Non-Linear Pricing
and Optimal Shipping Policies

Áron Tóbiás

Maxwell School of Citizenship and Public Affairs
Syracuse University

December 12, 2016

Abstract A monopolist produces a good for sale to a buyer with uncertain valuation. The seller seeks to implement a profit-maximizing non-linear pricing scheme, which includes the time at which the good is shipped to the consumer. If the buyer discounts future payoffs and the seller does not, then delayed shipments act as an almost-perfect substitute for complete information and the monopolist can extract almost all of the surplus from trade. Shipping policies thus serve as a powerful tool of enhancing price discrimination. However, if the seller is as patient as the buyer, then she cannot benefit from delaying allocations.

JEL Classification D42, D82, L12

Keywords monopoly; non-linear pricing; mechanism design; private information; surplus extraction

Acknowledgements I am indebted to Dirk Bergemann for the continuous support, guidance, and encouragement with which he provided me throughout this research project. I thank also Dávid Krisztián Nagy and Yiqing Xing for their useful comments and suggestions. All errors and omissions are mine.

1 Introduction

Shipping policies are an important component of pricing in electronic commerce. Many online retailers offer a variety of shipment options in terms of the time of delivery, and the rates are often steeply increasing for earlier deliveries. It is common also to offer consumers free shipping over a given threshold value of purchases. In some cases, free delivery is embodied in a two-part-tariff construction. Amazon, for example, offers its customers free two-day shipping via its Prime program in exchange for a flat annual membership fee. Allegations concerning promotional offers that include earlier or more affordable shipping have stirred controversy about the role delivery policies play in sellers' revenue-management strategies. Recently, two plaintiffs separately claimed that Amazon had been actually tending to charge Prime customers the same amount as non-members by including regular shipping costs in the price offered for members, advertising the inflated price as involving "free" shipping.[1] The prevalence of such business practices suggests that shipping policies may, in fact, represent a subtle form of price discrimination—those consumers who exhibit higher valuations of a good can be enticed into paying a higher charge for it in exchange for earlier delivery.

In this paper, I argue that delayed deliveries constitute an extremely effective way of surplus extraction, indeed. I revisit the classic adverse-selection model of Mussa and Rosen (1978), in which a monopolist seeks to maximize her profits by offering a non-linear pricing schedule to a buyer whose valuation is private information, and I add a third component to the schedule beyond price and quantity: the seller can determine the time at which the good is delivered to the buyer. Under the assumption that the buyer discounts future payoffs but the seller does not, I demonstrate that the monopolist can extract *almost all* of the surplus from trade, while offering the same menu of efficient allocations as though she had complete information about the buyer's valuation.[2]

The key to almost-full surplus extraction lies in the monopolist's ability to offer such a schedule that the time of shipment is decreasing in the buyer's valuation: lower types obtain less of the good and they obtain it later. Delayed deliveries thus discourage higher-type buyers from purchasing menus designed for lower types. Since the seller is perfectly patient, she is not affected adversely by late delivery and the corresponding delay in payments. To the contrary: this additional dimension of price discrimination provides her with so much effective market power as to be an almost-perfect substitute for complete information about the buyer's valuation.

This sharp conclusion partially generalizes to the case in which the seller, too, discounts delayed payoffs. I show that if the monopolist is substantially more patient than the buyer, then the former can still exploit the difference between time prefer-

[1] *Burke v. Amazon Services LLC* and *Ekin v. Amazon Services LLC*. The court dockets corresponding to these lawsuits are available at https://dockets.justia.com/docket/washington/wawdce/2:2014cv00335/199396 and https://dockets.justia.com/docket/washington/wawdce/2:2014cv00244/198873, respectively. Date of access: December 12, 2016.

[2] Formally, if π^* denotes the optimal level of the seller's expected profits under complete information, then she can obtain at least $\pi^* - \varepsilon$ for any $\varepsilon > 0$ by designing a suitable delivery schedule, even if she is unaware of the buyer's valuation.

ences and improve upon the benchmark standard mechanism derived by Mussa and Rosen (1978), even if she is not perfectly patient. However, if the monopolist is less patient than or as patient as the buyer (as measured by the relative magnitude of their discount factors), then she cannot do better than the standard mechanism. Manufacturers of durable goods, such as electronics, appliances, and furniture, plausibly exhibit a greater degree of patience relative to the buyers of these goods than do retailers of non-durable goods such as perishable foodstuffs. Correspondingly, the optimal design of shipping policies bears more relevance to enhancing price discrimination in the sale of durables.

My paper fits into two main strands of literature. First, a variety of studies on mechanism design have exhibited distinct channels through which a revenue-maximizing principal can partially or fully enhance the extraction of surplus from trade, including valuations correlated among multiple buyers (Crémer and McLean, 1985, 1988; McAfee and Reny, 1992),[3] the seller's control over the private information the buyer possesses (Lewis and Sappington, 1994; Bergemann and Pesendorfer, 2007; Eső and Szentes, 2007; Bergemann and Wambach, 2015), refund contracts (Courty and Li, 2000; Akan *et al.*, 2015), and segmenting the market via third-degree price discrimination (Bergemann *et al.*, 2015).[4] The current paper highlights temporal delays in the shipment of allocations as not only another powerful channel of surplus extraction, but also one that does not presuppose any additional capability on the seller's part—such as possessing supplementary information or controlling that of the buyer—beyond simply being able to select and commit to the time at which the allocation is delivered.[5]

Second, a recurrent and robust prediction of the literature on bilateral bargaining is that relative patience represents a strategic advantage. Under the unique subgame-perfect Nash equilibrium outcome in the bargaining situation presented by Rubinstein (1982), both the buyer's and the seller's payoffs are increasing in their own discount factors and decreasing in their respective counterparties'. Chatterjee and Samuelson (1987) extend this result to a two-sided incomplete-information setting and show that the buyer is likelier to benefit from trade the more patient he is or the less patient the seller is. My results reinforce this key intuition about the strategic power of patience in an environment in which the monopolist makes unilateral offers to the buyer and no bargaining is involved: Perfect indifference between current payoffs and future ones endows the seller with an almost-perfect ability to extract the full surplus, whereas she cannot benefit at all from the opportunity to delay shipments if she is at most as patient as the buyer.

The paper proceeds as follows. Section 2 describes the model and defines the notion of an optimal mechanism. In Section 3, I show that the seller is able to expropriate almost all of the surplus from trade provided that she is perfectly patient. Section 4 analyzes the case in which the monopolist is impatient. The nature of the optimal mecha-

[3] However, Neeman (2004) and Heifetz and Neeman (2006) raise caveats concerning how much the principal can really profit from correlation between the agents' types.

[4] For a comprehensive literature review on price discrimination, see Stole (2007).

[5] Hörner and Samuelson (2011) study a related problem in which it is the *buyers* who strategically choose the time of purchase.

nism then depends on whether she is less or more patient than the buyer. In the former case, the optimal mechanism involves no delays and coincides with the standard mechanism derived by Mussa and Rosen (1978). However, if the seller is sufficiently more patient than the buyer (yet not perfectly so), then the standard mechanism ceases to be optimal: she can achieve greater profits by introducing delays in the delivery of the allocation at the lower end of the type distribution. Section 5 proposes an alternative to posting non-linear pricing schemes in a setting that resembles Dutch auctions and circumvents the direct screening of the buyer for his valuation of the good. In addition, it is here that I discuss the problem of time inconsistency associated with such a dynamic auction-like procedure. Section 6 concludes. A number of the formal proofs are relegated to the Appendix, while others appear in the main text.

2 Model

2.1 Setup

Consider a mechanism-design problem in which a principal (she) seeks to supply the agent (he) with a consumption good in order to maximize her revenues net of production costs (*cf.* Mussa and Rosen, 1978).[6] The extent to which the buyer values the good is subject to uncertainty from the principal's point of view. Specifically, the space of the agent's possible types is $\Theta \equiv [\underline{\theta}, \overline{\theta}]$, where $\overline{\theta} > \underline{\theta} > 0$. The agent's type is private information and its prior distribution is given by the density function $f : \Theta \to \mathbb{R}_+$. I assume that f is continuous and $f(\theta) > 0$ for all $\theta \in \Theta$. Let $F : \Theta \to [0, 1]$ denote the corresponding cumulative distribution function.

The agent values the good according to the twice-continuously-differentiable utility function $u : \mathbb{R}_+ \times \Theta \to \mathbb{R}_+$. Pecuniary transfers affect the consumer's utility in a quasi-linear way. Therefore, if the agent is of type $\theta \in \Theta$, consumes $q \in \mathbb{R}_+$ units, and is required to pay an overall charge $p \in \mathbb{R}$, then his net utility is given as $u(q, \theta) - p$. For the sake of tractability, I assume that $u(q, \theta) \equiv \theta v(q)$ for each $(q, \theta) \in \mathbb{R}_+ \times \Theta$, where $v : \mathbb{R}_+ \to \mathbb{R}_+$ is a twice-continuously-differentiable function satisfying $v' > 0$, $v'' < 0$, $v(0) = 0$, and $\lim_{q \to \infty} v(q) = \infty$.[7]

The good is supplied by the principal, who acts as a monopolist. The cost of producing $q \in \mathbb{R}_+$ units is given as $c(q)$, where $c : \mathbb{R}_+ \to \mathbb{R}_+$ is a twice-continuously-differentiable function satisfying $c'(q) > 0$ for $q > 0$, $c'' > 0$, and $c(0) = c'(0) = 0$. If the principal charges the agent a total of $p \in \mathbb{R}$ in exchange for $q \in \mathbb{R}_+$ units of the good, then her net payoff is $p - c(q)$.

The principal has discretion over deciding *when* the agent receives the allocation. The agent prefers obtaining the allocation sooner rather than later and discounts future

[6] The nouns "principal," "monopolist," and "seller," as well as the nouns "agent," "consumer," and "buyer," will be used synonymously throughout the paper.

[7] This utility specification satisfies the usual single-crossing condition:

$$\frac{\partial^2 u(q, \theta)}{\partial q \partial \theta} = v'(q) > 0 \quad \forall (q, \theta) \in \mathbb{R}_+ \times \Theta.$$

payoffs at a rate $\delta \in (0,1]$. Likewise, the principal discounts future payoffs at a rate $\beta \in (0,1]$. The set of admissible dates at which the good can be delivered is given as $T \equiv [0, \bar{t}]$, where $\bar{t} > 0$.[8] Consequently, if an agent of type $\theta \in \Theta$ receives $q \in \mathbb{R}_+$ units of the good at time $t \in T$ in exchange for a transfer payment of $p \in \mathbb{R}$ units of money, then his discounted utility is given as $\delta^t[\theta v(q) - p]$, and the principal's discounted payoff is $\beta^t[p - c(q)]$.[9]

2.2 Mechanisms with Delayed Allocations

Taking account of the fact that the principal and the agent discount future payoffs, I consider a concept of mechanism according to which the principal seeks to implement not only a collection of transfer–allocation menus $(p(\theta), q(\theta))_{\theta \in \Theta}$, but also a schedule $(t(\theta))_{\theta \in \Theta}$ determining when different types of the agent receive the prescribed allocation.

