Unidirectional incentive compatibility

Daniel Krähmer and Roland Strausz

August 25, 2023

Abstract

We study unidirectional incentive compatibility which incentivizes an agent to report truthfully when she can misrepresent private information in one direction only. In the canonical setting with continuous, one-dimensional private information, and quasi-linear utility, unidirectional incentive compatibility imposes no restrictions on the allocation rule and holds if and only if the change of the agent's information rent function respects a lower bound that is based on the allocation rule's monotone envelope. In monopolistic screening models with strong interdependent values or with countervailing incentives, optimal contracts differ from optimal bidirectionally incentive compatible contracts, possibly displaying non-monotone allocations.

Keywords: Screening, Verifiability, Implementability, Optimal Contracting, Countervailing Incentives
JEL: D82

1 Introduction

Incentive compatibility is the key concept in studies of private information as it captures the economic constraints that result from private information. The standard notion of incentive compatibility is built on the premise that a privately informed agent can misrepresent information arbitrarily because any claim about private information is, by assumption, non-verifiable. In many

Krähmer: Universität Bonn (kraehmer@hcm.uni-bonn.de); Strausz: Humboldt Universität Berlin (roland.strausz@hu-berlin.de); We much thank Gorkem Celik for highly inspiring discussions. We further thank Georg Nöldeke, Larry Samuelson. We gratefully acknowledge financial support from the German Research Foundation (DFG) through Germany's Excellence Strategy EXC 2126/1–390838866 and CRC TR 224 (Krähmer) and CRC TR 190 (Strausz).

applications, this assumption seems too stark, however. For instance, employers can use proficiency tests to expose applicants making excessive claims about their productivity. Similarly, a taxpayer who makes excessive claims about her wealth can be exposed by demanding access to financial accounts as proof of these claims. Likewise, a firm that wins a contract by understating its costs, runs the risk of going bankrupt, thus revealing the understatement of costs. Finally, a buyer who overstates her ability to pay will be busted if she does not have sufficient funds for the purchase.

Motivated by these examples, this paper studies a notion of incentive compatibility where an agent can misrepresent her private information in one direction only. We refer to this notion as unidirectional incentive compatibility.

We study unidirectional incentive compatibility in the canonical setting with continuous, one-dimensional private information ("types") and quasi-linear utility. This setting not only constitutes the work horse model for studying private information in applications such as optimal non-linear pricing, auctions, procurement, regulation, and many others. It is also the framework that offers a clean characterization of the standard notion of (bidirectional) incentive compatibility, allowing us to precisely identify the differences between uni- and bidirectional incentive compatibility.

The first contribution of the paper is to provide a characterization of unidirectional incentive compatibility. Recall that in the one-dimensional, quasi-linear framework, the standard (bidirectional) notion of incentive compatibility is characterized by monotonicity of the allocation rule and a payoff-equivalence formula that uniquely pins down the change of the agent's information rents in her type as a function of the allocation rule. Our characterization of unidirectional incentive compatibility departs in two ways: First, any allocation rule is implementable. Second, payoff-equivalence fails, and the change of information rents has to respect only a lower bound, but can be freely chosen otherwise. Moreover, unlike information rents in the bidirectional case, the lower bound is pinned down not by the allocation rule itself, but by its smallest monotone envelope.[1] In particular, different allocation rules may lead to the same lower bound.

That any allocation rule can be implemented can be seen as follows. To incentivize truth-telling by some type θ, it is sufficient to pay this type a transfer that is substantially larger than the transfers to those types $\hat{\theta}$ which she can mimic. With bidirectional incentive compatibility this large transfer may then induce some type $\hat{\theta}$ to mimic type θ. With unidirectional incentive compatibility, this issue does not occur simply because, by definition, the types $\hat{\theta}$ that can be mimicked by type θ cannot themselves mimic type θ. This argument also indicates why uni-

[1]The smallest monotone envelope of a function is the smallest monotone function that lies above the original function.

directional incentive compatibility only implies a lower, but not an upper bound on the change of transfers/information rents. The more difficult part of our characterization is to establish the lower bound, and we leave a careful explanation to the main text.

The second contribution of the paper is to use our characterization to study implications of unidirectional incentive compatibility for optimal contracting in the otherwise classical monopolistic screening model. Note that under our notion of unidirectional incentive compatibility, exactly those constraints are omitted which, under typical regularity assumptions, are *not* binding under an optimal contract with bidirectional incentive constraints. To identify differences between optimal uni- and bidirectionally incentive compatible contracts, we therefore consider non-regular environments such as with interdependent values or non-monotone hazard rates, as well as environments with countervailing incentives where the agent has a type-dependent outside option.

In the first environment, the key issue with bidirectional incentive compatibility is that monotonicity constraints might be binding, leading to "bunching". Because monotonicity is not a constraint with unidirectional incentive compatibility, the question arises whether optimal contracts exploit this slack and display non-monotone allocations as a result. We show that this is the case only when the degree of interdependency in the principal's and agent's preferences is sufficiently strong and the first-best allocation rule is not monotone, and we derive a sufficient condition for it. In contrast, when the preference interdependency is rather weak, which includes private values environments, optimal unidirectionally incentive compatible contracts remain monotone, and in fact, coincide with their bidirectionally incentive compatible counterparts. Thus, in these cases, the optimal allocation is monotone not for feasibility but for optimality reasons.

The driving force behind these results is that our lower bound on information rents implies that the minimum information rents needed to implement any allocation and its smallest monotone hull are identical. Therefore, because the principal extracts the difference between surplus and information rents, for a non-monotone allocation to be optimal, it needs to generate a larger surplus than its smallest monotone hull. A necessary condition for this is that the first-best allocation rule is not monotone.

To study countervailing incentives, we adopt the framework of Lewis and Sappington (1989), who consider a strictly concave outside option. First, we show that optimal uni- and bidirectionally incentive compatible contracts always differ, and that an optimal contract always exploits the failure of payoff-equivalence under unidirectional incentive compatibility, and there are no upward distortions, unlike in the case with bidirectional incentive compatibility. Second , we derive conditions under which the optimal contract displays a non-monotone allocation rule and show that these conditions are satisfied in a large class of cases.

To see how the failure of payoff-equivalence drives the first result, recall that with bidirectional incentive constraints, payoff-equivalence implies that the agent's information rent function is convex in type. With a concave outside option, this implies that under the optimal bidirectionally incentive compatible contract, individual rationality binds at only a single critical type, where the convex information rent function and the concave outside option are tangent. Moreover, the standard rent-efficiency trade-off implies upward distortions for types larger than the critical type.

By contrast, because unidirectional incentive constraints only imply a *lower* bound on information rents, the optimal bidirectionally incentive compatible allocation can be implemented with any information rent function that is *steeper* than its bidirectionally incentive compatible convex counterpart. Therefore, individual rationality can be made binding for all types larger than the critical type, and this constitutes a strict improvement over the optimal bidirectionally incentive compatible contract. Moreover, since for these larger types the change of information rents does not take on the lower bound, the rent-efficiency trade-off that drives upward distortions in the bidirectional case simply disappears with unidirectional incentive constraints. Finally, the optimal contract with unidirectional incentive constraints may display a non-monotone allocation because, as we show, this allows to extend the range of types for which individual rationality is binding.

1.1 Related Literature

The papers most closely related to our work are Celik (2006) and Sher and Vohra (2015). In a monopolistic screening model with interdependent values, Celik (2006) also contrasts unidirectional with bidirectional incentive constraints. Celik (2006) focuses on optimal contracting, showing for settings with discrete types, that if the first best is monotone, then the optimal contract with uni- and bidirectional incentive constraints coincide. We confirm that this result also holds in our setting with continuous types. We go beyond Celik (2006) by characterizing the set of implementable outcomes with unidirectional incentive constraints and derive optimal contracts when the first best is not monotone and with countervailing incentives. Moreover, the arguments from the setting with discrete types do not straightforwardly carry over to our setting with continuous types.

Sher and Vohra (2015) provide a graph-theoretical analysis of a monopolistic screening model with discrete types when private information is partially verifiable. Their analysis is complementary to ours in that they study more general verification structures in a specific monopolistic screening model with private values and linear preferences. By contrast, we study a specific verification structure in more general screening models with non-linear preferences, interdependent values and type-dependent outside options. To see the connections, note that in their discrete setup, our unidirectional incentive constraints correspond to an acyclical incentive graph in which evidence is "hierarchical" and representable by a tree with only a single branch. Their observation that for an acyclical incentive graph all allocations are incentive compatible is the discrete counterpart of our result that any allocation rule is implementable. For the case of a tree with a single branch, the lower bound on information rents they derive collapses to a discrete version of ours.

In a monopolistic screening setup with a risk averse buyer, Moore (1984) shows how one may solve for the optimal contract by explicitly exploiting the fact that, in a one-dimensional screening problem with private values, the bidirectional nature of incentive constraints is often inconsequential. In particular, Moore (1984) first solves for the optimal contract with only unidirectional incentive constraints and then invokes an assumption on risk preferences that ensures the solution also remains feasible with bidirectional incentive constraints.

Unidirectional incentive constraints have been studied in many applications of multi-dimensional screening where there is one dimension of private information that the agent can only either under- or overstate but not both. In addition to arising often naturally, the presence of unidirectional incentive constraints renders such problems more tractable. (See, for example, Che and Gale (2000), Iyengar and Kumar (2008), Beaudry et al. (2009), Pai and Vohra (2013).) While these papers differ from ours because of their focus on multi-dimensional screening, they illustrate the ubiquitous nature of unidirectional incentive constraints.

2 Setup

To fix ideas, we consider a principal and an agent contract over the production of a good by the agent. The agent's costs to produce the quantity $x \geq 0$ of the good is $c(x, \theta)$ where the cost parameter $\theta \in \Theta \equiv [\underline{\theta}, \bar{\theta}]$ is the agent's private information. The function c is twice continuously differentiable and satisfies $c(0, \theta) = 0$ for all θ, $c_\theta \geq 0, c_x \geq 0, c_{xx} \geq 0, c_{x\theta} \geq 0$. It is common knowledge that θ is distributed with the cdf F and a strictly positive pdf $f = F'$ on the support Θ.

The terms of trade are a quantity x and a transfer t from the principal to the agent. The agent's utility from the terms of trade (x, t) is $t - c(x, \theta)$.

The principal commits to a contract that conditions the terms of trade on communication by the agent. A contract (x, t) specifies a (measurable) allocation function $x : \Theta \to \mathbb{R}_+$ and a (measurable) transfer function $t : \Theta \to \mathbb{R}$.

Before we state the principal's problem, we first characterize in the next section the set of contracts that are incentive compatible in our setting. As motivated in the introduction, we focus on environments in which the agent's costs are partially verifiable in that the agent can prove that her costs are not smaller than her actual costs but can (falsely) exaggerate her costs.[2] Consequently, a contract is incentive compatible if the agent has no incentive to report a cost that is larger than her true cost. Formally, let $\tilde{U}(\tilde{\theta};\theta) = t(\tilde{\theta}) - c(x(\tilde{\theta}),\theta)$ be agent type θ's utility from reporting $\tilde{\theta}$, and let $U(\theta) = \tilde{U}(\theta;\theta)$. $U(\theta)$ represents the information rent of a type θ. With abuse of notation, we also refer to (U, x) as a contract. Thus, a contract (U, x) is incentive compatible if $U(\theta) \geq \tilde{U}(\hat{\theta};\theta)$ for all $\hat{\theta} \geq \theta$, which is equivalent to

$$U(\theta) - U(\hat{\theta}) \geq c(x(\hat{\theta}),\hat{\theta}) - c(x(\hat{\theta}),\theta) \quad \forall \hat{\theta} \geq \theta. \tag{1}$$

Notice that the constraints in (1) require the agent only not to report higher than her true costs, and we therefore refer to them as unidirectional incentive constraints.

By comparison, standard bidirectional incentive constraints take the form:

$$U(\theta) - U(\hat{\theta}) \geq c(x(\hat{\theta}),\hat{\theta}) - c(x(\hat{\theta}),\theta) \quad \forall \hat{\theta}, \theta. \tag{2}$$

3 Implementability

Our first result is a characterization of the unidirectional incentive constraints (1). To express this characterization, define for all $\tau \leq \hat{\theta}$

$$\bar{x}(\tau \mid \hat{\theta}) = \sup_{\rho \in [\tau, \hat{\theta}]} x(\rho) \tag{3}$$

as the smallest decreasing (upper) envelope of x on the interval $[\underline{\theta}, \hat{\theta}]$. Our characterization states that, with unidirectional incentive constraints, *any* allocation x is implementable and establishes a lower bound on the change of the information rent that has to be conceded to implement x.

[2]We comment below on the symmetric case that the agent can only understate her costs.

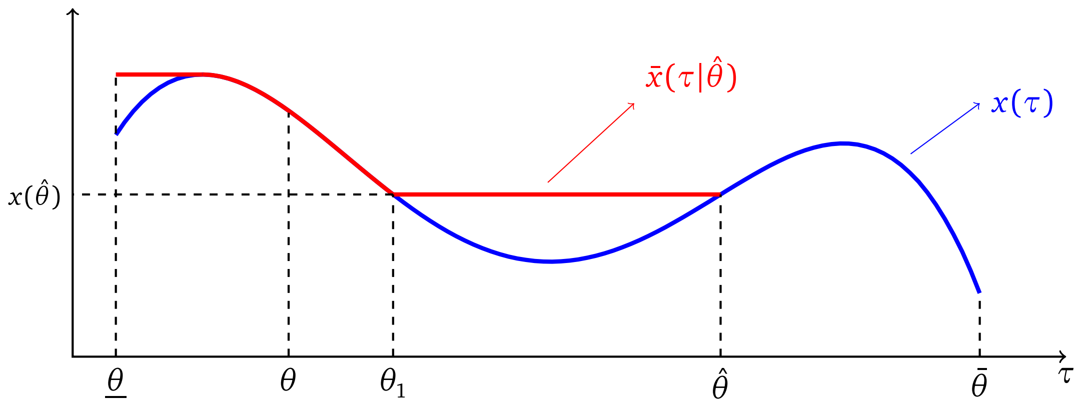

Figure 1: $x(\cdot)$ and $\bar{x}(\cdot|\hat{\theta})$

Proposition 1 *(x, U) satisfies the unidirectional incentive constraints (1) if and only if*

$$U(\theta) - U(\hat{\theta}) \geq \int_{\theta}^{\hat{\theta}} c_\theta(\bar{x}(\tau \mid \hat{\theta}), \tau)\, d\tau \quad \forall \theta \leq \hat{\theta}. \tag{4}$$

We relegate the proof to the appendix, where we show that the difficult part of the proposition is the only-if part. To gain intuition, suppose that the allocation x is continuous. This implies that at a point where $\bar{x}(\cdot \mid \hat{\theta})$ is strictly decreasing, it coincides with x, and at a point where $\bar{x}(\cdot \mid \hat{\theta})$ is flat, it coincides with a local maximum of x over the range $[\theta, \hat{\theta}]$. Figure 1 illustrates one such case where the interval $[\theta, \hat{\theta}]$ can be partitioned into an interval $[\theta, \theta_1]$ where $\bar{x}(\cdot \mid \hat{\theta})$ coincides with x, and an interval $(\theta_1, \hat{\theta}]$ where $\bar{x}(\cdot \mid \hat{\theta})$ is flat and equal to $x(\hat{\theta})$. On the basis of Figure 1, we next argue that incentive compatibility implies (4).

Now, incentive compatibility implies that no type τ in $[\theta, \theta_1)$ has an incentive to mimic a marginally higher type $\tau + d\tau$, that is,

$$\frac{U(\tau) - U(\tau + d\tau)}{d\tau} \geq \frac{c(x(\tau + d\tau), \tau + d\tau) - c(x(\tau + d\tau), \tau)}{d\tau} \tag{5}$$

$$\underset{d\tau \to 0}{\longrightarrow} \quad c_\theta(x(\tau), \tau) \tag{6}$$

$$= \quad c_\theta(\bar{x}(\tau \mid \hat{\theta}), \tau), \tag{7}$$

where the last equality is due to fact that $x(\tau) = \bar{x}(\tau \mid \hat{\theta})$ for $\tau \in [\theta, \theta_1)$. Hence,

$$U(\theta) - U(\theta_1) \geq \int_{\theta}^{\theta_1} c_\theta(\bar{x}(\tau \mid \hat{\theta}), \tau)\, d\tau. \tag{8}$$

Moreover, type θ_1 has no incentive to mimic type $\hat{\theta}$, that is,

$$U(\theta_1) - U(\hat{\theta}) \;\; \geq \;\; c(x(\hat{\theta}), \hat{\theta}) - c(x(\hat{\theta}), \theta_1). \tag{9}$$

Because for all $\tau \in [\theta_1, \hat{\theta}]$, we have that $\bar{x}(\tau \mid \hat{\theta})$ is constant equal to $x(\hat{\theta})$, we can replace $x(\hat{\theta})$ in the right hand side and use integration to obtain

$$U(\theta_1) - U(\hat{\theta}) \;\; \geq \;\; \int_{\theta_1}^{\hat{\theta}} c_\theta(\bar{x}(\tau \mid \hat{\theta}), \tau)\, d\tau. \tag{10}$$

Putting the inequalities (8) and (10) together now delivers expression (4).

