Theoretical Economics 16 (2021), 1281–1312 1555-7561/20211281

# Payoff implications of incentive contracting

Daniel F. Garrett
Department of Economics, University of Essex

In the context of a canonical agency model, we study the payoff implications of introducing optimally structured incentives. We do so from the perspective of an analyst who does not know the agent's preferences for responding to incentives, but does know that the principal knows them. We provide, in particular, tight bounds on the principal's expected benefit from optimal incentive contracting across feasible values of the agent's expected rents. We thus show how economically relevant predictions can be made robustly given ignorance of a key primitive.

Keywords. Asymmetric information, mechanism design, robustness, procurement.

JEL classification. D82.

## 1. Introduction

Economists often emphasize the virtues of incentives across settings from regulation and procurement to worker and executive compensation. Nonetheless, moves to introduce explicit incentives are often criticized for leaving large rents to agents. This is particularly true when the principal in the relationship appears to gain little from offering incentives. To give an example, reforms in the United Kingdom in the 1980s led public utilities to be privatized and subjected to regulation, part of an effort to harness the efficiency advantages of financial incentives. Later, the Blair government introduced the "windfall tax" on utility companies, a response to negative public sentiment surrounding the earlier reforms. The negative sentiment was fed by *both* the magnitude of corporate profits and a perception that the public (or, more directly, the government as principal) had failed to benefit from the changes.[1]

Daniel F. Garrett: dg21984@essex.ac.uk; dfgarrett@gmail.com
This work was completed while at Toulouse School of Economics, University of Toulouse Capitole. For invaluable research assistance, I am grateful to Stefan Pollinger. For helpful comments, I am grateful to Mike Abito, Nikhil Agarwal, Gani Aldashev, Dirk Bergemann, Mehmet Ekmekci, Volker Nocke, Alessandro Pavan, Nicola Pavoni, Patrick Rey, Jean Tirole, Alex Wolitzky, and to seminar participants at Bocconi, Boston College, ECARES, the Frankfurt School of Finance and Management and Goethe University Frankfurt, MIT, the University of Bonn, the Toulouse School of Economics Winter IO meeting, and the Workshop on New Directions in Mechanism Design at Stony Brook. The project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement 714147), as well as from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) programme (Grant ANR-17-EURE-0010).

[1] See Chennells (1997) for a discussion.

 https://doi.org/10.3982/TE4293

Economic theory offers a possible lens through which to examine the distribution of welfare that results from the introduction of incentives. Yet, putting incentive theory to work, say to make predictions on welfare implications, is difficult. In particular, determining the fundamentals of the economic environment is often challenging. It is, therefore, natural to ask what predictions are possible when details of the economic environment are not well understood.

This paper is concerned with the predictions available when ambiguity concerning the environment persists due to a lack of experience with incentives. We consider a canonical single-agent procurement framework where incentive contracts are to be newly introduced. We then determine the predictions available to an analyst who is ignorant regarding the agent's preferences (equivalently, technology) for responding to incentives. We suppose the analyst does, however, correctly anticipate the contracting model, and that the principal will choose the incentive contract to minimize expected payments given knowledge of the agent's preferences. That is, while the principal solves a well known procurement model, the analyst is tasked with predicting outcomes in the absence of a key primitive.

Apart from regularity conditions on the agent's preferences, our analyst has no way to determine the agent's willingness to respond to incentives. The agent might be highly responsive to any incentives offered or not responsive at all. For this reason, the analyst has no hope of making informative predictions on the *absolute* level of the agent's performance, or on the gains to the principal of incentive contracting (say relative to an alternative regime where no incentives are offered). However, it turns out that statements on *relative* payoffs of the players are possible. Our main contribution is to obtain a lower bound on the principal's gains from incentives given possible values of the agent's expected rents. This bound is negligible conditional on the agent earning negligible rent in an optimal incentive contract, but it increases with the agent's rents. Consider then a policy maker or other interested party who has the same information as the analyst and is concerned about the possibility of large agent rents under an optimal contract. For instance, in public procurement where the agent is a private contractor, the concern may be about enriching wealthy shareholders and exacerbating inequality. The analyst can offer the prediction that if the agent's preferences (or technology) turn out to be such that he can expect high rents, then the expected gains from incentive contracting will lie above a known bound. Depending on the value of the bound, such a prediction might offer some reassurance.

### *Details of the setting*

We specialize in this paper to a procurement model where the principal obtains a fixed number of units from the agent. Realized production costs are public, so payments to the agent can be conditioned on them; i.e., the setting is one of cost-based procurement. The cost of supplying these units without effort—often termed the agent's innate cost—is the agent's private information. The agent can privately choose effort to reduce the publicly observed production cost below his innate cost. The agent's preferences for cost-reducing effort are characterized by a disutility of effort function, taken to be

increasing, convex, and independent of the innate cost. The principal, having a prior on the innate costs and knowing the disutility function, offers an optimal contract. Optimal contracts can be determined using a mechanism design approach, as in Laffont and Tirole (1986).

As noted, the analyst's problem is to determine welfare predictions for optimal contracts. These predictions are made knowing that the above model of cost-based procurement applies and given the prior on innate costs, but without knowledge of the agent's disutility function. Availability of a prior on innate costs is in line with the analyst having observations on past cost performance under cost-plus contracting. Since cost-plus contracts pay the agent only the observed production cost, these contracts provide no incentives for effort and so induce a production cost equal to the innate cost. One interpretation of the analyst's problem is that she is tasked with informing a policy decision to introduce incentive contracts given a history of cost-plus contracting. While the analyst is ignorant of the disutility function, she anticipates the information will become available to the principal if a decision to implement incentive contracting proceeds (say, because implementation is accompanied by further study of the agent's technology or by the hiring of external expertise).

### *Main results*

Our main results characterize the expected payoffs from optimal incentive contracting across all permitted agent preferences for cost-reducing effort. A range of values for expected agent rents is possible in an optimal contract, depending on the disutility function. We find a lower bound on the principal's gains from incentives for each level of agent rents that is not only increasing with agent rents (as mentioned above), but is also convex. Convexity provides a stronger sense in which the principal's relative guarantee improves with the welfare of the agent.

We then investigate how the relative guarantee on the principal's expected gains from incentives depends on the distribution of innate costs. When the innate cost is uniformly distributed, the guarantee is exactly the size of agent expected rents. In other words, the principal is guaranteed at least half the efficiency gains from incentive contracting. More generally, we provide sufficient conditions on the distribution of innate costs for the guarantee to be greater than one half and conditions for the guarantee to be less (i.e., for the principal to obtain less than half of the efficiency gains for some realization of agent preferences). The focus on the possibility of a 50/50 split of surplus is not only analytically convenient, but may be relevant for assessing "fairness" considerations surrounding the introduction of incentives. For instance, as discussed by Lopomo and Ok (2016, p. 263), a regularity in laboratory experiments where one party holds a clear strategic advantage (such as ultimatum games) is surplus division around the 50/50 split, presumably because subjects consider such a split as fair. A guarantee that the principal will obtain at least half the surplus from the introduction of incentives is in this sense a guarantee that the expected outcome of contracting will not be unfair on the principal (or in public procurement settings, on the general public on whose behalf the principal might be presumed to act).

*Analytical approach*

In terms of our analytical approach, the main novelty lies in the problem of mapping, as a function of agent expected rents, the innate cost distribution to a tight or sharp lower bound for the principal's gains from incentives.[2] This takes place in two main steps. The first step involves determining a lower bound on the principal's expected gains from incentives for each level of agent expected rents. The second step involves showing that the bound is tight.

*First step.* Following a characterization of mechanisms that solve the principal's problem (see Section 3), we determine an expression for the principal's expected gains from incentives that depends on the agent's marginal disutility of effort at each value of the innate cost. A key property of any optimal mechanism is that the agent's marginal disutility of effort is monotone decreasing in the innate cost. Since the agent's expected rents are also determined by the marginal disutility of effort, a lower bound for the principal's expected gains given agent rents is obtained by minimizing them over the agent's marginal disutility of effort subject to a constraint determined by the agent's rents and subject to the aforementioned monotonicity. While optimization subject to monotonicity constraints has often presented analytical challenges in the literature (see Hellwig (2008) for a discussion), it turns out to be tractably solved in our case by reference to a "convexification argument" explained below. The solution to the minimization problem provides a lower bound on the principal's gains from incentives.

*Second step.* The remaining step establishes tightness. This involves exhibiting disutility functions for which the principal's expected gains are equal to, or at least arbitrarily close to, the aforementioned bound. This is done by finding disutility functions such that, for the solution to the principal's mechanism design problem, the agent's marginal disutility of effort corresponds to the solution to the minimization problem in the previous step. In this sense, it justifies our focus on the agent's marginal disutility of effort in an optimal mechanism.

*Relationship to empirical literature on procurement*

Our results connecting the shape of the innate cost distribution to the relative welfare of the players may be helpful for understanding existing empirical work on procurement. For instance, we argue (in Section 5) that efficiency gains guaranteed for the principal tend to be larger when the agent's innate costs are more concentrated at lower values. This seems to be the prevalent case in the empirical literature, where there is often a long tail of firms with high costs.

The fact that the shape of the innate cost distribution plays a critical role in determining the possible welfare implications of incentive contracting is, in fact, anticipated by empirical work. A case in point is Abito (2015), who uses a version of Laffont and Tirole's (1986) model in his study of electric utilities, and where each firm's innate cost efficiency is determined by its type. He explains that his counterfactual predictions turn on the shape of the type distribution:

[2]The terms "tight" and "sharp" are applied variously to bounds that cannot be improved on.

> The shape of the type distribution is an important determinant in the design of the optimal mechanism and the welfare gains it delivers. The gains from the optimal mechanism are not simply about getting all firms to exert more effort, but rather more about effectively mitigating the cost of the regulator's informational disadvantage. The cost of the informational disadvantage is determined by the shape of the type distribution which therefore importantly affects the measure of welfare gains. Thus, as in most studies on asymmetric information, the key challenge is to estimate this distribution.

The results in the present paper could turn out to be useful in such empirical settings, because they provide a way to make relevant predictions without data to inform agent preferences for effort. The usual approach in empirical work using the Laffont and Tirole framework (such as in Abito's paper, but also in other important contributions such as Gagnepain and Ivaldi (2002)) is to estimate disutility functions using data that go beyond data on firms' innate costs.[3]

### *Layout*

The layout of the paper is as follows. The rest of this section discusses further related literature. Section 2 then introduces the cost-based procurement model, and Section 3 provides an analysis of optimal contracting in this model. Section 4 derives our characterization of expected welfare under optimal contracts. Section 5 shows how the set of feasible expected payoffs depends on the distribution of innate costs. Section 6 concludes. Formal proofs not included in the main text are provided in the Appendix.

### *Further related literature*

At a conceptual level, the value in obtaining "robust predictions" on welfare in our environment is related to a broader interest in the theory literature for obtaining robust predictions on economic variables. Notably, work such as Bergemann and Morris (2013, 2016) and Bergemann et al. (2015, 2017) explore the predictions that can be made by an outside observer to an interaction, given information on certain fundamentals, but lacking other pertinent details. The pertinent details in these papers relate to the information structure, i.e., players' information on the payoff-relevant state or payoff types, and, where relevant, their higher-order beliefs.[4] An important part of their motivation is that in many settings, "the information structure will generally be very hard [for an outsider] to observe, as it is in the agents' minds and does not necessarily have an observable counterpart" (Bergemann and Morris (2013, p. 1252)). Our motivation is similar, although the economic objects are different. Our interest is in contracting settings where certain information—especially the distribution of innate costs—may be readily observed (or at least inferred from data) and at the same time, other information—especially regarding the agent's preferences for effort—is not.

[3]For instance, Abito uses firm performance in between regulatory rate cases, when incentives for cost reduction are strongest. Gagnepain and Ivaldi use the performance of those firms subject to high-powered fixed-price contracts.

[4]Interest in making robust predictions is clearly more widespread in the theory literature. An example is Segal and Whinston (2003), who determine predictions on outcomes that hold across a broad class of contracting games with a single principal and many agents.

Our work is also connected to the developing literature on robust incentive contracts, in that our focus is on the lower bound of the principal's performance across possible realizations of ambiguous preferences. The robust contracting literature takes a worst-case perspective to evaluating incentive contracts (in terms of levels or regret-type criteria) and includes work such as Hurwicz and Shapiro (1978), Chassang (2013), Garrett (2014), Carroll (2015), and Dai and Toikka (2017). If we view "adversarial Nature" as choosing the agent's preferences or technology, then the main difference between this literature and the present paper is the timing of Nature's move. In this paper, Nature moves before the principal determines an optimal contract, whereas in the robust contracting literature, Nature's move comes after. There is some similarity in proof approach to the paper by Garrett, who considers a similar procurement model, but for a principal who is ignorant of the agent's disutility function. The main similarity is the need to construct adversarial disutility functions (i.e., disutility functions chosen so that the principal obtains a low payoff).