Definition 1 *A **mechanism with delayed allocations** (or simply a **mechanism**) is a Borel-measurable function $(p, q, t) : \Theta \to \mathbb{R} \times \mathbb{R}_+ \times T$, where p is a **transfer policy**, q is an **allocation policy**, and t is a **schedule policy**.*

Let $\mathcal{M}$ denote the space of all mechanisms with delayed allocations. In order to implement a mechanism with delayed allocations, the principal must ensure that the incentive-compatibility and voluntary-participation constraints are satisfied. That is, if the agent is of type $\theta \in \Theta$, then he must not have an incentive to choose the transfer–allocation–schedule triple $(p(\widehat{\theta}), q(\widehat{\theta}), t(\widehat{\theta}))$ designed for a different type $\widehat{\theta} \in \Theta \setminus \{\theta\}$ over the one $(p(\theta), q(\theta), t(\theta))$ designed for him. Moreover, each type's net payoff must be at least as large as the agent's outside option (which is normalized to 0, relying on the implicit assumption that the agent can always refuse to purchase anything).[10]

Definition 2 *A mechanism with delayed allocations $(p, q, t) \in \mathcal{M}$ is **implementable** if the following incentive-compatibility and participation constraints are satisfied:*

$$\delta^{t(\theta)}[\theta v(q(\theta)) - p(\theta)] \geq \delta^{t(\widehat{\theta})}[\theta v(q(\widehat{\theta})) - p(\widehat{\theta})] \quad \forall \theta, \widehat{\theta} \in \Theta, \tag{1}$$

$$\delta^{t(\theta)}[\theta v(q(\theta)) - p(\theta)] \geq 0 \quad \forall \theta \in \Theta. \tag{2}$$

From now on, let $\mathcal{I} \subseteq \mathcal{M}$ denote the space of those mechanisms with delayed allocations that are implementable. The following lemma characterizes implementability in terms an envelope condition, a monotonicity condition, and participation of the lowest-type agent:

[8] The exogenous upper bound on the set of admissible delivery times—meaning that the principal is not allowed to postpone shipment indefinitely—is imposed chiefly for technical reasons. The only requirement is that the cap $\bar{t}$ be finite; apart from that, it can be set as large as desired.

[9] I assume that the agent has to pay for the allocation at the same he receives it and the principal has to pay for the inputs used in production at the time the product is delivered. Relaxing this assumption might give rise to opportunistic waiting for prices to drop on the agent's part, a complication well-known in online mechanism design (Friedman and Parkes, 2003).

[10] By the revelation principle, there is no loss of generality in restricting attention to direct-revelation mechanisms.

Lemma 1 *Consider a mechanism with delayed allocations* $(p, q, t) \in \mathcal{M}$ *and let*

$$U(\theta) \equiv \delta^{t(\theta)}[\theta v(q(\theta)) - p(\theta)] \quad \forall \theta \in \Theta \tag{3}$$

denote the agent's utility when he reports his type truthfully. Then, (p, q, t) *is implementable if and only if the following conditions hold:*

(i) (Envelope condition) The function U *satisfies*

$$U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \delta^{t(s)} v(q(s))\, \mathrm{d}s \quad \forall \theta \in \Theta. \tag{4}$$

(ii) (Monotonicity) The function $\theta \mapsto \delta^{t(\theta)} v(q(\theta))$ *is non-decreasing.*

(iii) (Participation of the lowest-type agent) $U(\underline{\theta}) \geq 0$*.*

Proof See the Appendix. ■

If $\delta = 1$ (that is, the agent is perfectly patient), then the requirement that the function $\theta \mapsto \delta^{t(\theta)} v(q(\theta))$ be non-decreasing is equivalent to the usual monotonicity condition imposed on the allocation policy q, given that the utility function v is strictly increasing. However, if $\delta < 1$ (that is, the agent is impatient), then the monotonicity of the allocation policy q is no longer guaranteed *a priori*: what matters in this case is that the *discounted* utility of consuming the good (normalized by the type parameter) must be monotone.[11]

It is worth emphasizing that the principal must be able to commit to the mechanism she seeks to implement. This is because once the principal learns the agent's type, waiting to deliver the allocation is inefficient *ex post*. Consider an implementable mechanism $(p, q, t) \in \mathcal{I}$ and suppose that $t(\theta) > 0$ for some $\theta \in \Theta$. If the principal and the agent are impatient ($\beta < 1$ and $\delta < 1$), then they would both benefit from having the transfer–allocation pair $(p(\theta), q(\theta))$ delivered at an earlier date $\tau \in [0, t(\theta))$, provided that both the principal's profits $p(\theta) - c(q(\theta))$ and the agent's net utility from the good $\theta v(q(\theta)) - p(\theta)$ are positive. If *ex-post* renegotiation of the particular outcome $(p(\theta), q(\theta), t(\theta))$ of the mechanism between the principal and the agent is permitted, then the concept of implementability becomes precarious.[12] For this reason, I assume that the principal has the power to commit to delivering the allocation precisely at the date specified by the mechanism and not sooner.

With this assumption on commitment in mind, the timeline of the model can be described as follows:

---

[11] Indeed, there exist incentive-compatible mechanisms such that $\theta > \widehat{\theta}$, but $q(\theta) < q(\widehat{\theta})$, so that some agent of a higher type receives less than another agent of a lower type—although the higher type must necessarily obtain the good sooner in order to conform to the monotonicity of normalized discounted utility: $t(\theta) < t(\widehat{\theta})$. Likewise, it is possible in principle for an agent of higher type to be offered consumption at a later date: $t(\theta) > t(\widehat{\theta})$, but he must then be offered more: $q(\theta) > q(\widehat{\theta})$.

[12] This problem is reminiscent of the conjecture of Coase (1972), the connection to which I will revisit *infra* in Subsection 4.3.

(1) Nature draws a type $\theta \in \Theta$ according to the density f and reveals it privately to the agent.

(2) The principal offers the agent a mechanism $(p, q, t) \in \mathcal{M}$.

(3) The agent reports a type $\widehat{\theta} \in \Theta$ to the principal.

(4) At time $t(\widehat{\theta})$, the principal delivers the agent $q(\widehat{\theta})$ units of the good and the agent pays the principal $p(\widehat{\theta})$ units of money. Discounted payoffs are realized.

2.3 Optimal Mechanisms

In order to find the mechanism that yields the greatest *ex-ante* expected discounted profits among all implementable mechanisms with delayed allocations, the principal seeks to solve the following problem:

$$\sup_{(p,q,t)\in\mathcal{I}} \left\{ \int_{\underline{\theta}}^{\overline{\theta}} \beta^{t(\theta)}[p(\theta) - c(q(\theta))] \, \mathrm{d}F(\theta) \right\} \tag{5}$$

Definition 3 *An implementable mechanism $(p^\circ, q^\circ, t^\circ) \in \mathcal{I}$ is said to be **optimal** if it solves the problem (5) and the supremum of the objective is attained.*

3 Patient Principal

In this section, I show that if the principal is perfectly patient ($\beta = 1$), but the agent is not ($\delta < 1$), then the principal can implement the same allocation as though she *exactly* knew the agent's type and, at the same time, she can expropriate *almost all* of the surplus from trade. Intuitively, the principal can delay allocations designed to lower types, so that the desirability of higher types' endogenous alternative options decreases and almost all of their information rents are "inflated away." This stark result reveals the extreme power with which consumers' impatience can endow the monopolist, who can exploit this impatience by introducing dynamic shipping policies.

3.1 Efficiency

For each given type, an efficient level of the provision of the good is characterized by the property that it maximizes the surplus from trade between the principal and the agent. This surplus stems from overall social welfare brought about by consumption and production: the agent's utility from enjoying the good net of the principal's costs of producing it.

Definition 4 *A Borel-measurable function $q^* : \Theta \to \mathbb{R}_+$ is said to be an **efficient allocation** if*

$$q^*(\theta) \in \arg\max_{q\in\mathbb{R}_+} \{\theta v(q) - c(q)\} \quad \forall \theta \in \Theta.$$

In order to ensure the existence and interiority of efficient allocations, I impose the following assumption:

Assumption 1 *The functions v and c are such that*

$$\lim_{q\to\infty}\left\{\frac{c'(q)}{v'(q)}\right\} = \infty.$$

Under this assumption, one can establish the following:

Proposition 1 *There exists a unique efficient allocation q^*. Moreover, $q^*(\theta) > 0$ for all $\theta \in \Theta$ and q^* is strictly increasing.*

Proof Fix $\theta \in \Theta$. Take $\overline{q} > 0$ so large that

$$\frac{c'(q)}{v'(q)} > \theta$$

for all $q \geq \overline{q}$, which is possible by Assumption 1. Since

$$\theta v'(q) - c'(q) < 0 \quad \forall q \geq \overline{q}, \tag{6}$$

one has that $\theta v(q) - c(q) < \theta v(\overline{q}) - c(\overline{q})$ for all $q > \overline{q}$. Since $q \mapsto \theta v(q) - c(q)$ is continuous on the compact interval $[0, \overline{q}]$, it has a maximum on this interval, which thus must be a maximum on $\mathbb{R}_+$ as well. By (6), no maximum can occur at $q = \overline{q}$. Similarly, $\theta v'(0) - c'(0) = \theta v'(0) > 0$ implies that no maximum occurs at $q = 0$. Therefore, the search for maxima can be restricted to the open interval $(0, \overline{q})$, on which the function is strictly concave. Hence, the maximum is unique and characterized by the first-order condition

$$\theta v'(q^*(\theta)) - c'(q^*(\theta)) = 0.$$

To see that q^* is strictly increasing, use the implicit-function theorem:

$$v'(q^*(\theta)) + \left[\theta v''(q^*(\theta)) - c''(q^*(\theta))\right]\frac{\mathrm{d}q^*(\theta)}{\mathrm{d}\theta} = 0,$$

from which it follows that

$$\frac{\mathrm{d}q^*(\theta)}{\mathrm{d}\theta} = -\frac{v'(q^*(\theta))}{\theta v''(q^*(\theta)) - c''(q^*(\theta))} > 0,$$

given that $v' > 0$, $v'' < 0$, and $c'' > 0$. ■

3.2 Full Surplus Extraction with Complete Information

Suppose for a moment that both the principal and the agent are perfectly patient ($\beta = \delta = 1$) and that the principal is able to directly observe the agent's type. Because of perfect patience, the timing of the allocation no longer plays a role. Moreover, given the principal's complete information about the agent's type, the incentive-compatibility constraints are no longer relevant and can be omitted—only voluntary participation must be ensured. Formally, the principal seeks to solve the following problem:

$$\sup_{(p,q):\Theta\to\mathbb{R}\times\mathbb{R}_+}\left\{\int_{\underline{\theta}}^{\overline{\theta}}[p(\theta) - c(q(\theta))]\,\mathrm{d}F(\theta)\right\} \tag{7}$$

$$\text{s.t.} \quad \theta v(q(\theta)) - p(\theta) \geq 0 \quad \forall \theta \in \Theta,$$
$$(p, q) : \Theta \to \mathbb{R} \times \mathbb{R}_+ \text{ is Borel measurable.}$$

The allocation implemented by the principal is then the efficient allocation and the transfer payments are such that no surplus is left for the agent of any type.

Proposition 2 *The efficient allocation q^*, together with the transfer policy*

$$p^*(\theta) \equiv \theta v(q^*(\theta)) \quad \forall \theta \in \Theta$$

that leads to the expropriation of all of the surplus from trade by the principal, is a solution to (7).

Proof In optimum, the participation constraint must bind for (almost) all $\theta \in \Theta$. Plugging the binding participation constraint into the objective leads to

$$\int_{\underline{\theta}}^{\overline{\theta}} [\theta v(\theta) - c(q(\theta))] \, \mathrm{d}F(\theta).$$

Pointwise maximization of the integrand leads to the desired conclusion. ■

Now suppose that the principal offers the efficient allocation q^* but, instead of expropriating *all* of the surplus, she leaves the agent with a small net utility of $\varepsilon \geq 0$ uniformly across all types. I call such transfer–allocation policies *ε-optimal*:

Definition 5 *Let $\varepsilon \geq 0$ and q^* be the efficient allocation, and define*

$$p^*_\varepsilon(\theta) = \theta v(q^*(\theta)) - \varepsilon.$$

*Then, the transfer–allocation policy $(p^*_\varepsilon, q^*) : \Theta \to \mathbb{R} \times \mathbb{R}_+$ is said to be **ε-optimal**.*

3.3 Approximate Surplus Extraction with Incomplete Information

If the principal had complete information about the agent's type, then—as Proposition 2 has shown—she could provide the efficient allocation q^* while expropriating all of the surplus. To wit, she would be able to implement a 0-optimal policy according to the terminology introduced in Definition 5. The next result reveals that consumers' impatience endows a fully patient monopolist with almost as much effective power as complete information: even if the agent's type is private information, the principal can get arbitrarily close to expropriating all of the surplus from all agent types by relying on the agent's intertemporal preferences.

