DISCUSSION PAPER SERIES

DP14725

PAYOFF IMPLICATIONS OF INCENTIVE CONTRACTING

Daniel Garrett

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

PAYOFF IMPLICATIONS OF INCENTIVE CONTRACTING

Daniel Garrett

Discussion Paper DP14725
Published 07 May 2020
Submitted 04 May 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

PAYOFF IMPLICATIONS OF INCENTIVE CONTRACTING

Abstract

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.

JEL Classification: D82

Keywords: mechanism design, robustness, Procurement

Daniel Garrett - dfgarrett@gmail.com
Toulouse School of Economics, University of Toulouse Capitole and CEPR

Payoff Implications of Incentive Contracting*

Daniel F. Garrett

April 29, 2020

Abstract

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.

JEL classification: D82

Keywords: asymmetric information, mechanism design, robustness

*Toulouse School of Economics, University of Toulouse Capitole.** E-mail: daniel.garrett@tse-fr.eu. I am grateful for comments from Nikhil Agarwal, Gani Aldashev, Dirk Bergemann, Mehmet Ekmekci, Alessandro Pavan, Nicola Pavoni, 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 Toulouse School of Economics Winter IO meeting, the University of Bonn, and the Workshop on New Directions in Mechanism Design at Stony Brook. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 714147).

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. To give an example, reforms in the UK 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 the magnitude of corporate profits, as well as a perception that the public had failed to benefit from the changes.[1]

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 that might be made when ambiguity concerning the environment persists due to a lack of experience with incentives. Formally, we consider a principal-agent framework where incentive contracts are to be newly introduced. We then determine the predictions available to an analyst who is ignorant about details of the environment; in particular, who is ignorant regarding the agent's preferences (equivalently, technology) for responding to incentives.

Although facing ambiguity in the environment, the analyst is taken to have certain information. First, the distribution of performance absent incentives. This may be based on agent performance prior to the possibility of incentive contracting. Second, knowledge that the principal will introduce incentive contracts optimally given an accurate understanding of the agent's preferences. This may be based on an understanding that the principal will have more intimate knowledge of the contracting problem (say due to further study of the problem, or due to specialized expertise), as well as the freedom to design optimal contracts. Third, certain restrictions on these preferences.

We focus, for concreteness, on a model of cost-based procurement. The agent is tasked with supplying a fixed number of units to the principal. The cost of supplying these units

[1] To give another example, earlier, in the US, the Renegotiation Act of 1951 established the Renegotiation Board with the objective of "renegotiating" contracts deemed to have delivered excessive profits to government contractors (see Burns, 1970, for a description of the historical context).

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 (thus here, "good performance" is synonymous with a low production 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 (one that minimizes the expected total payment). Optimal contracts can be determined using a mechanism design approach, as in Laffont and Tirole (1986).

The problem of the analyst in this setting is to determine welfare predictions for optimal contracts. These predictions are made given the prior on the innate cost, but without knowledge of the agent's disutility function. This basis for predictions is in line with a setting where the analyst has observations on cost performance under cost-plus contracting, but has no experience with incentive 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, and provides analysis while ignorant of the disutility function, but anticipates information on this will become available 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).

Our main result is then a characterization of the possible expected payoffs from optimal incentive contracting, across all permitted agent preferences for cost-reducing effort. These expected payoffs can be measured relative to the status quo of no incentives (i.e., cost-plus contracting). Thus we consider the expected "gains from incentive contracting" for the principal as well as the expected rents of the agent (expected agent rents are zero in a cost-plus contract that provides no incentives to the agent).

A range of values for expected agent rents is possible in an optimal contract, depending on the disutility function. Our characterization of expected payoffs follows from determining a tight lower bound on the principal's gains from incentives for each level of expected agent rents. This lower bound turns out to be increasing in the expected rents, and convex. In other words, the principal is guaranteed at least a certain share of the expected efficiency improvements associated with optimal incentive contracting, and the guaranteed share increases with the size of the improvements.

We show how the 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 on the principal's expected gains 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). We argue that the share of efficiency gains guaranteed for the principal is smaller when the agent's innate costs are more concentrated at higher values.

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, Brooks and Morris (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 — players' information on the payoff-relevant state or payoff types, and where relevant their higher-order beliefs.[2] 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); at the same time other information, especially regarding the agent's preferences for effort, is not.

The value for robust predictions in our particular setting relate to a number of applications. Our predictions are of interest in the context of public-sector reform as discussed above, where the previous mode of production was government provision, usually associated with weak incentives to produce at low costs.[3] The introduction of incentives is also of interest in empirical work on procurement and regulation. For instance, Gagnepain and Ivaldi (2002) study data on contracts for transportation services written by local municipalities in France, with a quarter of firms subject to cost-plus contracts. Abito (2017) studies electric utilities subject to rate-of-return regulation. One aim of these studies (as well as the empirical literature on procurement and regulation more broadly) is counterfactual analysis on the

[2]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.