Another connection to the robust contracting literature is the observation that high payoffs for the agent can imply a good outcome for the principal. This idea is exploited in the analysis of linear contracts by Chassang (2013) and Carroll (2015), where linear contracts turn out to guarantee the principal a payoff that is proportional to the agent's rents. The present analysis also shows that a high value of expected agent rents can imply a high guarantee on the principal's expected gains from incentive contracting. This guarantee is obtained under the hypothesis of optimal contracting by the principal, rather than given an arbitrary linear incentive scheme.

A final strand of literature to be mentioned is econometric analyses of incentive design in regulation and procurement. For instance, Perrigne and Vuong (2011) show how one can identify (in their case, nonparametrically) structural parameters of the Laffont and Tirole (1986) model using data on observables such as realized demand, realized cost, and payments to the agent. A connection to the present work is the objective of drawing implications from a combination of weak assumptions on model primitives together with the hypothesis of optimal contracting.

## 2. The model

### *The procurement model*

We introduce our ideas in a standard procurement framework that is a simplified version of Laffont and Tirole ((1986, 1993); henceforth, LT). The model we consider is popular in the literature; see, for instance, Rogerson (2003) and Chu and Sappington (2007).

The principal is responsible for procuring a fixed quantity of a good from an agent who is the supplier. We normalize the quantity to a single unit. The principal aims to procure this unit while minimizing total payments to the agent.

The agent is associated with an "innate cost" $\beta$ (that we sometimes refer to as his type), and a cost-reduction technology. The latter is characterized by a disutility function $\psi : \mathbb{R} \to \mathbb{R}_+$. If the agent exerts effort $e$ to reduce costs, then he incurs a private disutility $\psi(e)$. This disutility could represent the inconvenience of putting measures in place to lower costs or could represent physical costs incurred by the agent that are

not direct costs accounted for in the contract. After effort $e$, the realized production cost is $C = \beta - e \in \mathbb{R}$. While the principal knows the function $\psi$ and observes the realized production cost $C$, both the innate cost $\beta$ and the effort $e$ are the agent's private information.

The environment permits transfers between the principal and agent. Following LT, we adopt the accounting convention that the realized production cost $C$ is paid by the principal. In addition, the agent receives a transfer $y$. Payoffs are quasi-linear in money, so that the agent's Bernoulli utility (in case of effort $e$ and transfer $y$) is $y - \psi(e)$. In case the agent refuses the contract, he does not produce and earns payoff zero. Procurement of the unit is taken to be essential for the principal. Subject to the constraint of ensuring the unit is supplied, the principal's objective is then to minimize the expectation of total expenditure $y + C$.

The disutility function $\psi$ satisfies the following requirements. It is taken to be nondecreasing and convex, with $\psi$ strictly increasing on $\mathbb{R}_+$ and constant at zero on $\mathbb{R}_-$. We take $\psi$ to be Lipschitz continuous with the best Lipschitz constant strictly greater than 1. We then let $\Psi$ be the set of all disutility functions $\psi$ that satisfy these conditions.

That the agent incurs positive disutility from positive effort ensures that the innate cost $\beta$ has the intended interpretation: the agent chooses zero effort when incentives are absent. We assume the agent can costlessly inflate the production cost above the innate cost by choosing negative effort, although this does not occur in equilibrium.[5] The possibility for the agent to inflate the production cost above his innate cost ensures that every type can attain the production cost prescribed for every other type. This in turn readily permits a characterization of incentive compatibility through application of a certain envelope theorem (that of Carbajal and Ely (2013), as described below).[6] Monotonicity and convexity of $\psi$ are standard shape restrictions. It is natural to expect that higher effort is more costly (monotonicity) and oftentimes additionally that there are diminishing returns to cost reductions (convexity). Diminishing returns would also imply the best Lipschitz constant being greater than 1. This is a kind of Inada condition that guarantees efficient effort is bounded and that plays a role in the existence of optimal mechanisms. Lipschitz continuity itself is a technical condition, which, given convexity of $\psi$, is a restriction on this function only at large values of effort $e$ that are not chosen in equilibrium. It is again helpful for permitting application of the envelope theorem.

Note in addition that the agent's preferences for effort are independent of the innate cost (i.e., $\psi$ does not depend on $\beta$). While this assumption is common in the procurement literature, its applicability depends on the circumstances at hand. For instance, independence describes well a scenario where the agent's private information on $\beta$ relates to the cost of obtaining a fixed input to production, where the quantity of this input does not depend on the amount of effort exerted.[7]

[5] In particular, conditional on the agent reporting truthfully, an optimal mechanism never asks the agent to generate a production cost strictly above his innate cost.

[6] See Garrett and Pavan (2012), who also permit such cost inflation to facilitate application of an envelope theorem in an LT-type model.

[7] Note that our analysis is still informative about the set of expected payoffs for broader classes of preferences, since the payoff set for these broader preferences must nest the set that we characterize below (for

The agent's innate cost $\beta$ is drawn from a cumulative distribution function (cdf) $F$ that is twice continuously differentiable and has density $f$. We take $F$ to have full support on a bounded interval $[\underline{\beta}, \bar{\beta}]$, where it seems natural to require $\underline{\beta} > 0$. Finally, we assume throughout that $F(\beta)/f(\beta)$ is strictly increasing (equivalently, $F$ is strictly log concave) and Lipschitz continuous, denoting its first derivative by $h(\beta)$.

Since our view is that the analyst knows the distribution of innate costs $F$, the above assumptions can at least be verified on a case-by-case basis. Log concavity of $F$ is a common restriction in the literature.

The timing of the game is then the same as in LT. First, the agent learns his private type $\beta$, which is drawn from $F$. Then the principal offers a mechanism, which prescribes payments to the agent as a function of any messages sent by the agent and the realized cost, which is observable and contractible. Next, the agent determines whether to accept the mechanism. If he does not, the agent earns payoff zero. If he does accept, then he sends a message to the principal and then makes his effort choice. The production cost is realized and the principal makes a payment to the agent as prescribed by the mechanism.

Without loss of generality, we can consider incentive-compatible and individually rational direct mechanisms. The agent makes a report of his type $\hat{\beta}$ to the mechanism. The mechanism then prescribes a "production cost target" $C(\hat{\beta})$. If the agent reports his innate cost $\beta$ truthfully, then meeting the cost target requires effort $e(\beta) = \beta - C(\beta)$, which can therefore be understood as the effort recommendation of the mechanism for type $\beta$. If the agent achieves the target, i.e., $C = C(\hat{\beta})$, then he is paid $y(\hat{\beta})$. Otherwise, if $C \neq C(\hat{\beta})$, the payment to the agent is negative. Since the mechanism is individually rational, a choice to report $\widehat{\beta}$ and select cost $C \neq C(\hat{\beta})$ is never optimal for the agent. This observation is enough to transform the principal's problem from one of both moral hazard and adverse selection into one of only adverse selection.

### *Objective of the analysis*

The aim of our analysis is to understand the payoff implications of introducing incentive contracts. As discussed in the Introduction, we consider an analyst who understands that the cost-based procurement model above is the correct description of the environment and who has a reliable prior belief $F$ regarding the innate cost $\beta$. However, she does not know the agent's preferences for effort, only that they are described by a function in $\Psi$. She *does know* that the principal, who eventually designs and implements an incentive contract to minimize the expected total payment to the agent, has the same distribution $F$ in mind for the innate cost, knows the disutility function $\psi$ precisely, and chooses mechanisms optimally. We ask, what expected payoff implications does the analyst consider possible?

instance, when the principal is guaranteed only a small fraction of the expected surplus under the imposed restrictions on preferences, the guarantee can only be smaller for more admissive restrictions). For a more precise characterization, one would need to adapt the steps in our analysis for the broader preferences (which may be more or less tractable depending on the restrictions in question).

## 3. Preliminaries

### *Analysis of the principal's contracting problem*

We begin by extending analysis familiar from LT to the present environment. The main point of difference is that we are more permissive in the restrictions on $\psi$; for instance, we do not require $\psi$ to be differentiable. Fix the mechanism offered by the principal (as described above). Note that if the agent makes a report $\hat{\beta}$, then the mechanism prescribes a production cost target $C(\hat{\beta})$. By the observation regarding individual rationality of the mechanism in the previous section, we may presume the agent meets the cost target. Then the agent's payoff, if his true innate cost is $\beta$, is

$$y(\hat{\beta}) - \psi\big(\beta - C(\hat{\beta})\big).$$

Let $\partial_- \psi$ denote the left derivative of $\psi$. We argue (see Appendix A.1) that we can consider mechanisms where the agent's rents are given, as a function of his true innate cost $\beta$, by

$$\int_{\beta}^{\bar{\beta}} [\partial_- \psi]\big(e(x)\big)\, dx. \tag{1}$$

This follows from incentive compatibility of the mechanism, after applying the envelope result of Carbajal and Ely (2013) for nondifferentiable objective functions, and from considering mechanisms that maximize the principal's expected payoff for a given effort policy $e(\cdot)$.[8] Taking expectations and integrating by parts, we can express the agent's expected rents as

$$\mathbb{E}\left[\frac{F(\tilde{\beta})}{f(\tilde{\beta})}[\partial_- \psi]\big(e(\tilde{\beta})\big)\right]. \tag{2}$$

The principal's expected total payment in a mechanism that optimally implements an effort policy $e(\cdot)$ is

$$\mathbb{E}\big[\tilde{\beta} - \mathrm{VG}\big(e(\tilde{\beta}), \tilde{\beta}\big)\big], \tag{3}$$

where

$$\mathrm{VG}(e, \beta) = e - \psi(e) - \frac{F(\beta)}{f(\beta)}[\partial_- \psi](e) \tag{4}$$

(we leave the dependence of VG on $\psi$ and $F$ implicit). Here, $\mathrm{VG}(e, \beta)$ is the virtual gain from incentives that induce effort $e$ for innate cost $\beta$, comprising efficiency gains $e - \psi(e)$ from effort less a term accounting for agent rents.

Considering minimization of (3) by choice of the effort policy, we have the following result.

[8]Note that one might be tempted to believe that precisely the same analysis usually performed when $\psi$ is differentiable should carry through, given that a convex disutility function $\psi$ is differentiable except at countably many points. The difficulty, however, is that effort is endogenous, since it is chosen by the principal and, hence, may be chosen at kinks in the disutility with positive probability (in spite of the continuous distribution of innate costs). As Carbajal and Ely point out, this necessitates alternative arguments.

PROPOSITION 3.1. *Any effort policy $e^*(\cdot)$ for an optimal mechanism solves, for almost all innate costs $\beta$,*

$$W(\beta) = \max_e \mathrm{VG}(e, \beta).$$

*Optimal effort policies $e^*(\cdot)$ are essentially unique and nonincreasing. Also, $[\partial_- \psi](e^*(\beta)) < 1$ for almost all $\beta$.*[9]

The result shows that there is an optimal effort policy that maximizes virtual gains from incentives pointwise; also, the optimal policy is essentially unique (in what follows, we restrict attention to versions of the optimal policy $e^*(\beta)$ that maximize virtual gains at all values of $\beta$, not merely almost all). In other words, the validity of the "first-order" or "relaxed program" approach to solving the design problem is established. While such a result is readily anticipated from earlier work (including LT), it is obtained under weaker conditions than usually assumed. Because the first-order approach is valid, no additional restrictions on the shape of $\psi$ are needed to justify restriction to deterministic effort policies (see Strausz (2006) for this observation in a related model).

The properties obtained for optimal effort $e^*(\cdot)$ follow from examining the virtual gains $\mathrm{VG}(e, \beta)$. Effort is weakly downward distorted (note that we may have $[\partial_- \psi](e^{FB}) < 1$ at an efficient effort level $e^{FB}$ if there is a kink in $\psi$ at $e^{FB}$; hence, unlike the case for differentiable disutility functions, an optimal mechanism may specify efficient effort for a positive measure of innate costs). Downward distortions in effort are due to the familiar reason that they reduce the rents the agent can expect in an incentive-compatible and individually rational mechanism. Distortions are larger for higher values of $\beta$, which can be understood in part from examining the expression for agent rents in (1): in particular, the agent's rents for a given innate cost depends on the effort induced from all higher innate costs. It is worth emphasizing here that monotonicity of effort therefore comes from optimization of the virtual gains in (4) given log concavity of $F$. The condition for an effort policy to be implementable in an incentive-compatible mechanism is weaker, since all that is required is that $C(\beta) = \beta - e(\beta)$ is nondecreasing with $\beta$ (see the discussion immediately preceding Lemma A.2 in the Appendix).