The actual proof of Proposition 1 is significantly more involved than this simple heuristic argument suggests. The reason is that Proposition 1 is true for *any* allocation x. For an arbitrary allocation, it need neither be the case that at a point where $\bar{x}(\cdot|\hat{\theta})$ is strictly decreasing, it coincides with x, nor that at a point where $\bar{x}(\cdot|\hat{\theta})$ is flat, it coincides with a local maximum of x over the range $[\theta, \hat{\theta}]$. In fact, it is possible to construct functions x such that at all points τ, $\bar{x}(\tau|\hat{\theta})$ is strictly larger than $x(\tau)$. To prove Proposition 1, we show that, for any allocation x, we can approximate its associated $\bar{x}(\cdot|\hat{\theta})$ by a sequence of step functions each of which coincides with x at its jump points. For such step functions, (4) can be easily established, and the proof shows that (4) carries over in the limit.

It is instructive to compare the characterization of unidirectional incentive compatibility (1) to the characterization of the standard bidirectional incentive constraints (2). It is a classic result that bidirectional incentive compatibility is equivalent to monotonicity of the allocation function x and payoff equivalence with respect to the agent's information rents:

Proposition 2 *(x, U) satisfies the bidirectional incentive constraints (2) if and only if x is monotone (decreasing) and*

$$U(\theta) - U(\bar{\theta}) = \int_{\theta}^{\bar{\theta}} c_\theta(x(\tau), \tau)\, d\tau \quad \forall \theta \leq \bar{\theta}. \tag{11}$$

Comparing Proposition 2 to Proposition 1 clarifies that, relative to the standard case, unidirectional incentive constraints relax implementability with regard to both the allocations as well as to the information rents that can be implemented. First, while bidirectional incentive constraints imply that only monotone allocations are implementable, unidirectional incentive constraints do not imply any restrictions on the allocation x. Second, whenever x is monotone decreasing, $x(\tau)$ coincides with $\bar{x}(\tau|\hat{\theta})$ for any $\hat{\theta} \geq \tau$. In this case, we may equivalently express (11) as

$$U(\theta) - U(\hat{\theta}) = \int_{\theta}^{\hat{\theta}} c_{\theta}(\bar{x}(\tau \mid \hat{\theta}), \tau)\, d\tau \quad \forall \theta \leq \hat{\theta}. \tag{12}$$

This shows that while bidirectional incentive constraints imply that the slope of the information rents is fully pinned down by the allocation x, that is, $U'(\theta) = -c_{\theta}(x(\theta), \theta)$, unidirectional incentive constraints imply only upper bounds on this slope: $U'(\theta) \leq -c_{\theta}(\bar{x}(\theta, \hat{\theta}), \theta)$ for any $\hat{\theta} > \theta$.[3] In other words, payoff equivalence obtains under bidirectional incentive constraints, but not under unidirectional constraints.

To understand why monotonicity of the allocation is a necessary condition for bidirectional, yet not for unidirectional incentive compatibility, suppose some low type obtained a smaller quantity than some high type. Then for this low type to not mimic, and thus utilize her cost advantage relative to, the high type, he would need to obtain such a high rent (or transfer) that (because of single crossing preferences) it would then actually become profitable for the high type to mimic the low type. This implies that the allocation needs to be monotone if the high type can mimic the low type, but not if he cannot.

The reason why, with bidirectional information constraints, the marginal change in the information rent of a type τ is fully pinned down by $c_{\theta}(x(\tau), \tau)$ can be seen from (5). With bidirectional incentive constraints, type τ has to be deterred not only from mimicking a marginally higher type $\tau + d\tau$, but also from mimicking a marginally lower type $\tau - d\tau$, and this provides not only an upper but also a lower bound on how the information rent paid to type τ can change. More intuitively, if the change of information rent was strictly smaller than the right hand side of (11), then a high type would find it profitable to mimic a low type. The latter concern is absent when there are only unidirectional incentive constraints which is why it implies only an upper bound on the change in information rent.

Remark Proposition 1 characterizes unidirectional incentive compatibility for the case that the agent can only overstate her costs. The case that the agent can only understate her costs can be treated analogously. In particular, consider the unidirectional "downward" incentive constraints

$$U(\theta) - U(\hat{\theta}) \geq c(x(\hat{\theta}), \hat{\theta}) - c(x(\hat{\theta}), \theta) \quad \forall \hat{\theta} \leq \theta. \tag{13}$$

[3] Note that since U is decreasing in both cases, the derivative exists almost everywhere.

Define for all $\tau \geq \hat{\theta}$

$$\underline{x}(\tau \mid \hat{\theta}) = \inf_{\rho \in [\hat{\theta}, \tau]} x(\rho) \tag{14}$$

as the largest decreasing (lower) envelope of x on the interval $[\underline{\theta}, \hat{\theta}]$. Analogously to Proposition 1, we obtain:

Proposition 3 *(x, U) satisfies (13) if and only if*

$$U(\theta) - U(\hat{\theta}) \geq -\int_{\hat{\theta}}^{\theta} c_\theta(\underline{x}(\tau \mid \hat{\theta}), \tau)\, d\tau \quad \forall \hat{\theta} \leq \theta. \tag{15}$$

4 Optimal Contracting

We now turn to the question how *optimal* contracts are affected when the principal faces only uni- instead of bidirectional incentive constraints. We consider a general principal agent problem allowing for interdependent values and a type-dependent outside option for the agent. We shall argue that in private values environments and with type-independent outside option, optimal contracts are unaffected. In contrast, in interdependent values environments, the principal may do better with unidirectional incentive constraints because of the extra flexibility to implement a non-monotone allocation. And when the optimal contract with bidirectional incentive constraints exhibits countervailing incentives, the principal always does strictly better with unidirectional incentive constraints because of the extra flexibility to structure information rents. In this case, the optimal contract may display a non-monotone allocation.

Formally, the principal's utility from quantity x and transfer t is denoted $v(x, \theta) - t$. The function v is twice continuously differentiable and satisfies $v(0, \theta) = 0$ for all θ, $v_x \geq 0$ and $v_{xx} \leq 0$. The model exhibits "private values" if $v(x, \theta) = v(x)$ and "interdependent values" otherwise.

The first best allocation maximizes the total surplus is thus given by

$$x_0^*(\theta) = \arg\max_x v(x, \theta) - c(x, \theta) \tag{16}$$

Note that x_0^* is decreasing if the cross partial $v_{x\theta}$ is positive (which is the case with private values). But in general, x_0^* may not be decreasing.

Moreover, the agent has a possibly type-dependent outside option, yielding type θ the reservation utility $u_R(\theta)$ when no agreement with the principal is reached. Accordingly, a contract (U, x) is "individually rational" if it yields any type at least her reservation utility, that is, if

$$U(\theta) \geq u_R(\theta) \quad \forall \theta. \tag{17}$$

Because the principal's payoff equals aggregate surplus minus the agent's information rent, the principal's optimal contract (x_1^*, U_1^*) with unidirectional incentive constraints is a solution to

$$P_1: \qquad \max_{x,U} \int_{\underline{\theta}}^{\bar{\theta}} v(x(\theta),\theta) - c(x(\theta),\theta) - U(\theta)\, dF(\theta) \quad s.t. \quad \text{(1) and (17).} \tag{18}$$

By contrast, the principal's optimal contract (x_2^*, U_2^*) with bidirectional incentive constraints is a solution to

$$P_2: \qquad \max_{x,U} \int_{\underline{\theta}}^{\bar{\theta}} v(x(\theta),\theta) - c(x(\theta),\theta) - U(\theta)\, dF(\theta) \quad \text{s.t.} \quad \text{(2) and (17).} \tag{19}$$

We call a combination (x, U) feasible with respect to P_1 if it satisfies the constraints (1) and (17), and feasible with respect to problem P_2 if it satisfies the constraints (2) and (17). Because (2) includes all the constraints in (1) a combination that is feasible with respect to P_2 is also feasible with respect to P_1 but not vice versa.

4.1 Interdependent values

As shown in the previous section, when there are only unidirectional incentive constraints, the principal can implement a non-monotone allocation. This suggests that if the monotonicity constraint in the solution to problem P_2 with bidirectional constraints is binding, then the solutions to P_1 and P_2 may differ. In this section, we show that this intuition is only partially true. In fact, even if the monotonicity constraint in P_2 is binding, the solution to P_1 is different only if the environment exhibits strong interdependent values in the sense that the first best allocation is non-monotone.

To isolate the effect of interdependent values on optimal contracting, we assume in this section that the agent has a type-independent outside option, i.e. $u_R(\theta) = 0$. It is well known that in this case the maximizer of P_2 is a pair (x_2^*, U_2^*) such that x_2^* maximizes the expected virtual surplus

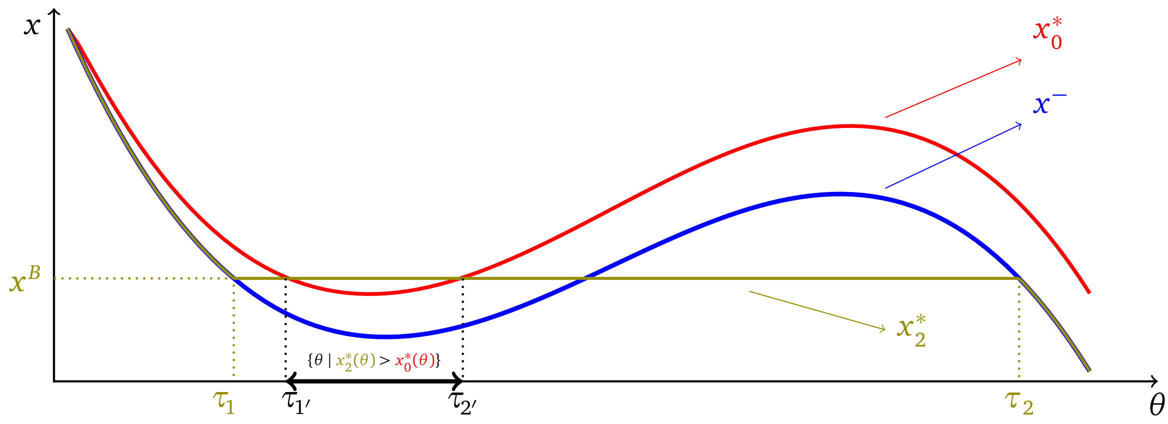

Figure 2: Screening with interdependent values, where the optimal contract with bidirectional incentive compatibility, x_2^*, displays bunching for $\theta \in [\tau_1, \tau_2]$ but, by Proposition 5, is not optimal under unidirectional incentive compatibility as the non-monotone first best x_0^* cuts x_2^* and we have $\{\theta \mid x_2^*(\theta) > x_0^*(\theta)\} = (\tau_{1'}, \tau_{2'}) \neq \emptyset$.

subject to a monotonicity constraint:

$$x_2^* = \arg\max_x \quad \int_{\underline{\theta}}^{\bar{\theta}} v(x(\theta), \theta) - c(x(\theta), \theta) - \frac{F(\theta)}{f(\theta)} c_\theta(x(\theta), \theta)\, dF(\theta) \quad s.t. \tag{20}$$

$$x(\theta) \text{ is decreasing in } \theta, \tag{21}$$

and that U_2^* satisfies (11) with $U_2^*(\bar{\theta}) = 0$.

To identify conditions under which the value of P_1 differs from P_2, it is helpful to define

$$x^-(\theta) = \arg\max_x v(x, \theta) - c(x, \theta) - \frac{F(\theta)}{f(\theta)} c_\theta(x, \theta) \tag{22}$$

as the point-wise maximizer of the virtual surplus in (20). In so-called regular settings, where x^- is decreasing in θ, x^- represents a solution to P_2. If the setting is not regular, then the solution x_2^* to P_2 involves "ironing" the non-monotonic allocation x^-, leading to a "bunching" of types θ which are all assigned the same allocation x^B. Figure 2 illustrates such bunching for the interval $[\tau_1, \tau_2]$.

To solve P_1, note first that since (1) implies that $U(\theta)$ is decreasing in θ, and since in this section the agent's outside option is assumed to be type-independent, individual rationality is equivalent to individual rationality only for the highest type $\bar{\theta}$:

$$U(\bar{\theta}) \geq 0. \tag{23}$$

Moreover, Proposition 1 provides a lower bound for the information rent that the principal has to pay to the agent in order to implement an allocation x. In fact, in order to dissuade a type θ to report the highest type $\hat{\theta} = \bar{\theta}$, (4) implies that the rent paid to type θ has to satisfy:

$$U(\theta) \geq U(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} c_\theta(\bar{x}(\tau \mid \bar{\theta}), \tau)\, d\tau \quad \forall \theta. \tag{24}$$

We now relax the principal's problem P_1 by only considering the constraints (23) and (24). Because U enters the principal's objective negatively, (23) is binding, and (24) is binding for all θ at an optimum of this relaxed problem, yielding information rent:

$$U(\theta) = \int_{\theta}^{\bar{\theta}} c_\theta(\bar{x}(\tau \mid \bar{\theta}), \tau)\, d\tau \quad \forall \theta. \tag{25}$$

This expression says that (at an optimum) information rents are pinned down by the smallest decreasing envelope $\bar{x}(\cdot \mid \bar{\theta})$ of x. More specifically, due the single-crossing property of c, information rents are larger, the larger is $\bar{x}(\cdot \mid \bar{\theta})$. Thus, the principal faces a modified, less stringent, rent-efficiency trade-off compared to the case with bidirectional incentive constraints. In the latter case, payoff equivalence pins down information rents by the allocation x. Thus, increasing the allocation necessarily comes at the cost of higher information rents. In contrast, with unidirectional incentive constraints, increasing the allocation comes at the cost of higher information rents only if this also increases $\bar{x}(\cdot \mid \bar{\theta})$.

To capture the modified rent-efficiency trade-off more formally, we shall now relax the problem further by treating $\bar{x}(\cdot \mid \bar{\theta})$ as an independent choice variable y which inherits from $\bar{x}(\cdot \mid \bar{\theta})$ the constraints that y be decreasing and $x(\theta) \leq y(\theta)$ for all θ. Thus, we arrive at the problem:

$$R_1: \qquad \max_{x,y,U} \quad \int_{\underline{\theta}}^{\bar{\theta}} v(x(\theta), \theta) - c(x(\theta), \theta) - U(\theta)\, dF(\theta) \quad s.t. \tag{26}$$

$$U(\theta) = \int_{\theta}^{\bar{\theta}} c_\theta(y(\tau), \tau)\, d\tau \quad \forall \theta, \tag{27}$$

$$y(\theta) \text{ is decreasing in } \theta, \tag{28}$$

$$x(\theta) \leq y(\theta) \quad \forall \theta. \tag{29}$$

Lemma 1 *Let (x, y, U) be solution to R_1. Then (x, U) is a solution to P_1.*

We are now in a position to identify conditions under which the principal does, and does not, benefit from the fact that incentive constraints are uni- rather than bidirectional.

Proposition 4 *Let the first best x_0^* or the virtual surplus maximizer x^- be decreasing. Then the second best (x_2^*, U_2^*) is a solution to both P_1 and P_2.*

To understand the result, consider first the case that the first-best x_0^* is decreasing. Note first that at an optimum of R_1, the allocation x is smaller than the first-best allocation x_0^*. Otherwise, it would be feasible and profitable to lower x to the first-best level and maintain y, since lowering x would improve surplus and maintaining y would keep information rents the same. Second, at an optimum of R_1, y is equal to the smallest decreasing envelope of x, because this minimizes the information rents to implement the allocation x.

Now, because x_0^* is above the optimal allocation x, x_0^* is thus a decreasing envelope of x, and since y is the smallest decreasing envelope of x, we also have that y is smaller than the first best level. But since x and y are smaller than the first best level, x and y must be the same, because otherwise, and this is the key observation, if x were strictly smaller than y for some values of θ, increasing the allocation x would increase the surplus without affecting rents. But, when x and y are the same, information rents are pinned down by the allocation x, and x must be decreasing. Thus, problem R_1 becomes equal to the second best problem P_2 to maximize the virtual surplus subject to monotonicity.

Note that this argument is true even if the monotonicity constraint is binding in P_2 so that the second best x_2^* displays bunching. The reason is that whenever both x and y are below the first best, then the modified rent efficiency trade-off in R_1 plays out in exactly the same way as the standard rent efficiency trade-off with bidirectional incentive constraints.