[3]Public sector reform also led to the introduction of explicit incentives for top executives at state-owned enterpises. Examples include reforms in China (see Mengistae and Xu, 2004), and in New Zealand (see Scott, Bushnell and Sallee, 1990).

introduction of optimal incentives. A key conceptual difference to this paper is that their analysis is informed by data on costs in settings *both* absent and with incentives (for instance, Gagnepain and Ivaldi's analysis is informed by data on both cost-plus and high-powered "fixed price" contracts). Their analysis also leverages functional form assumptions on the disutility of effort. Another setting where incentives can be freshly introduced is in labor contracts; for instance, Lazear (2000) documents the effects of a transition from low-powered fixed-wage contracts to piece-rate incentive schemes.[4]

The rest of this paper is as follows. Section 2 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 then shows how the set of feasible expected payoffs depends on the distribution of innate costs, as well as outlining an application to managerial compensation. Section 6 discusses related literature before Section 7 concludes. Proofs not in the main text are contained in the Appendix.

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; henceforth, LT). The model we consider has been 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" β, 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 in place measures 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 ψ and observes the realized production cost C, both the innate cost β and the effort e are the agent's private information.

[4] The model we consider is based on that of Laffont and Tirole (1986), where the application is cost-based regulation and procurement. This model often been applied in other settings, especially settings for worker or executive compensation; see Edmans and Gabaix (2011), Edmans, Gabaix, Sadzik and Sannikov (2012), Garrett and Pavan (2012, 2015), and Carroll (2016). We explain how our analysis can be adapted to these kinds of applications in Section 5.1.

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 ψ takes non-negative values and satisfies the following requirements. It is taken to be non-decreasing and convex; with ψ strictly increasing on $\mathbb{R}_+$ and constant at zero on $\mathbb{R}_-$. We take ψ to satisfy the Inada condition $\lim_{e \to +\infty} \{e - \psi(e)\} = -\infty$ and to be Lipschitz continuous. We then let Ψ be the set of all disutility functions ψ satisfying these conditions.

That the agent incurs positive disutility from positive effort ensures that the innate cost β 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 will not occur in equilibrium. Monotonicity and convexity of ψ 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 Inada condition; this condition will play a role in guaranteeing the existence of efficient and optimal policies. Lipschitz continuity is a technical condition, which, given convexity of ψ, is a restriction on this function only at large values of effort e that will not be chosen in equilibrium.[5]

Note in addition that the agent's preferences for effort are independent of the innate cost (i.e., ψ does not depend on β). While this assumption has been common in the procurement literature, its applicability would depend on the circumstances at hand. For instance, independence describes well a scenario where the agent's private information on β 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.[6]

[5]While we expect the assumption of Lipschitz continuity can be dispensed with, it facilitates application of an appropriate envelope theorem (in particular, Carbajal and Ely, 2013), used in the derivation of the principal's optimal policy for given ψ.

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

The agent's innate cost β is drawn from a cdf F that is twice continuously differentiable, with 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, throughout we assume 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 has frequently been a restriction in the literature, often justified by a claim that many commonly-considered distributions satisfy this property.

The timing of the game is then the same as in LT. First, the agent learns his private type β, 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" $C\left(\hat{\beta}\right)$. If the agent reports his innate cost β 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 β. If the agent achieves the target — i.e., $C = C\left(\hat{\beta}\right)$ — then he is paid $y\left(\hat{\beta}\right)$. Otherwise, if $C \neq C\left(\hat{\beta}\right)$, the payment to the agent is negative. Since the mechanism is individually rational, a choice $C \neq C\left(\hat{\beta}\right)$ 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 has a reliable prior belief F regarding the innate cost β. However, she does not know the agent's preferences for effort, only that they are described by a function in Ψ. 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, will know the disutility function ψ precisely, and will choose mechanisms optimally. We ask, what expected payoff implications does the analyst

may be more or less tractable depending on the restrictions in question).

consider possible?

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 ψ; for instance, we do not require ψ 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" $C\left(\hat{\beta}\right)$ that the agent finds it optimal to meet. Hence, the agent's payoff, if his true innate cost is β and he chooses effort optimally, is

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

Let $\partial_{-}\psi$ denote the left derivative of ψ. We argue in the Appendix that we can consider mechanisms where the agent's rents are given, as a function of his true innate cost β, by

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

This follows from incentive compatibility of the mechanism, after applying the envelope result of Carbajal and Ely (2013) for non-differentiable objective functions, and from considering mechanisms that maximize the principal's expected payoff for a given effort policy $e(\cdot)$.