### *Quadratic disutility*

To shed further light on the properties of an optimal mechanism, it is useful to consider the case of quadratic disutility, which often receives attention in the literature. To be precise, given the restrictions in the model setup, consider functions that satisfy $\psi(e) = ke^2/2$ on an interval $[0, \bar{e}]$, with $\bar{e} > 1/k$ and $k > 0$ (such functions can be chosen in $\Psi$ for any $k > 0$). Optimal effort then satisfies, for all $\beta$,

$$e^*(\beta) = \max\{0, 1/k - F(\beta)/f(\beta)\}.$$

It is then easy to see that the expected rents in (2) vary continuously with $k$. As $k$ grows large, optimal effort equals 0 with a probability approaching 1. Hence, expected rents

[9] Note that the analysis here assumes the agent resolves indifferences over reports by reporting truthfully. This is the usual approach in the literature.

shrink to 0 as $k \to \infty$. Conversely, as $k$ is taken to 0, optimal effort is large and positive uniformly across types $\beta$ (since $F/f$ is bounded), and the agent's marginal disutility of effort $ke^*(\beta)$ is uniformly close to 1. Expected rents are then close to

$$\bar{R} \equiv \int_{\underline{\beta}}^{\bar{\beta}} F(\beta)\, d\beta, \tag{5}$$

which by Proposition 3.1 is an upper bound on rents that holds across all disutility functions in $\Psi$, given the distribution $F$. We can conclude that expected rents can take any value in $(0, \bar{R})$ for some value of $k$.

### *Defining the analyst's problem*

We now define the objects of interest for the analyst: the principal's expected gains from incentives and agent expected rents under an optimal mechanism. Given a cdf $F$ for innate costs, for any $\psi \in \Psi$, the principal implements an optimal mechanism with essentially unique effort $e^*(\cdot)$. Agent expected rents in the optimal mechanism are given by (2), evaluated at the optimal effort policy. We denote these expected rents by $R(\psi; F)$. Alternatively, the principal's expected gains from incentives are

$$G(\psi; F) = \mathbb{E}\big[W(\tilde{\beta})\big].$$

Our interest is in characterizing, for each $F$, the set

$$\mathcal{U} \equiv \big\{\big(R(\psi; F), G(\psi; F)\big) \in \mathbb{R}^2_+ : \psi \in \Psi\big\}.$$

## 4. Analysis

### *Further preliminary observations on the analyst's problem*

Recalling Proposition 3.1 and the discussion in the previous section, the set of possible agent rents is $[0, \bar{R})$ with $\bar{R}$ given by (5). To see this, recall that expected rents in $(0, \bar{R})$ are obtained by the quadratic disutility functions considered in the previous section. Expected rents equal to $\bar{R}$ cannot occur, due to the final claim in Proposition 3.1 and the expression for expected rents in (2). Expected rents equal to 0 occur if and only if optimal effort is constant (almost surely) at 0. The reason is that $\partial_- \psi$ is strictly positive at positive effort values. Therefore, the case with zero rents occurs only when the right derivative of $\psi$ is above 1 at 0 (i.e., $[\partial_+ \psi](0) \geq 1$). In such cases, the principal's expected gains from incentives are 0.

Given these observations, our interest is to determine the expected gains from incentives when the expected agent rents $R$ are in $(0, \bar{R})$. We characterize the function

$$G^{\inf}(R) \equiv \inf_{\psi \in \Psi} \big\{G(\psi; F) : \psi \in \Psi, R(\psi; F) = R\big\}$$

on $(0, \bar{R})$. This function determines the lower boundary of the set $\mathcal{U}$. We show by Propositions 4.1 and 4.2 below that it is strictly increasing and weakly convex (properties that were emphasized in the Introduction).

Finally, note that while $G^{\inf}(R)$ defines the infimum of expected gains from incentives for each level of agent expected rents $R \in (0, \bar{R})$, arbitrarily higher gains from incentives can occur depending on the disutility function. We formalize this in Corollary 4.1 below. The argument is based on the following idea. For a disutility function $\psi \in \Psi$ associated with a point close to the boundary of $\mathcal{U}$, we can consider another disutility function of the form

$$\bar{\psi}(e; a, \varepsilon) = \begin{cases} 0 & \text{if } e \leq 0, \\ \varepsilon e & \text{if } e \in (0, a], \\ \varepsilon a + \psi(e - a) & \text{if } e > a \end{cases} \tag{6}$$

for $\varepsilon, a > 0$. These parameters can be chosen so that expected gains from incentives under an optimal mechanism take values above $G(\psi; F)$, while expected rents are close to $R(\psi; F)$.[10] The idea behind considering disutility functions of this form is that the agent is permitted to achieve cost reduction $a$ almost for free when $\varepsilon$ is small, implying an increase in the surplus that can be generated from incentives. For such a disutility function with small enough $\varepsilon$, optimal effort is at least $a$ for all innate costs. Also, for an innate cost $\beta$ assigned effort $e^*(\beta) > 0$ in an optimal mechanism for $\psi$, optimal effort can be set to $e^*(\beta) + a$ in a mechanism that is optimal for $\bar{\psi}(\cdot; a, \varepsilon)$. It follows that expected surplus increases by at least $a(1 - \varepsilon)$ in a mechanism optimal for $\bar{\psi}(\cdot; a, \varepsilon)$, while any additional expected rents vanish as $\varepsilon \to 0$. Thus, once we have determined disutility functions associated with points at or arbitrarily close to the boundary of $\mathcal{U}$, it is possible to modify these functions to attain points with higher expected gains from incentives.

### *Main arguments*

A key step in determining $G^{\inf}(R)$ (given the innate cost distribution $F$) is to recognize that the virtual gains from incentives can be represented by an envelope formula. Given $F$ and $\psi$, the virtual gains are $W(\beta) = \max_e \mathrm{VG}(e, \beta)$ (where recall VG is defined in (4)). Because $\psi$ is Lipschitz, and because $F/f$ is differentiable and Lipschitz, the conditions for the envelope theorem of Milgrom and Segal (2002) are satisfied. We can conclude that

$$W(\beta) = W(\bar{\beta}) + \int_{\beta}^{\bar{\beta}} h(s)[\partial_- \psi]\big(e^*(s)\big)\, ds,$$

where recall $h(\beta) = \frac{d}{d\beta}[\frac{F(\beta)}{f(\beta)}]$. Note that $W(\bar{\beta})$ is nonnegative and may be strictly positive depending on the disutility function. Also, $W(\cdot)$ is nonincreasing. This can be understood by observing that the term that accounts for rents in (4), i.e., $-(F(\beta)/f(\beta))[\partial_- \psi](e)$, is nonincreasing in $\beta$ for any effort $e$ (as $F/f$ is strictly increasing). Put simply, the virtual gains are larger for lower innate costs because the expected rent the principal must give to the agent as a result of raising the efforts for these innate

[10] For the disutility functions $\psi$ that we show are close to the boundary of $\mathcal{U}$, the modified disutility $\bar{\psi}(\cdot; a, \varepsilon)$ remains convex and, hence, in $\Psi$ provided $\varepsilon$ is small enough.

costs is smaller (recall that, by (1), the rents earned for an agent with innate cost $\beta$ are determined by the effort asked for all higher innate costs).

We can now find a convenient expression for the expected gains from incentives for the principal. We have

$$\begin{aligned} G(\psi; F) &= \mathbb{E}\big[W(\tilde{\beta})\big] \\ &= W(\bar{\beta}) + \mathbb{E}\left[\frac{F(\tilde{\beta})}{f(\tilde{\beta})} h(\tilde{\beta})[\partial_- \psi]\big(e^*(\tilde{\beta})\big)\right], \end{aligned} \tag{7}$$

where the second equality follows from integration by parts. One way to think about the second term in (7) is to note that a reduction in $\beta$ increases the term in (4) that accounts for agent rents $(-(F(\beta)/f(\beta))[\partial_- \psi](e))$. The marginal effect, given an optimal effort policy, is $h(\beta)[\partial_- \psi](e^*(\beta))$. This effect can be viewed as cumulative; the marginal effect accrues to all lower innate costs, which have probability $F(\beta)$, and this is weighted in the expectation by the density $f(\beta)$ for type $\beta$.

The rest of our argument can be understood in relation to the two main steps outlined in the Introduction. Completing the first step involves determining a lower bound on the principal's gains from incentives given agent expected rents $R$. To do so, we set the principal's virtual gains from incentives for type $\bar{\beta}$ (that is, $W(\bar{\beta})$) equal to its minimum possible value, 0. We then consider minimizing the second term of (7) by choice of the marginal disutility of effort function $[\partial_- \psi](e^*(\cdot))$, subject to the agent's expected rents in (2) being equal to $R$. We know from Proposition 3.1 that the marginal disutility of effort can be assumed to be non-increasing, so we impose this monotonicity constraint in the optimization. The constrained minimization problem delivers a lower bound on the principal's expected gains $G(\psi; F)$ over disutility functions $\psi \in \Psi$, given that agent expected rents $R(\psi; F)$ must be equal to $R$. Our second step then involves showing the bound is tight. This requires demonstrating the existence of a disutility function in $\Psi$ such that the principal's expected gains from incentives coincide with the bound or at least can be taken arbitrarily close. This involves reverse engineering admissible disutilities $\psi$ such that the marginal disutility of effort function $[\partial_- \psi](e^*(\cdot))$ is equal, or arbitrarily close to, the one determined in the minimization problem of the first step; that is, it involves showing that the solution to the minimization problem or some nearby function is actually the marginal disutility of effort for the agent in the principal's optimal mechanism for some disutility function $\psi$.

We now state the minimization problem of the first step, where we determine our lower bound $L^*(R)$ as a function of agent rents $R$.

Problem I. *Let $\Gamma$ be the set of functions $\gamma : [\underline{\beta}, \bar{\beta}] \to [0, 1]$ such that $\gamma$ is nonincreasing. For any $R \in (0, \bar{R})$, determine*

$$L^*(R) = \min_{\{\gamma \in \Gamma : \int_{\underline{\beta}}^{\bar{\beta}} F(\beta)\gamma(\beta)\, d\beta = R\}} \int_{\underline{\beta}}^{\bar{\beta}} F(\beta) h(\beta) \gamma(\beta)\, d\beta. \tag{8}$$

One way to understand Problem I is to consider the linear functional

$$P(\gamma) \equiv \left( \int_{\underline{\beta}}^{\bar{\beta}} F(\beta)\gamma(\beta)\, d\beta, \int_{\underline{\beta}}^{\bar{\beta}} F(\beta)h(\beta)\gamma(\beta)\, d\beta \right).$$

For an interpretation of $P$, suppose there is a disutility function such that $\gamma(\beta)$ is the agent's marginal disutility of effort for each type $\beta$ in an optimal mechanism. Suppose also that type $\bar{\beta}$ generates zero virtual gains for the principal (i.e., that $W(\bar{\beta}) = 0$). Then the first component of $P(\gamma)$ is the agent's expected rent, while the second component is the principal's expected gains from optimal incentives. Then function $L^*(\cdot)$ can be obtained from the lower boundary of the set

$$\{P(\gamma) : \gamma \in \Gamma\}.$$

As a basis for functions in $\Gamma$, we consider step functions

$$\gamma_x(\beta) = \begin{cases} 1 & \text{if } \beta \in [\underline{\beta}, x), \\ 0 & \text{if } \beta \in [x, \bar{\beta}], \end{cases}$$

where $x \in [\underline{\beta}, \bar{\beta}]$. We show (see Step 1 in the proof of Proposition 4.1 in the Appendix) that $\{P(\gamma) : \gamma \in \Gamma\}$ is equal to the convex hull of $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$.[11] Because $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$ is a closed curve (see further explanation following Proposition 4.1), the convex hull is also closed.

A pair $(R, L^*(R))$ for $R \in (0, \bar{R})$ is a point on the lower boundary of this convex hull. It is then immediate that $L^*(\cdot)$ is strictly increasing (since $h$ is strictly positive) and weakly convex. In addition (by an application of Carathéodory's theorem), any point $(R, L^*(R))$ for $R \in (0, \bar{R})$ is a convex combination of points $P(\gamma_x)$ for at most two values of $x$. Hence (by linearity of $P$), there is a solution to Problem I that can be written as a convex combination of step functions $\gamma_x$ for two values of $x$. To summarize, we have the following result.