Consider next the case that the virtual surplus maximizer x^- is decreasing, and thus is equal to x_2^*. Consider the relaxed version of R_1 where y is not required to be decreasing. Then, clearly, for any allocation x, choosing $y = x$ minimizes information rents. With $x = y$ and absent the monotonicity constraint, the problem reduces to the relaxed version of the second best problem (20) where the monotonicity is ignored. By definition, the solution to this problem is x^-. Hence, $x = y = x^-$ is a solution to the relaxed version of R_1. Consequently, if x^- is decreasing, $y = x^-$ is decreasing and automatically satisfies the monotonicity constraints in the original problem R_1.

By Proposition 4, unidirectional incentive compatibility can make a difference only in cases where x_0^* and x^- are not decreasing. As explained, whenever x^- is not decreasing, x_2^* involves bunching of some types θ. If there is a strong degree of interdependent values so that the first best x_0^* is sufficiently non-monotone, then x_0^* intersects with x_2^*, as illustrated in Figure 2. As the next proposition shows this is a sufficient condition for the solutions to problem P_1 and P_2 to be different.[4,5]

Proposition 5 *Suppose the set* $\{\theta \mid x_2^*(\theta) > x_0^*(\theta)\}$ *has strictly positive measure. Then the principal's value from problem* P_1 *is strictly larger than his value from problem* P_2*. Moreover, any solution* (x_1^*, U_1^*) *to* P_1 *exhibits a non-monotone allocation* x_1^**.*

Figure 2 provides an intuition behind this result. Whenever x^- is non-monotone, the optimal allocation with bidirectional incentive constraints, x_2^*, involves bunching, as illustrated for the range $[\tau_1, \tau_2]$ at level x^B. Clearly, x_2^* together with U_2^* is also feasible with unidirectional constraints. The principal can however improve on the pair (x_2^*, U_2^*) by choosing the combination (x', U_2^*) where x' coincides with x_0^* in the region $[\tau_{1'}, \tau_{2'}]$—where x_0^* lies below x_2^*—and x' is equal to x_2^* otherwise. In line with our characterization, the combination (x', U_2^*) remains incentive compatible, as $x_2^*(\theta) = \bar{x}'(\theta|\hat{\theta}) \equiv \sup_{\tau\in[\theta,\hat{\theta}]} x'(\tau)$, and thus the minimal information rent to implement x' is U_2^* according to (25). The combination (x', U_2^*) yields the principal a strictly higher payoff, because x' is closer to the first best and therefore yields a higher surplus, while the agent's information rents do not change.

Remark If the agent can only understate her costs, then Proposition 3 implies that the principal can implement the first-best allocation x_0^* without leaving the agent any information rents, that is, $U(\theta) = 0$ for all θ. To see this, note that $U(\theta) = 0$ for all θ satisfies (15) because the right hand side is negative.

4.2 Countervailing incentives

In this section, we show how the extra flexibility in structuring the agent's information rents when there are only unidirectional incentive constraints alleviates optimal contracting when the agent has a type-dependent outside option. Indeed, when incentive constraints are bidirectional, the presence of a type-dependent outside option may lead to "countervailing incentives" (Lewis

[4]The quantifier in Proposition 5 that $x_2^*(\theta) > x_0^*(\theta)$ for some θ, implies that x_0^* and x^- are both non-monotone so that Proposition 4 and 5 describe mutually exclusive cases. To see this, note first that if x^- is decreasing, then $x^- = x_2^*$ and $x_0^* \geq x_2^*$ then follows from the fact that $x_0^* \geq x^- = x_2^*$. Hence, if $x_2^*(\theta) > x_0^*(\theta)$ for some θ, then x^- is non-decreasing. To see that also x_0^* is non-decreasing, observe that since x^- is non-decreasing, x_2^* exhibits bunching for some interval, and it can only be in such a bunching interval that $x_2^*(\theta) > x_0^*(\theta)$. The fact that $x_0^* \geq x^-$ then implies that there is some θ_1 such that $x_2^*(\theta_1) > x_0^*(\theta_1) \geq x^-(\theta_1)$, but optimal bunching implies that there is some $\theta_2 > \theta_1$ so that $x^-(\theta_2) > x_2^*(\theta_2) = x_2^*(\theta_1)$. But if x_0^* were decreasing, this would lead to the contradiction that $x^-(\theta_2) > x_2^*(\theta_2) = x_2^*(\theta_1) > x_0^*(\theta_1) \geq x_2^*(\theta_2)$. Hence, x_0^* must also be non-monotone.

[5]The remaining case not covered by either proposition is the one where x_0^* and x^- are non-monotone, but $x_0^*(\theta) > x_2^*(\theta)$ almost anywhere. We provide examples in Appendix 2 that for this remaining case, the solutions to P_1 and P_2 may or may not coincide. A full characterization of this remaining case is non-obvious.

and Sappington, 1989). Such countervailing incentives obtain when, at the optimum, there are types whose incentive constraints are binding in one direction, as well as some other types whose incentive constraints are binding in the other direction. Because with unidirectional incentive constraints, incentive constraints can be binding only in one direction, this suggests that the solutions to the problems with and uni- and bidirectional incentive constraints, P_1 and P_2, differ when there are countervailing incentives.

To isolate the effect of the type-dependency of the outside option, we consider a model with private values, $v(\theta,x) = v(x)$, and adopt the specification in Lewis and Sappington (1989) with linear costs $c(\theta,x) = \theta x$, and a decreasing and concave outside option $u_R(\theta)$ for the agent: $u_R'(\theta) < 0, u_R''(\theta) < 0$. We also follow Lewis and Sappington (1989) in their assumptions that $u_R'(\theta) \in (-x_0^*(\underline{\theta}), -x_0^*(\bar{\theta}))$ for all θ.[6]

Moreover, we assume that the (inverse) hazard rate $F(\theta)/f(\theta)$ is increasing, while the (inverse) hazard rate $1-F(\theta)/f(\theta)$ is decreasing so that, in line with Lewis and Sappington (1989), the point-wise maximizer, $x^-(\theta)$, of the (downward) virtual surplus,

$$\phi^-(x,\theta) \equiv v(x) - \theta x - \frac{F(\theta)}{f(\theta)}x \tag{30}$$

and the point-wise maximizer, $x^+(\theta)$, of the (upward) virtual surplus

$$\phi^+(x,\theta) \equiv v(x) - \theta x + \frac{1-F(\theta)}{f(\theta)}x \tag{31}$$

are both decreasing.

We first recall the characterization of the optimal contract with bidirectional incentive constraints.

Lemma 2 (Lewis and Sappington, 1989) *There are a bunching level x_2^B and thresholds $\theta_2^- < \theta_2^R < \theta_2^+$ such that a solution (x_2^*, U_2^*) to P_2 exhibits $x_2^B = x_2^-(\theta_2^-) = x^+(\theta_2^+) = -u_R'(\theta_2^R)$ and*

$$x_2^*(\theta) = \begin{cases} x^-(\theta) & \text{if } \theta \le \theta_2^- \\ x_2^B & \text{if } \theta_2^- < \theta < \theta_2^+ \\ x^+(\theta) & \text{if } \theta \ge \theta_2^+ \end{cases}, \qquad U_2^*(\theta) = \int_\theta^{\theta_2^R} x_2^*(\tau)\, d\tau + u_R(\theta_2^R). \tag{32}$$

Thus, there is a unique interior type θ_2^R for whom the individual rationality constraint is bind-

[6] Lewis and Sappington (1989) interpret $u_R(\theta)$ as type θ's fixed costs of production. Because we motivate unidirectional incentive constraints as resulting from the principal 's ability to verify lower bounds on costs, it is more natural in our context to interpret $u_R(\theta)$ as type θ's opportunity cost of an alternative project or self-employment.

ing ($U_2^*(\theta_2^R) = u_R(\theta_2^R)$), whereas for all other types θ the individual rationality constraint is slack, ($U_2^*(\theta) > u_R(\theta)$). The optimal contract (x_2^*, U_2^*) exhibits countervailing incentives in the sense that for types smaller than θ_2^R, incentive constraints are locally binding upwards, whereas for types larger than θ_2^R, incentive constraints are locally binding downwards.

Intuitively, with a type-dependent outside option, the agent has two sources of rent. As with a type-independent outside option the agent can cash in on the cost advantage relative to a higher cost type by overstating her cost. Moreover, by understating her cost, and thus overstating her outside option, the agent can cash in on her higher willingness (or lower opportunity cost) to accept the contract relative to a lower cost type. The binding incentive constraints under the optimal contract are determined by the interplay between these two forces. With a concave outside option, the second force dominates for large cost types $\theta > \theta_2^R$. The reason is that in this range, the outside option declines fastest.

Since large cost types $\theta > \theta_2^R$ can secure information rents (relative to their outside option) by understating their type, the principal pays lower information rents the more steeply the function U declines in the range $\theta > \theta_2^R$ and is thus more closely aligned with the outside option u_R. By payoff equivalence, however, the slope of U is pinned down by the quantity ($U' = -x$), and thus the principal can implement a more steeply declining U only by increasing the quantity x appropriately. As a result, in the range $\theta > \theta_2^R$, the rent efficiency trade-off is optimally resolved by distorting the quantity upward beyond the first best.

In contrast, when there are only unidirectional incentive constraints, payoff equivalence for cost types in the range $\theta > \theta_2^R$ is not a constraint. The principal can therefore implement an information rent function U that declines faster than $-x$. We now first illustrate that adapting only U in this way while maintaining the allocation x_2^* allows the principal to attain a strict improvement over (x_2^*, U_2^*).

Lemma 3 *The following adapted contract (x_1, U_1) is feasible in problem P_1 and yields strictly more than V_2^*:*

$$x_1(\theta) = x_2^*(\theta), \qquad U_1(\theta) = \begin{cases} U_2^*(\theta) & \text{if } \theta < \theta_2^R \\ u_R(\theta) & \text{if } \theta \geq \theta_2^R \end{cases}. \tag{33}$$

The previous modification only adapts the information rents and is the most "obvious" way to exploit the absence of payoff equivalence to obtain a payoff improvement for the principal when there are only unidirectional incentive constraints. We now ask whether a further improvement can be obtained by also adapting the allocation. Recall from the previous section that when the outside option is type-independent, then the optimal allocation is monotone if the first-best is monotone. This may suggest that in the current setting, where the first-best is monotone, a monotone allocation is also optimal. Interestingly, however, this may not be the case.

To show this, we now first derive the optimal monotone contract with unidirectional incentive constraints. We then consider a local modification of this contract where we introduce a (feasible) non-monotonicity in the allocation and then state sufficient conditions under which this modification is profitable.

The optimal *monotone* contract (x_1^m, U_1^m) with unidirectional incentive constraints solves the version of P_1 with the additional constraint that x be monotone decreasing:

$$P_1^m: \qquad \max_{x,U} \int_{\underline{\theta}}^{\bar{\theta}} v(x(\theta)) - \theta x(\theta) - U(\theta)\, dF(\theta) \quad s.t. \quad (1), (17),\ x \text{ is decreasing}\ .$$

Lemma 4 *The solution (x_1^m, U_1^m) to P_1^m is characterized by a bunching level x_m^B and thresholds $\mu^- < \mu^R < \mu_0$ given as the solution to the four equations*

$$\begin{aligned} x_m^B &= x^-(\mu^-) = x_0^*(\mu_0) = -u_R'(\mu^R), \\ \int_{\mu-}^{\mu^R} v'(x_m^B) - \theta &- \frac{F(\theta)}{f(\theta)}\, dF + \int_{\mu^R}^{\mu_0} v'(x_m^B) - \theta\ dF = 0. \end{aligned} \tag{34}$$

Moreover,

$$x_1^m(\theta) = \begin{cases} x^-(\theta) & \text{if } \theta \le \mu^- \\ x_m^B & \text{if } \mu^- < \theta < \mu_0 \\ x_0^*(\theta) & \text{if } \mu^0 \le \theta \end{cases}, \quad U_1^m(\theta) = \begin{cases} u_R(\theta) & \text{if } \theta \ge \mu^R \\ u_R(\mu^R) + \int_\theta^{\mu^R} x_1^m(\tau)\, d\tau & \text{if } \theta < \mu^R. \end{cases} \tag{35}$$

Figure 3 illustrates the optimal monotone contract (x_1^m, U_1^m). The incentive constraints are (locally) binding up to type μ^R, and the individual rationality constraint is binding for all types larger than μ^R. Intuitively, by overstating her cost, a type θ can secure the utility of the higher type, $U(\hat{\theta})$, plus the cost advantage in producing the higher type's quantity, $(\hat{\theta} - \theta)x(\hat{\theta})$. Now, under a monotone contract, the decreasing quantity schedule intersects with the increasing (negative) slope of the reservation utility, $-u_R'$, at some unique type μ^R, as illustrated in the left panel of Figure 3. This implies that for types $\theta > \mu^R$, the individual rationality constraint is binding at an optimum. The reason is that if all types $\theta > \mu^R$ get their reservation utility, $U(\theta) = u_R(\theta)$, overstating one's type is not worthwhile as the utility falls faster than the cost advantage ($u_R' + x <$

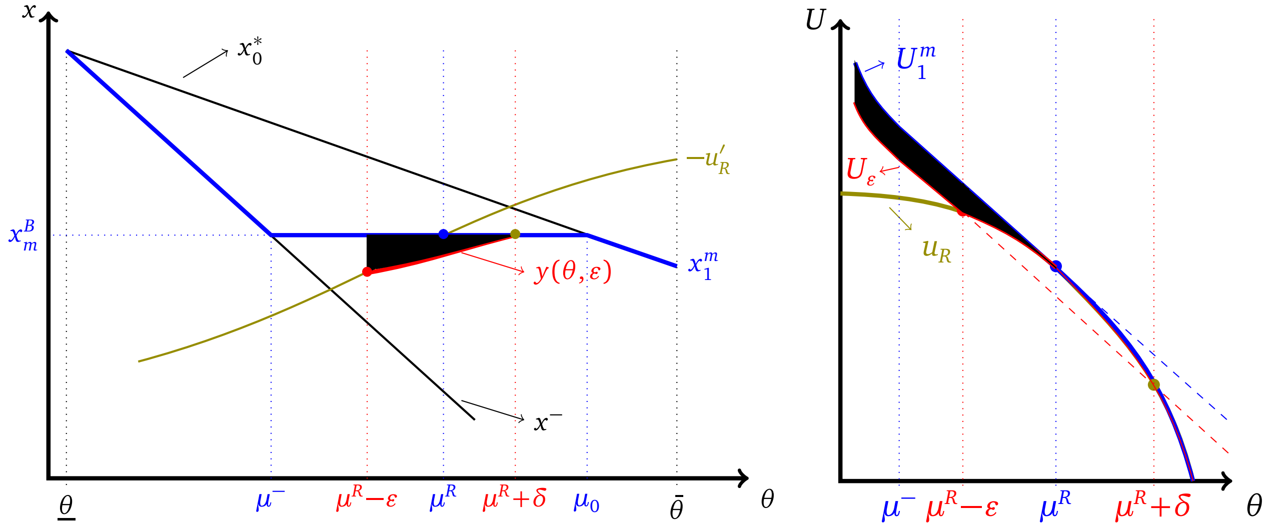

Figure 3: The optimal monotone schedule x_1^m (left panel) and associated rent U_1^m (right panel) and their modifications $y(\theta,\varepsilon)$ and U_ε. The shaded area in the left panel represents the loss, ΔS, in surplus of the modification, while the shaded area in the right panel represents its gain in the reduction of rents, ΔU.

0). Therefore, incentive constraints are slack in the range $\theta > \mu^R$. Analogously, for types $\theta < \mu^R$ incentive constraints are binding and individual rationality is slack, because in this range, the reservation utility falls more slowly than the cost advantage from mimicking a higher type.

We now derive conditions under which the principal improves over the monotone contract by a contract with a non-monotone quantity schedule. We do this by considering a marginal modification of the optimal monotone contract where we adapt the quantity schedule over an interval $[\mu^R-\varepsilon, \mu^R+\delta]$ so that this remains incentive compatible. Figure 3 illustrates these modifications in red.

Intuitively, introducing a non-monotone allocation relaxes the previously binding local incentive constraints in the range $\theta \in [\mu^R-\varepsilon, \mu^R]$ and instead of compensating a type in this range for not overstating her cost, the principal has to compensate such a type only for her outside option, implying lower information rents overall. Of course, since the allocation has been made smaller, there is a surplus loss. Hence, whether the modification is profitable depends on the trade–off between the reduction in surplus and the reduction in information rents. We analyze this trade-off next and, subsequently, derive a sufficient condition so that the rent effect dominates.