We next follow familiar steps to write the principal's expected total payment in a mechanism that optimally implements an effort policy $e(\cdot)$. This is

$$\mathbb{E}\left[\tilde{\beta} - VG\left(e\left(\tilde{\beta}\right), \tilde{\beta}\right)\right], \tag{2}$$

where

$$VG(e,\beta) = e - \psi(e) - \frac{F(\beta)}{f(\beta)}\left[\partial_{-}\psi\right](e) \tag{3}$$

(we leave the dependence of VG on ψ and F implicit). Here $VG(e,\beta)$ is the "virtual gain" from incentives inducing effort e for innate cost β, comprising efficiency gains $e - \psi(e)$ from effort less a term accounting for agent rents. Considering maximization of (2) by choice of the effort policy, we have the following result.

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

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

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

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 β, not merely almost all). In other words, the "first-order" or "relaxed program" approach to solving the design problem is established to be valid. While such a result is readily anticipated from earlier work (including LT), it is obtained under weaker conditions than usually assumed.[7] Because the first-order approach is valid, no additional restrictions on the shape of ψ 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 $VG(e,\beta)$. Effort is weakly downward distorted (note that we may have $[\partial_-\psi]\left(e^{FB}\right) < 1$ at an efficient effort level e^{FB} if there is a kink in ψ 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 β, which can be understood in part from examining the expression for agent rents in Equation (1): in particular, the agent's rents for a given innate cost depends on the effort induced from all higher innate costs.

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 satisfying the restrictions of the model set-up, and for any $\psi \in \Psi$, the principal implements an optimal mechanism with essentially unique effort $e^*(\cdot)$. Agent expected rents are then[8]

$$R(\psi;F) \equiv \mathbb{E}\left[\frac{F\left(\tilde{\beta}\right)}{f\left(\tilde{\beta}\right)}[\partial_-\psi]\left(e^*\left(\tilde{\beta}\right)\right)\right] \tag{4}$$

while

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

[7] Part of the additional generality relates to the possible non-differentiability of ψ. As noted above, this is handled by applying the envelope result of Carbajal and Ely (2013). One might be tempted to believe that precisely the same analysis as usually performed when ψ is differentiable should carry through, given that a convex disutility function ψ is diffentiable 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.

[8] The expression follows by taking the expectation of rents expressed in (1) and integrating by parts.

denotes the principal's "expected gains from incentives" (making the dependence of R and G on ψ and F explicit). Our interest will be in characterizing, for each F, the set

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

4 Analysis

Preliminary observations on the analyst's problem. We begin by determining the rents that the agent can be expected to obtain in an optimal mechanism. Proposition 3.1 implies that, irrespective of $\psi \in \Psi$, expected rents satisfy $R\left(\psi;F\right) < \bar{R} \equiv \int_{\underline{\beta}}^{\bar{\beta}} F\left(\beta\right) d\beta$. We can conclude that the set of feasible agent rents can be no larger than $[0, \bar{R})$; and indeed it is easy to verify that any level of rents in this set can occur under optimal contracting for *some* disutility $\psi \in \Psi$.[9]

Consider then the case where ψ and F are such that $R\left(\psi;F\right) = 0$. Given that $\partial_{-}\psi$ is strictly positive at positive effort values, we deduce that the agent exerts effort zero with probability one. Hence, $G\left(\psi;F\right) = 0$, and this holds irrespective of $\psi \in \Psi$. Our interest then is to determine the expected gains from incentives when the expected agent rents R are in $(0, \bar{R})$. Given F, our characterization of $\mathcal{U}$ will then follow from determining a function

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

on $[0, \bar{R})$. This function defines the lower boundary of the set $\mathcal{U}$.

Finally, note that while, for each level of agent expected rent $R \in (0, \bar{R})$, $G^{\inf}\left(R\right)$ defines the infimum of expected gains from incentives, 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

$$\bar{\psi}\left(e;a,\varepsilon\right) = \begin{cases} 0 & \text{if } e \leq 0 \\ \varepsilon e & \text{if } e \in (0,a], \\ \varepsilon a + \psi\left(e - a\right) & \text{if } e > a \end{cases} \tag{5}$$

for $\varepsilon, a > 0$. These parameters can be chosen so that expected gains from incentives under an optimal mechanism take values above $G\left(\psi;F\right)$, while expected rents are close to $R\left(\psi;F\right)$.[10]

[9] One way to see this is to consider disutility functions that are quadratic over the relevant range, i.e. with $\psi\left(e\right) = \frac{k}{2}e^2$ over $[0, \bar{e}]$ for some $\bar{e} > 1/k$, with $k > 0$.