Proposition 3 *Suppose that $\beta = 1$ and $\delta < 1$. For a given $\varepsilon > 0$, let (p^*_ε, q^*) be an ε-optimal policy. Then, there exists a Borel-measurable schedule policy $t^*_\varepsilon : \Theta \to T$ such that $(p^*_\varepsilon, q^*, t^*_\varepsilon)$ is an implementable mechanism with delayed allocations. Such a schedule policy is given as follows:*[13]

$$t^*_\varepsilon(\theta) \equiv \int_\theta^{\overline{\theta}} \frac{v(q^*(s))}{(-\log \delta)\varepsilon} \, \mathrm{d}s \quad \forall \theta \in \Theta. \tag{8}$$

[13] In order to ensure that t^*_ε takes values in $T \equiv [0, \bar{t}]$, one would potentially need to increase the exogenous cap $\bar{t}$ for smaller values of ε.

Proof Fix $\varepsilon > 0$ and a ε-optimal transfer–allocation policy (p^*_ε, q^*). By the definition of ε-optimality, one has that

$$\theta v(q^*(\theta)) - p^*_\varepsilon(\theta) = \varepsilon > 0,$$

so that the participation constraint (2) is satisfied.

Now fix $\theta, \widehat{\theta} \in \Theta$ and consider the right-hand side of the incentive-compatibility constraint (1) for $(p^*_\varepsilon, q^*, t^*_\varepsilon)$:

$$\delta^{t^*_\varepsilon(\widehat{\theta})}[\theta v(q^*(\widehat{\theta})) - p^*_\varepsilon(\widehat{\theta})] = \exp\left[-\frac{1}{\varepsilon}\int_{\widehat{\theta}}^{\overline{\theta}} v(q^*(s))\,\mathrm{d}s\right]\left[(\theta - \widehat{\theta})v(q^*(\widehat{\theta})) + \varepsilon\right].$$

Differentiating this quantity with respect to $\widehat{\theta}$ yields, after some rearrangements, the following:

$$\exp\left[-\frac{1}{\varepsilon}\int_{\widehat{\theta}}^{\overline{\theta}} v(q^*(s))\,\mathrm{d}s\right](\theta - \widehat{\theta})\left[\frac{1}{\varepsilon}v(q^*(\widehat{\theta}))^2 + v'(q^*(\widehat{\theta}))\frac{\mathrm{d}q^*(\widehat{\theta})}{\mathrm{d}\widehat{\theta}}\right].$$

Given that $v' > 0$ and that the derivative of q^* is strictly positive (see Proposition 1), this expression is positive if $\widehat{\theta} < \theta$, is negative if $\widehat{\theta} > \theta$, and vanishes if $\widehat{\theta} = \theta$. This implies that the right-hand side of (1),

$$\delta^{t^*_\varepsilon(\widehat{\theta})}[\theta v(q^*(\widehat{\theta})) - p^*_\varepsilon(\widehat{\theta})]$$

is maximized at $\widehat{\theta} = \theta$, so that incentive compatibility is satisfied. ■

From (8), it is evident that $t^*_\varepsilon(\overline{\theta}) = 0$, so that the highest type receives the allocation immediately. Observe also that t^*_ε is non-increasing, so that lower types obtain the allocation later. Delaying the provision of the efficient allocation q^* for lower types is a convenient way for the principal to make downward misreports unprofitable, thereby "inflating away" almost all of higher types' information rents and increasing the amount of virtual surplus she can extract from them.[14]

The economic intuition behind the formula (8) can be understood as follows. Suppose for a moment that the ε-optimal transfer–allocation policy (p^*_ε, q^*) is subject to no delay. Then, by construction, an agent of type $\theta \in \Theta$ gains a net utility level of ε. If he misreports his type to be $\widehat{\theta} \in \Theta$, then his net utility is

$$\theta v(q^*(\widehat{\theta})) - p^*_\varepsilon(\widehat{\theta}) = (\theta - \widehat{\theta})v(q^*(\widehat{\theta})) + \widehat{\theta}v(q^*(\widehat{\theta})) - p^*_\varepsilon(\widehat{\theta}) = (\theta - \widehat{\theta})v(q^*(\widehat{\theta})) + \varepsilon,$$

[14] A caveat ought to be mentioned at this point. If the principal and the agent discount future payoffs differently because the former is patient and the latter is not, then the implicit assumption of transferable utility underlying in the notion of total social surplus is precarious—a payment made at a given future date affects the two parties' discounted utilities differently. For this reason, it is worth emphasizing that what the notion of "efficiency" of the allocation q^* represents in this context is that this is the *profit-maximizing* allocation that the principal would implement if she had complete information about the agent's type. It is *not* meant to imply that this allocation is *welfare-maximizing*, all the more so because the principal's welfare and the agent's welfare are no longer directly comparable due to the differences between their intertemporal preferences.

which is strictly greater than ε provided that $\widehat{\theta} < \theta$. Hence, the ε-optimal transfer–allocation policy (p^*_ε, q^*) is not implementable without delays, as all types (except for the lowest one) have an incentive to distort their type reports downwards. Suppose now that the principal tries implementing the ε-optimal transfer–allocation policy (p^*_ε, q^*) using a schedule policy $t^*_\varepsilon : \Theta \to T$. Letting $X^*_\varepsilon(\theta) \equiv \delta^{t^*_\varepsilon(\theta)}$, the utility of type θ is $X^*_\varepsilon(\theta)\varepsilon$ if he tells the truth, whereas it is

$$X^*_\varepsilon(\widehat{\theta})[(\theta - \widehat{\theta})v(q^*(\widehat{\theta})) + \varepsilon]$$

if he reports $\widehat{\theta} < \theta$ instead. Letting $\Delta(\widehat{\theta}, \theta)$ denote the gain from distorting the report downward, the benefit of misreport normalized by the size of the misreport is given as

$$\frac{\Delta(\widehat{\theta}, \theta)}{\theta - \widehat{\theta}} = \frac{X^*_\varepsilon(\widehat{\theta})[(\theta - \widehat{\theta})v(q^*(\widehat{\theta})) + \varepsilon] - X^*_\varepsilon(\theta)\varepsilon}{\theta - \widehat{\theta}} = X^*_\varepsilon(\widehat{\theta})v(q^*(\widehat{\theta})) - \frac{X^*_\varepsilon(\theta) - X^*_\varepsilon(\widehat{\theta})}{\theta - \widehat{\theta}}\varepsilon.$$

Now, "local" incentive compatibility requires that this "per-unit" benefit of misreport be of "second order" and vanish as $\widehat{\theta}$ draws close to θ.[15] Assuming that X^*_ε is differentiable, this requirement can be expressed as

$$\lim_{\widehat{\theta} \uparrow \theta} \frac{\Delta(\widehat{\theta}, \theta)}{\theta - \widehat{\theta}} = X^*_\varepsilon(\theta)v(q^*(\theta)) - \frac{\mathrm{d}X^*_\varepsilon(\theta)}{\mathrm{d}\theta}\varepsilon = 0.$$

This differential equation has the solution

$$X^*_\varepsilon(\theta) = X^*_\varepsilon(\overline{\theta}) \exp\left[-\frac{1}{\varepsilon}\int_\theta^{\overline{\theta}} v(q^*(s))\,\mathrm{d}s\right].$$

Since there is no reason to distort the highest type's date of delivery (this would only result in a uniform upward shift in the schedule policy without affecting reporting incentives), one can set $t^*_\varepsilon(\overline{\theta}) = 0$, so that $X^*_\varepsilon(\overline{\theta}) = 1$. Taking logarithms and recalling that $X^*_\varepsilon(\theta) = \delta^{t^*_\varepsilon(\theta)}$ yield (8).

4 Impatient Principal

Now I relax the assumption imposed on the patience of the principal and consider again the general case in which $\beta \in (0, 1]$ and $\delta \in (0, 1]$. In order to represent the problem of finding the optimal mechanism (according to Definition 3) in a more tractable form, I first define an auxiliary problem that provides an alternative characterization of optimality.

4.1 Auxiliary Problem

I begin by defining a function $J : \mathbb{R}_+ \to \mathbb{R}_+$ as $J(x) \equiv c \circ v^{-1}(x)$ for all $x \geq 0$.[16] For later reference, it is useful to record several properties of this function:

---

[15]Using Landau's order notation, it must be the case that $\Delta(\widehat{\theta}, \theta) = o(\theta - \widehat{\theta})$.

[16]Since v is strictly increasing, $v(0) = 0$, and $\lim_{q\to\infty} v(q) = \infty$, it follows that v is bijective and hence invertible.

Lemma 2 *The function $J \equiv c \circ v^{-1}$ satisfies the following conditions:*

(i) It is twice-continuously-differentiable, $J(0) = J'(0) = 0$, $J'(x) > 0$ for $x > 0$, and $J'' > 0$.

(ii) For any $w \geq 0$, the function

$$y \mapsto y \times J\left(\frac{w}{y}\right)$$

is non-increasing on the interval $(0, \infty)$. In fact, it is constant if $w = 0$ and it is strictly decreasing if $w > 0$.

(iii) If $w > 0$, then

$$\lim_{y \downarrow 0} \left\{ y \times J\left(\frac{w}{y}\right) \right\} = \infty. \tag{9}$$

Proof See the Appendix. ■

I now introduce an alternative space of decision variables for the principal. Let[17]

$$\begin{aligned} \mathcal{W} \equiv \{(W, t) \mid W : \Theta \to \mathbb{R}_+ \text{ is non-decreasing,} \\ t : \Theta \to T \text{ is Borel measurable}\}. \end{aligned}$$

The main idea behind this alternative space is that for an implementable mechanism $(p, q, t) \in \mathcal{I}$, W substitutes for the agent's normalized discounted utility from the good:

$$W(\theta) \equiv \delta^{t(\theta)} v(q(\theta)) \quad \forall \theta \in \Theta,$$

which must be non-decreasing by Lemma 1. With this alternative space of decision variables, I define an *auxiliary problem* as follows:

Definition 6 *The **auxiliary problem** is defined as the following maximization program:*

$$\sup_{(W,t) \in \mathcal{W}} \left\{ \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta) \left[\theta W(\theta) - \int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s - \delta^{t(\theta)} J\left(\frac{W(\theta)}{\delta^{t(\theta)}}\right) \right] \mathrm{d}\theta \right\}.$$

*A pair $(W^\circ, t^\circ) \in \mathcal{W}$ is said to **solve the auxiliary problem** if it attains the supremum.*

The following proposition establishes that solving the auxiliary problem and finding the optimal implementable mechanism are one and the same.

Proposition 4 *The optimal mechanism according to Definition 3 and the auxiliary problem according to Definition 6 are related as follows:*

[17] Note that the measurability of the function W need not be stipulated separately, as a monotone function is always Borel measurable (Folland, 1999, p. 48).

(i) Suppose that $(W^\circ, t^\circ) \in \mathcal{W}$ *solves the auxiliary problem and define*

$$q^\circ(\theta) \equiv v^{-1}\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}}\right) \quad \forall \theta \in \Theta,$$
$$p^\circ(\theta) \equiv \frac{1}{\delta^{t^\circ(\theta)}}\left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s\right] \quad \forall \theta \in \Theta.$$

Then, $(p^\circ, q^\circ, t^\circ) \in \mathcal{I}$ *and* $(p^\circ, q^\circ, t^\circ)$ *is an optimal mechanism.*

(ii) Conversely, suppose that $(p^\circ, q^\circ, t^\circ) \in \mathcal{I}$ *is an optimal mechanism and define*

$$W^\circ(\theta) \equiv \delta^{t^\circ(\theta)} v(q^\circ(\theta)) \quad \forall \theta \in \Theta.$$

Then, $(W^\circ, t^\circ) \in \mathcal{W}$ *and* (W°, t°) *solves the auxiliary problem.*

Proof See the Appendix. ■

Before proceeding, it is useful to rewrite the maximand corresponding to the auxiliary problem into a slightly different form that will prove useful later.