Specifically, suppose that all types $\theta \in [\mu^R-\varepsilon, \bar{\theta}]$ receive their reservation utility $u_R(\theta)$. Define $\delta = \delta(\varepsilon)$ so that if type $\mu^R+\delta$ produces quantity x_m^B, then type $\mu^R-\varepsilon$ is indifferent between truth-telling and mimicking type $\mu^R + \delta$:[7]

$$u_R(\mu^R - \varepsilon) - u_R(\mu^R + \delta) = (\delta + \varepsilon) x_m^B. \tag{36}$$

Consistent with (36), define for types $\theta \in [\mu^R - \varepsilon, \bar{\theta}]$ the quantity $y(\theta, \varepsilon)$ as the largest quantity that type θ can produce so that type $\mu^R - \varepsilon$ is indifferent between truth-telling and mimicking type θ:

$$u_R(\mu^R - \varepsilon) - u_R(\theta) = (\theta - \mu^R + \varepsilon) y(\theta, \varepsilon). \tag{37}$$

We now define the modification in such a way that all types in $\theta \in [\mu^R - \varepsilon, \mu^R + \delta]$ are assigned the quantity $y(\theta, \varepsilon)$, and all other types are assigned the same quantity as under the optimal monotone contract:

$$y_\varepsilon(\theta) \equiv \begin{cases} y(\theta, \varepsilon) & \text{if } \theta \in [\mu^R - \varepsilon, \mu^R + \delta] \\ x_1^m(\theta) & \text{otherwise.} \end{cases} \tag{38}$$

The left panel in Figure 3 illustrates the modification for small ε. Starting from $\underline{\theta}$, the schedule y_ε first follows the optimal monotone schedule x_1^m up to type $\mu^R - \varepsilon$, at which point there is a downward jump to $-u_R'(\mu^R - \varepsilon)$.[8] It then increases up to the level x_m^B at type $\mu^R + \delta$ from which on it follows the optimal monotone schedule again.[9]

Moreover, and as illustrated in the right panel of Figure 3, we define the information rent U_ε associated with y_ε so that for all types lower than $\mu^R - \varepsilon$, the IC constraint is locally binding, and all types larger than $\mu^R - \varepsilon$ receive their reservation utility:

$$U_\varepsilon(\theta) \equiv \begin{cases} u_R(\mu^R - \varepsilon) + \int_\theta^{\mu^R - \varepsilon} y_\varepsilon(\tau)\, d\tau & \text{if } \theta < \mu^R - \varepsilon \\ u_R(\theta) & \text{if } \theta \geq \mu^R - \varepsilon \end{cases}. \tag{40}$$

We next show that the non-monotone contract $(y_\varepsilon, U_\varepsilon)$ is feasible for small ε.

Lemma 5 *For any $\varepsilon > 0$ so that $\mu^- < \mu^R - \varepsilon$ and $\mu^R + \delta(\varepsilon) < \mu_0$, the contract $(x_1^\varepsilon, U_1^\varepsilon)$ is feasible.*

[7] Note that because u_R is strictly concave, δ is well-defined and we have that $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$.

[8] To see this note that

$$y(\mu^R - \varepsilon, \varepsilon) = \lim_{\theta \downarrow \mu^R - \varepsilon} \frac{u_R(\mu^R - \varepsilon) - u_R(\theta)}{\theta - \mu^R + \varepsilon} = -u_R'(\mu^R - \varepsilon) < x_m^B \tag{39}$$

[9] Note that $y(\mu^R + \delta, \varepsilon) = x_m^B$ by definition of δ. That y is increasing in θ follows from the concavity of u_R.

Having established the feasibility of the modified contract $(y_\varepsilon, U_\varepsilon)$, we next address whether it is more profitable than the optimal monotone contract. As illustrated by the shaded area in the left panel of Figure 3, the modification has a negative surplus effect because its modified schedule y_ε is more distorted than the original schedule x_1^m. On the other hand, the modified rent schedule U_ε leaves less rents to the agent than the schedule U_1^m, as illustrated by the shaded area in the right panel of the figure. Hence, the principal gains from the modification if the reduction in rents dominates the negative surplus effect. The next lemma states a necessary and sufficient condition for the marginal modification to be profitable.

Lemma 6 *There is $\varepsilon > 0$ so that the principal's payoff from the modification $(y_\varepsilon, U_\varepsilon)$ is strictly larger than that from the optimal monotone contract (x_1^m, U_1^m) if and only if*

$$v'(x_m^B) - \mu^R - \frac{F(\mu^R)}{f(\mu^R)} + v'(x_m^B) - \mu^R < 0. \tag{41}$$

The proof of Lemma 6 is tedious because the first order effects of the modification on both surplus and rents are zero and, hence, we need to consider the second order effects. As it turns out, the left hand side of (41) signs the second order effect.

It is instructive to see the role that the strict concavity of the outside option plays for the profitability of the modification. Recall that under the optimal monotone contract, incentive constraints are binding for types $\theta < \mu^R$, and individual rationality is binding for types $\hat{\theta} > \mu^R$. Because the outside option is strictly concave, this implies that a type $\theta < \mu^R$ is worse off when mimicking a high type $\hat{\theta} > \mu^R$. This slackness in the incentive constraints is exploited by the non-monotone modification. In fact, the modification ensures that the utility for types $\hat{\theta} > \mu^R$ remains the same, but type $\mu^R - \varepsilon$ is made indifferent between truth-telling and mimicking type $\mu^R + \delta$. Thus, type $\mu^R - \varepsilon$ (and all smaller types) receive a strictly lower information rent after the modification.

This would be different if the outside option was linear and, say, followed the dashed blue line in the right panel of Figure 3. In this case, the optimal monotone contract would be unchanged, but any type $\theta < \mu^R$ would now be indifferent between truth-telling and mimicking a high type $\hat{\theta} > \mu^R$. This is because if the outside option is linear with slope $-x_m^1$, then type θ's cost advantage in producing type $\hat{\theta}$'s quantity and the lower utility of type $\hat{\theta}$ exactly offset each other. Consequently, if a non-monotone modification of the form $(y_\varepsilon, U_\varepsilon)$ is performed in this case, this has only a (negative) surplus effect but no rent effect. Graphically, the dashed blue and red line then coincide and the shaded area is degenerate, implying that the modification does not save any information rents.[10]

While the previous lemma states a condition that is both necessary and sufficient for the marginal modification to be profitable, the condition is not directly defined in terms of the primitives of the model. The next proposition shows that an increasing density f is sufficient for condition (41) to hold. In particular, for a uniform distribution, the optimal contract is non-monotone.

Proposition 6 *Suppose the density f is increasing and that the agent's reservation utility is decreasing and strictly concave. Then a solution to P_1 displays a non-monotone schedule x_1^*.*

The proof of Proposition 6 shows that under an increasing density, equality (34) which defines the endogenous values x_m^B and μ^R, implies condition (41). To see an intuition and the role of the density, note that equality (34) states that the weighted (negative) area, weighted by f, under the marginal virtual surplus curve, $v'(x_m^B) - \theta - F(\theta)/f(\theta)$, from μ^- to μ^R equals the weighted (positive) area under the marginal surplus curve, $v'(x_m^B) - \theta$, from μ^R to μ_0. Now if condition (41) is false, then, in absolute terms, the marginal virtual surplus is smaller than the marginal surplus at $\theta = \mu^R$. Together with the defining values of μ^- and μ_0, this implies that the unweighted area under the marginal virtual surplus curve from μ^- to μ^R is smaller than the unweighted (positive) area under the marginal surplus curve from μ^R to μ_0. Hence, for the uniform distribution, where the weighted and unweighted areas trivially coincide, (34) therefore implies condition (41). Because an increasing density puts more weights on larger values of θ, the result under a uniform distribution extends to any increasing distribution.

Remark When there are countervailing incentives, and the agent can only understate her costs, then we obtain the mirror image of the case just discussed. Intuitively, when the agent agent can only understate her costs, is then the case that the individual rationality constraints are binding and the first best is implemented for types below a critical type. The reason is that in this range the outside option is relatively flat and therefore lies above the (negative) bound on the right hand side in (15). As types get larger, the outside option becomes steeper and hits the bound in (15). From then on, incentive constraints are binding, and the principal has to introduce an upward distortion to satisfy (15).

[10] A similar argument explains the difference between the case with and without type-dependent outside option. Recall that with private values and type-independent outside option, the optimal allocation is monotone, and a non-monotone modification of the type of $(y_\varepsilon, U_\varepsilon)$ is never profitable. The reason is that with type-independent outside option, the individual rationality constraint is binding only for the highest type $\bar{\theta}$, and the utility for all other types is determined by the binding incentive constraints. Intuitively, the non-monotone modification is then not profitable because it does not affect the set of incentive constraints that determine the agent's information rents.

5 Conclusion

In many natural settings, agents can misreport their private information in one direction only. In this paper, we characterize the corresponding notion of unidirectional incentive compatibility. Our main result shows that unidirectional incentive compatibility imposes no restriction on the allocation rule that can be implemented but only a lower bound on the change of the agent's information rent function. We show that in settings with strong interdependent values or countervailing incentives, optimal contracts differ from the traditional optimal bidirectionally incentive compatible contracts. In these settings, the principal therefore has a demand for a verification technology that allows him to verify claims by the agent that deviate from the truth in one direction. Reversely, we also identify settings in which optimal contracts with uni- and bidirectional incentive constraints do not differ, and thus a principal would not be willing to pay for such a verification technology.

Appendix A

In this appendix, we provide a formal proof of the characterization of unidirectional incentive compatibility as expressed in Proposition 1.

Proof of Proposition 1

"if": We have to show that (4) implies (1) Note that $\bar{x}(\tau \mid \hat{\theta}) = \sup_{\rho \in [\tau, \hat{\theta}]} \geq x(\hat{\theta})$. Therefore, (4) and the single-crossing condition for c and imply

$$
\begin{aligned}
U(\theta) - U(\hat{\theta}) &\geq \int_{\theta}^{\hat{\theta}} c_{\theta}(\bar{x}(\tau \mid \hat{\theta}), \tau)\, d\tau && (42) \\
&\geq \int_{\theta}^{\hat{\theta}} c_{\theta}(x(\hat{\theta}), \tau)\, d\tau && (43) \\
&= c(x(\hat{\theta}), \hat{\theta}) - c(x(\hat{\theta}), \theta), && (44)
\end{aligned}
$$

which is (1), as we wanted to show.

"only if": Fix θ and $\hat{\theta}$. To simplify notation, we write $\bar{x}$ for the function $\bar{x}(\cdot \mid \hat{\theta})$. We first prove the statement for the case that $\bar{x}$ is left-continuous, and in a second step extend it to general $\bar{x}$.

STEP 1: Let $\bar{x}$ be left-continuous.

We will use the following auxiliary lemma.

Lemma 7 *Let $\{\tau_R, \ldots, \tau_J\} \subset [\underline{\theta}, \bar{\theta}]$ be a set of increasing points $\tau_R < \tau_1 < \ldots < \tau_J$. Then, we have*

$$
U(\tau_R) - U(\tau_J) \geq \sum_{j=0}^{J-1} c(\max_{k \in \{j+1,\ldots,J\}} x(\tau_k), \tau_{j+1}) - c(\max_{k \in \{j+1,\ldots,J\}} x(\tau_k), \tau_j). \tag{45}
$$

Proof of Lemma 7 We prove the claim by induction over J:

Initial case: For $J = 1$, the claim is immediate from (1).

Induction step: Suppose the claim is true for J. We show that it is true for $J+1$. Let $\{\tau_R, \ldots, \tau_{J+1}\} \subset [\underline{\theta}, \bar{\theta}]$ be given. Let $k^* = \arg\max\{x(\tau_1), \ldots, x(\tau_{J+1})\}$. Note that this also implies:

$$
x(\tau_{k^*}) = \max_{k \in \{j+1,\ldots,J+1\}} x(\tau_k) \quad \forall j = 0, \ldots, k^* - 1. \tag{46}
$$

Then, we have

$$
U(\tau_R) - U(\tau_{J+1}) = U(\tau_R) - U(\tau_{k^*}) + U(\tau_k^*) - U(\tau_{J+1}) \tag{47}
$$

Now, observe first that by (1):

$$\begin{aligned} U(\tau_R)-U(\tau_{k^*}) &\geq c(x(\tau_{k^*}),\tau_{k^*})-c(x(\tau_{k^*}),\tau_0) && (48)\\ &= \sum_{j=0}^{k^*-1} c(x(\tau_{k^*}),\tau_{j+1})-c(x(\tau_{k^*}),\tau_j) && (49)\\ &= \sum_{j=0}^{k^*-1} c(\max_{k\in\{j+1,\ldots,J+1\}} x(\tau_k),\tau_{j+1})-c(\max_{k\in\{j+1,\ldots,J+1\}} x(\tau_k),\tau_j), && (50) \end{aligned}$$

where the two equalities follow from re-arranging terms and (46). Observe second that by induction hypothesis:

$$U(\tau_{k^*})-U(\tau_{J+1}) \geq \sum_{j=k^*}^{J} c(\max_{k\in\{j+1,\ldots,J+1\}} x(\tau_k),\tau_{j+1})-c(\max_{k\in\{j+1,\ldots,J+1\}} x(\tau_k),\tau_j). \qquad (51)$$

Putting the two observations together delivers (45) for $J+1$, and this completes the proof of Lemma 7. QED.

To continue with the proof of STEP 1, let $T=\{\tau_R,\ldots,\tau_J\}$ be a partition with $\theta=\tau_R<\tau_1<\ldots<\tau_J=\hat{\theta}$ be a partition of $[\theta,\hat{\theta}]$. Define for $\tau\in[\theta,\hat{\theta}]$ the function

$$\xi(\tau \mid T)=\max_{\tau_i\geq\tau,\tau_i\in T} x(\tau_i). \qquad (52)$$

Notice that $\xi(\cdot \mid T)$ is a decreasing step function. The next lemma states that if $\bar{x}$ is left-continuous, then it can be approximated through a sequence of such step functions.

Lemma 8 *Let $\bar{x}$ be left-continuous. There is a sequence $T^{(n)}=\{\tau_R^{(n)},\ldots,\tau_{J(n)}^{(n)}\}, n=1,2,\ldots$ of partitions of $[\theta,\hat{\theta}]$ so that*

$$\xi(\tau \mid T^{(n)}) \to \bar{x}(\tau) \text{ as } n\to\infty \text{ for all } \tau\in[\theta,\hat{\theta}]. \qquad (53)$$

Before we prove Lemma 8, we show it implies (4) for left-continuous $\bar{x}$. Let $T^{(n)}, n=1,2,\ldots$

be a sequence of partitions as in Lemma 8. By Lemma 7 and the definition of ξ:

$$\begin{aligned}
U(\theta)-U(\hat{\theta}) &= U(\tau_R^{(n)})-U(\tau_{J^{(n)}}^{(n)}) && (54)\\
&\geq \sum_{j=0}^{J^{(n)}-1} c(\xi(\tau_{j+1}^{(n)} \mid T^{(n)}),\tau_{j+1}^{(n)})-c(\xi(\tau_{j+1}^{(n)} \mid T^{(n)}),\tau_j^{(n)}) && (55)\\
&= \sum_{j=0}^{J^{(n)}-1} \int_{\tau_j^{(n)}}^{\tau_{j+1}^{(n)}} c_\theta(\xi(\tau_{j+1}^{(n)} \mid T^{(n)}),\tau)\,d\tau && (56)\\
&= \sum_{j=0}^{J^{(n)}-1} \int_{\tau_j^{(n)}}^{\tau_{j+1}^{(n)}} c_\theta(\xi(\tau \mid T^{(n)}),\tau)\,d\tau && (57)\\
&= \int_{\theta}^{\hat{\theta}} c_\theta(\xi(\tau \mid T^{(n)}),\tau)\,d\tau, && (58)
\end{aligned}$$

where the equality (57) in the second to last line follows, because $\xi(\cdot \mid T^{(n)})$ is a step function that is constant equal to $\xi(\tau_{j+1}^{(n)} \mid T^{(n)})$ for all $\theta \in (\tau_j^{(n)}, \tau_{j+1}^{(n)}]$.

Because the sequence $\xi(\cdot \mid T^{(n)})$ converges to $\bar{x}$ almost surely (by Lemma 8)and since c_θ is continuous and thus bounded, the dominated convergence theorem implies that the right hand side converges to the (Lebesgue) integral over $c_\theta(\bar{x}(\tau),\tau)$ as n goes to infinity. This implies (4) as we wanted to show. (Note that since c_θ is continuous and $\bar{x}$ is continuous almost everywhere, the Lebesgue integral exist and, in fact, coincides with the Riemann integral.)

