}(\ddot{q}(\omega),\theta(\omega)\big] - \max_{\omega'\in\Omega}\mathbb{E}^{\mu}_{\omega}\big[\big\{u\big(\ddot{q}(\omega'),\theta(\omega)\big) - u\big(\ddot{q}(\omega'),\theta(\omega')\big)\big\}^{+}\big]$$
$$\leq \mathbb{E}^{f}_{\theta}\big[\mathcal{S}(\tilde{q}(\theta),\theta\big] - \max_{\theta'\in\Theta}\mathbb{E}^{f}_{\theta}\big[\big\{u\big(\tilde{q}(\theta'),\theta\big) - u\big(\tilde{q}(\theta'),\theta'\big)\big\}^{+}\big]. \tag{58}$$

Moreover, the inequality is strict if there exists a positive measure of states $\omega \in \Omega$ such that $\ddot{q}(\omega) \neq \tilde{q}(\theta(\omega))$.[27] Therefore, we must have that for almost every $\omega \in \Omega$, $\ddot{q}(\omega) = \tilde{q}(\theta(\omega))$.

Because we showed in Lemma 5 that (12) has a unique solution, it must be that $\tilde{q}(\theta)$ is the (essentially unique) solution to (12). Otherwise, (58) evaluated at a different set of quantities would be strictly lower than (12). Thus, we have that for almost every $\omega \in \Omega$,

$$q\big(\sigma^*(\omega)\big) \cdot o^*(\omega) = q^*\big(\theta(\omega)\big).$$

This proves (26).

We previously proved that the quantities implemented by any optimal mechanism must solve (12) (i.e., must be equal to $q^*(\theta)$). Note that for all $\theta'' \in \Theta$ such that $q^*(\theta'') < q^s(\theta'')$,

$$\max_{\theta'\in\Theta}\mathbb{E}^{f}_{\theta}\big[\big\{u\big(q^*(\theta'),\theta\big) - u\big(q^*(\theta'),\theta'\big)\big\}^{+}\big] = \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q^*(\theta''),\theta\big) - u\big(q^*(\theta''),\theta''\big)\big\}^{+}\big]. \tag{59}$$

In other words, constraint (12b) is satisfied with equality for all $\theta'' \in \Theta$ such that $q^*(\theta'') < q^s(\theta'')$. Suppose there exists $\omega \in \Omega$ such that $q^*(\theta(\omega)) < q^s(\theta(\omega))$ and $x(\sigma^*(\omega)) < u(q^*(\theta(\omega)),\theta(\omega))$ (i.e., suppose (27) is violated). Then

$$\mathbb{E}^{f}_{\theta'}\big[\big\{u\big(q^*\big(\theta(\omega)\big),\theta'\big) - u\big(q^*\big(\theta(\omega),\theta(\omega)\big)\big\}^{+}\big] < \mathbb{E}^{f}_{\theta'}\big[\big\{u\big(q^*\big(\theta(\omega),\theta'\big) - x^*\big(\sigma(\omega)\big)\big\}^{+}\big]. \tag{60}$$

As we showed in Section 4.1, in this case, the agent can guarantee himself rents at least as high as $\mathbb{E}^{f}_{\theta}[\{u(q^*(\omega),\theta) - x^*(\sigma(\omega))\}^{+}]$. Because the principal's profits are equal to the total surplus minus the agent's rents, the principal's profits will be less than

$$\Pi < \mathbb{E}^{f}_{\theta}\big[\mathcal{S}\big(q^*(\theta),\theta\big)\big] - \mathbb{E}^{f}_{\theta}\big[\big\{u\big(q^*(\omega),\theta\big) - x^*\big(\sigma(\omega)\big)\big\}^{+}\big].$$

It follows from (59) and (60) that the principal's profits will be strictly lower than

$$\Pi < \mathbb{E}^{f}_{\theta}\big[\mathcal{S}\big(q^*(\theta),\theta\big)\big] - \max_{\theta'\in\Theta}\mathbb{E}^{f}_{\theta}\big[\big\{u\big(q(\theta'),\theta\big) - u\big(q(\theta'),\theta'\big)\big\}^{+}\big].$$

Thus the principal's profits will be lower than the optimal mechanism. Therefore, it must be the case that $x(\sigma^*(\omega)) = u(q^*(\theta(\omega)),\theta(\omega))$ for all ω such that $q^*(\theta(\omega)) < q^s(\theta(\omega))$. This proves the result. □

[27] The strict inequality follows from the fact that $\mathcal{S}(q,\theta)$ is strictly quasi-concave. Therefore, for every $\theta \in \Theta$, if there exists $\omega,\omega' \in \Omega$ such that $\theta(\omega) = \theta(\omega')$ and $\ddot{q}(\omega) \neq \ddot{q}(\omega')$, then $\mathcal{S}(\tilde{q}(\theta),\theta) > \min\{\mathcal{S}(\ddot{q}(\omega),\theta),\mathcal{S}(\ddot{q}(\omega'),\theta)\}$.