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

The idea behind considering disutility functions of this form is that the agent is permitted to achieve cost reduction a almost for free when ε is small, implying an increase in the surplus that can be generated from incentives. For such a disutility function with small enough ε, optimal effort is at least a for all innate costs. Also, for an innate cost β assigned effort $e^*(\beta) > 0$ in an optimal mechanism for ψ, 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 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 ψ, the virtual gains are $W(\beta) = \max_e VG(e, \beta)$ (where recall VG is defined in Equation (3)). Because ψ 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

$$W(\beta) = W(\bar{\beta}) + \int_{\beta}^{\bar{\beta}} h(s) [\partial_{-}\psi](e^*(s)) \, ds, \tag{6}$$

where recall $h(\beta) = \frac{d}{d\beta}\left[\frac{F(\beta)}{f(\beta)}\right]$. Note that $W(\bar{\beta})$ is non-negative, and may be strictly positive depending on the disutility function. Also, $W(\cdot)$ is non-increasing. This can be understood by observing that the term that accounts for rents in Equation (3), i.e. $-\frac{F(\beta)}{f(\beta)}[\partial_{-}\psi](e)$, is non-increasing in β for any effort e (as $\frac{F(\beta)}{f(\beta)}$ 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 costs is smaller (recall that, by Equation (1), the rents earned for an agent with innate cost β 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

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

where the second equality follows from integration by parts.

One way to think about the integrand of the second term of Equation (7) is to note that a reduction in β increases the term in Equation (3) that accounts for agent rents

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

It also seems of interest to write $G(\psi; F)$ as

$$W(\bar{\beta}) + \mathbb{E}\left[\frac{F\left(\tilde{\beta}\right)}{f\left(\tilde{\beta}\right)} h\left(\tilde{\beta}\right)[\partial_{-}\psi]\left(e^{*}\left(\tilde{\beta}\right)\right)\right],$$

which is more easily compared to the expression for expected agent rents in Equation (4). Indeed, this comparison suggests expected rents should be informative about the value $G(\psi; F)$, even without knowledge of ψ. From Proposition 3.1, we may view $[\partial_{-}\psi](e^{*}(\cdot))$ as non-increasing and taking values in the unit interval. This suggests determining a lower bound on the expected gains from incentives, given expected agent rents, as a solution to the following problem.

Problem I. *Let Γ be the set of functions $\gamma : [\underline{\beta}, \bar{\beta}] \to [0,1]$ such that γ is non-increasing. For any F satisfying the conditions of the model set-up, any $R \in (0, \bar{R})$, determine*

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

In Problem I, $\gamma(\beta)$ can be viewed as representing values that the marginal disutility of effort $[\partial_{-}\psi](e^{*}(\beta))$ might take at innate cost β in an optimal mechanism for some disutility function ψ. The function $Z^{*}(\cdot)$ will turn out to define the lower boundary of $\mathcal{U}$. To show this, having obtained a solution γ^{*} to Problem I, we will show below that there is a $\psi \in \Psi$ such that, for an optimal effort policy e^{*}, the marginal disutility of effort $[\partial_{-}\psi](e^{*}(\cdot))$ either coincides with, or is arbitrarily close to γ^{*}.

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

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

and the set

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

We aim to obtain $Z^{*}(\cdot)$ as determined by the lower boundary of this set.

Consider now the right-continuous 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 in the Appendix that $\{P(\gamma) : \gamma \in \Gamma\}$ is equal to the convex hull of $\{P(\gamma_x) : x \in [\underline{\beta}, \bar{\beta}]\}$, which is a closed curve.[11] Hence, the convex hull is also closed.

Observe then that the pair $(R, Z^*(R))$ for $R \in (0, \bar{R})$ is a point on the lower boundary of this convex hull. It is then immediate that $Z^*(\cdot)$ is strictly increasing (since h is strictly positive) and weakly convex. In addition (by an application of Carathéodory's Theorem), any point $(R, Z^*(R))$ 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 γ_x for two values of x. To summarize, we have the following result.

Proposition 4.1. *Fix F satisfying the conditions of the model set-up. 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 $Z^*(\cdot)$ is strictly increasing and weakly convex. For any $R \in (0, \bar{R})$, there is a solution to Problem I for which the following is true. There are two cut-offs β_l and β_u, with $\underline{\beta} \leq \beta_l \leq \beta_u \leq \bar{\beta}$, such that $\gamma^*(\beta) = 1$ on $[\underline{\beta}, \beta_l)$, $\gamma^*(\beta)$ is constant and strictly between zero and one on $[\beta_l, \beta_u)$, and $\gamma^*(\beta) = 0$ on $[\beta_u, \bar{\beta}]$.*

To understand the properties of solutions to Problem I, we need to examine the curve $\{P(\gamma_x) \mid x \in [\underline{\beta}, \bar{\beta}]\}$. This can be represented as $(R, \kappa(R))$ for $R \in [0, \bar{R}]$, where