We complete the proof of STEP 1 by providing the

Proof of Lemma 8 Note first that the lemma holds trivially if $\bar{x}$ is constant over $[\theta,\hat{\theta}]$, because we then have $\bar{x}(\tau)=x(\tau_J)=\xi(\tau|T^{(n)})$ for any τ and any partition $T^{(n)}$. Hence, suppose that $\bar{x}$ is not constant, implying that $\bar{x}(\theta)-\bar{x}(\hat{\theta})>0$, because $\bar{x}$ is decreasing. For $n \geq 0$ and $j=0,\ldots,2^n$, let

$$\varepsilon_n = \frac{\bar{x}(\theta)-\bar{x}(\hat{\theta})}{2^n}, \qquad y_j^{(n)} = \bar{x}(\theta)-j\cdot\varepsilon_n, \quad \theta_j^{(n)} = \sup\{\theta|\bar{x}(\theta)\geq y_j^{(n)}\}. \tag{59}$$

By the definition of the supremum, for all $\delta>0$, there is $\tau_j^{(n)} \geq \theta_j^{(n)}$ so that

$$|\bar{x}(\theta_j^{(n)})-x(\tau_j^{(n)})|<\delta. \tag{60}$$

We prove the claim for the two cases that 1) $\bar{x}$ is continuous and 2) $\bar{x}$ is left-continuous, where 2) builds on 1).

Case 1): Let $\bar{x}$ be continuous. We show (53) for $T^{(n)}=\{\tau_0^{(n)},\ldots,\tau_{2^n}^{(n)}\}$.

For this $T^{(n)}$, convergence of $\xi(\tau|T^{(n)})$ to $\bar{x}(\tau)$ holds trivially for $\tau = \hat{\theta}$, as for any n we have $\bar{x}(\hat{\theta}) = x(\hat{\theta}) = \xi(\hat{\theta} \mid T^{(n)}) = \xi(\tau_{2^n}^{(n)} \mid T^{(n)})$. So consider $\tau \neq \hat{\theta}$. Indeed, let $\eta > 0$ and $\tau \in [\theta, \hat{\theta})$ be given. We show that there is N so that for all $n \geq N$, we have $|\bar{x}(\tau) - \xi(\tau \mid T^{(n)})| < \eta$. Since, by definition, $\bar{x}(\tau) - \xi(\tau \mid T^{(n)}) \geq 0$, it is enough to show:

$$\bar{x}(\tau) - \xi(\tau \mid T^{(n)}) < \eta. \tag{61}$$

To see this, choose N and $\delta > 0$ such that $\varepsilon_N + \delta < \eta$. Now consider $n \geq N$. Let $j = j(n)$ be the index so that $\tau \in [\theta_{j-1}^{(n)}, \theta_j^{(n)})$. Because $\bar{x}$ is continuous by assumption, we have

$$\bar{x}(\theta_i^{(n)}) = y_i^{(n)} \quad \forall i. \tag{62}$$

By definition of $\theta_j^{(n)}$, and since $\bar{x}$ is decreasing, this implies

$$\bar{x}(\tau) \in [y_j^{(n)},\ y_{j-1}^{(n)}]. \tag{63}$$

We now distinguish two cases:

- Case 1a): $\tau \in [\theta_{j-1}^{(n)}, \tau_{j-1}^{(n)})$ (Notice that we may have that $\theta_j^{(n)} < \tau_{j-1}^{(n)}$ in which case this is the only case.)

Then by definition of ξ, and since $\tau < \tau_{j-1}^{(n)}$, we have

$$\xi(\tau \mid T^{(n)}) = \max_{\tau_i^{(n)} \geq \tau, \tau_i^{(n)} \in T^{(n)}} x(\tau_i^n) \geq x(\tau_{j-1}^{(n)}) \geq \bar{x}(\theta_{j-1}^{(n)}) - \delta = y_{j-1}^{(n)} - \delta, \tag{64}$$

where the final inequality follows from (60) and the final equality from (62). Together with (63), the definition of ε_n and our choice of N and δ, this implies

$$\bar{x}(\tau) - \xi(\tau \mid T^{(n)}) \leq y_{j-1}^{(n)} - \xi(\tau \mid T^{(n)}) \leq y_{j-1}^{(n)} - (y_{j-1}^{(n)} - \delta) = \delta < \varepsilon_N + \delta < \eta, \tag{65}$$

which is what we wanted to show.

- Case 1b): $\tau \in [\tau_{j-1}^{(n)}, \theta_j^{(n)})$.

Then by definition of ξ, and since $\tau < \theta_j^{(n)} \leq \tau_j^{(n)}$, we have

$$\xi(\tau \mid T^{(n)}) = \max_{\tau_i^{(n)} \geq \tau, \tau_i^{(n)} \in T^{(n)}} x(\tau_i^n) \geq x(\tau_j^{(n)}) \geq \bar{x}(\theta_j^{(n)}) - \delta = y_j^{(n)} - \delta, \tag{66}$$

where the final inequality follows from (60) and the final equality from (62). Thus, together with

(63), the definitions of $y_j^{(n)}$ and ε_n, and our choice of N, δ and n, we obtain:

$$\bar{x}(\tau)-\xi(\tau \mid T^{(n)}) \leq y_{j-1}^{(n)}-\xi(\tau \mid T^{(n)}) \leq y_{j-1}^{(n)}-(y_j^{(n)}-\delta)=\varepsilon_n+\delta \leq \varepsilon_N+\delta<\eta, \tag{67}$$

which is what we wanted to show.

Case 2): Let $\bar{x}$ be left-continuous.

Because $\bar{x}$ is decreasing, it can only have jump discontinuities with at most countably many jump points. If $\bar{x}$ has a jump point at τ, then, for n large enough, there will be finitely many points $\theta_j^{(n)}, \theta_{j+1}^{(n)}, \ldots, \theta_{j+K}^{(n)}$ as defined in (59) which are all equal the jump point τ. Moreover, left-continuity of $\bar{x}$, the definition of the supremum, and the definition of $y_j^{(n)}$ in (59) imply for any of such jump points $\theta_{j+k}^{(n)}$ that

$$x(\theta_{j+k}^{(n)})=\bar{x}(\theta_{j+k}^{(n)}) \geq y_j^{(n)}. \tag{68}$$

In addition, for all j such that $\theta_{j-1}^{(n)} \neq \theta_j^{(n)}$, left-continuity of $\bar{x}$ still implies

$$\tau \in (\theta_{j-1}^{(n)}, \theta_j^{(n)}] \quad \Rightarrow \quad \bar{x}(\tau) \in (y_j^{(n)}, y_{j-1}^{(n)}]. \tag{69}$$

The partition $T^{(n)}$ is now constructed as follows:

⋄ If $\bar{x}$ has a jump at $\theta_j^{(n)}$, we define $\tau_j^{(n)}=\theta_j^{(n)}$ so that by (68) we have

$$x(\tau_j^{(n)})=\bar{x}(\tau_j^{(n)}) \geq y_j^{(n)}. \tag{70}$$

⋄ Otherwise, $\bar{x}$ is continuous at $\theta_j^{(n)}$, and we then define $\tau_j^{(n)}$ as in (60) above.

To show (61) for given $\eta>0$ and $\tau \in [\theta, \hat{\theta})$, choose again N and $\delta>0$ such that $\varepsilon_n+\delta<\eta$ for all $n \geq N$, and let $j=j(n)$ so that $\tau \in [\theta_{j-1}^{(n)}, \theta_j^{(n)})$ where $\theta_{j-1}^{(n)} \neq \theta_j^{(n)}$. We distinguish four cases:

• Case 2a): $\bar{x}$ is continuous at $\theta_{j-1}^{(n)}$ and $\theta_j^{(n)}$. The argument is then as in the continuity case.

• Case 2b): $\bar{x}$ is continuous at $\theta_{j-1}^{(n)}$ and has a jump at $\theta_j^{(n)}$.

The argument is then the same as in the continuity case with the only difference that, because (70) implies $x(\tau_j^{(n)})=\bar{x}(\theta_j^{(n)}) \geq y_j^{(n)}$ in case 1b), (66) simplifies to $\xi(\tau \mid T^{(n)}) \geq y_j^{(n)}$. This does not affect the conclusion.

• Case 2c): $\bar{x}$ has a jump at $\theta_{j-1}^{(n)}$ and is continuous at $\theta_j^{(n)}$.

Then, by construction $\tau_{j-1}^{(n)}=\theta_{j-1}^{(n)}$, and the claim is thus trivially true for $\tau=\theta_{j-1}^{(n)}$ (recall (70)). For $\tau>\theta_{j-1}^{(n)}$, the argument is the same as for the continuity case with the simplification that we only have case 1b).

• Case 2d): $\bar{x}$ has a jump at $\theta_{j-1}^{(n)}$ and at $\theta_j^{(n)}$.

Then, by construction $\tau_{j-1}^{(n)} = \theta_{j-1}^{(n)}$, and the claim is thus trivially true for $\tau = \theta_{j-1}^{(n)}$ (recall (70)). For $\tau > \theta_{j-1}^{(n)}$, we have by construction that

$$\xi(\tau \mid T^{(n)}) = x(\theta_j^{(n)}) = x(\tau_j^{(n)}). \tag{71}$$

Hence,

$$\bar{x}(\tau) - \xi(\tau \mid T^{(n)}) \leq y_{j-1}^{(n)} - \xi(\tau \mid T^{(n)}) = y_{j-1}^{(n)} - x(\tau_j^{(n)}) \leq y_{j-1}^{(n)} - y_j^{(n)} \leq \varepsilon_n < \eta, \tag{72}$$

where the first inequality follows from (69) and the penultimate inequality from (70). This completes the proof of Lemma 8. QED

STEP 2 We now extend the claim to general $\bar{x}$ that is not left-continuous. In what follows we shall write $g(\tau^-) = \lim_{\rho\uparrow\tau} g(\rho)$ for the left limit of a function g at τ. Recall also that a decreasing function has a left limit at any point. Since $\bar{x}$ is decreasing, it has at most countably many jump points. Define the function $z(\tau)$ by

$$z(\tau) = \begin{cases} \bar{x}(\tau^-) & \text{if} \quad \tau \text{ is jump point of } \bar{x} \\ x(\tau) & \text{else} \end{cases}. \tag{73}$$

Note that because $\bar{x}$ is decreasing, it follows that $z(\tau) \geq x(\tau)$ and thus also $\bar{z}(\tau) \geq \bar{x}(\tau)$ for all τ. We have:

Lemma 9 *(i)* (x, U) *is incentive compatible if and only if* (z, U) *is incentive compatible.*
(ii) $\bar{z}$ *is left-continuous.*
(iii) $\bar{x} = \bar{z}$ *almost everywhere.*

Before proving Lemma 9, we show that it implies (4) for general x. If (x, U) is incentive compatible, so is (z, U) and since $\bar{z}$ is left-continuous, we have

$$U(\theta) - U(\hat{\theta}) \;\geq\; \int_\theta^{\hat{\theta}} c_\theta(\bar{z}(\tau), \tau)\, d\tau = \int_\theta^{\hat{\theta}} c_\theta(\bar{x}(\tau), \tau)\, d\tau, \tag{74}$$

where the equality follows from the fact that $\bar{x} = \bar{z}$ almost everywhere. This shows (4) as desired.

Proof of Lemma 9 As to (i). To see the if-part, let (z, U) be incentive compatible. Since $z \geq x$, it follows from single-crossing, $c_{x\theta} \geq 0$, that $c_\theta(z(\hat{\theta}), \tau) \geq c_\theta(x(\hat{\theta}), \tau)$. That (x, U) is incentive compatible then follows from

$$
\begin{aligned}
U(\theta)-U(\hat{\theta}) &\geq c(z(\hat{\theta}),\hat{\theta})-c(z(\hat{\theta}),\theta) && (75)\\
&= \int_{\theta}^{\hat{\theta}} c_{\theta}(z(\hat{\theta}),\tau)d\tau && (76)\\
&\geq \int_{\theta}^{\hat{\theta}} c_{\theta}(x(\hat{\theta}),\tau)d\tau = c(x(\hat{\theta}),\hat{\theta})-c(x(\hat{\theta}),\theta). && (77)
\end{aligned}
$$

To see the only-if-part, let (x,U) be incentive compatible, and consider $\theta,\hat{\theta}$ with $\theta<\hat{\theta}$. By inspection, condition (1) differs for (z,U) and (x,U) only if $z(\hat{\theta})$ differs from $x(\hat{\theta})$. Hence, by definition of z, (z,U) trivially satisfies (1) if $\hat{\theta}$ is not a jump point of $\bar{x}$.

Hence, consider the case that $\hat{\theta}$ is a jump point of $\bar{x}$ so that $z(\hat{\theta})=\bar{x}(\hat{\theta}^{-})$. Moreover, by Lemma 10 below, there is a sequence $\tau_n\to\hat{\theta}$, with $\tau_n\leq\hat{\theta}$, so that $\lim_{n\to\infty}x(\tau_n)=\bar{x}(\hat{\theta}^{-})$. Taken together

$$
\lim_{n\to\infty} x(\tau_n)=z(\hat{\theta}). \tag{78}
$$

Recall that incentive compatibility of (x,U) implies that U is decreasing. Since $\tau_n\leq\hat{\theta}$, this together with (1) yields:

$$
U(\theta) \geq U(\tau_n)+c(x(\tau_n),\tau_n)-c(x(\tau_n),\theta)\geq U(\hat{\theta})+c(x(\tau_n),\tau_n)-c(x(\tau_n),\theta). \tag{79}
$$

Since this holds for all n, and since c is continuous, the inequality carries over to the limit. Thus, by (78),

$$
U(\theta) \geq U(\hat{\theta})+c(z(\hat{\theta}),\hat{\theta})-c(z(\hat{\theta}),\theta). \tag{80}
$$

Accordingly, (z,U) satisfies (1) and is thus incentive compatible.

To see (ii), suppose to the contrary that $\bar{z}$ is not left-continuous. Then there is τ so that $\bar{z}(\tau^{-})>\bar{z}(\tau)$. Since τ is thus a jump point of $\bar{z}$, Lemma 10 implies that there is a sequence $\tau_n, n=1,2,\ldots$, with $\tau_n\leq\tau$ and converging to τ so that

$$
\lim_{n\to\infty} z(\tau_n)=\bar{z}(\tau^{-})>\bar{z}(\tau). \tag{81}
$$

Consider first the case that infinitely many τ_n's are not jump points of $\bar{x}$. Re-label the subsequence of these points as $\tau_{n_k}, k = 1, 2, \ldots$. By the definition of z, we have $z(\tau_{n_k}) = x(\tau_{n_k})$. Since $\bar{x} \geq x$ and $\bar{z} \geq \bar{x}$, (81) implies that

$$\bar{x}(\tau^-) = \lim_{k\to\infty} \bar{x}(\tau_{n_k}) \geq \lim_{k\to\infty} x(\tau_{n_k}) = \lim_{k\to\infty} z(\tau_{n_k}) = \bar{z}(\tau^-) > \bar{z}(\tau) \geq \bar{x}(\tau). \tag{82}$$

Hence, (82) implies that τ is also a jump point of $\bar{x}$. Thus, by definition of z, we have $z(\tau) = \bar{x}(\tau^-)$, and thus from (82):

$$z(\tau) = \bar{x}(\tau^-) \geq \bar{z}(\tau^-), \tag{83}$$

which contradicts the hypothesis that $\bar{z}(\tau) < \bar{z}(\tau^-)$.

Consider next the other case that only finitely many τ_n's are not jump points of $\bar{x}$. Then for large enough n_0, all points $\tau_n, n = n_0, n_0 + 1, \ldots$ are jump points of $\bar{x}$. By the definition of z, we have $z(\tau_n) = \bar{x}(\tau_n^-)$. By Lemma 10, we can for all $n \geq n_0$ find a $\hat{\tau}_n \leq \tau_n$ so that $|x(\hat{\tau}_n) - \bar{x}(\tau_n^-)| < 1/n$. Since $z(\tau_n) = \bar{x}(\tau_n^-)$ and since $\lim_{n\to\infty} z(\tau_n) = \bar{z}(\tau^-)$ by the first equality in (81), this implies that

$$\lim_{n\to\infty} x(\hat{\tau}_n) = \lim_{n\to\infty} \bar{x}(\tau_n^-) = \lim_{n\to\infty} z(\tau_n) = \bar{z}(\tau^-). \tag{84}$$

Since $\bar{x} \geq x$, this, together with $\tau_n \to \tau$ with $\tau_n \leq \tau$, (81), and $\bar{z} \geq \bar{x}$, implies that

$$\bar{x}(\tau^-) = \lim_{n\to\infty} \bar{x}(\tau_n) \geq \lim_{n\to\infty} x(\hat{\tau}_n) = \bar{z}(\tau^-) > \bar{z}(\tau) \geq \bar{x}(\tau). \tag{85}$$

We can now use the identical arguments from the first case that follow after (82) to show that $\bar{z}$ is left-continuous, and this concludes the proof of (ii).