Lemma 3 *For any* $(W, t) \in \mathcal{W}$*, the objective of the auxiliary problem can be rewritten as follows:*

$$\int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta)\left[\theta W(\theta) - \int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s - \delta^{t(\theta)} J\left(\frac{W(\theta)}{\delta^{t(\theta)}}\right)\right]\mathrm{d}\theta.$$
$$= \int_{\underline{\theta}}^{\overline{\theta}} \left\{\left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta)\left[\theta W(\theta) - \delta^{t(\theta)} J\left(\frac{W(\theta)}{\delta^{t(\theta)}}\right)\right] - \left[\int_{\theta}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(s)} f(s)\,\mathrm{d}s\right] W(\theta)\right\}\mathrm{d}\theta. \quad (10)$$

Proof Lebesgue – Stieltjes integration by parts for absolutely continuous functions (see Folland, 1999, p. 108)[18] yields that

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[\int_{\theta}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(s)} f(s)\,\mathrm{d}s\right] W(\theta)\,\mathrm{d}\theta =$$
$$= \underbrace{\left[\int_{\theta}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(s)} f(s)\,\mathrm{d}s\right]\left[\int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s\right]\Bigg|_{\theta=\underline{\theta}}^{\theta=\overline{\theta}}}_{=0} + \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta)\left[\int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s\right]\mathrm{d}\theta,$$

[18] Note that both of the functions

$$\theta \mapsto \int_{\theta}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(s)} f(s)\,\mathrm{d}s \quad \text{and} \quad \theta \mapsto \int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s$$

are absolutely continuous on Θ by Proposition 3.35 in Folland (1999). This is because the integrand $s \mapsto W(s)$ is measurable and bounded (given that it is non-decreasing on a compact interval), whereas the integrand $s \mapsto (\beta/\delta)^{t(s)} f(s)$ is measurable and integrable, given the upper bound on the delay $\overline{t}$.

which completes the proof. ■

It is intuitively clear that the optimal mechanism cannot involve cross-subsidization, with or without delay. That is, it is not optimal for the principal to offer such allocations that she runs losses on some types in order to extract greater profits from other types. After all, the principal would prefer shutting down unprofitable transfer–allocation pairs rather than running losses on them. Technically speaking, this condition amounts to requiring that the maximand of the auxiliary problem in Definition 6 be *pointwise* non-negative, which is equivalent to the lack of cross-subsidization given the duality result embodied in Proposition 4. The following claim formalizes this intuition:

Proposition 5 *Suppose that the pair $(W^\circ, t^\circ) \in \mathcal{W}$ solves the auxiliary problem. Then, the pointwise profit function*

$$\theta \mapsto \left(\frac{\beta}{\delta}\right)^{t^\circ(\theta)} \left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s - \delta^{t^\circ(\theta)} J\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}}\right)\right]$$

is non-negative for every $\theta \in (\underline{\theta}, \overline{\theta}]$.

Proof See the Appendix. ■

4.2 Standard Mechanism

If one assumes that $\beta = \delta = 1$ (that is, both the principal and the agent are perfectly patient), then the timing of the transfer–allocation pair loses importance, and the problem of finding the optimal allocation becomes the classical model of Mussa and Rosen (1978), which I call the *standard mechanism*. Given the equivalence result of Proposition 4, the function in the auxiliary problem W can be identified with $v \circ q$, and the monotonicity condition on W becomes equivalent to the usual monotonicity condition on the allocation policy q. The particular choice of the schedule policy is irrelevant, so that it is without loss of generality to assume that delivery occurs immediately. Keeping these simplifications in mind and using the alternative formula for the maximand of the auxiliary problem in Lemma 3, the standard mechanism can be characterized as follows:

Definition 7 *The **standard mechanism** $(p^{\mathrm{st}}, q^{\mathrm{st}}, t^{\mathrm{st}})$ is the optimal mechanism in the case in which $\beta = \delta = 1$. It satisfies*

$$q^{\mathrm{st}} \in \arg\max_{q} \left\{ \int_{\underline{\theta}}^{\overline{\theta}} f(\theta) \left[\theta v(q(\theta)) - c(q(\theta)) - \frac{1 - F(\theta)}{f(\theta)} v(q(\theta))\right] \mathrm{d}\theta \right\} \tag{11}$$
$$\text{s.t.} \quad q : \Theta \to \mathbb{R}_+ \text{ is non-decreasing,}$$

with

$$p^{\mathrm{st}}(\theta) = \theta v(q^{\mathrm{st}}(\theta)) - \int_{\underline{\theta}}^{\theta} v(q^{\mathrm{st}}(s))\,\mathrm{d}s \quad \forall \theta \in \Theta,$$

and $t^{\mathrm{st}}(\theta) = 0$ for all $\theta \in \Theta$.

As usual, if the agent's "virtual type" $\theta \mapsto \theta - [1 - F(\theta)]/f(\theta)$ is non-decreasing, then the monotonicity constraint on q^{st} can be omitted and the optimal allocation can be found through pointwise maximization of the integrand in (11). Otherwise, the optimal allocation policy involves bunching (that is, multiple types being offered the same allocation) and the monotonicity constraint must be explicitly incorporated via ironing techniques (Baron and Myerson, 1982; Myerson, 1981; Mussa and Rosen, 1978; Toikka, 2011).

4.3 Very Impatient Principal

In this subsection, I characterize the optimal mechanism in the case in which the principal is no more patient than the agent, that is $\beta \leq \delta \leq 1$. Intuition suggests that since the principal discounts future payoffs to a greater extent than does the agent, the former cannot benefit from delaying the allocation. This turns out to be precisely the case:

Proposition 6 *If $\beta \leq \delta \leq 1$, then the standard mechanism of Definition 7 is an optimal mechanism.*

Proof Assume first that $\beta = \delta$ and consider any pair $(W, t) \in \mathcal{W}$ for the auxiliary problem. Using Lemma 3, the payoff to (W, t) is given as

$$\int_{\underline{\theta}}^{\overline{\theta}} \left\{ f(\theta) \left[\theta W(\theta) - \delta^{t(\theta)} J\left(\frac{W(\theta)}{\delta^{t(\theta)}} \right) \right] - \left(\int_{\theta}^{\overline{\theta}} f(s)\, \mathrm{d}s \right) W(\theta) \right\} \mathrm{d}\theta. \tag{12}$$

This expression is maximized pointwise by setting $\delta^{t(\theta)} J(W(\theta)/\delta^{t(\theta)})$ as low as possible, and, for a given $W(\theta) \geq 0$, this is achieved by setting $t(\theta) = 0$ by Lemma 2. Identifying W with $v \circ q$ (see Proposition 4) and using the definition of $J = c \circ v^{-1}$, the expression in (12) becomes

$$\int_{\underline{\theta}}^{\overline{\theta}} f(\theta) \left[\theta v(q(\theta)) - c(q(\theta)) - \frac{1 - F(\theta)}{f(\theta)} v(q(\theta)) \right] \mathrm{d}\theta.$$

Maximizing this expression with respect to q subject to the restriction that W and thus q must be non-decreasing gives precisely the standard mechanism.

Now suppose that $\beta < \delta$ and let $(W^\circ, t^\circ) \in \mathcal{W}$ be a solution to the auxiliary problem. One then has that the maximum value of the auxiliary problem satisfies

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta} \right)^{t^\circ(\theta)} f(\theta) \left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\, \mathrm{d}s - \delta^{t^\circ(\theta)} J\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}} \right) \right] \mathrm{d}\theta \\
&\quad \leq \int_{\underline{\theta}}^{\overline{\theta}} f(\theta) \left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\, \mathrm{d}s - \delta^{t^\circ(\theta)} J\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}} \right) \right] \mathrm{d}\theta \\
&\quad \leq \int_{\underline{\theta}}^{\overline{\theta}} f(\theta) \left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\, \mathrm{d}s - J(W^\circ(\theta)) \right] \mathrm{d}\theta \\
&\quad \leq \int_{\underline{\theta}}^{\overline{\theta}} f(\theta) \left[\theta W^{\text{st}}(\theta) - \int_{\underline{\theta}}^{\theta} W^{\text{st}}(s)\, \mathrm{d}s - J(W^{\text{st}}(\theta)) \right] \mathrm{d}\theta,
\end{aligned}$$

where $W^{\mathrm{st}} \equiv v \circ q^{\mathrm{st}}$ corresponds to the agent's type-normalized utility from consuming the good in the standard mechanism. The first inequality stems from the facts that (i) $\beta/\delta < 1$; and (ii) a solution to the auxiliary problem must involve non-negative profits *pointwise* (except possibly at $\underline{\theta}$) by Proposition 5. The second inequality is due to Lemma 2. The third inequality follows from the definition of the standard mechanism. Hence, the pair $(W^{\mathrm{st}}, 0)$ corresponding to the standard mechanism in the auxiliary problem does at least as well as (W°, t°). Clearly, $(W^{\mathrm{st}}, 0)$ is feasible. Therefore, it must be a solution to the auxiliary problem and thus the standard mechanism is optimal. ■

The intuition behind this proposition is fairly clear. The only reason the principal would want to delay the delivery of an allocation is to induce high-type agents not to choose allocations designed for low-type agents. Delayed allocations relax incentive constraints and inflate high-type agents' information rents away. However, if the principal is less patient than the agent, then she would inflate away her own monopoly profits even more quickly by delaying the allocation. Hence, the best a very impatient principal can do is stick with the standard mechanism and immediate delivery.

The result that the optimal allocation involves no delay bears a passing resemblance to the result by Stokey (1981); however, the underlying intuition is quite different. Stokey (1981) corroborates the conjecture of Coase (1972): A monopolist loses her market power if she cannot commit to refraining from selling additional units of a durable good once she has already supplied the profit-maximizing quantity. As a result, the only equilibrium that involves buyers having perfectly rational expectations (in the sense that the monopolist maximizes profits by fulfilling buyers' expectations even off the equilibrium trajectory) that satisfy a continuity requirement is one in which the monopolist saturates the market immediately.[19] The conclusions of Stokey (1981) are driven by lack of commitment: the monopolist effectively plays a game not only against the buyers but also against her future selves, who would find it optimal to keep selling the good beyond the profit-maximizing level. The buyers foresee this pattern, and refuse to buy the good above the price that would prevail in a competitive market. By contrast, the monopolist has perfect commitment by assumption in the current model. If she is at most as patient as the agent, then the benefits of enhancing price discrimination by devaluing the endogenous alternative options of higher-type agents and depreciating their information rents are more than offset by the costs of delaying the allocation in terms of postponed profits. It is simply because of the unprofitability of delays—and not the lack of commitment—that an impatient principal cannot improve upon the standard mechanism and chooses to deliver the allocation immediately for all types instead.[20]

[19] Gul *et al.* (1986) reach essentially the same conclusion in a model in which the monopolist posts prices as opposed to determining quantities.

[20] Also related is the work of Gul and Sonnenschein (1988), who consider a bilateral alternating-offers bargaining problem, in which the buyer's valuation is private information and the seller and the buyer exhibit the same intertemporal preferences. They show that no significant delays in reaching an agreement occur as long as the time elapsed between offers is short.

4.4 Moderately Impatient Principal

The construction presented in Subsection 3.3 has shown that a perfectly patient monopolist selling to an impatient buyer can achieve almost-full expropriation of the surplus from trade by suitably delaying the allocation for lower types, whereas Subsection 4.3 has revealed that she cannot benefit from postponing the allocation if she is no more patient than the buyer. An interesting intermediate case is one in which the principal is more patient than the agent, but not perfectly patient; that is, $\delta < \beta < 1$. The purpose of this subsection is to provide a characterization of the optimal mechanism in this intermediate case and to show that a sufficiently, but not perfectly, patient principal can still benefit from delaying the allocation.