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

with $x(\cdot)$ defined implicitly by $R = \int_{\underline{\beta}}^{x(R)} F(s)\, ds$. We have $\kappa'(R) = h(x(R))$ (this can be seen using differentiability of $x(\cdot)$, with $x'(R) = \frac{1}{F(x(R))}$ for $R \in (0, \bar{R})$, as follows from the implicit function theorem). Thus, when h is strictly increasing (which occurs when F/f is strictly convex), the function $\kappa(\cdot)$ is convex. In this case, $Z^*(R) = \kappa(R)$ for all $R \in (0, \bar{R})$. Also, for all $R \in (0, \bar{R})$, there is a solution to Problem I given by $\gamma^* = \gamma_{x(R)}$ (i.e., $\beta_l = \beta_u = x(R)$). When h is strictly decreasing (which occurs when F/f is strictly concave), the curve $\kappa(R)$ is strictly concave. In this case, points $(R, Z^*(R))$ lie on the lower boundary of the convex hull of $\{P(\gamma_x) \mid x \in [\underline{\beta}, \bar{\beta}]\}$, and hence are convex combinations of $(0,0)$ and $(\bar{R}, \kappa(\bar{R}))$. In particular, we have $Z^*(R) = \frac{R}{\bar{R}}\kappa(\bar{R})$ for all $R \in (0, \bar{R})$. Also, for all $R \in (0, \bar{R})$, there is a solution to Problem I given by $\gamma^*(\beta) = \frac{R}{\bar{R}}$ for $\beta \in [\underline{\beta}, \bar{\beta}]$ (in this case $\beta_l = \underline{\beta}$ while $\beta_u = \bar{\beta}$). We provide sufficient conditions for F/f to be convex and concave in the following section.

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

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

Proposition 4.2. *Fix F satisfying the conditions of the model set-up, and fix any $R \in \left(0, \bar{R}\right)$. For any $\varepsilon > 0$, there exists $\psi \in \Psi$ such that*

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

and

$$G\left(\psi; F\right) < Z^{*}\left(R\right) + \varepsilon.$$

Hence, $G^{\inf}\left(R\right) = Z^{*}\left(R\right)$.

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. $\left[\partial_{-}\psi\right]\left(e^{*}\left(\cdot\right)\right)$, approaches a fixed solution γ^{*} to Problem I. We find it easiest to focus on solutions that can be described by cut-offs β_l and β_u, as introduced in Proposition 4.1. While the Appendix considers the case where the cut-offs in Proposition 4.1 satisfy $\underline{\beta} < \beta_l < \beta_u$, we consider here, in sequence, the cases with $\underline{\beta} = \beta_l < \beta_u$, and with $\underline{\beta} < \beta_l = \beta_u$ (this exhausts the relevant possibilities). These cases occur for instance when F/f is concave and when F/f is convex, respectively (as explained above).

Consider then the case with $\underline{\beta} = \beta_l < \beta_u$, so that the fixed solution to Problem I, γ^{*}, is constant at $\frac{R}{\int_{\underline{\beta}}^{\beta_u} F(s)ds} \in (0,1)$ on an interval $(\underline{\beta}, \beta_u)$. We aim to find a disutility function $\psi \in \Psi$ such that, at an optimal effort policy $e^{*}\left(\cdot\right)$, (a) the left derivative of disutility of effort $\left[\partial_{-}\psi\right]\left(e^{*}\left(\beta\right)\right)$ is constant and equal to $\frac{R}{\int_{\underline{\beta}}^{\beta_u} F(s)ds}$ for innate costs β below β_u, and is zero (with zero effort exerted) for higher innate costs, and (b) virtual gains from incentives $VG\left(e^{*}\left(\beta\right), \beta\right)$ 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 $Z^{*}\left(R\right)$.[12]

To determine an appropriate disutility function, let

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

and $k > 1$, and put

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

[12] Conditions (a) and (b) are not only sufficient for this to be true, but will also be necessary provided that the solution γ^{*} to Problem I is essentially unique (e.g., if F/f is strictly concave). This can be seen from Equations (6), (7) and (8) above.

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 β above β_u, if any. This shows that the infimum of expected gains from incentives (conditional on expected rents R) is attained.

For the case where $\underline{\beta} < \beta_l = \beta_u \equiv \beta^*$, 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 will exert positive effort for any innate cost below some threshold β_n, but zero effort for any higher innate cost. When positive effort is chosen, the left derivative of disutility will be close to one; precisely, we will ensure it is equal to $1 - \frac{\eta}{n}$ for a small but positive value η. In order that, for every n, the expected rent is equal to $R \in (0, \bar{R})$, we will require (recalling Equation (4) for expected rents) that

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

Taking η small enough, this equation determines a decreasing sequence $(\beta_n)_{n=1}^{\infty}$ in $(\beta^*, \bar{\beta})$, convergent to β^*, 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

$$\psi_n(e) \equiv \begin{cases} 0 & \text{if } e \leq 0 \\ \left(1 - \frac{\eta}{n}\right) e & \text{if } 0 < e \leq b_n \\ \left(1 - \frac{\eta}{n}\right) b_n + k(e - b_n) & \text{if } e > b_n \end{cases}$$

for some $k > 1$, and for each positive integer n. For each n, ψ_n belongs to Ψ, and an optimal mechanism features effort b_n for innate costs below the threshold β_n; effort for innate costs above β_n is zero. We thus obtain $R(\psi_n; F) = R$ for each n, and can verify that

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

as $n \to +\infty$.