Proposition 4.1. *For any $R \in (0, \bar{R})$, a solution $\gamma^* : [\underline{\beta}, \bar{\beta}] \to [0, 1]$ to the minimization in Problem I exists. The minimum function $L^*(\cdot)$ is strictly increasing and weakly convex. For any $R \in (0, \bar{R})$, there is a solution described by two cutoffs $\beta_l$ and $\beta_u$, with $\underline{\beta} \le \beta_l \le \beta_u \le \bar{\beta}$. In particular, $\gamma^*(\beta) = 1$ on $[\underline{\beta}, \beta_l)$, $\gamma^*(\beta)$ is constant and strictly between 0 and 1 on $[\beta_l, \beta_u)$, and $\gamma^*(\beta) = 0$ on $[\beta_u, \bar{\beta}]$.*

To understand better Proposition 4.1, consider the curve $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$, which is parametrically defined through the thresholds $x$ of the step functions $\gamma_x$ and defined in the space of players' welfare. To define this curve explicitly as a function of the agent's rents $R$, let $x(R)$ be the threshold for the step function associated with rent $R$. This threshold is implicitly defined by

$$R = \int_{\underline{\beta}}^{x(R)} F(s)\, ds.$$

[11] The convex hull of the set $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$ is the smallest convex set that contains it.

Then let

$$p(R) = \int_{\underline{\beta}}^{x(R)} F(s)h(s)\,ds$$

so that

$$\big\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\big\} = \big\{\big(R,\, p(R)\big) : R \in [0, \bar{R}]\big\}.$$

To give an interpretation to $p(R)$, it is the expected gains for the principal when the agent has expected rent $R$, when the agent's marginal disutility of effort is given by a step function and, hence, by $\gamma_{x(R)}$, and when virtual gains for the highest type are equal to zero (i.e., $W(\bar{\beta}) = 0$). Using the implicit function theorem, $x'(R) = 1/F(x(R))$ for all $R$. Therefore, $p'(R) = h(x(R))$. Since $x(\cdot)$ is an increasing function, $p(\cdot)$ is convex when $h$ is increasing (i.e., when $F/f$ is strictly convex) and concave when $h$ is decreasing (i.e., when $F/f$ is strictly concave).[12]

Now consider the value of the minimization problem. If $h$ is increasing, then $L^*(R) = p(R)$ for all $R \in (0, \bar{R})$, since the lower boundary of the aforementioned convex hull is given by the curve $p(\cdot)$ itself. That is, the fact $L^*(R) = p(R)$ follows because, when $F/f$ is convex, the solution to Problem I is a step function given by $\gamma_{x(R)}$ for each value of expected rents $R$. An example of this case is displayed on the left side of Figure 1. The thresholds in Proposition 4.1 are then $\beta_l = \beta_u = x(R)$. If $h$ is decreasing, then points on the lower boundary of the convex hull are convex combinations of $(0, 0)$ and $(\bar{R},\, p(\bar{R}))$. An example of this case is displayed on the right side of Figure 1. A solution to Problem I is then $\gamma^* = (1 - R/\bar{R})\gamma_{\underline{\beta}} + (R/\bar{R})\gamma_{\bar{\beta}}$, which is constant at $R/\bar{R}$ on $(0, \bar{R})$. Note that $L^*(R) = (R/\bar{R})\,p(\bar{R})$. The thresholds in the proposition are $\beta_l = \underline{\beta}$ and $\beta_u = \bar{\beta}$. Finally, note that cases where $F/f$ is neither convex nor concave can also be handled by considering points on the lower boundary of the convex hull of $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$, as explained above.

We now show that the lower bound on gains from incentives given by $L^*$ is tight and, hence, coincides with the function $G^{\inf}$.

Proposition 4.2. *Fix a distribution $F$ and fix any $R \in (0, \bar{R})$. For any $\varepsilon > 0$, there exists $\psi \in \Psi$ such that*

$$R(\psi; F) = R$$

*and*

$$G(\psi; F) < L^*(R) + \varepsilon.$$

*Hence, $G^{\inf}(R) = L^*(R)$.*

[12] If $F$ is thrice differentiable, we have that $F/f$ strictly convex over $[\underline{\beta}, \bar{\beta}]$ if, for all $\beta$,

$$f'(\beta) < \frac{F(\beta)}{f(\beta)^2}\big(2f'(\beta)^2 - f''(\beta)f(\beta)\big),$$

while $F/f$ is strictly concave when the reverse inequality holds. Mierendorff (2016) discusses the convexity/concavity of $(1 - F)/f$ and gives an analogous condition.

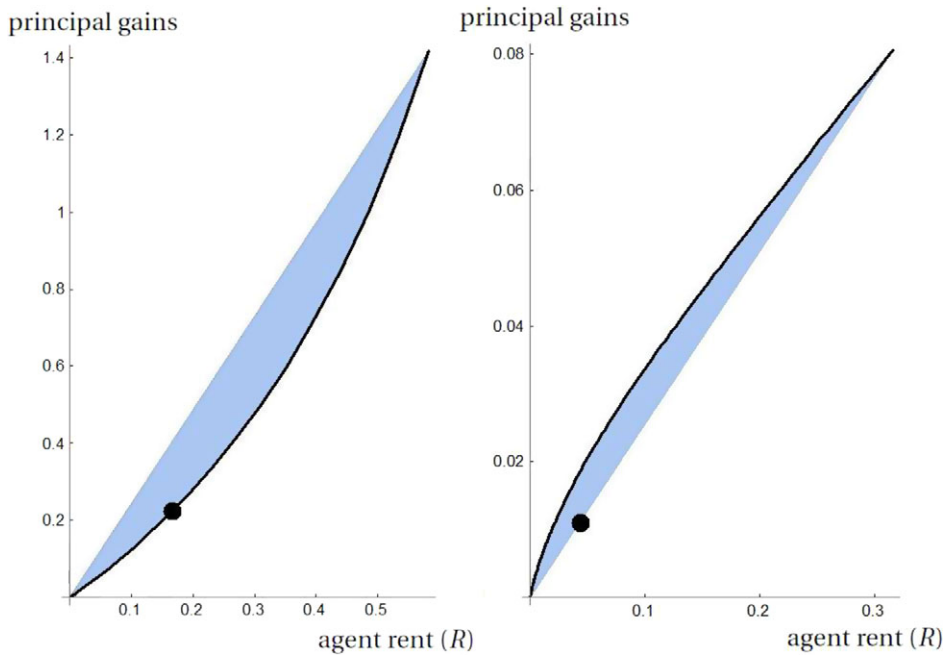


FIGURE 1. Examples with $F/f$ convex and concave. Left side: $\beta$ distributed with density $f(\beta) = 5/2 - \beta$ on $[1, 2]$ so that $F/f$ is convex. Right side: $\beta$ from a shifted and truncated standard normal distribution so that $F/f$ is concave. Truncation is to $[-3, -2]$ for the standard normal, and subsequent shifting yields a support $[1, 2]$. On each side, the black curve is $p(\cdot)$ and the shaded area is its convex hull. The black circles lie on the lower boundary of the convex hulls $((R, p(R)) = (0.167, 0.224) = P(\gamma_{1.5})$ for the left side and $(R, L^*(R)) = (0.044, 0.011)$ for the right side, where note $R = \int_1^{1.5} F(s)\,ds)$.

The proof of Proposition 4.2 involves finding disutility functions $\psi \in \Psi$ such that the left derivative of disutility at optimal effort levels, i.e., $[\partial_- \psi](e^*(\cdot))$, approaches a fixed solution $\gamma^*$ to Problem I (as mentioned in the outline of our approach above). To illustrate the nature of the argument, we consider in the main text the apparently simplest case where the thresholds in Proposition 4.1 satisfy $\underline{\beta} = \beta_l < \beta_u \leq \bar{\beta}$, which occurs, for instance, if $F/f$ is concave (in this case, recall that $\beta_u = \bar{\beta}$). The corresponding solution to Problem I, $\gamma^*$, is constant at $R/\int_{\underline{\beta}}^{\beta_u} F(s)\,ds \in (0, 1)$ on the interval $[\underline{\beta}, \beta_u)$.

We aim to find a disutility function $\psi \in \Psi$ such that, at an optimal effort policy $e^*(\cdot)$, (a) the left derivative of disutility of effort $[\partial_- \psi](e^*(\beta))$ is constant and equal to $R/\int_{\underline{\beta}}^{\beta_u} F(s)\,ds$ for innate costs $\beta$ below $\beta_u$, and is zero (with zero effort exerted) for higher innate costs, and (b) virtual gains from incentives $\mathrm{VG}(e^*(\beta), \beta)$ are equal to zero for $\beta = \beta_u$. For such a disutility function, the agent must obtain expected rents $R$, and the principal's expected gains from incentives must equal $L^*(R)$.[13]

[13]Conditions (a) and (b) are not only sufficient for this to be true, but are also necessary provided that the solution $\gamma^*$ to Problem I is essentially unique (e.g., if $F/f$ is strictly concave). This can be seen from (7) and (8) above.

To determine an appropriate disutility function, let

$$b = \frac{F(\beta_u)R}{f(\beta_u)\left(\int_{\underline{\beta}}^{\beta_u} F(s)\,ds - R\right)}$$

and $k > 1$, and put

$$\psi(e) = \begin{cases} 0 & \text{if } e \le 0, \\ \dfrac{R}{\int_{\underline{\beta}}^{\beta_u} F(s)\,ds}\, e & \text{if } 0 < e \le b, \\ \dfrac{Rb}{\int_{\underline{\beta}}^{\beta_u} F(s)\,ds} + k(e-b) & \text{if } e > b. \end{cases}$$

Then an optimal policy for the principal is to specify $e^*(\beta) = b$ for $\beta \in [\underline{\beta}, \beta_u]$ and $e^*(\beta) = 0$ for $\beta$ above $\beta_u$, if any. This shows that the infimum of expected gains from incentives (conditional on expected rents $R$) is attained.

The above derivation may be of interest because it allows us to understand the kind of disutility function for which the lower bound is attained. For instance, when $F/f$ is strictly concave, the solution $\gamma^*$ to Problem I is essentially unique. Then points on the lower boundary of the expected payoffs are obtained *only* by disutility functions of the above form; specifically, with disutility of effort linear from zero up to some value $b$, as determined above. The form of disutility functions that attain or approach the boundary may be of interest if some functions are considered more plausible than others. In this context, it is worth emphasizing that the bound obtained in Proposition 4.1 continues to apply if we admit disutility functions only from some strict subset of $\Psi$ (naturally, in this case, the bound may no longer be tight).

It may also be of interest to observe that the optimal mechanism corresponding to the above disutility function is arguably rather simple. This mechanism has an indirect implementation where the agent simply produces at cost $C$ (without making any report of type) and receives payment

$$y = \max\left\{0, (\beta_u - C)\frac{R}{\int_{\underline{\beta}}^{\beta_u} F(s)\,ds}\right\}$$

in addition to reimbursement of the production cost $C$. This can be viewed as a menu comprising a cost-reimbursement contract (i.e., $y = 0$ for all production costs) and a "linear cost-sharing rule" where the agent gets $R/\int_{\underline{\beta}}^{\beta_u} F(s)\,ds$ per unit of cost savings. See Chu and Sappington (2007) for an analysis of the performance of such menus more generally. Because the agent's marginal disutility of effort is $R/\int_{\underline{\beta}}^{\beta_u} F(s)\,ds$ up to effort $b$, the agent is indifferent across efforts in $[0, b]$ under the linear cost-sharing rule. The

agent puts no effort and earns payoff zero under the cost-reimbursement contract. It is then immediate that the agent prefers the linear cost-sharing rule if and only if $\beta \leq \beta_u$. Since in this case the agent is willing to choose effort equal to $b$, the effort policy $e^*(\beta)$ described above is optimal for the agent, and this mechanism is optimal for the principal, assuming the agent chooses effort according to $e^*(\beta)$.

Let us conclude this section by considering expected gains from incentives above the lower boundary $G^{\inf}$. The following corollary can be established using disutility functions of the form introduced in (6).

COROLLARY 4.1. *For any $R \in (0, \bar{R})$, any $G > G^{\inf}(R)$, and any $\varepsilon > 0$, there exists $\psi \in \Psi$ such that $|G(\psi; F) - G| < \varepsilon$ and $|R(\psi; F) - R| < \varepsilon$.*

As discussed above, the result is useful as it indicates that the characterization of possible values of expected welfare reduces to a characterization of the lower bound $G^{\inf}$.