Observe that the equation (10) presented in Lemma 3 provides two equivalent formulae for the expected profits of the principal in the auxiliary problem. These two formulae can be used in tandem to characterize the optimal mechanism (invoking the equivalence result of Proposition 4). If $(W^\circ, t^\circ) \in \mathcal{W}$ solves the auxiliary problem, then, having t° fixed, the function W° must maximize the integral on the right-hand side of (10) among all non-decreasing functions $W : \Theta \to \mathbb{R}_+$. Also, having W° fixed, the function t° must maximize the integral on the left-hand side of (10) among all Borel-measurable functions $t : \Theta \to \mathbb{R}_+$. Still, pointwise maximization of the integrand on the right-hand side with respect to $W(\theta)$ with having $t^\circ(\theta)$ fixed is preposterous, because this procedure ignores the monotonicity constraint. However, since no monotonicity constraint is imposed on the schedule policy, the optimal policy t° can be found by maximizing the integrand on the left-hand side of (10) pointwise with respect to $t(\theta)$, having $W^\circ(\theta)$ fixed. The following proposition summarizes this conclusion.

Proposition 7 *Suppose that $\delta < \beta < 1$. If $(W^\circ, t^\circ) \in \mathcal{W}$ solves the auxiliary problem, then the following must hold:*

$$\begin{aligned}
&\beta^{t^\circ(\theta)}(-\log\delta)\frac{1}{\delta^{t^\circ(\theta)}}\left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s\right] \\
&\quad \lesseqgtr \beta^{t^\circ(\theta)}(-\log\delta) J'\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}}\right)\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}} \\
&\quad + \beta^{t^\circ(\theta)}(-\log\beta)\left\{\frac{1}{\delta^{t^\circ(\theta)}}\left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s\right] - J\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}}\right)\right\}
\end{aligned} \tag{13}$$

according as $t^\circ(\theta) = 0$, $t^\circ(\theta) \in (0, \bar{t})$, or $t^\circ(\theta) = \bar{t}$, respectively, for almost every $\theta \in \Theta$.

Proof Since no monotonicity restriction is imposed on the schedule policy, the integral on the left-hand side of (10) is maximized if and only if $t(\theta)$ maximizes the integrand for almost every $\theta \in \Theta$ given $W^\circ(\theta)$. Hence, the task is to maximize the function

$$t \mapsto \left(\frac{\beta}{\delta}\right)^{t}\left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s - \delta^t J\left(\frac{W^\circ(\theta)}{\delta^t}\right)\right] \tag{14}$$

on the compact interval $t \in [0, \bar{t}]$, for almost every $\theta \in \Theta$. Clearly, a maximum exists, given that the function in (14) is continuous. Since it is also differentiable, the necessary first-order condition must hold for the optimal value $t^\circ(\theta)$. This first-order condition can be computed to be (13) after some algebraic manipulations. ■

Intuitively, the first-order condition (13) expresses the unprofitability of a particular type of deviation from the optimal mechanism. In order to understand the underlying economic forces at play, it is useful to recall the equivalence results of Proposition 4, the definition of $J \equiv c \circ v^{-1}$, and the formula (24) for J' to rewrite (13) as follows:

$$\begin{aligned}&\beta^{t^\circ(\theta)}(-\log\delta)p^\circ(\theta)\\&\quad\le \beta^{t^\circ}(\theta)(-\log\delta)\frac{c'(q^\circ(\theta))}{v'(q^\circ(\theta))}v(q^\circ(\theta)) + \beta^{t^\circ(\theta)}(-\log\beta)[p^\circ(\theta)-c(q^\circ(\theta))]\end{aligned} \tag{15}$$

assuming that $t^\circ(\theta) \in [0,\bar{t})$, so that the exogenous upper bound $\bar{t}$ is locally irrelevant. Recall from Proposition 4 that $W^\circ(\theta) = \delta^{t^\circ(\theta)}v(q^\circ(\theta))$ is the agent's normalized discounted utility from consuming the good in the optimal mechanism. Now consider increasing $t^\circ(\theta)$ slightly while keeping the normalized discounted utility level $W^\circ(\theta)$ constant. Then, $q^\circ(\theta)$ must increase accordingly: to keep the normalized discounted utility level constant, a delayed delivery must be accompanied by an increased quantity of the good (*cf.* n. 11). This increased quantity makes it possible for the principal to charge a higher price, the marginal benefit of which is reflected by the left-hand side of (15).[21]

The marginal cost of introducing a slight delay is twofold. First, the principal's profits decrease directly due to discounting, which is reflected by the second term on the right-hand side of (15). Moreover, as argued above, a delay must be accompanied by an increased quantity of the good, producing which is costly: this is reflected in the first term on the right-hand side of (15).[22] In optimum, such a deviation from the

[21] Indeed, note that the agent's overall utility in the optimal mechanism is given as

$$U^\circ(\theta) = \delta^{t^\circ(\theta)}[\theta v(q^\circ(\theta)) - p^\circ(\theta)].$$

If this overall utility were to be unchanged by a small deviation from $t^\circ(\theta)$ to $t^\circ(\theta)+\Delta t$, then, since $W^\circ(\theta) = \delta^{t^\circ(\theta)}v(q^\circ(\theta))$ is unaffected, it must be the case that $\delta^{t^\circ(\theta)}p^\circ(\theta)$ is unchanged as well. A differential argument reveals that this requirement amounts to the following:

$$\delta^{t^\circ(\theta)}(\log\delta)p^\circ(\theta)\Delta t + \delta^{t^\circ(\theta)}\Delta p \approx 0,$$

or $\Delta p/\Delta t \approx (-\log\delta)p^\circ(\theta)$. This shows that the marginal benefit from increasing sales and introducing further delays at the same time while keeping $W^\circ(\theta)$ constant is approximately $(-\log\delta)p^\circ(\theta)$, whose discounted value for the principal is precisely the left-hand side of (15).

[22] If $W^\circ(\theta) = \delta^{t^\circ(\theta)}v(q^\circ(\theta))$ were to be kept constant, then the infinitesimal changes Δt and Δq ought to be related as follows:

$$\frac{\Delta q}{\Delta t} \approx (-\log\delta)\frac{v(q^\circ(\theta))}{v'(q^\circ(\theta))}. \tag{16}$$

Moreover, the corresponding change in the cost, originally at the level of $c(q^\circ(\theta))$, is approximately

$$c'(q^\circ(\theta))\Delta q. \tag{17}$$

schedule policy must be unprofitable, which implies that (15) must hold. In addition, it must hold with equality if $t^\circ(\theta) > 0$, because an analogous decrease in delays must be unprofitable as well.

The first-order condition given in Proposition 7 can be used to establish that the principal can benefit from introducing delays into the optimal mechanism if she is sufficiently patient, even if she is not perfectly patient. More specifically, the next result yields that if the standard mechanism satisfies some regularity conditions, then it cannot be optimal for a principal who is patient enough.

Proposition 8 *Suppose that the standard mechanism introduced in Definition 7 is such that the agent's "virtual type"*

$$\theta \mapsto \theta - \frac{1 - F(\theta)}{f(\theta)}$$

is non-decreasing and the principal obtains positive profits from every type. Then, there exists some $\beta_0 \in [\delta, 1)$ such that the standard mechanism is not optimal for a principal with any discount factor $\beta \in (\beta_0, 1)$.

Proof Define the virtual-type function $g : \Theta \to \mathbb{R}$ as

$$g(\theta) \equiv \theta - \frac{1 - F(\theta)}{f(\theta)} \quad \forall \theta \in \Theta,$$

and assume it is non-decreasing. Then, the monotonicity constraint corresponding to the standard mechanism can be omitted and the integrand of (11) can be maximized pointwise. Since the principal obtains positive profits from each type by assumption, it must be the case that $q^{\mathrm{st}}(\theta) > 0$ for each $\theta \in \Theta$, so that the first-order condition corresponding to pointwise maximization yields

$$g(\theta) v'(q^{\mathrm{st}}(\theta)) = c'(q^{\mathrm{st}}(\theta)) \quad \forall \theta \in \Theta.$$

Defining $W^{\mathrm{st}}(\theta) \equiv v(q^{\mathrm{st}}(\theta))$ and using the formula (24) for J', this first-order condition can be expressed as

$$g(\theta) = J'(W^{\mathrm{st}}(\theta)) \quad \forall \theta \in \Theta. \tag{18}$$

Since g is non-decreasing and J' is strictly increasing, this confirms that W^{st} is non-decreasing as well, so that the monotonicity constraint is redundant, indeed.

Define the profit function of the standard mechanism $\pi^{\mathrm{st}} : \Theta \to \mathbb{R}$ as

$$\pi^{\mathrm{st}}(\theta) \equiv p^{\mathrm{st}}(\theta) - c(q^{\mathrm{st}}(\theta)) = \theta W^{\mathrm{st}}(\theta) - \int_{\underline{\theta}}^{\theta} W^{\mathrm{st}}(s)\,\mathrm{d}s - J(W^{\mathrm{st}}(\theta)) \quad \forall \theta \in \Theta. \tag{19}$$

By assumption, $\pi^{\mathrm{st}}(\theta) > 0$ for all $\theta \in \Theta$. Fix $\delta < 1$ and let

$$\beta_0 \equiv \max\left\{\delta, \exp\left[\frac{W^{\mathrm{st}}(\underline{\theta})(\log \delta)}{f(\underline{\theta})\pi^{\mathrm{st}}(\underline{\theta})}\right]\right\}.$$

Combining (16) and (17) and discounting the resulting change in costs according to the principal's time preferences yield the first term on the right-hand side of (15).

Note that $\beta_0 < 1$, given that $\log \delta < 0$. I claim that if $\beta > \beta_0$, then the standard mechanism is not optimal.

To see this, suppose, for the sake of contradiction, that it is optimal. Then, by Proposition 4, $(W^{\mathrm{st}}, 0)$ solves the auxiliary problem. Applying Proposition 7 to $W^\circ = W^{\mathrm{st}}$, $t^\circ \equiv 0$, and $\theta = \underline{\theta}$, and using (18) and (19), one can see that the following inequality must hold:

$$(-\log \delta)\underline{\theta} W^{\mathrm{st}}(\underline{\theta}) \leq (-\log \delta) g(\underline{\theta}) W^{\mathrm{st}}(\underline{\theta}) + (-\log \beta)\pi^{\mathrm{st}}(\underline{\theta}).$$

Given that $g(\underline{\theta}) = \underline{\theta} - 1/f(\underline{\theta})$, one can further conclude that

$$\frac{(-\log \delta)}{f(\underline{\theta})} W^{\mathrm{st}}(\underline{\theta}) \leq (-\log \beta)\pi^{\mathrm{st}}(\underline{\theta}) < (-\log \beta_0)\pi^{\mathrm{st}}(\underline{\theta}) < \frac{(-\log \delta)}{f(\underline{\theta})} W^{\mathrm{st}}(\underline{\theta}),$$

which is a contradiction.

That is, the inequality (13) is violated for $(W^\circ, t^\circ) = (W^{\mathrm{st}}, 0)$ at $\theta = \underline{\theta}$. By the continuity of W^{st}, J, and J', this inequality will also be violated in some small open interval around $\underline{\theta}$. Hence, setting $t(\theta) = 0$ is *not* optimal for any $\theta \in \Theta$ in this interval, yielding that pointwise profits can be raised on a set of positive measure by introducing delays. This means that the standard mechanism cannot be optimal. ■

This result reveals that if the principal is sufficiently, but not necessarily perfectly, patient, then she can improve upon the standard mechanism, provided that the distribution of the agent's type is such that his virtual type is non-decreasing (so that ironing is not required) and the standard mechanism serves every type, yielding positive profits to the principal pointwise. The intuition for this result is as follows. In order for higher types not to choose cheaper allocations designed for lower types, the principal must keep quantity inefficiently scanty at the lower end of the type distribution in the standard mechanism. With the possibility of delayed deliveries, however, the principal can do better: she can increase quantity at the lower end of the distribution and postpone delivery at the same time in such a way that the overall utility from enjoying the good is unchanged for each type and incentive compatibility is preserved. Since payments are made at delivery, the price of such an increased allocation can be raised, too. If the principal is patient enough, then the additional charges she can extract from low-type consumers dominate the losses both due to the overall delay in profits she must endure and due to increased costs of production.