Let us conclude this section by considering expected gains from incentives above the lower boundary $G^{\text{inf}}(R)$. The following can be established using disutility functions of the form introduced in Equation (5).

Corollary 4.1. *Fix F satisfying the conditions of the model set-up. For any $R \in (0, \bar{R})$, any $G > G^{\text{inf}}(R)$, and any $\varepsilon > 0$, there exists $\psi \in \Psi$ such that $|G(\psi; F) - G| < \varepsilon$ and $|R(\psi; F) - R| < \varepsilon$.*

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 one (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, Laffont and Sharkey, 1997, and Rogerson, 2003). Building on the observation for uniform distributions, we show the following.

Corollary 5.1. *Fix a distribution F satisfying the conditions of the model set-up.*

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

2. *If $\frac{F(\beta)}{f(\beta)}$ is convex, and if $\mathbb{E}\left[\tilde{\beta}|\tilde{\beta} \leq \beta\right] \leq \frac{\underline{\beta}+\beta}{2}$ 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 $\mathbb{E}\left[\tilde{\beta}|\tilde{\beta} \leq \beta\right] < \frac{\underline{\beta}+\beta}{2}$ for all $\beta \in (\underline{\beta}, \bar{\beta}]$.*

Part 1 of this result implies that, if F is symmetric, while F/f is concave, then the infimum of the expected gains from incentives for a given level of agent rents is less than these rents. The result is also informative about asymmetric distributions. For instance, provided F/f is concave, the mean of the innate costs being above the midpoint $\frac{\underline{\beta}+\bar{\beta}}{2}$ is sufficient to conclude $G^{\inf}(R) \leq R$. The condition is thus 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, 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 Equation (1).

Next, to understand better when Corollary 5.1 applies, consider when F/f is convex or concave. Assuming for a moment that F is thrice differentiable, we have that F/f is strictly concave over $[\underline{\beta}, \bar{\beta}]$ if, for all β,

$$f'(\beta) > \frac{F(\beta)}{f(\beta)^2}\left(2f'(\beta)^2 - f''(\beta) f(\beta)\right)$$

while F/f is strictly convex when the reverse inequality holds. Mierendorff (2016) discusses the convexity/concavity of $(1-F)/f$ and gives an analogous condition. Evaluating this condition permits one to verify the following examples.

Example 1. Let $k \in (0,1)$ and suppose $0 < \underline{\beta} < \bar{\beta}$. The distribution with cdf $F(\beta) = (1-k)\frac{\beta-\underline{\beta}}{\bar{\beta}-\underline{\beta}} + k\frac{\left(\beta-\underline{\beta}\right)^2}{\left(\bar{\beta}-\underline{\beta}\right)^2}$ satisfies the conditions of Part 1 of Corollary 5.1. The inequality is strict; i.e., $G^{\inf}(R) < R$ for all $R \in (0, \bar{R})$.

Example 2. Let $k \in (0,1)$ and suppose $0 < \underline{\beta} < \bar{\beta}$. The distribution with cdf $F(\beta) = (1-k)\frac{\beta-\underline{\beta}}{\bar{\beta}-\underline{\beta}} + k\frac{\left(\bar{\beta}-\underline{\beta}\right)^2-\left(\bar{\beta}-\beta\right)^2}{\left(\bar{\beta}-\underline{\beta}\right)^2}$ satisfies the conditions of Parts 2 of Corollary 5.1. The inequality is strict; i.e., $G^{\inf}(R) > R$ for all $R \in (0, \bar{R})$.

A related question 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 3. Consider innate cost distributions with cdf $F(\beta) = \frac{\left(k\left(\beta-\underline{\beta}\right)\right)^{1/k}}{\left(k\left(\bar{\beta}-\underline{\beta}\right)\right)^{1/k}}$ for $k > 0$. The distribution F satisfies all our conditions, and $\frac{F(\beta)}{f(\beta)} = k\left(\beta - \underline{\beta}\right)$, so that $h(\beta) = k$. Therefore $G^{\inf}(R) = kR$ for $R \in \left(0, \bar{R}\right)$; this can be taken arbitrarily large or small with k.

The intuition for Example 3 is much the same as the one provided above in relation to Corollary 5.1, Part 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.

The notion that firms will be ceded little rent when the distribution of the innate cost is concentrated towards lower values perhaps has some support in the empirical literature. Wolak (1994) and Brocas, Chan and Perrigne (2006) find the distribution of the productivity of firms (here, regulated water utilities) is left-skewed; i.e. there are many fairly efficient firms and a tail of a few inefficient ones.[13] Brocas, Chan and Perrigne suggest that the regulator (in their case, the California Public Utilities Commission) "tends to be cautious in the rents

[13] A similarly skewed distribution is found by Gagnepain and Ivaldi (2002) for urban transportation contracts.

given to firms". Our result suggests that this is a robust feature of optimal contracting. In particular, our findings suggest that the principal would extract a relatively large share of the surplus in such cases, robustly across different specifications for agent effort preferences.