## 5. PROPERTIES OF THE PAYOFF REGION

We now consider how the principal's guaranteed gains from incentives depend on the shape of the innate cost distribution. First note that when $F$ is any uniform distribution, $h$ is constant and equal to 1 (since $F(\beta)/f(\beta) = \beta - \underline{\beta}$), and so $G^{\inf}(R) = R$ for all $R \in (0, \bar{R})$. In other words, when the expected surplus from incentive contracting is not too large (precisely, when it is below $2\bar{R}$), the smallest share of this surplus that the principal may earn is one half. This observation itself could be of interest for applications, as several papers have drawn conclusions based on uniformly distributed innate costs (see, for instance, Gasmi et al. (1997) and Rogerson (2003)). Building on the observation for uniform distributions, we show the following corollary.

COROLLARY 5.1. *Fix a distribution $F$. Then the following statements hold:*[14]

*(i) If $F/f$ is concave and $\mathbb{E}[\tilde{\beta}] \geq (\underline{\beta} + \bar{\beta})/2$, then $G^{\inf}(R) \leq R$ for all $R \in (0, \bar{R})$; the inequality is strict if either concavity is strict or if $\mathbb{E}[\tilde{\beta}] > (\underline{\beta} + \bar{\beta})/2$.*

*(ii) If $F/f$ is convex and if $\mathbb{E}[\tilde{\beta}|\tilde{\beta} \leq \beta] \leq (\underline{\beta} + \beta)/2$ for all $\beta \in (\underline{\beta}, \bar{\beta}]$, then $G^{\inf}(R) \geq R$ for all $R \in (0, \bar{R})$; the inequality is strict if either convexity is strict or if $\mathbb{E}[\tilde{\beta}|\tilde{\beta} \leq \beta] < (\underline{\beta} + \beta)/2$ for all $\beta \in (\underline{\beta}, \bar{\beta}]$.*

Consider first part (i). Provided $F/f$ is concave, the mean of the innate costs being above the midpoint $(\underline{\beta} + \bar{\beta})/2$ is sufficient to conclude $G^{\inf}(R) \leq R$. The condition is a sense in which the distribution is negatively skewed. The reason for the result is related to the observation that when innate costs are concentrated at higher values, the principal's optimal policy, for a fixed disutility function, calls for relatively small distortions for high innate costs. In particular, the principal's policy calls for positive effort,

[14]Note that both cases can occur. Case (i) applies to the distribution considered for the right side of Figure 1, while case (ii) applies to the distribution considered for the left side of Figure 1.

even when the surplus generated from this effort is relatively small. In turn, this permits the agent to earn high expected rents even for disutility functions that permit only relatively small increases in surplus through cost-reducing effort. That the agent obtains high rents when the principal specifies positive effort at high innate costs follows from considering the expression for rents in (1).

Intuition for part (ii) is then the reverse. When innate costs are more concentrated at lower values, the optimal effort policy tends to be much more distorted at higher values of the innate cost. Examining the equation for agent rents in (1), close to efficient effort can be asked of types close to $\underline{\beta}$, inducing marginal disutility of effort close to 1, without granting too large rents to the agent (even for the lowest type $\underline{\beta}$). When innate costs are concentrated at lower values, even if the agent chooses close to efficient effort with high probability, it can be guaranteed that the agent's expected rents are not too large.

Another question related to the above discussion is whether any predictions on the magnitude of the bound $G^{\inf}(R)$ can be made without any restrictions on the cost distributions $F$. The answer is negative as the following example attests.

EXAMPLE 1. Consider innate cost distributions with cdf $F(\beta) = (k(\beta - \underline{\beta}))^{1/k}/(k(\bar{\beta} - \underline{\beta}))^{1/k}$ for $k > 0$. The distribution $F$ satisfies all our conditions and $F(\beta)/f(\beta) = k(\beta - \underline{\beta})$, so that $h(\beta) = k$. Therefore, $G^{\inf}(R) = kR$ for $R \in (0, \bar{R})$; this can be taken arbitrarily large or small with $k$.

The intuition for Example 1 is much the same as the one provided above in relation to Corollary 5.1. When $k$ is small, the cdf $F$ is convex and the distribution is concentrated on high values of the innate cost. The principal's optimal policy then asks high effort for high values of the innate cost, even if the surplus generated through effort is small. Conversely, when $k$ is large, the cdf $F$ is concave and the distribution is concentrated on low values of the innate cost, so the reverse is true: the principal is unwilling to ask high effort for high values of the innate cost unless the surplus generated through effort is large.

As mentioned in the Introduction, empirical work on procurement often finds a skewed distribution for firm costs, with many firms having similar cost performance but with a long right tail of less efficient firms. Such an observation was made by Wolak (1994) and Brocas et al. (2006) for regulated water utilities, and in the context of Laffont and Tirole's model by Gagnepain and Ivaldi (2002) for urban transport and by Abito (2015) for electric utilities. Figure 2 presents a graphical analysis for a distribution estimated in the study of Gagnepain and Ivaldi. Specifically, they estimate the log of innate cost to be distributed (up to scaling) according to a Beta distribution, with parameters 0.59 and 2.15 (see their Table 3).[15] The density, plotted on the left side of Figure 2, is sharply downward sloping, capturing the concentration of firms at a baseline level of the innate cost. The ratio of worst-case gains to rents $G^{\inf}(R)/R$ is plotted on the right side of the figure. This is a setting where the principal, under an optimal contract, can

[15]It should be kept in mind that these parameter estimates are for a somewhat different underlying model than the one we treat here, in particular because Gagnepain and Ivaldi's model incorporates elastic quantity choices by the firm.

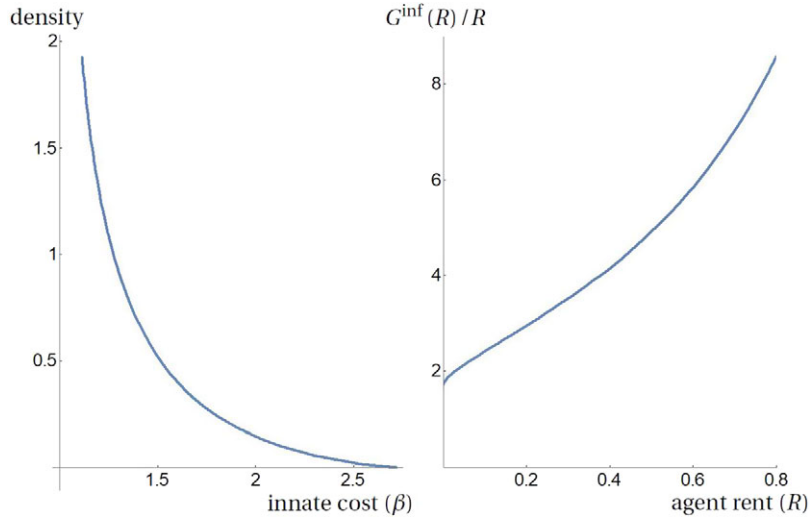


FIGURE 2. Example based on Gagnepain and Ivaldi (2002, Table 3). Left side: Density of the variable whose natural logarithm is distributed according to a Beta distribution with parameters 0.59 and 2.15. Right side: Corresponding ratio of infimum expected gains to expected agent rents (values of rent $R$ shown below 0.8).

be sure to extract a large fraction of the surplus generated through incentives. The ratio $G^{\inf}(R)/R$ is below 2 for small values of agent rents $R$, but grows with the value of $R$.

## 6. CONCLUSIONS

This paper considered the problem of an analyst tasked with predicting equilibrium outcomes of a principal–agent relationship, while possessing limited information about the environment. In particular, we assumed that while the analyst has good grounds for determining the distribution of (cost) performance absent incentives, she is ignorant of the feasible agent technologies or preferences for responding to incentives. Given this lack of information, we made only weak assumptions on agent preferences: monotonicity and convexity of the disutility of effort as well as separability from the innate cost. We then showed how to obtain sharp predictions on the set of expected payoffs that can arise in equilibrium.

The analysis is informative regarding the relationship between agent and principal rents in well designed incentive contracts under restrictions on the environment that can be guided by theory (rather than resulting from, say, ad hoc functional form assumptions on the technology or agent preferences). The findings could perhaps be helpful in further clarifying and refining a message on which economists seem to agree: in many agency relationships, the presence of asymmetric information implies agent rents are in expectation strictly positive, and sometimes sizable, even if incentive contracts are well designed. Large agent rents need not be indicative of incentive contracts performing poorly: we uncovered a tight positive relationship between the expected payoff of the agent and the expected gains to the principal in optimal incentive contracts.

In addition, this paper has developed a novel approach to determining the relationship between principal and agent rents, which seems likely to be useful in other settings. Most immediately, the Laffont–Tirole model has been applied in settings of executive compensation in Edmans and Gabaix (2011), Edmans et al. (2012), Garrett and Pavan (2012, 2015), and Carroll and Meng (2016). Given the proximity of these models to the procurement model that we studied here, our results can be mapped almost directly. The approach in this paper can, therefore, provide informative predictions on welfare, say in a setting where managerial incentives are to be newly introduced (for instance, in the context of a state-owned enterprise which had previously been run by bureaucrats in the absence of direct financial incentives). Other settings where the results above might perhaps be readily adapted include auctions for incentive contracts (as in Laffont and Tirole (1987)) and dynamic incentive contracts with stochastically evolving types (as in Garrett and Pavan mentioned above). More speculatively, there may exist ways to adapt the ideas of the paper to other settings with adverse selection generally, such as second-degree price discrimination models in the style of Mussa and Rosen (1978). For instance, consider a firm that has practiced linear pricing of quantity with a fixed price per unit, generating historical quantity data. What predictions can be made regarding welfare if the firm, while still facing one-dimensional asymmetric information, departs from linear pricing to choose an optimal nonlinear schedule? Note that a reasonably successful application of our ideas might involve only finding bounds that are not tight (or cannot be shown to be tight). In other words, it could be that only the first step in our methodology can be mimicked or repurposed in some contexts.

## Appendix: Proofs of all results

### A.1 *Proof of Proposition 3.1*

We begin by finding a lower bound on the principal's expected payoff in a mechanism with the production cost target given by $C(\cdot)$.

Lemma A.1. *Fix an integrable function $C : [\underline{\beta}, \bar{\beta}] \to \mathbb{R}$ that prescribes production costs to each innate cost $\beta$. A lower bound on the principal's expected total payment in an incentive-compatible and individually rational mechanism is given by*

$$\mathbb{E}\big[C(\tilde{\beta}) + y(\tilde{\beta})\big] = \mathbb{E}\big[\tilde{\beta} - \mathrm{VG}\big(e(\tilde{\beta}), \tilde{\beta}\big)\big],$$

*where $e(\beta) = \beta - C(\beta)$ for all $\beta$ and where* VG *is given by (4).*

Proof. Let the agent of type $\beta$ have payoff, when producing at realized cost $C$, equal to $v(C, \beta) = -\psi(\beta - C)$ plus the transfer received from the principal. Here we can view the cost target $C$ as drawn from a set $\mathcal{C} = \mathbb{R}$ (the "allocation set" in the language of Carbajal and Ely (2013)). We seek to apply Theorem 1 of Carbajal and Ely to this setting.