To see (iii), recall $\bar{z}(\tau) \geq \bar{x}(\tau)$ for all τ. Therefore, and because the set of jump points of $\bar{x}$ is a null set, it is enough to show that

$$\bar{z}(\tau) \leq \bar{x}(\tau) \text{ for all } \tau \text{ which are not jump points of } \bar{x}. \tag{86}$$

Suppose to the contrary that τ is not a jump point of $\bar{x}$ and $\bar{x}(\tau) < \bar{z}(\tau)$. By definition of $\bar{x}$, there is thus $B > 0$ so that

$$x(\rho) < \bar{z}(\tau) - B \text{ for all } \rho \geq \tau. \tag{87}$$

Let $\rho^* \geq \tau$ be an element of $\arg\sup_{\rho \geq \tau} z(\rho)$. By definition of the supremum, there is a $\hat{\tau} \geq \rho^*$ so that

$$|\bar{z}(\tau) - z(\hat{\tau})| < B/4. \tag{88}$$

Moreover, by (87), since $\hat{\tau} \geq \rho^* \geq \tau$, we also have

$$x(\hat{\tau}) < \bar{z}(\tau) - B. \tag{89}$$

This implies that $\hat{\tau}$ is a jump point of $\bar{x}$ because otherwise $x(\hat{\tau}) = z(\hat{\tau})$ (by definition of z) which would contradict (88) and (89) jointly, since $\bar{z}(\tau) \geq z(\hat{\tau})$. Since $\hat{\tau}$ is a jump point of $\bar{x}$, we have that $z(\hat{\tau}) = \bar{x}(\hat{\tau}^-)$ by definition of z and that $\hat{\tau} > \tau$, since τ is not a jump point by assumption.

By Lemma 10, there is therefore $\tilde{\tau} \in (\tau, \hat{\tau}]$ so that $|x(\tilde{\tau}) - \bar{x}(\hat{\tau}^-)| < B/4$. Together with (88), this yields

$$\begin{aligned} |\bar{z}(\tau) - x(\tilde{\tau})| &\leq |\bar{z}(\tau) - z(\hat{\tau})| + |z(\hat{\tau}) - x(\tilde{\tau})| && (90)\\ &= |\bar{z}(\tau) - z(\hat{\tau})| + |\bar{x}(\hat{\tau}^-) - x(\tilde{\tau})| && (91)\\ &< B/4 + B/4 = B/2. && (92) \end{aligned}$$

But this is a contradiction to (87) for $\rho = \hat{\tau} \geq \tau$.

Lemma 10 *Let τ be a jump point of $\bar{x}$. Then there is a sequence $\tau_n, n = 1, 2, \ldots$ with $\tau_n \leq \tau$ converging to τ so that $x(\tau_n)$ converges to the left limit of $\bar{x}$ at τ:*

$$x(\tau_n) \to \bar{x}(\tau^-). \tag{93}$$

Proof of Lemma 10 Note first that since τ is a jump point and $\bar{x}$ is decreasing, there is $B > 0$ so that $\bar{x}(\tau^-) - \bar{x}(\rho) > B$ for all $\rho > \tau$. Since $\bar{x}(\rho) \geq x(\rho)$, we also have

$$\bar{x}(\tau^-) - x(\rho) > B \quad \text{for all } \rho > \tau. \tag{94}$$

Now take a sequence τ'_n, $n = 1, 2, \ldots$, converging to τ from the left. By the definition of $\bar{x}$, we can find for all $\alpha > 0$ a $\tau_n \geq \tau'_n$ so that

$$|x(\tau_n) - \bar{x}(\tau'_n)| < \alpha. \tag{95}$$

Moreover, because $\bar{x}$ is decreasing and $\tau_n' \leq \tau$, (94) implies

$$\bar{x}(\tau_n') - x(\rho) > B \quad \text{for all } \rho > \tau. \tag{96}$$

This, together with the previous inequality implies that $\tau_n \leq \tau$ whenever $\alpha \leq B$.

We are now in the position to show that the sequence τ_n has the desired properties. Indeed, let $\alpha \leq B$, then $\tau_n \in [\tau_n', \tau]$, and hence $\tau_n \to \tau$, since $\tau_n' \to \tau$. To see that $x(\tau_n)$ converges to $\bar{x}(\tau^-)$, let $\varepsilon > 0$ and $\alpha = \min\{\varepsilon/2, B\}$. Moreover, by the definition of the left limit, we can choose N such that $|\bar{x}(\tau_n') - \bar{x}(\tau^-)| < \varepsilon/2$ for all $n \geq N$. Together with (95), we obtain:

$$|x(\tau_n) - \bar{x}(\tau^-)| \leq |x(\tau_n) - \bar{x}(\tau_n')| + |\bar{x}(\tau_n') - \bar{x}(\tau^-)| \leq \alpha + \varepsilon/2 \leq \varepsilon. \tag{97}$$

This shows that $x(\tau_n) \to \bar{x}(\tau^-)$, as desired. QED

Appendix B

In this appendix, we collect the proofs of the two applications in Section 4.

Proof of Lemma 1 Observe first that at a solution (x,y,U) to R_1, we have that $y=\bar{x}(\cdot\mid\theta)$. The reason is that y is a decreasing function larger than x. Hence, choosing y equal to the smallest decreasing function larger than x minimizes information rents $U(\theta)$ the principal has to pay.

It remains to show that for $y=\bar{x}(\cdot\mid\theta)$, (27) implies (4). Indeed, suppose that (x,y,U) with $y=\bar{x}(\cdot\mid\theta)$ satisfies (27). Note that for all $\tau\le\hat{\theta}\le\bar{\theta}$, we have

$$\bar{x}(\tau\mid\hat{\theta})=\sup_{\rho\in[\tau,\hat{\theta}]}x(\rho)\le\sup_{\rho\in[\tau,\bar{\theta}]}x(\rho)=\bar{x}(\tau\mid\bar{\theta}). \tag{98}$$

Thus, by (27) and the single-crossing condition for c:

$$U(\theta)-U(\hat{\theta})=\int_{\theta}^{\bar{\theta}}c_{\theta}(\bar{x}(\tau\mid\bar{\theta}),\tau)\,d\tau\ge\int_{\theta}^{\bar{\theta}}c_{\theta}(\bar{x}(\tau\mid\hat{\theta}),\tau)\,d\tau. \tag{99}$$

But this is inequality (4) which we wanted to show. QED

Proof of Proposition 4 By Lemma 1, it is sufficient to show that $x=y=x_2^*$ and $U=U_2^*$ is a solution to R_1.

Suppose first, x_0^* is decreasing. Let $\tilde{x},\tilde{y},\tilde{U}$ be feasible for R_1. Then $x(\theta)=\min\{\tilde{y}(\theta),x_0^*(\theta)\}$ together with $\tilde{y},\tilde{U}$ is also feasible. Moreover, $(x,\tilde{y},\tilde{U})$ is a (weak) improvement over $(\tilde{x},\tilde{y},\tilde{U})$, since we have that $x(\theta)\in[\tilde{x}(\theta),x_0^*(\theta)]$, and hence the total surplus $v(x(\theta),\theta)-c(x(\theta),\theta)$ goes up due to concavity. In addition, the information rent $\tilde{U}$ is unchanged.

Given x, $y(\theta)=\min\{\tilde{y}(\theta),x_0^*(\theta)\}$ with U defined by (25) is feasible, because (a) $x=y$, and (b) $y(\theta)$ is decreasing in θ, because $\tilde{y}(\theta)$ is decreasing (since it is feasible), and $x_0^*(\theta)$ is decreasing by assumption. Moreover, (x,y,U) is a (weak) improvement over $(x,\tilde{y},U)$ because the total surplus is unchanged, and the information rent U is (weakly) reduced since $y\le\tilde{y}$ (recall that $c_\theta(y,\theta)$ is increasing in y due to single-crossing).

We conclude that there is a solution (x,y,U) to R_1 that satisfies $y=x$ with y decreasing. If we insert U into the objective and perform a standard integration by parts step, we obtain that problem R_1 reduces to

$$\max_{x}\quad\int_{\underline{\theta}}^{\bar{\theta}}v(x(\theta),\theta)-c(x(\theta),\theta)-\frac{F(\theta)}{f(\theta)}c_{\theta}(x(\theta),\theta)\,dF(\theta)\quad s.t. \tag{100}$$

$$x(\theta)\text{ is decreasing in }\theta. \tag{101}$$

But this is identical to the problem that we obtain after performing the corresponding integration of parts step in P_2. Hence, $x = y = x_2^*$ with $U = U_2^*$ is a solution to R_1, which is what we wanted to show.

Suppose next that $x^-(\theta)$ is decreasing. Relax problem R_1 by disregarding the monotonicity constraint (28), and call this problem $\tilde{R}_1$. Because the information rent U is increasing in y, it is optimal to set y as small as possible. Hence, in light of constraint (29), a solution (x, y) to $\tilde{R}_1$ satisfies $x = y$ and it follows as in the previous paragraph that x is thus given as the solution to

$$\max_x \quad \int_{\underline{\theta}}^{\bar{\theta}} v(x(\theta),\theta) - c(x(\theta),\theta) - \frac{F(\theta)}{f(\theta)} c_\theta(x(\theta),\theta)\, dF(\theta). \tag{102}$$

The solution to this problem is x^-, and hence, $x = y = x^-$ with U given by (27) is a solution to $\tilde{R}_1$. Because x^- is decreasing by assumption, the neglected constraint (28) is also satisfied, and hence $x = y = x^-$ is also a solution to R_1. Finally, note that since x^- is decreasing, we have that $x^- = x_2^*$ and thus also $U = U_2^*$. Hence, $x = y = x_2^*$ with $U = U_2^*$ is a solution to R_1, and this is what we wanted to show. QED

Proof of Proposition 5 Recall that (x_2^*, U_2^*) is a solution to P_2 and feasible in P_1. Adapting (x_2^*, U_2^*), we construct a contract (x, U) that is feasible for P_1 and yields the principal a strictly higher utility than the value of P_2.

To do so, let $x = \min\{x_2^*, x_0^*\}$, and define U by (25). Note that inequality (98) applies, and thus (x, U) is feasible for P_1. Moreover, since $x_2^*(\theta) > x_0^*(\theta)$ for some positive measure of θ's, x is more efficient than x_2^* and yields a strictly higher surplus than x_2^*. Furthermore, $x \le x_2^*$ implies that $\bar{x}(\cdot \mid \bar{\theta}) \le x_2^*$, and hence $U \le U_2^*$. Therefore, (x, U) both yields a strictly higher surplus and goes along with (weakly) lower information rents than (x_2^*, U_2^*). Thus, the value from problem P_1 is strictly larger than that from problem P_2.

The second statement that a solution to P_1 exhibits a non-monotone allocation follows by contradiction. Indeed, if P_1 has a solution (x^*, U^*) with a monotone allocation x^*, then U^* must satisfy $U^*(\bar{\theta}) = 0$ and (25) because, by Proposition 1, this is a lower bound on information rents needed to implement x^*. Because x^* is monotone, U^* then also satisfies (2), because the monotonicity of x^* implies that $\bar{x}^*(\theta|\bar{\theta}) = x^*(\theta)$. Hence, (x^*, U^*) is feasible for P_2, and, since P_2 is a more constrained problem than P_1, (x^*, U^*) is, in fact optimal for P_2. This however contradicts the first statement of the proposition. QED

Proof of Lemma 3 Because U_1 equals U_2^* for types $\theta \le \theta_2^R$ but is strictly lower for all types $\theta > \theta_2^R$, it is clear that (x_1, U_1) yields the principal a strictly higher payoff than (x_2^*, U_2^*). It also clear from construction that (x_1, U_1) is individually rational. To show that (x_1, U_1) is unidirectional incentive compatible, we show (4). Let $\theta < \hat{\theta}$. Since $x_1 = x_2^*$ is decreasing, we have that $x_1(\tau) = \bar{x}_1(\tau \mid \hat{\theta})$. Hence, the right hand side in (4) is

$$RHS = \int_{\theta}^{\hat{\theta}} x_1(\tau)\, d\tau = \int_{\theta}^{\hat{\theta}} x_2^*(\tau)\, d\tau = U_2^*(\theta) - U_2^*(\hat{\theta}).$$

The left hand side in (4) is

$$LHS = U_1(\theta) - U_1(\hat{\theta})$$

Thus, $RHS = LHS$ for $\theta, \hat{\theta} \leq \theta_2^R$. Consider next the case that $\theta_2^R \leq \theta, \hat{\theta}$. Note first that for $\tau \geq \theta_2^R$, we have that $-u_R'(\tau) \geq x_2^*(\tau)$ by Lemma 2. Indeed, at θ_2^R, we have $-u_R'(\theta_2^R) = x_2^B = x_2^*(\theta_2^R)$, and so the claim follows from the fact that $-u_R'$ is increasing (by concavity of u_R) and x_2^* is decreasing. Hence,

$$LHS = u_R(\theta) - u_R(\hat{\theta}) = \int_{\hat{\theta}}^{\theta} u_R'(\tau)\, d\tau = \int_{\theta}^{\hat{\theta}} -u_R'(\tau)\, d\tau \geq \int_{\theta}^{\hat{\theta}} x_2^*(\tau)\, d\tau = RHS.$$

The argument for the final case where $\theta \leq \theta_2^R \leq \hat{\theta}$ is analogous. QED

Proof of Lemma 4 For $\omega \in [\underline{\theta}, \bar{\theta}]$ consider the following relaxation of P_1^m:

$$R_1^m(\omega): \qquad \max_{x,U} \quad \int_{\underline{\theta}}^{\bar{\theta}} v(x(\theta)) - \theta x(\theta) - U(\theta)\, dF(\theta) \quad s.t. \tag{103}$$

$$U(\theta) - U(\omega) \geq \int_{\theta}^{\omega} \bar{x}(\tau \mid \omega)\, d\tau \quad \forall \theta < \omega \tag{104}$$

$$U(\theta) \geq u_R(\theta) \quad \forall \theta \geq \omega \tag{105}$$

x decreasing.

$R_1^m(\omega)$ relaxes P_1^m in two ways. First, after replacing (1) by the equivalent conditions (4), it imposes these constraints only for $\theta < \omega$, where ω is fixed. Second, it requires IR only for types $\theta \geq \omega$. We first derive a solution to $R_1^m(\omega)$ and then find a specific ω so that the solution also solves the original problem P_1^m.

Because x is decreasing, we have $\bar{x}(\tau \mid \omega) = x(\tau)$ in the IC constraints z(104), and since the objective is decreasing in information rents, both (104) and (105) are binding at an optimum.

After inserting them into the objective and performing a standard integration by parts step, $R_1^m(\omega)$ simplifies to

$$\max_x \quad \int_{\underline{\theta}}^{\omega} v(x(\theta)) - [\theta + \frac{F(\theta)}{f(\theta)}]x(\theta)\, dF(\theta) + \int_{\omega}^{\bar{\theta}} v(x(\theta)) - \theta x(\theta)\, dF(\theta) - K$$
$$s.t. \quad x \text{ decreasing}$$

where $K = F(\omega)u_R(\omega) + \int_{\omega}^{\bar{\theta}} u_R(\theta)\, dF(\theta)$ is a constant term that does not depend on x.

We relax the problem one step further and replace the monotonicity constraint by the weaker constraint

$$x(\theta) \geq x(\omega) \quad \forall \theta < \omega, \qquad x(\omega) \geq x(\theta) \quad \forall \theta > \omega.$$

For a fixed value of $x(\omega)$, the solution to this problem is as follows: for $\theta < \omega$ the objective is maximized by setting $x(\theta)$ equal to the maximizer of the virtual surplus, $x^-(\theta)$, if $x^-(\theta) > x(\omega)$, and to set $x(\theta)$ equal $x(\omega)$ otherwise. Likewise, for $\theta > \omega$ the objective is maximized by setting $x(\theta)$ equal to maximizer of the surplus, $x_0^*(\theta)$, if $x_0^*(\theta) < x(\omega)$, and to set $x(\theta)$ equal $x(\omega)$ otherwise:

$$\tilde{x}(\theta; x(\omega), \omega) = \begin{cases} \max\{x^-(\theta), x(\omega)\} & \text{if} \quad \theta < \omega \\ \min\{x_0^*(\theta), x(\omega)\} & \text{if} \quad \theta > \omega. \end{cases}$$

Notice that $\tilde{x}$ is decreasing in θ and thus satisfies the original monotonicity constraint. Therefore, problem R_1^m boils down to a maximization problem over $x(\omega)$:

$$R_1^m(\omega): \quad \max_{x(\omega)} \quad \int_{\underline{\theta}}^{\omega} v(\tilde{x}(\theta)) - [\theta + \frac{F(\theta)}{f(\theta)}]\tilde{x}(\theta)\, dF(\theta) + \int_{\omega}^{\bar{\theta}} v(\tilde{x}(\theta)) - \theta \tilde{x}(\theta)\, dF(\theta) \tag{106}$$

Denote the solution to this problem by $x^B(\omega)$. It is easy to see that $x^B(\omega)$ must lie in the interval $[x_0^*(\bar{\theta}), x^-(\underline{\theta})]$ (since otherwise the objective could be improved by either lowering or increasing $x(\omega)$). Therefore, because both x^- and x_0^* are decreasing, there are (unique) $\omega^- = \omega^-(x^B(\omega))$ and $\omega_0 = \omega_0(x^B(\omega))$ so that

$$x^-(\omega^-) = x^B(\omega) \quad \text{and} \quad x_0^*(\omega_0) = x^B(\omega), \tag{107}$$

and thus the optimal schedule becomes

$$x_\omega(\theta) \equiv \tilde{x}(\theta; x^B(\omega), \omega) = \begin{cases} x^-(\theta) & \text{if} \quad \theta < \omega^- \\ x^B(\omega) & \text{if} \quad \theta \in [\omega^-, \omega_0] \\ x_0^*(\theta) & \text{if} \quad \omega_0 < \theta. \end{cases} \tag{108}$$

If we insert x_ω into the objective, we obtain:

$$\int_{\underline{\theta}}^{\omega^-} v(x^-(\theta)) - [\theta + \frac{F(\theta)}{f(\theta)}]x^-(\theta)\, dF(\theta) + \int_{\omega^-}^{\omega} v(x^B(\omega)) - [\theta + \frac{F(\theta)}{f(\theta)}]x^B(\omega)\, dF(\theta)$$
$$+ \int_{\omega}^{\omega_0} v(x^B(\omega)) - \theta x^B(\omega)\, dF(\theta) + \int_{\omega_0}^{\bar{\theta}} v(x_0^*(\theta)) - \theta x_0^*(\theta)\, dF(\theta).$$

A straightforward but tedious calculation yields the first order condition for the maximizer $x^B(\omega)$:[11]

$$\int_{\omega-}^{\omega} v'(x^B(\omega)) - \theta - \frac{F(\theta)}{f(\theta)}\, dF + \int_{\omega}^{\omega_0} v'(x^B(\omega)) - \theta\, dF = 0. \tag{109}$$

In summary, the solution (x_ω, U_ω) to $R_m^1(\omega)$ is characterized by the conditions (107), (108), and (109), and the information rent U_ω is pinned down by the binding constraints (104) and (105).