5.1 Application to managerial compensation

We now discuss how our ideas can be extended models of managerial compensation, observing that the LT model has been put to use in such settings; see Edmans and Gabaix (2011), Edmans, Gabaix, Sadzik and Sannikov (2012), Garrett and Pavan (2012, 2015), and Carroll (2016). The agent is a manager of the firm who works to generate high output, or cash flows, for the firm. To illustrate, let the agent's "type" or "innate productivity" θ be drawn from a distribution F, say on $[\underline{\theta}, \bar{\theta}]$, with $0 < \underline{\theta} < \bar{\theta} < +\infty$. This can be taken to satisfy all our regularity conditions, except we require $1 - F$ to be strictly log concave, rather than F. The cash flow is given by $\pi = \theta + e$, where e is agent effort. The cash flow π is observable and contractible, though the effort e and type θ are agent private information.

The agent's payoff is $y - \psi(e)$, where y is the transfer and ψ is a disutility function satisfying the same conditions as in our version of the LT model above. The principal's payoff is $\pi - y$.

The principal (the firm, or its board) can be viewed as choosing an incentive-compatible direct mechanism which asks the agent to generate observable cash flow $\pi\left(\hat{\theta}\right)$ and pays the agent $y\left(\hat{\theta}\right)$ in case the target is met. The agent, after learning θ, has the option to reject the contract and earn payoff zero or accept it, report $\hat{\theta}$, and then choose an effort $e \in \mathbb{R}$.

The structure of optimal contracts in this setting is analogous to those for the LT model, and the derivation follows the same lines. Our approach can then be applied almost directly to this setting.

6 Literature review

This paper relates to several active literatures in contract theory and mechanism design. Our focus on the infimum of the "expected gains from incentives" for each level of agent rent is evocative of the developing literature on robustness in incentive contracts. For instance, Hurwicz and Shapiro (1978) studied a moral hazard problem in which agent disutility of effort is ambiguous to the principal, but drawn from a class of quadratic disutilities. They show that a 50/50 split of output between the principal and agent maximizes the infimal value of an "efficiency" measure, which is the ratio of the principal's realized performance to the payoff under knowledge of the disutility. In a dynamic context, Chassang (2013) similarly motivates linear contracts for a regret-based criterion. Rogerson (2003) and Chu and Sappington (2007) employ regret-type criteria to assess the performance of certain simple procurement contracts (the benchmark here is the "fully-optimal" contract, as opposed to the simple contract). Other work, such as Garrett (2014), Carroll (2015), and Dai and Toikka (2017), provided a different rationale for simple incentive contracts, by exhibiting settings in which such contracts maximize the principal's worst-case payoff, where the worst case for the principal is taken again over information that the principal does not know. That is, the contracts are optimal for a principal that is ambiguity averse.

Of course, the objective of the present paper is quite different from the earlier robustness analyses of incentive contracts, because the Bayesian principal maximizes her expected payoff (i.e., minimizes the total expected procurement cost). We are concerned here with drawing robust implications for the payoffs that emerge from such contracting. Nonetheless, there are inherently similarities in the proof approach. In particular, Garrett (2014) considers a principal who *does not know* the agent's disutility function, and knows only a broad feasible set for the possibilities. He shows that a simple incentive scheme is max-min optimal. One can view "adversarial Nature" as determining, for each proposed incentive scheme, a disutility function that yields a high total procurement cost for the principal. In the present paper, "adversarial Nature" can again be viewed as playing a role, but this time in generating disutility functions that permit the principal only a small share of the surplus in an optimal mechanism.

A further connection to the existing literature on robustness in incentive contracts 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 it is noted that linear contracts can guarantee the principal a payoff that is proportional to the agent's rents. Our 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.

As noted in the Introduction, our analysis is related to work on "robust predictions" by an analyst ignorant of key details of an interaction. To illustrate further the connection to this work, consider Bergemann, Brooks and Morris (2015) on the limits of price discrimination. In the language introduced above, their paper posits an analyst who wants to understand the welfare implications of third-degree price discrimination by a monopolist. The analyst shares the same view of the marginal distribution over buyer values as the monopolist, but does not know the additional information the monopolist has on demand in identifiable sub-markets (or even what these sub-markets might be). Their result is a characterization of all possible values of producer and consumer surplus under optimal third-degree price discrimination by the monopolist. The parallel between their paper and the present one is that the present analysis seeks to evaluate welfare implications over feasible cost-reduction technologies, while positing optimal contracting by the principal, whereas their analysis considers all feasible "segmentations" of demand into different markets, positing optimal price discrimination by the monopolist.

Finally, our work relates to 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 to draw implications from a combination of weak assumptions on model primitives together with the hypothesis of optimal contracting.