5 Implementation without Screening

Recall from Subsection 2.2 that the timeline of the model is such that the principal first announces a mechanism with delayed allocations, then she next screens the agent for his type, and finally the good is delivered at the time prescribed by the particular mechanism that the principal has chosen and to which she has committed herself. In the present section, I demonstrate an alternative way of implementation, in which screening is replaced with the principal offering a flow of transfer–allocation pairs over time, and the agent is free to accept the prevailing offer at any given time without revealing his type directly. In other words, the principal designs what is reminiscent of a Dutch auction of transfer–allocation pairs (*cf.* Hörner and Samuelson, 2011).

More concretely, consider any implementable mechanism $(p, q, t) \in \mathcal{I}$ (in the sense of Definition 2). For any instant of time $\tau \in T \equiv [0, \bar{t}]$, let

$$\Theta_\tau \equiv \{\theta \in \Theta \,|\, t(\theta) = \tau\}$$

be the set of those types for whom the schedule policy t stipulates that the good be shipped at time τ (maintaining the possibility that Θ_τ is empty). Define

$$M_\tau \equiv \{(p(\theta), q(\theta)) \,|\, \theta \in \Theta_\tau\} \subseteq \mathbb{R} \times \mathbb{R}_+,$$

with the convention that $M_\tau \equiv \varnothing$ if $\Theta_\tau = \varnothing$. That is, at each instant of time $\tau \in T$, M_τ is the collection of those transfer–allocation menus that are chosen by types for which the good is to be delivered at time τ according to the schedule policy t. The agent is free to choose any transfer–allocation menu from the set M_τ (if this set is non-empty), provided that he has not made a choice yet at an earlier time $\tau' \in [0, \tau)$, or choose nothing and let time pass further. According to this Dutch-auction-like setting of implementing the mechanism (p, q, t), the principal's problem reduces to designing a correspondence $M : T \to 2^{\mathbb{R} \times \mathbb{R}_+}$, according to which she makes it possible for the agent to choose any given menu from M_τ, or no menu at all, at any time $\tau \in T$. Given that the mechanism (p, q, t) is implementable, it is straightforward to see that an agent of type $\theta \in \Theta$ optimally picks the menu $(p(\theta), q(\theta))$ from $M_{t(\theta)}$ at time $t(\theta)$.

It continues to be crucial for the principal to have the power to commit herself to the Dutch-auction correspondence M from the outset in this modified setting, too. Indeed, as time passes, the principal can potentially learn more and more about the agent's type due to the sheer observation that he has not claimed the good yet. For example, if the agent has not bought the good up until time $\tau \in T$, then the principal can rule out types in the set $\bigcup_{\tau' \in [0,\tau)} \Theta_{\tau'}$ (as those types would have optimally claimed the good before) and update her prior. If the principal lacked commitment, then she would potentially have incentives to redesign the mechanism based on the posterior distribution of types in each instant of time, and implementation would unravel. In other words, the Dutch-auction correspondence associated with a given implementable mechanism may fail to be *time-consistent*, so that the principal's prior commitment is indispensable for implementation.

In certain cases, nonetheless, the principal is unable to derive substantial profits from redesigning the mechanism *ex post* and Dutch-auction-like implementation goes through even without commitment. One such case is the one in which the principal is perfectly patient. To show this, I first provide an operationalization of time consistency.

Let $(p^*, q^*, t^*) \in \mathcal{I}$ be an implementable, though not necessarily optimal, mechanism. For each $\tau \in T$, let

$$\overline{\Theta}^*_\tau \equiv \{\theta \in \Theta \,|\, t^*(\theta) \geq \tau\}$$

be the set of those types of the agent that have not claimed the good before time τ.[23]

[23]Note that $\overline{\Theta}^*_\tau$ is a Borel-measurable subset of Θ, as it is the pre-image of the interval $[\tau, \bar{t}]$ under t^*.

For any $\tau \in T$ for which the set $\overline{\Theta}^*_\tau$ has positive Lebesgue measure, let

$$f^*_\tau(\theta) \equiv \frac{f(\theta)}{\displaystyle\int_{\widehat{\theta}\in\overline{\Theta}^*_\tau} f(\widehat{\theta})\,\mathrm{d}\widehat{\theta}} \quad \forall\theta \in \overline{\Theta}^*_\tau$$

denote the posterior distribution of types from the point of view of the principal who knows that the agent has not claimed the good before τ. Furthermore, let $\mathcal{I}^*_\tau$ denote the set of those Borel-measurable functions $(p, q, t) : \overline{\Theta}^*_\tau \to \mathbb{R} \times \mathbb{R}_+ \times [\tau, \bar{t}]$ that satisfy incentive compatibility and voluntary participation when the type set is restricted to $\overline{\Theta}^*_\tau$;[24] that is,

$$\delta^{t(\theta)}[\theta v(q(\theta)) - p(\theta)] \geq \delta^{t(\widehat{\theta})}[\theta v(q(\widehat{\theta})) - p(\widehat{\theta})] \quad \forall\theta, \widehat{\theta} \in \overline{\Theta}^*_\tau,$$
$$\delta^{t(\theta)}[\theta v(q(\theta)) - p(\theta)] \geq 0 \quad \forall\theta \in \overline{\Theta}^*_\tau.$$

Definition 8 *For any $\varepsilon \geq 0$, an implementable mechanism $(p^*, q^*, t^*) \in \mathcal{I}$ is said to be **ε-consistent** if for each $\tau \in T$ for which $\overline{\Theta}^*_\tau$ has positive Lebesgue measure, one has that*

$$\int_{\theta\in\overline{\Theta}^*_\tau} \beta^{t(\theta)}[p(\theta) - c(q(\theta))]f^*_\tau(\theta)\,\mathrm{d}\theta \leq \int_{\theta\in\overline{\Theta}^*_\tau} \beta^{t^*(\theta)}[p^*(\theta) - c(q^*(\theta))]f^*_\tau(\theta)\,\mathrm{d}\theta + \varepsilon$$

for all $(p, q, t) \in \mathcal{I}^*_\tau$.

In words, the notion of ε-consistency captures the requirement that even if the principal learned additional information about the agent's type through the Dutch-auction-like construction outlined at the beginning of this section, she must not be able to earn more than ε on average upon redesigning the mechanism in an incentive-compatible way.

A particular case of ε-consistency arises when the principal is perfectly patient but the agent is not.

Proposition 9 *Suppose that $\beta = 1$ and $\delta < 1$. For a given $\varepsilon > 0$, let (p^*_ε, q^*) be an ε-optimal policy (in the sense of Definition 5). Moreover, let t^*_ε be the schedule policy as in Proposition 3. Then, the mechanism $(p^*_\varepsilon, q^*, t^*_\varepsilon)$ is ε-consistent.*

Proof From (8), the schedule policy t^*_ε is a non-increasing and continuous function of type. Therefore, for each $\tau \in T$,

$$\overline{\Theta}^*_\tau \equiv \{\theta \in \Theta \mid t^*_\varepsilon(\theta) \geq \tau\} = [\underline{\theta}, \theta^*_\tau]$$

is a closed interval whenever it is not empty, where

$$\theta^*_\tau \equiv \sup\{\theta \in \Theta \mid t^*_\varepsilon(\theta) \geq \tau\}.$$

It is easy to check also that $t^*_\varepsilon(\theta^*_\tau) = \tau$.

[24] Note that the adjusted space of schedule policies takes account of the fact that future deliveries can occur no sooner than τ.

Definitions 4 and 5 imply that neither p^*_ε nor q^* depends on the particular distribution f of types. Hence, when the principal learns that the agent has not claimed the good by time τ, she updates the space of potential types from $[\underline{\theta}, \overline{\theta}]$ to $[\underline{\theta}, \theta^*_\tau]$. Assuming that the latter interval is non-empty and non-degenerate (so that it has positive Lebesgue measure), Proposition 3 implies that the ε-optimal policy (p^*_ε, q^*) restricted to the interval $[\underline{\theta}, \theta^*_\tau]$ continues to be implementable by the (renormalized) schedule policy

$$\begin{aligned}\theta \mapsto \tau + \int_\theta^{\theta^*_\tau} \frac{v(q^*(s))}{(-\log\delta)\varepsilon}\,\mathrm{d}s &= t^*_\varepsilon(\theta^*_\tau) + \int_\theta^{\theta^*_\tau} \frac{v(q^*(s))}{(-\log\delta)\varepsilon}\,\mathrm{d}s \\ = \int_{\theta^*_\tau}^{\overline{\theta}} \frac{v(q^*(s))}{(-\log\delta)\varepsilon}\,\mathrm{d}s + \int_\theta^{\theta^*_\tau} \frac{v(q^*(s))}{(-\log\delta)\varepsilon}\,\mathrm{d}s &= \int_\theta^{\overline{\theta}} \frac{v(q^*(s))}{(-\log\delta)\varepsilon}\,\mathrm{d}s = t^*_\varepsilon(\theta)\end{aligned}$$

for all $\theta \in [\underline{\theta}, \theta^*_\tau]$. This observation entails that the principal cannot gain more than ε by redesigning the mechanism at any point of time. ■

Proposition 9 highlights an additional powerful feature of the lack of time-discounting on the principal's part. Not only can a perfectly patient principal expropriate almost all of the surplus from trade and destroy the agent's information rents, but she can also do so in a manner that is ε-consistent. Therefore, there is almost no need for the principal to commit herself to an ε-optimal policy beforehand: both she and the agent know that she will have negligible incentives to make adjustments to the almost-optimal menu of transfer–allocation pairs as her information about the agent's type becomes more refined over time if that menu is to be implemented in a Dutch-auction-like setting.

6 Conclusion

The fundamental trade-off in mechanism design consists in the fact that the principal has to resort to setting allocations offered to lower-type agents inefficiently meager in order to provide higher-type agents with improved incentives. This paper shows that quantity distortion on the lower end of the type distribution can be partially or even almost fully substituted for by time distortion. Lower-type agents are offered menus with higher allocations as compared to the standard mechanism, but they obtain them later. In this way, the principal is able to "inflate away" a part of higher-type agents' information rents. If the principal is more patient than the agent, then a decrease in the agent's discounted rents exceeds the decrease in the principal's discounted revenues, so that she can increase optimal profits above the second-best level of the standard mechanism with no delays. Moreover, if she is perfectly patient, then delaying the allocation involves no loss in terms of her discounted revenues, so that the trade-off between lowering the agent's information rents and lowering revenues disappears; accordingly, she is able to extract almost all of the surplus.[25]

[25]The only impediment to extracting *all* of the surplus is that allocations cannot be delayed indefinitely.

These insights suggest that shipping policies are not designed merely to price the separate, standalone service of dispatching a good from the warehouse to the consumer's home. Instead, they constitute an integral part of the management of revenues stemming from the good itself. Given that large corporations are likely to discount future payoffs to a lesser extent than individual consumers due to an imminent need for the good, credit or cash constraints, or mere psychological impatience, the optimal design of shipment policies provides firms with a subtle yet effective method of enhancing the differentiation between consumers in terms of the benefit they derive from the good. Such opportunities for accessory price discrimination bear important implications for revenue management, consumer vigilance, and regulatory policies alike.

References

Akan, M., B. Ata, and J. D. Dana, Jr. (2015): "Revenue Management by Sequential Screening," *Journal of Economic Theory*, 159, 728–774.

Baron, D. P. and R. B. Myerson (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50, 911–930.

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *The American Economic Review*, 105, 921–957.

Bergemann, D. and M. Pesendorfer (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137, 580–609.

Bergemann, D. and A. Wambach (2015): "Sequential Information Disclosure in Auctions," *Journal of Economic Theory*, 159, 1074–1095.