We now choose ω such that (x_ω, U_ω) is also a solution to the original problem P_1^m. To do so, let μ be the unique solution to[12]

$$-u_R'(\mu^R) = x^B(\mu^R). \tag{110}$$

Recall the definition of (x_1^m, U_1^m) in the statement of the lemma, and note that $(x_\mu, U_\mu^R) = (x_1^m, U_1^m)$. Since $R_m^1(\omega)$ is a relaxed version of P_1^m, we have completed the proof if we show that (x_μ, U_μ^R) satisfies all constraints of P_1^m that are not in $R_m^1(\omega)$.

We first verify for $\theta < \mu^R \leq \hat{\theta}$ the neglected IC constraint

$$U_\mu(\theta) - U_\mu(\hat{\theta}) \geq \int_\theta^{\hat{\theta}} x_\mu(\tau)\, d\tau.$$

[11] It is straightforward to check that the first order condition is sufficient.

[12] To see that such a μ exists and is unique, note that as mentioned above $x^B(\omega) \in [x_0^*(\bar{\theta}), x^-(\underline{\theta})] = [x_0^*(\bar{\theta}), x_0^*(\underline{\theta})]$. Moreover, we have $-u_R'(\theta) \in (x_0^*(\bar{\theta}), x_0^*(\underline{\theta}))$ by assumption. Hence, as $-u_R'(\omega)$ is strictly increasing, and $x^B(\omega)$ is decreasing in ω, there is a unique solution $\mu \in (\underline{\theta}, \bar{\theta})$ to (110).

Indeed, since x_μ is decreasing, and $-u_R'$ is strictly increasing, (110) implies that $-u_R'(\tau) > x_\mu(\tau)$ for all $\tau > \mu^R$. With this and the definition of U_μ^R, we infer

$$U_\mu(\theta) - U_\mu(\hat{\theta}) = \int_\theta^{\mu^R} x_\mu(\tau)\, d\tau + u_R(\mu^R) - u_R(\hat{\theta}) = \int_\theta^{\mu^R} x_\mu(\tau)\, d\tau - \int_{\mu^R}^{\hat{\theta}} u_R'(\tau)\, d\tau > \int_\theta^{\hat{\theta}} x_\mu(\tau)\, d\tau,$$

as desired. All other constraints can be verified in a similar way. We omit the details. QED

Proof of Lemma 5: To see individual rationality, note that, by construction, we have $U_\varepsilon(\theta) = u_R(\theta)$ for $\theta \geq \mu^R - \varepsilon$, while for $\theta < \mu^R - \varepsilon$, we have $y_\varepsilon(\theta) \geq -u_R'(\theta)$, which implies

$$U_\varepsilon(\theta) = u_R(\mu^R - \varepsilon) + \int_\theta^{\mu^R-\varepsilon} y_\varepsilon(\tau)\, d\tau \geq u_R(\mu^R - \varepsilon) + \int_\theta^{\mu^R-\varepsilon} -u_R'(\tau)\, d\tau = u_R(\theta). \tag{111}$$

Hence, $(y_\varepsilon, U_\varepsilon)$ is individually rational.

We next show that $(y_\varepsilon, U_\varepsilon)$ is also incentive compatible. Note that by (1) and because of constant marginal cost, we have to show

$$U_\varepsilon(\theta) - U_\varepsilon(\hat{\theta}) \geq (\hat{\theta} - \theta) y_\varepsilon(\hat{\theta}) \quad \forall \theta \leq \hat{\theta}. \tag{112}$$

To see this inequality holds, consider some pair $\theta < \hat{\theta}$. Given $\mu^R + \delta < \mu_0$, there is a total of six constellations to check:

Case 1: $\theta < \mu^R - \varepsilon$.

(a) $\hat{\theta} < \mu^R - \varepsilon$: In this case, we have by construction and incentive compatibility of (x_1^m, U_1^m) that

$$\begin{aligned} U_\varepsilon(\theta) - U_\varepsilon(\hat{\theta}) = \int_\theta^{\hat{\theta}} y_\varepsilon(\tau) d\tau \quad &= \quad \int_\theta^{\hat{\theta}} x_1^m(\tau) d\tau \\ &= \quad U_1^m(\theta) - U_1^m(\hat{\theta}) \geq (\hat{\theta} - \theta) x_1^m(\hat{\theta}) = (\hat{\theta} - \theta) x_\varepsilon(\hat{\theta}). \end{aligned}$$

(b) $\mu^R - \varepsilon \leq \hat{\theta} < \mu^R + \delta$: For this case, first note that for all $\tau \in [\theta, \mu^R - \varepsilon]$, we have $y_\varepsilon(\tau) \geq x_m^B \geq y(\hat{\theta}, \varepsilon) = y_\varepsilon(\hat{\theta})$ so that

$$U_\varepsilon(\theta) = u_R(\mu^R - \varepsilon) + \int_\theta^{\mu^R-\varepsilon} y_\varepsilon(\tau)\, d\tau \geq u_R(\mu^R - \varepsilon) + \int_\theta^{\mu^R-\varepsilon} y_\varepsilon(\hat{\theta})\, d\tau = (\mu^R - \varepsilon - \theta) y_\varepsilon(\hat{\theta}) + u_R(\mu^R - \varepsilon).$$

Because $U_\varepsilon(\hat{\theta}) = u_R(\hat{\theta})$, it follows from (37) that

$$\begin{aligned} U_\varepsilon(\theta) - U_\varepsilon(\hat{\theta}) &\geq (\mu^R - \varepsilon - \theta) y_\varepsilon(\hat{\theta}) + u_R(\mu^R - \varepsilon) - u_R(\hat{\theta}) \\ &= (\mu^R - \varepsilon - \theta) y_\varepsilon(\hat{\theta}) + (\hat{\theta} - \mu^R + \varepsilon) y(\hat{\theta}, \varepsilon) = (\hat{\theta} - \theta) y_\varepsilon(\hat{\theta}). \end{aligned}$$

(c) $\mu^R + \delta \leq \hat{\theta}$: Similarly to case (b) we have $U_\varepsilon(\theta) \geq (\mu^R - \varepsilon - \theta) y_\varepsilon(\hat{\theta}) + u_R(\mu^R - \varepsilon)$. Moreover, since $\mu^R + \delta \leq \hat{\theta}$, we have $y(\hat{\theta}, \varepsilon) \geq y_\varepsilon(\hat{\theta})$ so that $U_\varepsilon(\mu^R - \varepsilon) - U_\varepsilon(\hat{\theta}) = u_R(\mu^R - \varepsilon) - u_R(\hat{\theta}) = (\hat{\theta} - \mu^R + \varepsilon) y(\hat{\theta}, \varepsilon) \geq (\hat{\theta} - \mu^R + \varepsilon) y_\varepsilon(\hat{\theta})$. Hence, using $u_R(\mu^R - \varepsilon) = U_\varepsilon(\mu^R - \varepsilon)$, it follows

$$U_\varepsilon(\theta) - U_\varepsilon(\hat{\theta}) = U_\varepsilon(\theta) - U_\varepsilon(\mu^R - \varepsilon) + U_\varepsilon(\mu^R - \varepsilon) - U_\varepsilon(\hat{\theta}) \geq (\hat{\theta} - \theta) y_\varepsilon(\hat{\theta}). \tag{113}$$

Case 2: $\mu^R - \varepsilon \leq \theta < \mu^R + \delta$.

(a) $\mu^R - \varepsilon \leq \hat{\theta} < \mu^R + \delta$. In this case, the concavity of u_R and (37) imply

$$\frac{u_R(\theta) - u_R(\hat{\theta})}{\hat{\theta} - \theta} \geq \frac{u_R(\mu^R - \varepsilon) - u_R(\hat{\theta})}{\hat{\theta} - \mu^R + \varepsilon} = y(\hat{\theta}, \varepsilon) = y_\varepsilon(\hat{\theta}), \tag{114}$$

which implies $U_\varepsilon(\theta) - U_\varepsilon(\hat{\theta}) = u_R(\theta) - u_R(\hat{\theta}) \geq (\hat{\theta} - \theta) y_\varepsilon(\hat{\theta})$.

(b) $\mu^R + \delta \leq \hat{\theta}$: As in case (a), the concavity of u_R implies that

$$\frac{u_R(\theta) - u_R(\hat{\theta})}{\hat{\theta} - \theta} \geq y(\hat{\theta}, \varepsilon) \geq y_\varepsilon(\hat{\theta}), \tag{115}$$

where the second inequality follows because $\mu^R + \delta \leq \hat{\theta}$. Hence: $U_\varepsilon(\theta) - U_\varepsilon(\hat{\theta}) = u_R(\theta) - u_R(\hat{\theta}) \geq (\hat{\theta} - \theta) y_\varepsilon(\hat{\theta})$.

Case 3: $\mu^R + \delta \leq \theta < \hat{\theta}$: In this case, we have

$$\begin{aligned} U_\varepsilon(\theta) - U_\varepsilon(\hat{\theta}) = u_R(\theta) - u_R(\hat{\theta}) &= \int_\theta^{\hat{\theta}} -u_R'(\tau) d\tau \\ &\geq \int_\theta^{\hat{\theta}} y_\varepsilon(\hat{\theta}) d\tau = (\hat{\theta} - \theta) y_\varepsilon(\hat{\theta}), \end{aligned}$$

where the inequality follows because $-u_R'(\tau) \geq y_\varepsilon(\tau) \geq y_\varepsilon(\hat{\theta})$ for all $\tau \in [\theta, \hat{\theta}]$.

This completes the proof. QED

Proof of Lemma 6

Let $\Delta S(\varepsilon)$ and $\Delta U(\varepsilon)$, respectively, be the difference in surplus and rents, respectively, between the non-monotone contract $(y_\varepsilon, U_\varepsilon)$ and the optimal monotone contract (x_1^m, U_1^m). Hence, $\Delta V(\varepsilon) \equiv \Delta S(\varepsilon) - \Delta U(\varepsilon)$ expresses the principal's payoff difference between $(y_\varepsilon, U_\varepsilon)$ and (x_1^m, U_1^m). We will show that $\Delta S'(0) = \Delta U'(0) = \Delta V'(0) = 0$ so that marginally the modification does not affect the surplus, rents, and principal's payoff. However, we show that we have $\Delta V''(0) = \Delta S''(0) - \Delta U''(0) > 0$ if and only if (41) is true. Thus, $\Delta V(\varepsilon)$ attains a local minimum at $\varepsilon = 0$ in this case.

We begin with three auxiliary observations that we prove at the end of the proof. We have:

$$\frac{\partial y(\mu^R - \varepsilon, \varepsilon)}{\partial \varepsilon} = \frac{u_R''(\mu^R - \varepsilon)}{2}, \quad \delta'(0) = 1, \quad \lim_{\varepsilon \to 0} \frac{\partial y(\mu^R + \delta(\varepsilon), \varepsilon)}{\partial \varepsilon} = \frac{u_R''(\mu^R)}{2}. \tag{116}$$

We first compute $\Delta S(\varepsilon)$ for sufficiently small $\varepsilon > 0$:

$$\begin{aligned}\Delta S(\varepsilon) &= \int_{\underline{\theta}}^{\bar{\theta}} [v(y_\varepsilon(\theta)) - \theta y_\varepsilon(\theta)] - [v(x_1^m(\theta)) - \theta x_1^m(\theta))]\, dF \\ &= \int_{\mu^R - \varepsilon}^{\mu^R + \delta(\varepsilon)} [v(y(\theta, \varepsilon)) - \theta y(\theta, \varepsilon)] - [v(x_m^B) - \theta x_m^B]\, dF\end{aligned}$$

Hence,

$$\begin{aligned}\Delta S'(\varepsilon) &= [v(y(\mu^R + \delta, \varepsilon)) - (\mu^R + \delta) y(\mu^R + \delta, \varepsilon) - v(x_m^B) + (\mu^R + \delta) x_m^B] f(\mu^R + \delta) \delta'(\varepsilon) \\ &\quad + [v(y(\mu^R - \varepsilon, \varepsilon)) - (\mu^R - \varepsilon) y(\mu^R - \varepsilon, \varepsilon) - v(x_m^B) + (\mu^R - \varepsilon) x_m^B] f(\mu^R - \varepsilon) \\ &\quad + \int_{\mu^R - \varepsilon}^{\mu^R + \delta} [v'(y(\theta, \varepsilon)) - \theta] \frac{\partial y(\theta, \varepsilon)}{\partial \varepsilon} - y(\theta, \varepsilon) + x_m^B\, dF\end{aligned}$$

The first term vanishes because, as previously noted, $y(\mu^R + \delta, \varepsilon) = x_m^B$. Moreover, by (39), we can replace $y(\mu^R - \varepsilon, \varepsilon)$ by $-u_R'(\mu^R - \varepsilon)$ in the second line. Thus

$$\begin{aligned}\Delta S'(\varepsilon) &= [v(-u_R'(\mu^R - \varepsilon)) - v(x_m^B) + (\mu^R - \varepsilon)\{u_R'(\mu^R - \varepsilon) + x_m^B\}] f(\mu^R - \varepsilon) \\ &\quad + \int_{\mu^R - \varepsilon}^{\mu^R + \delta} [v'(y(\theta, \varepsilon)) - \theta] \frac{\partial y(\theta, \varepsilon)}{\partial \varepsilon} - y(\theta, \varepsilon) + x_m^B\, dF.\end{aligned}$$

Now observe that by definition of μ^R, we have $-u_R'(\mu^R) = x_m^B$. Therefore, when inserting $\varepsilon = 0$, we obtain $\Delta S'(0) = 0$.