Note that because $\psi$ is assumed Lipschitz continuous, $\psi(\beta - C)$ is equi-Lipschitz continuous in $\beta$ across $C \in \mathcal{C}$, with the Lipschitz constant the same as for $\psi$. This ensures

the satisfaction of Assumption A3 of Carbajal and Ely. Note that satisfaction of their Conditions A1 and A2 is immediate.[16]

Define, for each $\beta \in [\underline{\beta}, \bar{\beta}]$ and each $C \in \mathcal{C}$,

$$\begin{aligned}\bar{d}v(C, \beta) &\equiv \liminf_{r \searrow 0}\left[\frac{-\psi(\beta + r - C) + \psi(\beta - C)}{r}\right] \\ &= \lim_{r \searrow 0}\left[\frac{-\psi(\beta + r - C) + \psi(\beta - C)}{r}\right]\end{aligned}$$

and

$$\begin{aligned}\underline{d}v(C, \beta) &\equiv \limsup_{r \nearrow 0}\left[\frac{-\psi(\beta + r - C) + \psi(\beta - C)}{r}\right] \\ &= \lim_{r \nearrow 0}\left[\frac{-\psi(\beta + r - C) + \psi(\beta - C)}{r}\right],\end{aligned}$$

where the equalities follow from convexity of $\psi$. Hence, given $-\psi$ is concave, functions $\bar{d}v(C, \beta)$ and $\underline{d}v(C, \beta)$ are superderivatives of $-\psi(\cdot)$, evaluated at $\beta - C$. As a result, the correspondence $S : [\underline{\beta}, \bar{\beta}] \rightrightarrows \mathbb{R}$ given by

$$S(\beta) \equiv \left\{r \in \mathbb{R} : \bar{d}v\big(C(\beta), \beta\big) \leq r \leq \underline{d}v\big(C(\beta), \beta\big)\right\}$$

is nonempty. The correspondence $S(\beta)$ is single-valued in case the above limits are equal at $(C(\beta), \beta)$ and is a closed interval of positive length otherwise. By convexity of $\psi$, $\bar{d}v(-C, \beta)$ and $\underline{d}v(-C, \beta)$ are nonincreasing in $(C, \beta)$; hence, $\bar{d}v$ and $\underline{d}v$ are measurable functions, while $C(\cdot)$ is assumed measurable. Hence, $\bar{d}v(C(\cdot), \cdot)$ and $\underline{d}v(C(\cdot), \cdot)$ are measurable, verifying Ely and Carbajal's Assumption M. Note also that by the above definitions, $\bar{d}v(C(\beta), \beta)$ and $\underline{d}v(C(\beta), \beta)$ depend only on $e(\beta) = \beta - C(\beta)$ (and not $\beta$ and $C(\beta)$ individually).

Now recall that the payment rule can be chosen to ensure the agent always finds it optimal to set effort equal to $\beta - C(\hat{\beta})$ for any report $\hat{\beta}$. If the direct mechanism implementing production cost rule $C(\cdot)$ is incentive compatible, the agent's payoff can be denoted $V(\beta) = y(\beta) - \psi(\beta - C(\beta)) = \max_{\hat{\beta} \in [\underline{\beta}, \bar{\beta}]}\{y(\hat{\beta}) - \psi(\beta - C(\hat{\beta}))\}$. Since A1–A3 and M of Carbajal and Ely are satisfied, Theorem 1 of their paper applies. Hence, for any $\beta \in [\underline{\beta}, \bar{\beta}]$,

$$V(\beta) = V(\bar{\beta}) - \int_{\beta}^{\bar{\beta}} s(x)\, dx$$

for some measurable selection $s$ of $S$.

A lower bound on agent rents in an incentive-compatible and individually rational mechanism is provided by taking $s(\beta) = -[\partial_{-}\psi](e(\beta))$ for all $\beta$ (i.e., equal to the upper bound for $S$), and by setting $V(\bar{\beta}) = 0$ (since individual rationality requires $V(\bar{\beta}) \geq 0$).

[16]For Condition A1, we can pair $\mathcal{C}$ with the Borel sigma algebra on $\mathbb{R}$, since feasible production cost assignments are then measurable functions $C : [\underline{\beta}, \bar{\beta}] \to \mathbb{R}$.

For an agent of type $\beta$ to earn rents $\int_{\beta}^{\bar{\beta}}[\partial_{-}\psi](e(x))\,dx$ when truth-telling in the direct mechanism, it must be that $y(\beta) = \psi(e(\beta)) + \int_{\beta}^{\bar{\beta}}[\partial_{-}\psi](e(x))\,dx$. After integration by parts, we have

$$\mathbb{E}\big[C(\tilde{\beta}) + y(\tilde{\beta})\big] = \mathbb{E}\bigg[\tilde{\beta} - e(\tilde{\beta}) + \psi\big(e(\tilde{\beta})\big) + \frac{F(\tilde{\beta})}{f(\tilde{\beta})}[\partial_{-}\psi]\big(e(\tilde{\beta})\big)\bigg]$$

as desired. □

We now characterize effort policies that minimize the lower bound. Such policies maximize pointwise the virtual gains $\mathrm{VG}(e, \beta)$ by choice of $e \in \mathbb{R}$ for almost every $\beta$; in what follows, we omit the qualification that statements hold only for sets of innate costs $\beta$ that have probability 1, simply considering effort policies that maximize $\mathrm{VG}(e, \beta)$ for *every* value of $\beta$.

Because the best Lipschitz constant for $\psi$ is greater than 1, for each $\beta \in [\underline{\beta}, \bar{\beta}]$, there exists $u > 0$ such that $\mathrm{VG}(e, \beta) < 0$ for all $e < 0$ and all $e > u$. Note that because $\psi$ is convex, the left derivative of $\psi$, i.e., $\partial_{-}\psi$, is left continuous and nondecreasing. Hence, $\mathrm{VG}(\cdot, \beta)$ is upper semicontinuous for all $\beta$. This means that the maximizers $E^*(\beta) \equiv \arg\max[\mathrm{VG}(e, \beta)]$ are nonempty and closed for each $\beta$. Since $F(\beta)/f(\beta)$ is increasing, standard monotone comparative statics arguments (see Topkis (1978)) imply that $E^*(\beta)$ is nonincreasing in the strong set order. We can then consider monotone (nonincreasing) selections, denoted $e^*(\beta)$, of the correspondence $E^*$ (for instance, one can take $\max E^*(\beta)$ or $\min E^*(\beta)$).

We now show that effort policies that are monotone selections from $E^*$ can be implemented as part of an incentive-compatible and individually rational mechanism, with the principal's expected payment equal to the lower bound in Lemma A.1. For a monotone selection $e^*(\cdot)$, the cost target is given by $C^*(\beta) = \beta - e^*(\beta)$ for each $\beta$ (hence, $C^*(\cdot)$ is nondecreasing). Let then the payments to the agent when the cost target is met (in addition to the reimbursement of production costs) be given by $y^*(\beta) = \psi(e^*(\beta)) + \int_{\beta}^{\bar{\beta}}[\partial_{-}\psi](e^*(x))\,dx$. Take payments when the agent fails to meet the cost target to be small enough that this is never optimal for the agent.

Now let $U(\beta, \hat{\beta})$ be the payoff obtained by type $\beta$ when reporting $\hat{\beta}$ and choosing effort to meet the cost target. We have

$$\begin{aligned}
U(\beta, \hat{\beta}) &= y(\hat{\beta}) - \psi\big(\beta - C(\hat{\beta})\big) \\
&= U(\beta, \beta) + \int_{\hat{\beta}}^{\beta}[\partial_{-}\psi]\big(e(x)\big)\,dx - \big(\psi\big(\beta - C(\hat{\beta})\big) - \psi\big(\hat{\beta} - C(\hat{\beta})\big)\big) \\
&= U(\beta, \beta) - \int_{\hat{\beta}}^{\beta}\big([\partial_{-}\psi]\big(x - C(\hat{\beta})\big) - [\partial_{-}\psi]\big(x - C(x)\big)\big)\,dx \\
&\leq U(\beta, \beta).
\end{aligned}$$

The third equality follows using the concept that a convex function is differentiable except for at most countably many points (i.e., $\partial_{-}\psi = \psi'$, except at these points). The inequality follows because $C$ and $\partial_{-}\psi$ are nondecreasing functions. Given that the agent

finds it optimal to meet the cost target $C(\hat{\beta})$ for any report $\hat{\beta}$, the inequality implies incentive compatibility, as desired. Hence, the effort policy $e^*$ is implementable in an incentive-compatible mechanism where the principal's expected payment is given in Lemma A.1, as we wanted to show.

We now prove a result that establishes the final claim in the proposition.

LEMMA A.2. *Let $e^*(\cdot)$ be any measurable selection from $E^*$. For all $\beta > \underline{\beta}$, the left derivative of disutility at equilibrium effort, $[\partial_-\psi](e^*(\beta))$, must be strictly less than 1.*

PROOF. Let $e^{\min}(\underline{\beta})$ be the minimal element of $E^*(\underline{\beta})$. Note that $[\partial_-\psi](e^{\min}(\underline{\beta})) \leq 1$; if $[\partial_-\psi](e^{\min}(\underline{\beta})) > 1$, effort can be reduced from $e^{\min}(\underline{\beta})$ while increasing surplus, contradicting the definition of $e^{\min}(\underline{\beta})$. In addition, $[\partial_-\psi](e) < 1$ for all $e < e^{\min}(\underline{\beta})$. Given the first claim and convexity of $\psi$, the only way this can fail to be true is if $[\partial_-\psi](e^{\min}(\underline{\beta})) = [\partial_-\psi](e) = 1$ for some $e < e^{\min}(\underline{\beta})$. However, in this case, $\psi$ is linear on $[e, e^{\min}(\underline{\beta})]$ with gradient equal to 1, contradicting that $e^{\min}(\underline{\beta})$ is the minimum of the efficient effort choices.

Now fixing $\beta > \underline{\beta}$, we want to show that $[\partial_-\psi](e^*(\beta)) < 1$. Because $F/f$ is assumed strictly increasing, $[\partial_-\psi](e^*(\beta)) \leq [\partial_-\psi](e^{\min}(\underline{\beta}))$ follows from optimality of $e^{\min}(\underline{\beta})$ for type $\underline{\beta}$ and of $e^*(\beta)$ for type $\beta$. Hence, the only case we need to consider is where $[\partial_-\psi](e^{\min}(\underline{\beta})) = 1$. For this case, consider the effect on the virtual gain from incentives $\mathrm{VG}(e, \beta)$ when reducing effort to $e = e^{\min}(\underline{\beta}) - \varepsilon$ for $\varepsilon > 0$ from the efficient effort $e^{\min}(\underline{\beta})$. The change is

$$\begin{aligned}
&e^{\min}(\underline{\beta}) - \varepsilon - \psi\big(e^{\min}(\underline{\beta}) - \varepsilon\big) - \frac{F(\beta)}{f(\beta)}[\partial_-\psi]\big(e^{\min}(\underline{\beta}) - \varepsilon\big) \\
&\quad - \left(e^{\min}(\underline{\beta}) - \psi\big(e^{\min}(\underline{\beta})\big) - \frac{F(\beta)}{f(\beta)}[\partial_-\psi]\big(e^{\min}(\underline{\beta})\big)\right) \\
&= -\int_{e^{\min}(\underline{\beta})-\varepsilon}^{e^{\min}(\underline{\beta})} \big(1 - [\partial_-\psi](e)\big)\, de + \frac{F(\beta)}{f(\beta)}\big([\partial_-\psi]\big(e^{\min}(\underline{\beta})\big) - [\partial_-\psi]\big(e^{\min}(\underline{\beta}) - \varepsilon\big)\big) \\
&\geq \left(\frac{F(\beta)}{f(\beta)} - \varepsilon\right)\big(1 - [\partial_-\psi]\big(e^{\min}(\underline{\beta}) - \varepsilon\big)\big).
\end{aligned}$$

The equality follows because $\psi$ is convex and, hence, differentiable except at countably many points. The inequality follows because $\partial_-\psi$ is nondecreasing. The right-hand side of the inequality is strictly positive for $\varepsilon$ sufficiently small, since $\frac{F(\beta)}{f(\beta)}$ is strictly positive. This shows that, indeed, $e^*(\beta) < e^{\min}(\underline{\beta})$ and, hence, $[\partial_-\psi](e^*(\beta)) < 1$. □

We next determine further properties of optimal effort policies.

LEMMA A.3. *Optimal effort $e^*(\cdot)$ is essentially unique and essentially nonincreasing.*

PROOF. First, consider why any selection from optimal effort policies $E^*$ must be nonincreasing (the argument is closely related to the one in Topkis (1978, Theorem 6.3)). Consider, for a contradiction, an effort policy $e^*$ that maximizes virtual gains, but for which

there are $\beta', \beta'' \in [\underline{\beta}, \bar{\beta}]$ with $\beta' < \beta''$ and $e^*(\beta') < e^*(\beta'')$. From the previous lemma, $[\partial_-\psi](e^*(\beta'')) < 1$ and, hence, since $\psi$ is convex, we conclude that $e^*(\beta'') - \psi(e^*(\beta'')) > e^*(\beta') - \psi(e^*(\beta'))$. Hence, if $[\partial_-\psi](e^*(\beta'')) = [\partial_-\psi](e^*(\beta'))$, $e^*(\beta')$ does not maximize the virtual gains $\mathrm{VG}(e, \beta')$. Suppose then that $[\partial_-\psi](e^*(\beta'')) > [\partial_-\psi](e^*(\beta'))$ and note that

$$\begin{aligned} &e^*(\beta') - \psi(e^*(\beta')) - \frac{F(\beta')}{f(\beta')}[\partial_-\psi](e^*(\beta')) \\ &\geq e^*(\beta'') - \psi(e^*(\beta'')) - \frac{F(\beta')}{f(\beta')}[\partial_-\psi](e^*(\beta'')) \end{aligned}$$

because $e^*(\beta')$ maximizes virtual gains $\mathrm{VG}(e, \beta')$. Since $\frac{F(\beta'')}{f(\beta'')} > \frac{F(\beta')}{f(\beta')}$, we have

$$\begin{aligned} &e^*(\beta') - \psi(e^*(\beta')) - \frac{F(\beta'')}{f(\beta'')}[\partial_-\psi](e^*(\beta')) \\ &> e^*(\beta'') - \psi(e^*(\beta'')) - \frac{F(\beta'')}{f(\beta'')}[\partial_-\psi](e^*(\beta'')), \end{aligned}$$

which contradicts $e^*(\beta'')$ maximizing the virtual gains $\mathrm{VG}(e, \beta'')$. We conclude that $e^*(\beta'') \leq e^*(\beta')$.