Chatterjee, K. and L. Samuelson (1987): "Bargaining with Two-Sided Incomplete Information: An Infinite Horizon Model with Alternating Offers," *The Review of Economic Studies*, 54, 175–192.

Coase, R. H. (1972): "Durability and Monopoly," *The Journal of Law and Economics*, 15, 143–149.

Courty, P. and H. Li (2000): "Sequential Screening," *The Review of Economic Studies*, 67, 697–717.

Crémer, J. and R. P. McLean (1985): "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist When Demands Are Interdependent," *Econometrica*, 53, 345–361.

——— (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1257.

Eső, P. and B. Szentes (2007): "Optimal Information Disclosure in Auctions and the Handicap Auction," *The Review of Economic Studies*, 74, 705–731.

Folland, G. B. (1999): *Real Analysis: Modern Techniques and Their Applications*, New York: John Wiley & Sons.

Friedman, E. J. and D. C. Parkes (2003): "Pricing WiFi at Starbucks: Issues in Online Mechanism Design," in *Proceedings of the 4th ACM Conference on Electronic Commerce*, New York: Association for Computing Machinery, 240–241.

Gul, F. and H. Sonnenschein (1988): "On Delay in Bargaining with One-Sided Uncertainty," *Econometrica*, 56, 601–611.

Gul, F., H. Sonnenschein, and R. Wilson (1986): "Foundations of Dynamic Monopoly and the Coase Conjecture," *Journal of Economic Theory*, 39, 155–190.

Heifetz, A. and Z. Neeman (2006): "On the Generic (Im)Possibility of Full Surplus Extraction in Mechanism Design," *Econometrica*, 74, 213–233.

Hörner, J. and L. Samuelson (2011): "Managing Strategic Buyers," *Journal of Political Economy*, 119, 379–425.

Lewis, T. R. and D. E. M. Sappington (1994): "Supplying Information to Facilitate Price Discrimination," *International Economic Review*, 35, 309–327.

McAfee, R. P. and P. J. Reny (1992): "Correlated Information and Mechanism Design," *Econometrica*, 60, 395–421.

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Neeman, Z. (2004): "The Relevance of Private Information in Mechanism Design," *Journal of Economic Theory*, 117, 55–77.

Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97–109.

Stokey, N. L. (1981): "Rational Expectations and Durable Goods Pricing," *The Bell Journal of Economics*, 12, 112–128.

Stole, L. A. (2007): "Price Discrimination and Competition," in *Handbook of Industrial Organization*, ed. by M. Armstrong and R. Porter, Amsterdam: Elsevier, vol. 3, 2221–2299.

Toikka, J. (2011): "Ironing without Control," *Journal of Economic Theory*, 146, 2510–2526.

Appendix

Proof of Lemma 1 Assume that $(p, q, t) \in \mathcal{I}$ is implementable and let $U : \Theta \to \mathbb{R}$ be as given in (3). Fix $\theta, \widehat{\theta} \in \Theta$ and assume, without loss of generality, that $\theta > \widehat{\theta}$. Incentive compatibility implies that

$$U(\theta) = \delta^{t(\theta)}[\theta v(q(\theta)) - p(\theta)] \geq \delta^{t(\widehat{\theta})}[\theta v(q(\widehat{\theta})) - p(\widehat{\theta})] = U(\widehat{\theta}) + \delta^{t(\widehat{\theta})}(\theta - \widehat{\theta})v(q(\widehat{\theta}))$$

and that

$$U(\widehat{\theta}) = \delta^{t(\widehat{\theta})}[\widehat{\theta}v(q(\widehat{\theta})) - p(\widehat{\theta})] \geq \delta^{t(\theta)}[\widehat{\theta}v(q(\theta)) - p(\theta)] = U(\theta) + \delta^{t(\theta)}(\widehat{\theta} - \theta)v(q(\theta)).$$

In consequence,

$$0 \leq \delta^{t(\widehat{\theta})}v(q(\widehat{\theta})) \leq \frac{U(\theta) - U(\widehat{\theta})}{\theta - \widehat{\theta}} \leq \delta^{t(\theta)}v(q(\theta)). \tag{20}$$

This chain of inequalities shows that both U and $\theta \mapsto \delta^{t(\theta)}v(q(\theta))$ are non-decreasing. Having $\widehat{\theta}$ approach θ from below,

$$\limsup_{\widehat{\theta}\uparrow\theta} \frac{U(\theta) - U(\widehat{\theta})}{\theta - \widehat{\theta}} \leq \delta^{t(\theta)}v(q(\theta)). \tag{21}$$

A symmetric argument (in which $\theta < \widehat{\theta}$) reveals that

$$\liminf_{\widehat{\theta}\downarrow\theta} \frac{U(\widehat{\theta}) - U(\theta)}{\widehat{\theta} - \theta} \geq \delta^{t(\theta)}v(q(\theta)). \tag{22}$$

Since U is non-decreasing, it is differentiable almost everywhere (in terms of the Lebesgue measure on Θ) by Theorem 3.23(b) in Folland (1999). If $\theta \in \Theta$ is a point of differentiability of U, then (21) and (22) imply that

$$U'(\theta) = \delta^{t(\theta)}v(q(\theta)). \tag{23}$$

The inequality (20) and the fact that $\theta \mapsto \delta^{t(\theta)}v(q(\theta))$ is non-decreasing yield also that

$$\sup_{\substack{\theta,\widehat{\theta}\in\Theta \\ \theta\neq\widehat{\theta}}} \left| \frac{U(\theta) - U(\widehat{\theta})}{\theta - \widehat{\theta}} \right| \leq \delta^{t(\bar{\theta})}v(q(\bar{\theta})),$$

so that U is, in fact, Lipschitz continuous. Moreover, (23) implies that

$$|U'(\theta)| = \delta^{t(\theta)}v(q(\theta)) \leq \delta^{t(\bar{\theta})}v(q(\bar{\theta}))$$

for almost every $\theta \in \Theta$, so that U is absolutely continuous (Folland, 1999, p. 108). The envelope condition (4) now follows from the fundamental theorem of calculus for Lebesgue integrals (Proposition 3.35 in Folland, 1999). Finally, (2) obviously implies that $U(\underline{\theta}) \geq 0$, completing the proof of necessity.

As for sufficiency, suppose that (4) holds, the function $\theta \mapsto \delta^{t(\theta)}v(q(\theta))$ is non-decreasing, and that $U(\underline{\theta}) \geq 0$. By (4), the agent's utility under truth-telling is non-decreasing, which readily implies (2). As for incentive compatibility, pick any $\theta, \widehat{\theta} \in \Theta$ and suppose, without loss of generality, that $\theta > \widehat{\theta}$. Then,

$$\begin{aligned}\delta^{t(\theta)}[\theta v(q(\theta)) - p(\theta)] - \delta^{t(\widehat{\theta})}[\theta v(q(\widehat{\theta})) - p(\widehat{\theta})] &= U(\theta) - U(\widehat{\theta}) - \delta^{t(\widehat{\theta})}(\theta - \widehat{\theta})v(q(\widehat{\theta})) \\ &= \int_{\widehat{\theta}}^{\theta} [\delta^{t(s)}v(q(s)) - \delta^{t(\widehat{\theta})}v(q(\widehat{\theta}))]\,\mathrm{d}s \geq 0,\end{aligned}$$

so that (1) holds. ■

Proof of Lemma 2 Since $c(0) = v(0) = 0$, it is clear that $J(0) = 0$. Moreover, for any $x \geq 0$, it is not difficult to compute that

$$J'(x) = \frac{c'(v^{-1}(x))}{v'(v^{-1}(x))} \geq 0 \quad \text{(with equality if and only if } x = 0\text{)}, \tag{24}$$

$$J''(x) = \frac{1}{v'(v^{-1}(x))^2}\left[c''(v^{-1}(x)) - \frac{c'(v^{-1}(x))v''(v^{-1}(x))}{v'(v^{-1}(x))}\right] > 0.$$

Next, fix any $w \geq 0$. Note that for any $y > 0$,

$$\frac{\mathrm{d}}{\mathrm{d}y}\left[y \times J\left(\frac{w}{y}\right)\right] = J\left(\frac{w}{y}\right) - \frac{w}{y} \times J'\left(\frac{w}{y}\right) = \int_0^{w/y}\left[J'(\omega) - J'\left(\frac{w}{y}\right)\right]\mathrm{d}\omega \leq 0,$$

with equality if and only if $w = 0$ given that $J'' > 0$.

Finally, suppose that $w > 0$. Then, L'Hôpital's rule yields that

$$\lim_{y\downarrow 0}\left\{y \times J\left(\frac{w}{y}\right)\right\} = \lim_{y\to\infty}\left\{\frac{J(w \times y)}{y}\right\} = \lim_{y\to\infty}\{J'(w \times y) \times w\} = \infty,$$

given that $w > 0$, (24), and Assumption 1. This proves (9). ■

Proof of Proposition 4 Let $(W^\circ, t^\circ) \in \mathcal{W}$ solve the auxiliary problem. Consider any implementable mechanism $(p, q, t) \in \mathcal{I}$ and define $W : \Theta \to \mathbb{R}_+$ as follows:

$$W(\theta) \equiv \delta^{t(\theta)} v(q(\theta)).$$

Given that the mechanism is incentive-compatible, Lemma 1 implies that W is non-decreasing. Therefore, $(W, t) \in \mathcal{W}$.

Next, the definition of U in (3) and the envelope condition (4) imply that the principal's payoff under (p, q, t) can be rewritten as

$$\begin{aligned}
\int_{\underline{\theta}}^{\overline{\theta}} \beta^{t(\theta)} f(\theta)[p(\theta) - c(q(\theta))]\,\mathrm{d}\theta &= \int_{\underline{\theta}}^{\overline{\theta}} \beta^{t(\theta)} f(\theta)\Big[\theta v(q(\theta)) - \delta^{-t(\theta)}U(\theta) - c(q(\theta))\Big]\mathrm{d}\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta)\left[\theta\delta^{t(\theta)}v(q(\theta)) - U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta}\delta^{t(s)}v(q(s))\,\mathrm{d}s - \delta^{t(\theta)}c(q(\theta))\right]\mathrm{d}\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta)\left[\theta W(\theta) - \int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s - \delta^{t(\theta)}c\left(v^{-1}\left(\frac{W(\theta)}{\delta^{t(\theta)}}\right)\right)\right]\mathrm{d}\theta \\
&\quad \underbrace{- U(\underline{\theta})\int_{\underline{\theta}}^{\overline{\theta}}\left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta)\,\mathrm{d}\theta}_{\equiv \Xi} \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta)\left[\theta W(\theta) - \int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s - \delta^{t(\theta)}J\left(\frac{W(\theta)}{\delta^{t(\theta)}}\right)\right]\mathrm{d}\theta - \Xi
\end{aligned}$$

$$\leq \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t^\circ(\theta)} f(\theta) \left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\, \mathrm{d}s - \delta^{t^\circ(\theta)} J\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}}\right)\right] \mathrm{d}\theta, \tag{25}$$

given that (W°, t°) solves the auxiliary problem and $\Xi \geq 0$ because $U(\underline{\theta}) \geq 0$.