Proof that $\mathbb{E}_\theta^f[\{u(q,\theta)-u(q,\theta')\}^+]$ is increasing in q. First, note that $\partial u(q,\theta)/\partial\theta \geq 0$ and, hence,

$$\mathbb{E}_\theta^f\big[\{u(q,\theta)-u\big(q,\theta'\big)\}^+\big]=\big(1-F\big(\theta'\big)\big)\mathbb{E}_\theta^f\big[u(q,\theta)-u\big(q,\theta'\big)|\theta\geq\theta'\big].$$

Now note that

$$\begin{aligned}&\frac{\partial}{\partial q}\big(1-F\big(\theta'\big)\big)\mathbb{E}_\theta^f\big[\{u(q,\theta)-u\big(q,\theta'\big)\}^+|\theta\geq\theta'\big]\\&=\big(1-F\big(\theta'\big)\big)\mathbb{E}_\theta^f\left[\frac{\partial u(q,\theta)}{\partial q}-\frac{\partial u\big(q,\theta'\big)}{\partial q}\middle|\theta\geq\theta'\right]\\&=\big(1-F\big(\theta'\big)\big)\mathbb{E}_\theta^f\left[\int_{\theta'}^{\theta}\frac{\partial^2 u(q,\theta)}{\partial q\partial\theta}\middle|\theta\geq\theta'\right].\end{aligned}$$

Since $\partial^2 u(q,\theta)/\partial q\partial\theta>0$ and $\theta>\theta'$, we get the result. □

References

Akan, Mustafa, Barış Ata, and James D. Dana Jr. (2015), "Revenue management by sequential screening." *Journal of Economic Theory*, 159, 728–774. [344]

Ata, Barış and James D. Dana (2015), "Price discrimination on booking time." *International Journal of Industrial Organization*, 43, 175–181. [344]

Baron, David P. and Roger B. Myerson (1982), "Regulating a monopolist with unknown costs." *Econometrica*, 50, 911–930. [319, 345]

Bergemann, Dirk, Francisco Castro, and Gabriel Y. Weintraub (2018), "The scope of sequential screening with ex-post participation constraints." Unpublished paper, Graduate School of Business, Columbia University. [345]

Bergemann, Dirk and Martin Pesendorfer (2007), "Information structures in optimal auctions." *Journal of Economic Theory*, 137, 580–609. [320]

Bergemann, Dirk and Achim Wambach (2015), "Sequential information disclosure in auctions." *Journal of Economic Theory*, 159, 1074–1095. [341, 342, 343, 344]

Courty, Pascal and Li Hao (2000), "Sequential screening." *Review of Economic Studies*, 67, 697–717. [341]

Demsetz, Harold (1968), "Why regulate utilities?" *Journal of Law & Economics*, 11, 55–65. [341]

Esö, Péter and Balázs Szentes (2007), "Optimal information disclosure in auctions and the handicap auction." *Review of Economic Studies*, 74, 705–731. [341, 342]

Gershkov, Alex (2002), "Optimal auctions and endogenous information structure." Unpublished paper, Hebrew University. [341]

Green, Jerry R. and Jean-Jacques Laffont (1987), "Posterior implementability in a two person decision problem." *Econometrica*, 55, 69–94. [342]

Harris, Milton and Artur Raviv (1978), "Some results on incentive contracts with applications to education and employment, health insurance, and law enforcement." *American Economic Review*, 68, 20–30. [341]

Kamenica, Emir and Matthew Gentzkow (2011), "Bayesian persuasion." *American Economic Review*, 101, 2590–2615. [320]

Krähmer, Daniel and Roland Strausz (2015), "Optimal sales contracts with withdrawal rights." *Review of Economic Studies*, 82, 762–790. [320, 341, 344, 345]

Krähmer, Daniel and Roland Strausz (2016), "Optimality of sequential screening with multiple units and ex post participation constraints." *Economics Letters*, 142, 64–68. [345]

Li, Hao and Xianwen Shi (2017), "Discriminatory information disclosure." *American Economic Review*, 107, 3363–3385. [341]

Loeb, Martin and Wesley A. Magat (1979), "A decentralized method for utility regulation." *Journal of Law & Economics*, 22, 399–404. [341]

Myerson, Roger B. (1981), "Optimal auction design." *Mathematics of Operations Research*, 6, 58–73. [319]

Sappington, David (1983), "Limited liability contracts between principal and agent." *Journal of Economic Theory*, 29, 1–21. [343]

Co-editor Thomas Mariotti handled this manuscript.

Manuscript received 14 March, 2017; final version accepted 5 August, 2019; available online 8 August, 2019.