We thus showed, in the language of Topkis (1978), that the set of maximizers $E^*(\beta)$ is strongly descending ($\beta'' > \beta'$ implies $e^*(\beta'') \leq e^*(\beta')$). Every $E^*(\beta)$ that is not a singleton corresponds to an open interval, say $(e'(\beta), e''(\beta))$ for $e'(\beta), e''(\beta) \in E^*(\beta)$. That $E^*(\beta)$ is strongly descending implies that the collection of such intervals, $\{(e'(\beta), e''(\beta)) : \beta \in [\underline{\beta}, \bar{\beta}]\}$, is disjoint. Hence, essential uniqueness of optimal effort follows because there can be at most countably many disjoint open intervals in $\mathbb{R}$. □

This completes the proof of Proposition 3.1.

## A.2 *Proof of results in Section 4*

Proof of Proposition 4.1. The proof consists of three steps.

**Step 1:** $\{P(\gamma) : \gamma \in \Gamma\} = \mathrm{co}\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$**.** We first show that $\{P(\gamma) : \gamma \in \Gamma\}$ is equal to the convex hull of $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$, as claimed in the main text. Note that, by Carathéodory's theorem, any point in the convex hull of $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$ (a set in $\mathbb{R}^2$) can be written as the convex combination of points $P(\gamma_x)$ for at most three values of $x$. By linearity of $P$ and because any convex combination of step functions $\gamma_x$ is in $\Gamma$, this point must reside in $\{P(\gamma) : \gamma \in \Gamma\}$; i.e.,

$$\mathrm{co}\big\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\big\} \subseteq \big\{P(\gamma) : \gamma \in \Gamma\big\}.$$

Conversely, any point $P(\gamma)$, $\gamma \in \Gamma$, can be approximated arbitrarily closely by points $P(\gamma^k)$, with $\gamma^k$ being right-continuous step functions and, hence, convex combinations of the step functions $\gamma_x$. In particular, there exists a sequence $(\gamma^k)_{k=1}^{\infty}$ of such

step functions such that $P(\gamma^k) \in \text{co}\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$ for all $k$ and with $P(\gamma^k) \to P(\gamma)$ as $k \to \infty$. Since the convex hull of a compact set in $\mathbb{R}^2$ is itself compact, the convex hull of $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$ is compact. It therefore contains $P(\gamma)$. This establishes $\{P(\gamma) : \gamma \in \Gamma\} \subseteq \text{co}\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$, which implies the result.

**Step 2: $L^*$ strictly increasing and convex.** That $L^*$ is strictly increasing and convex follows immediately from observing that $(R, L^*(R))$ for $R \in (0, \bar{R})$ is a point on the lower boundary of the convex hull $\text{co}\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$.

**Step 3: Form of a solution.** The fact that there is a solution $\gamma^*$ described by the cutoffs $\beta_l$ and $\beta_u$ follows because points on the lower boundary of $\text{co}\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$ can be written as convex combinations of $P(\gamma_x)$ for at most two values of $x$. This follows again by Carathéodory's theorem. Consider the tangent line to the convex hull passing through the point $(R, L^*(R))$. This point belongs to the intersection of $\text{co}\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$ and the aforementioned tangent line; a set with dimension 1. Hence, by Carathéodory's theorem, it can be written as the convex combination of at most two points in the set. The claim in the proposition then follows, since there is then a solution to Problem I that can be written as a convex combination of the step functions $\gamma_x$ for two values of $x$.

This completes the proof of Proposition 4.1. □

Proof of Proposition 4.2. Recall that in case $W(\bar{\beta}) = 0$, the expected gains from incentives is equal to $\int_{\underline{\beta}}^{\bar{\beta}} F(s)h(s)[\partial_- \psi](e^*(s))\,ds$, where $e^*$ is an optimal effort policy. Given $F$, consider a solution to Problem I, $\gamma^*$, that can be described by cutoffs $\beta_l$ and $\beta_u$ as introduced in Proposition 4.1. We aim to select a sequence of disutility functions in $\Psi$ such that the left derivative of the agent's marginal disutility of effort in equilibrium, $[\partial_- \psi](e^*(\cdot))$, approaches $\gamma^*(\cdot)$, and where $W(\bar{\beta})$ is equal to 0.

While the case where $\underline{\beta} = \beta_l < \beta_u$ is considered in the main text, there are two remaining cases.

*First case:* $\underline{\beta} < \beta_l = \beta_u \equiv \beta^*$. Suppose there is a solution to Problem I with $\underline{\beta} < \beta_l = \beta_u \equiv \beta^*$. We consider a sequence of disutility functions $(\psi_n)_{n=1}^{\infty}$. Under an optimal mechanism for the $n$th disutility function of the sequence, the agent exerts positive effort for any innate cost below some threshold $\beta_n$, but zero effort for any higher innate cost. When positive effort is chosen, the left derivative of disutility is close to 1; precisely, we ensure it is equal to $1 - \frac{\eta}{n}$ for a small but positive value $\eta$. So that, for every $n$, the expected rent is equal to $R \in (0, \bar{R})$, we require (recalling (2) for expected rents) that

$$\int_{\underline{\beta}}^{\beta_n} F(x)\left(1 - \frac{\eta}{n}\right) dx = R.$$

Taking $\eta$ small enough, this equation determines a decreasing sequence $(\beta_n)_{n=1}^{\infty}$ in $(\beta^*, \bar{\beta})$, convergent to $\beta^*$, as well as a strictly positive sequence $(b_n)_{n=1}^{\infty}$ with

$$b_n = \frac{F(\beta_n)}{f(\beta_n)}\left(\frac{n}{\eta} - 1\right).$$

The latter is used to define disutility functions

$$\psi_n(e) \equiv \begin{cases} 0 & \text{if } e \le 0, \\ \left(1-\dfrac{\eta}{n}\right)e & \text{if } 0 < e \le b_n, \\ \left(1-\dfrac{\eta}{n}\right)b_n + 2(e-b_n) & \text{if } e > b_n \end{cases}$$

for each positive integer $n$. For each $n$, $\psi_n$ belongs to $\Psi$, and an optimal mechanism features effort $b_n$ for innate costs below the threshold $\beta_n$; effort for innate costs above $\beta_n$ is 0. We thus obtain $R(\psi_n; F) = R$ for each $n$ and can verify that

$$G(\psi_n; F) \to \int_{\underline{\beta}}^{\bar{\beta}} F(s)h(s)\gamma^*(s)\,ds = L^*(R)$$

as $n \to +\infty$.

*Second case:* $\underline{\beta} < \beta_l < \beta_u$. Suppose there is a solution to Problem I with $\underline{\beta} < \beta_l < \beta_u$. Hence, there is an interval on which $\gamma^*(\beta) = 1$, an interval on which $\gamma^*(\beta) = \gamma^{\text{mid}}$ for $\gamma^{\text{mid}} \in (0, 1)$, and possibly an interval on which $\gamma^*(\beta) = 0$.

Let $a = \frac{F(\beta_u)}{f(\beta_u)} \frac{\gamma^{\text{mid}}}{1-\gamma^{\text{mid}}}$. Let $\eta > 0$ and let it be small enough that an innate cost $\beta_n$ is defined implicitly by

$$\left(1-\gamma^{\text{mid}}\right) \int_{\beta_l}^{\beta_n} F(s)\,ds = \frac{\eta}{n} \int_{\underline{\beta}}^{\beta_n} F(s)\,ds,$$

with $(\beta_n)_{n=1}^{\infty}$ a decreasing sequence in $(\beta_l, \beta_u)$ (convergent to $\beta_l$). Let, for each $n = 1, 2, \ldots,$

$$b_n = a + \frac{F(\beta_n)}{f(\beta_n)} \left( \frac{n}{\eta}\left(1-\gamma^{\text{mid}}\right) - 1 \right).$$

We can consider $\eta$ to be small enough that $(b_n)_{n=1}^{\infty}$ takes values strictly greater than $a$ for every $n$.

Define a sequence of disutility functions in $\Psi$ as follows: for each $n = 1, 2, \ldots,$

$$\psi_n(e) \equiv \begin{cases} 0 & \text{if } e \le 0, \\ \gamma^{\text{mid}} e & \text{if } e \in (0, a], \\ \gamma^{\text{mid}} a + \left(1-\dfrac{\eta}{n}\right)(e-a) & \text{if } e \in (a, b_n], \\ \gamma^{\text{mid}} a + \left(1-\dfrac{\eta}{n}\right)(b_n - a) + 2(e-b_n) & \text{if } e \in (b_n, \infty). \end{cases}$$

Consider now effort levels that maximize the virtual gains $\text{VG}_n(e, \beta) \equiv e - \psi_n(e) - \frac{F(\beta)}{f(\beta)}[\partial_- \psi_n](e)$. For each $n$, these satisfy $e_n^*(\beta) \in \{0, a, b_n\}$. The virtual gains for these

levels of effort are, respectively, 0,

$$a - \gamma^{\text{mid}} a - \frac{F(\beta)}{f(\beta)}\gamma^{\text{mid}},$$
$$b_n - \gamma^{\text{mid}} a - \left(1 - \frac{\eta}{n}\right)(b_n - a) - \frac{F(\beta)}{f(\beta)}\left(1 - \frac{\eta}{n}\right).$$

We have that both $e_n^*(\beta) = 0$ and $e_n^*(\beta) = a$ are optimal in case $\beta = \beta_u$, and both $e_n^*(\beta) = a$ and $e_n^*(\beta) = b_n$ are optimal in case $\beta = \beta_n$ (these observations follow by choice of $a$ and $b_n$). Thus, given disutility $\psi_n$, the principal chooses effort $e_n^*(\beta) = 0$ in case $\beta > \beta_u$, effort $e_n^*(\beta) = a$ in case $\beta \in (\beta_n, \beta_u)$, and effort $e_n^*(\beta) = b_n$ in case $\beta < \beta_n$. Note then that expected agent rents are

$$\begin{aligned}
\int_{\underline{\beta}}^{\bar{\beta}} F(s)[\partial_- \psi_n]\big(e_n^*(s)\big)\, ds &= \left(1 - \frac{\eta}{n}\right)\int_{\underline{\beta}}^{\beta_n} F(s)\, ds + \gamma^{\text{mid}} \int_{\beta_n}^{\beta_u} F(s)\, ds \\
&= \int_{\underline{\beta}}^{\beta_l} F(s)\, ds + \gamma^{\text{mid}} \int_{\beta_l}^{\beta_u} F(s)\, ds \\
&\quad + \big(1 - \gamma^{\text{mid}}\big)\int_{\beta_l}^{\beta_n} F(s)\, ds - \frac{\eta}{n}\int_{\underline{\beta}}^{\beta_n} F(s)\, ds \\
&= \int_{\underline{\beta}}^{\beta_l} F(s)\, ds + \gamma^{\text{mid}} \int_{\beta_l}^{\beta_u} F(s)\, ds \\
&= R.
\end{aligned}$$

The third equality holds by choice of $\beta_n$, while the final equality holds as a property of the solution to Problem I, $\gamma^*$. The principal's expected payoff is

$$\int_{\underline{\beta}}^{\bar{\beta}} F(s)h(s)[\partial_- \psi_n]\big(e_n^*(s)\big)\, ds,$$

which approaches $L^*(R) = \int_{\underline{\beta}}^{\bar{\beta}} F(s)h(s)\gamma^*(s)\, ds$ as $n \to +\infty$. This convergence follows using that $F(\beta)h(\beta)$ remains bounded on all of $[\underline{\beta}, \bar{\beta}]$. □