Now, define

$$q^\circ(\theta) \equiv v^{-1}\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}}\right) \quad \forall \theta \in \Theta, \tag{26}$$

$$p^\circ(\theta) \equiv \frac{1}{\delta^{t^\circ(\theta)}} \left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\, \mathrm{d}s\right] \quad \forall \theta \in \Theta. \tag{27}$$

Clearly, $\delta^{t^\circ(\theta)} v(q^\circ(\theta)) = W^\circ(\theta)$, which is non-decreasing, so that the monotonicity condition of Lemma 1 is satisfied. Moreover, letting U° denote the truth-telling agent's utility under $(p^\circ, q^\circ, t^\circ)$, one has that

$$\begin{aligned} U^\circ(\theta) &= \delta^{t^\circ(\theta)}[\theta v(q^\circ(\theta)) - p^\circ(\theta)] = \theta W^\circ(\theta) - \left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\, \mathrm{d}s\right] \\ &= \int_{\underline{\theta}}^{\theta} \delta^{t^\circ(s)} v(q^\circ(s))\, \mathrm{d}s. \end{aligned}$$

This yields that $U^\circ(\underline{\theta}) = 0$ and that the envelope condition (4) is satisfied as well. Hence, the mechanism $(p^\circ, q^\circ, t^\circ)$ is implementable. In addition, by (25), its value for the principal achieves the supremum corresponding to the auxiliary problem, which is an upper bound on the value for *all* implementable mechanisms. Therefore, $(p^\circ, q^\circ, t^\circ)$ is an optimal mechanism, as claimed.[26]

Conversely, suppose that $(p^\circ, q^\circ, t^\circ) \in \mathcal{I}$ is an optimal mechanism. Define

$$W^\circ(\theta) \equiv \delta^{t^\circ(\theta)} v(q^\circ(\theta)) \quad \forall \theta \in \Theta.$$

By the monotonicity condition in Lemma 1, the function W° is non-decreasing, so that $(W^\circ, t^\circ) \in \mathcal{W}$. Next, consider any $(W, t) \in \mathcal{W}$ and define

$$\begin{aligned} q(\theta) &\equiv v^{-1}\left(\frac{W(\theta)}{\delta^{t(\theta)}}\right) \quad \forall \theta \in \Theta, \\ p(\theta) &\equiv \frac{1}{\delta^{t(\theta)}} \left[\theta W(\theta) - \int_{\underline{\theta}}^{\theta} W(s)\, \mathrm{d}s\right] \quad \forall \theta \in \Theta, \\ U(\theta) &\equiv \delta^{t(\theta)}[\theta v(q(\theta)) - p(\theta)] \quad \forall \theta \in \Theta. \end{aligned}$$

Using precisely the same argument as after (26)–(27), one can easily show that $(p, q, t) \in \mathcal{I}$ and that $U(\underline{\theta}) = 0$. Moreover, using (25), the remark made in n. 26, and the fact that $(p^\circ, q^\circ, t^\circ)$ is optimal, one has that

$$\int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t(\theta)} f(\theta) \left[\theta W(\theta) - \int_{\underline{\theta}}^{\theta} W(s)\, \mathrm{d}s - \delta^{t(\theta)} J\left(\frac{W(\theta)}{\delta^{t(\theta)}}\right)\right] \mathrm{d}\theta$$

[26] This argument and (25) imply also that the utility of the lowest-type agent must vanish in any optimal mechanism.

$$
\begin{aligned}
&= \int_{\underline{\theta}}^{\overline{\theta}} \beta^{t(\theta)} f(\theta)[p(\theta) - c(q(\theta))]\,\mathrm{d}\theta \\
&\leq \int_{\underline{\theta}}^{\overline{\theta}} \beta^{t^\circ(\theta)} f(\theta)[p^\circ(\theta) - c(q^\circ(\theta))]\,\mathrm{d}\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\beta}{\delta}\right)^{t^\circ(\theta)} f(\theta)\left[\theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s - \delta^{t^\circ(\theta)} J\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}}\right)\right]\mathrm{d}\theta.
\end{aligned}
$$

This implies that (W°, t°) solves the auxiliary problem, as claimed. ■

Proof of Proposition 5 Suppose that $(W^\circ, t^\circ) \in \mathcal{W}$ solves the auxiliary problem. For any pair $(W, t) \in \mathcal{W}$, define the discounted profit function $\rho(\cdot|W, t) : \Theta \to \mathbb{R}$ as

$$
\rho(\theta|W, t) \equiv \left(\frac{\beta}{\delta}\right)^{t(\theta)} \left[\theta W(\theta) - \int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s - \delta^{t(\theta)} J\left(\frac{W(\theta)}{\delta^{t(\theta)}}\right)\right] \quad \forall \theta \in \Theta \tag{28}
$$

and the "undiscounted" profit function (according to which the delivery of the allocation occurs immediately) $\pi(\cdot|W) : \Theta \to \mathbb{R}$ as

$$
\pi(\theta|W) \equiv \rho(\theta|W, 0) = \theta W(\theta) - \int_{\underline{\theta}}^{\theta} W(s)\,\mathrm{d}s - J(W(\theta)) \quad \forall \theta \in \Theta. \tag{29}
$$

If (W°, t°) were to be optimal, then $\pi(\theta|W^\circ) < 0$ if and only if $\rho(\theta|W^\circ, t^\circ) < 0$. To see this, note that for any $(W, t) \in \mathcal{W}$, one has $\pi(\theta|W) = \rho(\theta|W, t)$ if $t(\theta) = 0$. Since the optimal value of $t^\circ(\theta)$ can be chosen pointwise in the auxiliary problem (given that there are no monotonicity restrictions *a priori* on the function t beyond measurability), it follows that $\rho(\theta|W^\circ, t^\circ) \geq \pi(\theta|W^\circ)$. Hence, if $\pi(\theta|W^\circ) \geq 0$, then $\rho(\theta|W^\circ, t^\circ) \geq 0$. Conversely, if $\pi(\theta|W^\circ) < 0$, then

$$
\begin{aligned}
\left(\frac{\delta}{\beta}\right)^{t^\circ(\theta)} \rho(\theta|W^\circ, t^\circ) &= \theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s - \delta^{t^\circ(\theta)} J\left(\frac{W^\circ(\theta)}{\delta^{t^\circ(\theta)}}\right) \\
&\leq \theta W^\circ(\theta) - \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s - J(W^\circ(\theta)) = \pi(\theta|W^\circ) < 0,
\end{aligned}
$$

where the weak inequality follows from Lemma 2.

Next, I argue that there is no loss of generality in assuming that the function W° is left-continuous—that is,

$$
W^\circ(\theta) = \lim_{\widehat{\theta} \uparrow \theta} W^\circ(\widehat{\theta}) \quad \forall \theta \in (\underline{\theta}, \overline{\theta}]. \tag{30}
$$

(Note that the left-hand limit always exists, given that W° is monotone.) This is because the monotonicity of W° implies that W° is continuous except perhaps at a countable set of points—see Theorem 3.23(a) in Folland (1999). Whenever W° is discontinuous, one could redefine it according to (30). This does not affect monotonicity and given that countable sets have measure zero, both the discounted profit function ρ in (28) and the undiscounted profit function π in (29) are changed only at a set of points that has measure zero, leaving the objective function unaffected. From now on, assume that W° is left-continuous. Once can easily check that the left-continuity of W° implies also the left-continuity of the undiscounted profit function $\theta \mapsto \pi(\theta|W^\circ)$.

Let

$$\Theta_- \equiv \{\theta \in (\underline{\theta}, \overline{\theta}] \,|\, \rho(\theta|W^\circ, t^\circ) < 0\} = \{\theta \in (\underline{\theta}, \overline{\theta}] \,|\, \pi(\theta|W^\circ) < 0\}.$$

Suppose, for the sake of contradiction, that $\Theta_- \neq \varnothing$ and let $\theta_0 \in \Theta_-$. Since W° and hence $\pi(\cdot|W^\circ)$ are left-continuous, it follows that there exists some $\varepsilon > 0$ such that

$$\pi(\theta|W^\circ) < 0 \quad \forall \theta \in (\theta_0 - \varepsilon, \theta_0]. \tag{31}$$

Suppose first that $\pi(\theta|W^\circ) < 0$ for all $\theta \in (\underline{\theta}, \theta_0]$. One can then replace W° with

$$\widetilde{W}(\theta) \equiv \begin{cases} 0 & \text{if } \theta \in [\underline{\theta}, \theta_0], \\ W^\circ(\theta) & \text{if } \theta \in (\theta_0, \overline{\theta}]. \end{cases}$$

Obviously, $\widetilde{W}$ is monotonic and involves zero (discounted and undiscounted) profits for $\theta \in (\underline{\theta}, \theta_0]$ (as opposed to negative ones). Moreover, switching from W° to $\widetilde{W}$ does not decrease undiscounted profits for $\theta \in (\theta_0, \overline{\theta}]$, since

$$\int_{\underline{\theta}}^{\theta} \widetilde{W}(s)\,\mathrm{d}s = \int_{\theta_0}^{\theta} W^\circ(s)\,\mathrm{d}s \leq \int_{\underline{\theta}}^{\theta} W^\circ(s)\,\mathrm{d}s,$$

so that

$$\pi(\theta|\widetilde{W}) \geq \pi(\theta|W^\circ) \quad \forall \theta \in (\theta_0, \overline{\theta}].$$

This clearly implies that

$$\rho(\theta|\widetilde{W}, t^\circ) \geq \rho(\theta|W^\circ, t^\circ) \quad \forall \theta \in (\theta_0, \overline{\theta}].$$

The perturbation from (W°, t°) to $(\widetilde{W}, t^\circ)$ *strictly* increases the objective function (the expected value of discounted profits), given that negative (discounted and undiscounted) profits have been made vanish on an interval $\theta \in (\underline{\theta}, \theta_0]$ of positive measure (and the singleton set $\{\underline{\theta}\}$ is of measure zero).

Suppose now that there exists some $\theta \in (\underline{\theta}, \theta_0)$ such that $\pi(\theta|W^\circ) \geq 0$. The argument is similar as before. Let

$$\theta_+ \equiv \sup\{\theta \in (\underline{\theta}, \theta_0] \,|\, \pi(\theta|W^\circ) \geq 0\}.$$

Clearly, the supremum is taken on a non-empty bounded set, so it is well-defined. Also, (31) implies that

$$\theta_+ \leq \theta_0 - \varepsilon$$

and one has, by the definition of the supremum, that

$$\pi(\theta|W^\circ) < 0 \quad \forall \theta \in (\theta_+, \theta_0]. \tag{32}$$

By the left-continuity of $\pi(\cdot|W^\circ)$, one has also that $\pi(\theta_+|W^\circ) \geq 0$. This implies that $W^\circ(\theta) > W^\circ(\theta_+)$ for all $\theta \in (\theta_+, \theta_0]$, because if $W^\circ(\theta) = W^\circ(\theta_+)$ for some $\theta \in (\theta_+, \theta_0]$, then, as it is easy to check, one would have $\pi(\theta|W^\circ) = \pi(\theta_+|W^\circ) \geq 0$, contradicting (32). Now consider the following perturbation:

$$\widetilde{W}(\theta) \equiv \begin{cases} W^\circ(\theta) & \text{if } \theta \in [\underline{\theta}, \theta_+], \\ W^\circ(\theta_+) & \text{if } \theta \in (\theta_+, \theta_0], \\ W^\circ(\theta) & \text{if } \theta \in (\theta_0, \overline{\theta}]. \end{cases}$$

It is not difficult to verify that the new discounted profit levels satisfy:

$$\pi(\theta|\widetilde{W}) = \pi(\theta_+|W^\circ) \geq 0 \quad \theta \in (\theta_+, \theta_0],$$

which ensures a non-negative discounted profit level $\rho(\theta|\widetilde{W}, \widetilde{t})$ for a possibly different choice of the schedule $\widetilde{t}(\theta)$. Moreover, the interval $(\theta_+, \theta_0]$ is of positive measure given that it contains the interval $(\theta_0 - \varepsilon, \theta_0]$. Profits for $\theta \in [\underline{\theta}, \theta_+]$ are unaffected, whereas profits (both undiscounted and discounted) on $(\theta_0, \overline{\theta}]$ are weakly increased by the same argument as before, given that $\widetilde{W} \leq W^\circ$ pointwise. Clearly, monotonicity is satisfied, so that $(\widetilde{W}, \widetilde{t}) \in \mathcal{W}$ and it strictly dominates (W°, t°) in terms of expected discounted profits, contradicting the optimality of the latter, because strictly positive improvements have been carried out on the set $(\theta_+, \theta_0]$ of positive measure (and non-negative improvements elsewhere). This completes the proof that a solution to the auxiliary problem must involve non-negative profits pointwise, except possibly at $\underline{\theta}$. ■