To identify a second order effect, we calculate the second derivative:

$$\begin{aligned}
\Delta S''(\varepsilon) &= -[v(-u_R'(\mu^R-\varepsilon))-v(x_m^B)+(\mu^R-\varepsilon)\{u_R'(\mu^R-\varepsilon)+x_m^B\}]f'(\mu^R-\varepsilon) \\
&\quad +[v'(-u_R'(\mu^R-\varepsilon))u_R''(\mu^R-\varepsilon)-\{u_R'(\mu^R-\varepsilon)+x_m^B\}-(\mu^R-\varepsilon)u_R''(\mu^R-\varepsilon)]f(\mu^R-\varepsilon) \\
&\quad +\{[v'(y(\mu^R+\delta,\varepsilon))-\mu^R-\delta]\frac{\partial y(\mu^R+\delta,\varepsilon)}{\partial \varepsilon}-y(\mu^R+\delta,\varepsilon)+x_m^B\}f(\mu^R+\delta)\delta'(\varepsilon) \\
&\quad +\{[v'(y(\mu^R-\varepsilon,\varepsilon))-\mu^R+\varepsilon]\frac{\partial y(\mu^R-\varepsilon,\varepsilon)}{\partial \varepsilon}-y(\mu^R-\varepsilon,\varepsilon)+x_m^B\}f(\mu^R-\varepsilon) \\
&\quad +\int_{\mu^R-\varepsilon}^{\mu^R+\delta}\frac{\partial}{\partial\varepsilon}\left([v'(y(\theta,\varepsilon))-\theta]\frac{\partial y(\theta,\varepsilon)}{\partial\varepsilon}-y(\theta,\varepsilon)+x_m^B\right)dF \\
&= T_1+T_2+T_3+T_4+T_5.
\end{aligned}$$

We next evaluate the five different terms of $\Delta S''(\varepsilon)$ at $\varepsilon=0$. Recalling that $-u_R'(\mu^R)=x_m^B$, the first term vanishes at $\varepsilon=0$, and the second term becomes

$$T_2=\{v'(x_m^B)-\mu\}u_R''(\mu^R)f(\mu^R).$$

Moreover, the fact that $y(\mu^R+\delta,\varepsilon)=x_m^B$ together with (116) implies that the third term becomes

$$T_3=[v'(x_m^B)-\mu]\frac{u_R''(\mu^R)}{2}f(\mu^R).$$

Further, by (39), we have $y(\mu^R-\varepsilon,\varepsilon)=-u_R'(\mu^R-\varepsilon)$. Thus, with (116) and the fact that $-u_R'(\mu^R)=x_m^B$, the fourth term becomes at $\varepsilon=0$:

$$T_4=[v'(x_m^B)-\mu^R]\frac{u_R''(\mu^R)}{2}f(\mu^R).$$

Finally, the fifth term vanishes at $\varepsilon=0$, as $\delta(0)=0$. Collecting the five terms, we get:

$$\Delta S''(0) \quad = \quad 2[v'(x_m^B)-\mu^R]u_R''(\mu^R)f(\mu^R)<0,$$

where the strict inequality follows, because $u_R''<0$, as u_R is strictly concave, and $v'(x_m^B)>\mu^R$, as $x_m^B<x_0(\mu^R)$ is distorted downwards.

We next consider the difference in information rents:

$$\Delta U(\varepsilon) \quad \equiv \quad \int_{\underline{\theta}}^{\bar{\theta}} U_\varepsilon(\tau)-U_1^m(\tau)\,dF.$$

Note that for types $\tau \geq \mu^R$ we have $U_\varepsilon(\tau) = U_1^m(\tau) = u_R(\tau)$, while for types $\tau \in (\mu^R - \varepsilon, \mu^R)$, we have

$$\begin{aligned} U_\varepsilon(\tau) - U_1^m(\tau) &= u_R(\tau) - u_R(\mu^R) - \int_\tau^{\mu^R} x_1^m(\theta)\, d\theta \\ &= u_R(\tau) - u_R(\mu^R) - \int_\tau^{\mu^R} x_m^B\, d\theta \\ &= u_R(\tau) - u_R(\mu^R) - [\mu^R - \tau]x_m^B. \end{aligned}$$

Moreover, for types $\tau \leq \mu^R - \varepsilon$, the difference in information rents is

$$\begin{aligned} U_\varepsilon(\tau) - U_1^m(\tau) &= u_R(\mu^R - \varepsilon) - u_R(\mu^R) + \int_\tau^{\mu^R-\varepsilon} y_\varepsilon(\theta)\, d\theta - \int_\tau^{\mu^R} x_1^m(\theta)\, d\theta \\ &= u_R(\mu^R - \varepsilon) - u_R(\mu^R) - \int_{\mu^R-\varepsilon}^{\mu^R} x_m^B\, d\theta \\ &= u_R(\mu^R - \varepsilon) - u_R(\mu^R) - \varepsilon x_m^B. \end{aligned}$$

It therefore follows that

$$\Delta U(\varepsilon) = \int_{\underline{\theta}}^{\mu^R-\varepsilon} u_R(\mu^R - \varepsilon) - u_R(\mu^R) - \varepsilon x_m^B\, dF + \int_{\mu^R-\varepsilon}^{\mu^R} u_R(\tau) - u_R(\mu^R) - [\mu^R - \tau]x_m^B\, dF.$$

Taking the derivative with respect to ε, we get

$$\begin{aligned} \Delta U'(\varepsilon) &= \int_{\underline{\theta}}^{\mu^R-\varepsilon} [-u_R'(\mu^R - \varepsilon) - x_m^B]\, dF - [u_R(\mu^R - \varepsilon) - u_R(\mu^R) - \varepsilon x_m^B] f(\mu^R - \varepsilon) \\ &\quad + [u_R(\mu^R - \varepsilon) - u_R(\mu^R) - (\mu^R - \mu^R - \varepsilon)x_m^B] f(\mu^R - \varepsilon) \\ &= F(\mu^R - \varepsilon)[u_R'(\mu^R) - u_R'(\mu^R - \varepsilon)]. \end{aligned}$$

Hence, $\Delta U'(0) = 0$ so that there is no first order effect at $\varepsilon = 0$.

To derive the second order effect at $\varepsilon = 0$, observe:

$$\Delta U''(\varepsilon) = -f(\mu^R - \varepsilon)[u_R'(\mu^R) - u_R'(\mu^R - \varepsilon)] + F(\mu^R - \varepsilon)u_R''(\mu^R - \varepsilon).$$

Hence, $\Delta U''(0) = F(\mu^R)u_R''(\mu^R) < 0$.

Combining the expressions for ΔS and ΔU, we obtain for the principal's payoff difference that

$$\Delta V'(0) = \Delta S'(0) - \Delta U'(0) = 0,$$

and

$$\begin{aligned}\Delta V''(0) &= \Delta S''(0) - \Delta U''(0)\\ &= 2[v'(x_m^B) - \mu]u_R''(\mu^R)f(\mu^R) - F(\mu^R)u_R''(\mu^R)\\ &= \left[v'(x_m^B) - \mu^R + \left(v'(x_m^B) - \mu - \frac{F(\mu^R)}{f(\mu^R)}\right)\right] f(\mu^R)u_R''(\mu^R).\end{aligned}$$

Because $f(\mu^R)u_R''(\mu^R) < 0$, this expression is positive (and thus the marginal modification is profitable) if and only if the term in the square brackets is negative, i.e., if and only if condition (41) is true, and this establishes the first part of the proposition.

It remains to show the auxiliary observations (116) that

$$\frac{\partial y(\mu^R - \varepsilon, \varepsilon)}{\partial \varepsilon} = \frac{u_R''(\mu^R - \varepsilon)}{2}, \quad \delta'(0) = 1, \quad \lim_{\varepsilon \to 0} \frac{\partial y(\mu^R + \delta(\varepsilon), \varepsilon)}{\partial \varepsilon} = \frac{u_R''(\mu^R)}{2}.$$

To see the first and the third part, note that, by (37,) we have

$$\frac{\partial y(\theta, \varepsilon)}{\partial \varepsilon} = \frac{\partial}{\partial \varepsilon} \frac{u_R(\mu^R - \varepsilon) - u_R(\theta)}{\theta - \mu^R + \varepsilon} = \frac{-u_R'(\mu^R - \varepsilon)(\theta - \mu^R + \varepsilon) - (u_R(\mu^R - \varepsilon) - u_R(\theta))}{(\theta - \mu^R + \varepsilon)^2}.$$

By using l'Hopital's rule twice,

$$\begin{aligned}\frac{\partial y(\mu^R - \varepsilon, \varepsilon)}{\partial \varepsilon} &= \lim_{\theta \to \mu^R - \varepsilon} \frac{\partial y(\theta, \varepsilon)}{\partial \varepsilon}\\ &= \lim_{\theta \to \mu^R - \varepsilon} \frac{-u_R'(\mu^R - \varepsilon)(\theta - \mu^R + \varepsilon) - u_R(\mu^R - \varepsilon) + u_R(\theta)}{(\theta - \mu^R + \varepsilon)^2}\\ &= \lim_{\theta \to \mu^R - \varepsilon} \frac{-u_R'(\mu^R - \varepsilon) + u_R'(\theta)}{2(\theta - \mu^R + \varepsilon)}\\ &= \lim_{\theta \to \mu^R - \varepsilon} \frac{u_R''(\theta)}{2} = \frac{u_R''(\mu^R - \varepsilon)}{2}.\end{aligned}$$

Similarly,

$$\begin{aligned}
\lim_{\varepsilon\to 0}\frac{\partial y(\mu^R+\delta(\varepsilon),\varepsilon)}{\partial\varepsilon} &= \lim_{\varepsilon\to 0}\frac{-u_R'(\mu^R-\varepsilon)(\delta(\varepsilon)+\varepsilon)-[u_R(\mu^R-\varepsilon)-u_R(\mu^R+\delta(\varepsilon))]}{(\delta(\varepsilon)+\varepsilon)^2}\\
&= \lim_{\varepsilon\to 0}\frac{-u_R'(\mu^R-\varepsilon)(\delta(\varepsilon)+\varepsilon)-x_m^B(\delta(\varepsilon)+\varepsilon)}{(\delta(\varepsilon)+\varepsilon)^2}\\
&= \lim_{\varepsilon\to 0}\frac{-u_R'(\mu^R-\varepsilon)+u_R'(\mu^R)}{\delta(\varepsilon)+\varepsilon}\\
&= \lim_{\varepsilon\to 0}\frac{u_R''(\mu^R-\varepsilon)}{\delta'(\varepsilon)+1}=\frac{u_R''(\mu^R)}{2}.
\end{aligned}$$

Here, we have used (36) in the second line and l'Hopital's rule as well as the fact that $\delta'(0)=1$ in the final line.

Finally, to see $\delta'(0)=\lim_{\varepsilon\to 0}\delta'(\varepsilon)=1$, recall from (36) that $\delta(\varepsilon)$ is implicitly defined by the relationship

$$u_R(\mu^R-\varepsilon)-u_R(\mu^R+\delta)=(\delta+\varepsilon)x_m^B.$$

By the implicit function theorem and using $u_R'(\mu^R)=-x_m^B$, we obtain:

$$\delta'(\varepsilon)=\frac{u_R'(\mu^R)-u_R'(\mu^R-\varepsilon)}{u_R'(\mu^R+\delta(\varepsilon))-x_m^B}.$$

By l'Hopital's rule, it follows that

$$\delta'(0)=\lim_{\varepsilon\to 0}\delta'(\varepsilon)=\lim_{\varepsilon\to 0}\frac{u_R''(\mu^R-\varepsilon)}{u_R''(\mu^R+\delta(\varepsilon))\delta'(\varepsilon)}=\frac{1}{\delta'(0)}.$$

Hence, we have $\delta'(0)^2=1$, and since $\delta'(\varepsilon)\geq 0$, this implies $\delta'(0)=1$.

And this completes the proof. QED

Proof of Proposition 6: We show that an increasing density f implies (41). Suppose to the contrary that f is increasing and (41) is not true, that is,

$$-\left(v'(x_m^B)-\mu-\frac{F(\mu^R)}{f(\mu^R)}\right)\leq v'(x_m^B)-\mu^R. \tag{117}$$

We derive a contradiction to (34). Indeed, we show first that (117) implies

$$-\int_{\mu^-}^{\mu^R} v'(x_m^B)-\theta-\frac{F(\theta)}{f(\theta)}\,d\theta<\int_{\mu^R}^{\mu_0} v'(x_m^B)-\theta\,d\theta. \tag{118}$$

To see this, note that with the change of variables $\theta=2\mu^R-\tilde{\theta}$, we can write the first integral as

$$-\int_{\mu^-}^{\mu^R} v'(x_m^B)-\theta-\frac{F(\theta)}{f(\theta)}\,d\theta=\int_{\mu^R}^{2\mu^R-\mu^-}-\left(v'(x_m^B)-(2\mu^R-\tilde{\theta})-\frac{F(2\mu^R-\tilde{\theta})}{f(2\mu^R-\tilde{\theta})}\right)d\tilde{\theta}$$

By (117), we have for $\tilde{\theta}=\mu^R$:

$$-\left(v'(x_m^B)-(2\mu^R-\tilde{\theta})-\frac{F(2\mu^R-\tilde{\theta})}{f(2\mu^R-\tilde{\theta})}\Big|_{\tilde{\theta}=\mu^R}\right)=-\left(v'(x_m^B)-\mu-\frac{F(\mu^R)}{f(\mu^R)}\right)\leq v'(x_m^B)-\mu^R.$$

Moreover, since the hazard rate F/f is increasing, we have that

$$\frac{d}{d\tilde{\theta}}-\left(v'(x_m^B)-(2\mu^R-\tilde{\theta})-\frac{F(2\mu^R-\tilde{\theta})}{f(2\mu^R-\tilde{\theta})}\right)<-1=\frac{d}{d\tilde{\theta}}(v'(x_m^B)-\tilde{\theta}).$$

These two observations imply that for all $\tilde{\theta}>\mu^R$, we have

$$-\left(v'(x_m^B)-(2\mu^R-\tilde{\theta})-\frac{F(2\mu^R-\tilde{\theta})}{f(2\mu^R-\tilde{\theta})}\right)<v'(x_m^B)-\tilde{\theta}.$$

Because, by definition, $v'(x_m^B)-(2\mu^R-\tilde{\theta})-\frac{F(2\mu^R-\tilde{\theta})}{f(2\mu^R-\tilde{\theta})}$ equals zero at $\tilde{\theta}=2\mu^R-\mu^-$, and $v'(x_m^B)-\tilde{\theta}$ equals zero at $\tilde{\theta}=\mu_0$, this also implies that $\mu_0>2\mu^R-\mu^-$ and $v'(x_m^B)>\tilde{\theta}$ for $\tilde{\theta}\in(2\mu^R-\mu^-,\mu_0)$. Taken together, these observations imply:

$$\begin{aligned}-\int_{\mu^-}^{\mu^R} v'(x_m^B)-\theta-\frac{F(\theta)}{f(\theta)}\,d\theta &= \int_{\mu^R}^{2\mu^R-\mu^-}-\left(v'(x_m^B)-(2\mu^R-\tilde{\theta})-\frac{F(2\mu^R-\tilde{\theta})}{f(2\mu^R-\tilde{\theta})}\right)d\tilde{\theta}\\ &< \int_{\mu^R}^{2\mu^R-\mu^-} v'(x_m^B)-\tilde{\theta}\,d\tilde{\theta}\\ &\leq \int_{\mu^R}^{\mu_0} v'(x_m^B)-\tilde{\theta}\,d\tilde{\theta},\end{aligned}$$

which proves (118) whenever (117) is true. Since the density is increasing and all terms are positive, we conclude:

$$\begin{aligned}
-\int_{\mu^-}^{\mu^R}\left\{v'(x_m^B)-\theta-\frac{F(\theta)}{f(\theta)}\right\}f(\theta)\,d\theta &\le -\int_{\mu^-}^{\mu^R}\left\{v'(x_m^B)-\theta-\frac{F(\theta)}{f(\theta)}\right\}f(\mu^R)\,d\theta \\
&< \int_{\mu^R}^{\mu_0}\left\{v'(x_m^B)-\tilde{\theta}\right\}f(\mu^R)\,d\tilde{\theta} \\
&\le \int_{\mu^R}^{\mu_0}\left\{v'(x_m^B)-\tilde{\theta}\right\}f(\tilde{\theta})\,d\tilde{\theta},
\end{aligned}$$

but this contradicts (34) and completes the proof. QED

References

Beaudry, P., C. Blackorby, and D. Szalay. 2009. "Taxes and Employment Subsidies in Optimal Redistribution Programs", American Economic Review 99 (1), 216–42.

Carroll, G. (2012), "When are Local Incentive Constraints Sufficient?", Econometrica 80 (2), 661–686.

Celik, G. (2006), "Mechanism Design with Weaker Incentive Compatibility Constraints", Games and Economic Behavior 56, 37—44.

Che, Y.-K. and I. Gale (2000), "The optimal mechanism for selling to a budget-constrained buyer", Journal of Economic Theory 92, 198—233.

Deneckere, R. and S. Severinov (2008), "Mechanism Design with Partial State Verifiability", Games and Economic Behavior 64, 487—513.

Dizdar, D., A. Gershkov, B. Moldovanu (2011) "Revenue maximization in the dynamic knapsack problem", Theoretical Economics 6, 157—184.

Iyengar, G., Kumar, A. (2008) "Optimal procurement mechanisms for divisible goods with capacitated suppliers" Review of Economic Design 12, 129—154.

Lewis, T. R. and D. E. M. Sappington (1989), "Countervailing Incentives in Agency Problems", Journal of Economic Theory 49.2, 294–313.

Maggi, G. and A. Rodriguez-Clare (1995), "On countervailing incentives", Journal of Economic Theory 66.1, 238–263.

Moore, J. (1984), "Global Incentive Constraints in Auction Design", Econometrica 52, 1523–1535.

Pai, M. M., and R. Vohra, (2013), "Optimal Dynamic Auctions and Simple Index Rules", Mathematics of Operations Research, 38(4), 682—697.

Riley, J. and R. Zeckhauser (1983), "Optimal Selling Strategies: When to Haggle, when to Hold Firm.", Quarterly Journal of Economics 98.2, 267–289.

Sher, I. and R. Vohra (2015), "Price Discrimination through Communication", Theoretical Economics 10, 597–648.