Proof of Corollary 4.1. The result is a consequence of the following observation. Consider any disutility function $\psi \in \Psi$, with the right derivative at zero strictly positive (recall that the proof of Proposition 4.2 considered only such functions, so as to approach the boundary of $\mathcal{U}$). Let $a, \varepsilon > 0$, with $\varepsilon$ less than the aforementioned right derivative. Then consider the disutility function $\bar{\psi}(e; a, \varepsilon)$ as defined in (6). Given this disutility function, the principal's virtual gains for innate cost $\beta$ are zero for effort zero,

$$a(1 - \varepsilon) - \frac{F(\beta)}{f(\beta)}\varepsilon$$

for effort $a$, and

$$e - \varepsilon a - \psi(e-a) - \frac{F(\beta)}{f(\beta)}[\partial_- \psi](e-a)$$

for effort $e > a$. Note that the latter can be written as

$$e' + a(1-\varepsilon) - \psi(e') - \frac{F(\beta)}{f(\beta)}[\partial_- \psi](e')$$

for $e' = e - a > 0$. Holding $a$ fixed, provided $\varepsilon$ is small enough, optimal effort for the disutility $\bar{\psi}(e; a, \varepsilon)$ is at least $a$ for all $\beta$. Also, if the agent with innate cost $\beta$ takes effort $\check{e} > 0$ in the optimal policy for disutility function $\psi$, he takes effort $\check{e} + a$ in the optimal policy for $\bar{\psi}(\cdot; a, \varepsilon)$ and, hence, the (left) marginal disutility of effort is unchanged (i.e., $[\partial_- \psi](\check{e}) = [\partial_- \bar{\psi}(\cdot; a, \varepsilon)](\check{e} + a)$). The (left) marginal disutility of effort for disutility function $\bar{\psi}(e; a, \varepsilon)$ is $\varepsilon$ whenever the agent takes effort $a$. Also, the measure of $\beta$ for which the agent takes effort greater than $a$ for $\bar{\psi}(\cdot; a, \varepsilon)$ but zero under $\psi(\cdot)$ vanishes as $\varepsilon \to 0$. Therefore, the expected gains from incentives under $\bar{\psi}(\cdot; a, \varepsilon)$ are larger by an amount that approaches $a$ from below as $\varepsilon$ is taken to zero. The agent's expected rents are either the same as under $\psi$ (for instance, if the agent takes positive effort with probability 1 under $\psi$) or approach the value under $\psi$ from above as $\varepsilon \to 0$. □

### A.3 *Proof of Corollary 5.1*

To prove Corollary 5.1, first consider part (i). Hence, suppose $\frac{F(\beta)}{f(\beta)}$ is concave and $\mathbb{E}[\tilde{\beta}] \geq \frac{\underline{\beta}+\bar{\beta}}{2}$. Then, as observed in the main text, a solution to Problem I is $\gamma^*(\beta) = \frac{R}{\bar{R}}$ for all $\beta \in [\underline{\beta}, \bar{\beta}]$. Therefore, the result follows if we can show

$$\int_{\underline{\beta}}^{\bar{\beta}} F(\beta) h(\beta)\, d\beta - \int_{\underline{\beta}}^{\bar{\beta}} F(\beta)\, d\beta \leq 0,$$

and if we can show the inequality is strict when $\frac{F(\beta)}{f(\beta)}$ is strictly concave or if $\mathbb{E}[\tilde{\beta}] > \frac{\underline{\beta}+\bar{\beta}}{2}$.

Integrating by parts, we find

$$\int_{\underline{\beta}}^{\bar{\beta}} F(\beta) h(\beta)\, d\beta = \frac{1}{f(\bar{\beta})} - \int_{\underline{\beta}}^{\bar{\beta}} F(\beta)\, d\beta.$$

Hence, we have

$$\begin{aligned} &\int_{\underline{\beta}}^{\bar{\beta}} F(\beta) h(\beta)\, d\beta - \int_{\underline{\beta}}^{\bar{\beta}} F(\beta)\, d\beta \\ &\quad = 2\int_{\underline{\beta}}^{\bar{\beta}} \left( \frac{1}{2f(\bar{\beta})} - \frac{\beta - \underline{\beta}}{f(\bar{\beta})(\bar{\beta} - \underline{\beta})} - \left( \frac{F(\beta)}{f(\beta)} - \frac{\beta - \underline{\beta}}{f(\bar{\beta})(\bar{\beta} - \underline{\beta})} \right) \right) f(\beta)\, d\beta. \end{aligned} \tag{9}$$

Note then that $\int_{\underline{\beta}}^{\bar{\beta}} \frac{\beta-\underline{\beta}}{f(\bar{\beta})(\bar{\beta}-\underline{\beta})} f(\beta)\,d\beta \geq \frac{1}{2f(\bar{\beta})}$, because $\mathbb{E}[\tilde{\beta}] \geq \frac{\underline{\beta}+\bar{\beta}}{2}$, and the inequality is strict if $\mathbb{E}[\tilde{\beta}] > \frac{\underline{\beta}+\bar{\beta}}{2}$. Also, $\frac{F(\beta)}{f(\beta)}$ and $\frac{\beta-\underline{\beta}}{f(\bar{\beta})(\bar{\beta}-\underline{\beta})}$ are functions that take the same value at $\underline{\beta}$ and $\bar{\beta}$, while $\frac{F(\beta)}{f(\beta)}$ is concave; hence, $\frac{F(\beta)}{f(\beta)} \geq \frac{\beta-\underline{\beta}}{f(\bar{\beta})(\bar{\beta}-\underline{\beta})}$ on $(\underline{\beta}, \bar{\beta})$ and the inequality is strict in case $\frac{F(\beta)}{f(\beta)}$ is strictly concave. Part (i) of the corollary therefore follows.

Now consider part (ii). Hence, suppose $\frac{F(\beta)}{f(\beta)}$ is convex and $\mathbb{E}[\tilde{\beta}|\tilde{\beta} \leq \beta] \leq \frac{\underline{\beta}+\beta}{2}$ for all $\beta \in (\underline{\beta}, \bar{\beta}]$. For a given value $R \in (0, \bar{R})$, there is a solution $\gamma^*$ to Problem I such that $\gamma^*(\beta) = 1$ for $\beta < \beta^*$ and $\gamma^*(\beta) = 0$ for $\beta > \beta^*$. Then note that the conditional distribution defined on $[0, \beta^*]$ by $\bar{F}(\beta) \equiv F(\beta)/F(\beta^*)$ with density $\bar{f}$ satisfies $\frac{\bar{F}(\beta)}{\bar{f}(\beta)} = \frac{F(\beta)}{f(\beta)}$, which is convex. In addition, $\mathbb{E}_{\bar{F}}[\tilde{\beta}] \leq \frac{\underline{\beta}+\beta^*}{2}$. Hence, considering the expression in (9) evaluated for the distribution $\bar{F}$, with upper limit of the support $\beta^*$, we have

$$\int_{\underline{\beta}}^{\beta^*} F(\beta)h(\beta)\,d\beta - \int_{\underline{\beta}}^{\beta^*} F(\beta)\,d\beta \geq 0,$$

with strict inequality when either $\frac{F(\beta)}{f(\beta)}$ is strictly convex or $\mathbb{E}_{\bar{F}}[\tilde{\beta}] < \frac{\underline{\beta}+\beta^*}{2}$. This establishes the result.

## References


Abito, Jose Miguel (2020), "Measuring the welfare gains from optimal incentive regulation." *Review of Economic Studies*, 87, 2019–2048. [1284, 1299]

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris (2015), "The limits of price discrimination." *American Economic Review*, 105, 921–957. [1285]

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris (2017), "First-price auctions with general information structures: Implications for bidding and revenue." *Econometrica*, 85, 107–143. [1285]

Bergemann, Dirk and Stephen Morris (2013), "Robust predictions in games with incomplete information." *Econometrica*, 81, 1251–1308. [1285]

Bergemann, Dirk and Stephen Morris (2016), "Bayes correlated equilibrium and the comparison of information structures in games." *Theoretical Economics*, 11, 487–522. [1285]

Brocas, Isabelle, Kitty Chan, and Isabelle Perrigne (2006), "Regulation under asymmetric information in water utilities." *American Economic Review*, 96, 62–66. [1299]

Carbajal, Juan Carlos and Jeffrey C. Ely (2013), "Mechanism design without revenue equivalence." *Journal of Economic Theory*, 148, 104–133. [1287, 1289, 1301]

Carroll, Gabriel (2015), "Robustness and linear contracts." *American Economic Review*, 105, 536–563. [1286]

Carroll, Gabriel and Delong Meng (2016), "Robust contracting with additive noise." *Journal of Economic Theory*, 166, 586–604. [1301]

Chassang, Sylvain (2013), "Calibrated incentive contracts." *Econometrica*, 81, 1935–1971. [1286]

Chennells, Lucy (1997), "The windfall tax." *Fiscal Studies*, 18, 279–291. [1281]

Chu, Leon and David Sappington (2007), "Simple cost-sharing contracts." *American Economic Review*, 97, 419–428. [1286, 1297]

Dai, Tianjiao and Juuso Toikka (2017), "Robust incentives for teams." Report, MIT and U Penn. [1286]

Edmans, Alex and Xavier Gabaix (2011), "Tractability in incentive contracting." *Review of Financial Studies*, 24, 2865–2894. [1301]

Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, and Yuliy Sannikov (2012), "Dynamic CEO compensation." *Journal of Finance*, 67, 1603–1647. [1301]

Gagnepain, Philippe and Marc Ivaldi (2002), "Incentive regulatory policies: The case of public transit systems in France." *RAND Journal of Economics*, 33, 605–629. [1285, 1299, 1300]

Garrett, Daniel (2014), "Robustness of simple menus of contracts in cost-based procurement." *Games and Economic Behavior*, 87, 631–641. [1286]

Garrett, Daniel and Alessandro Pavan (2012), "Managerial turnover in a changing world." *Journal of Political Economy*, 120, 879–925. [1287, 1301]

Garrett, Daniel and Alessandro Pavan (2015), "Dynamic managerial compensation: A variational approach." *Journal of Economic Theory*, 159, 775–818. [1301]

Gasmi, Farid, Jean-Jacques Laffont, and William Sharkey (1997), "Incentive regulation and the cost structure of the local telephone exchange network." *Journal of Regulatory Economics*, 12, 5–25. [1298]

Hellwig, Martin F. (2008), "A maximum principle for control problems with monotonicity constraints." Unpublished working paper, Max Planck Institute for Research on Collective Goods. [1284]

Hurwicz, Leonid and Leonard Shapiro (1978), "Incentive structures maximizing residual gain under incomplete information." *Bell Journal of Economics*, 9, 180–191. [1286]

Laffont, Jean-Jacques and Jean Tirole (1986), "Using cost observation to regulate firms." *Journal of Political Economy*, 94, 614–641. [1283, 1284, 1286]

Laffont, Jean-Jacques and Jean Tirole (1987), "Auctioning incentive contracts." *Journal of Political Economy*, 94, 921–937. [1301]

Laffont, Jean-Jacques and Jean Tirole (1993), *A Theory of Incentives in Procurement and Regulation*. MIT Press, Cambridge. [1286]

Lopomo, Giuseppe and Efe A. Ok (2001), "Bargaining, interdependence, and the rationality of fair division." *Rand Journal of Economics*, 32, 263–283.

Mierendorff, Konrad (2016), "Optimal dynamic mechanism design with deadlines." *Journal of Economic Theory*, 161, 190–222. [1283]

Milgrom, Paul and Ilya Segal (2002), "Envelope theorems for arbitrary choice sets." *Econometrica*, 70, 583–601. [1292]

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18, 301–317. [1301]

Perrigne, Isabelle and Quang Vuong (2011), "Nonparametric identification of a contract model with adverse selection and moral hazard." *Econometrica*, 79, 1499–1539. [1286]

Rogerson (2003), "Simple menus of contracts in cost-based procurement and regulation." *American Economic Review*, 93, 919–926. [1286, 1298]

Segal, Ilya and Michael Whinston (2003), "Robust predictions for bilateral contracting with externalities." *Econometrica*, 71, 757–791. [1285]

Strausz, Roland (2006), "Deterministic versus stochastic mechanisms in principal-agent models." *Journal of Economic Theory*, 128, 306–314. [1290]

Topkis, Donald (1978), "Minimizing a submodular function on a lattice." *Operations Research*, 26, 305–321. [1303, 1304, 1305]

Wolak, Frank A. (1994), "An econometric analysis of the asymmetric information, regulator-utility interaction." *Annales d'Economie et de Statistique*, 34, 13–69. [1299]

---

Co-editor Simon Board handled this manuscript.

Manuscript received 30 April, 2020; final version accepted 25 January, 2021; available online 10 February, 2021.