7 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, the analyst 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 sizeable, 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, the paper has developed a novel approach to determining the relationship between principal and agent rents, which seems likely to be useful in other settings. An earlier working paper version showed how our approach can be extended to make payoff predictions for dynamic incentive contracts, where the agent's innate cost evolves stochastically over time. Another setting where our approach would be directly applicable is in auctioning incentive contracts (see Laffont and Tirole, 1987). More speculatively, our approach may also hold relevance for problems in public finance where agents are citizens who have different labor/leisure preferences.

References

[1] Abito, Jose Miguel, 2012, 'Welfare Gains from Optimal Pollution Regulation,' mimeo Northwestern University.

[2] Bergemann, Dirk, Benjamin Brooks and Stephen Morris, 2015, 'The Limits of Price Discrimination,' *American Economic Review*, 105, 921-957.

[3] Bergemann, Dirk, Benjamin Brooks and Stephen Morris, 2017, 'First-Price Auctions with General Information Structures: Implications for Bidding and Revenue,' *Econometrica,* 85, 107-143.

[4] Bergemann, Dirk and Stephen Morris, 2013, 'Robust Predictions in Games with Incomplete Information,' *Econometrica*, 81, 1251-1308.

[5] Bergemann, Dirk and Stephen Morris, 2016, 'Bayes correlated equilibrium and the comparison of information structures in games,' *Theoretical Economics*, 11, 487-522.

[6] Brocas, Isabelle, Kitty Chan, and Isabelle Perrigne, 2006, 'Regulation under Asymmetric Information in Water Utilities,' *American Economic Review*, 96, 62-66.

[7] Burns, Arthur, 1970, 'The Tax Court and Profit Renegotiation,' *Journal of Law and Economics*, 13, 307-326.

[8] Carbajal, Juan Carlos and Jeffrey C. Ely, 2013, 'Mechanism design without revenue equivalence,' *Journal of Economic Theory*, 148, 104-133.

[9] Carroll, Gabriel, 2015, 'Robustness and Linear Contracts,' *American Economic Review*, 105, 536-563.

[10] Carroll, Gabriel and Delong Meng, 2016, 'Robust Contracting with Additive Noise,' *Journal of Economic Theory*, 166, 586-604.

[11] Chassang, Sylvain, 2013, 'Calibrated Incentive Contracts,' *Econometrica*, 81, 1935-1971.

[12] Chu, Leon and David Sappington, 2007, 'Simple Cost-Sharing Contracts,' *American Economic Review*, 97, 419-428.

[13] Dai, Tianjiao and Juuso Toikka, 2017, 'Robust Incentives for Teams,' mimeo MIT and U Penn.

[14] Edmans, Alex and Xavier Gabaix, 2011, 'Tractability in Incentive Contracting,' *Review of Financial Studies*, 24, 2865-2894.

[15] Edmans, Alex Xavier Gabaix, Tomasz Sadzik and Yuliy Sannikov, 'Dynamic CEO Compensation,' *Journal of Finance*, 67, 1603-1647.

[16] Gagnepain, Philippe and Marc Ivaldi, 2002, 'Incentive regulatory policies: the case of public transit systems in France,' *RAND Journal of Economics*, 33, 605-629.

[17] Garrett, Daniel, 2014, 'Robustness of Simple Menus of Contracts in Cost-Based Procurement,' *Games and Economic Behavior*, 87, 631-641.

[18] Garrett, Daniel and Alessandro Pavan, 2012, 'Managerial Turnover in a Changing World,' *Journal of Political Economy*, 120, 879-925.

[19] Garrett, Daniel and Alessandro Pavan, 2015, 'Dynamic Managerial Compensation: A Variational Approach,' *Journal of Economic Theory*, 159, 775-818.

[20] 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.

[21] Hurwicz, Leonid and Leonard Shapiro, 1978, 'Incentive Structures Maximizing Residual Gain under Incomplete Information,' *Bell Journal of Economics*, 9, 180-191.

[22] Laffont, Jean-Jacques and Jean Tirole, 1986, 'Using Cost Observation to Regulate Firms,' *Journal of Political Economy*, 94, 614-641.

[23] Laffont, Jean-Jacques and Jean Tirole, 1987, 'Auctioning Incentive Contracts,' *Journal of Political Economy*, 94, 921-937.

[24] Lazear, Edward P., 2000, 'Performance Pay and Productivity,' *American Economic Review*, 90, 1346-1361.

[25] Mengistae, Taye and Lixin Xu, 2004, 'Agency Theory and Executive Compensation: The Case of State-Owned Enterprises,' *Journal of Labor Economics*, 22, 615-637.

[26] Milgrom, Paul and Ilya Segal, 2002, 'Envelope Theorems for Arbitrary Choice Sets,' *Econometrica*, 70, 583-601.

[27] Perrigne, Isabelle and Quang Vuong, 2011, 'Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard,' *Econometrica*, 79, 1499-1539.

[28] Rogerson, 2003, 'Simple menus of contracts in cost-based procurement and regulation,' *American Economic Review*, 93, 919-926.

[29] Scott, Graham, Peter Bushnell and Nikitin Sallee, 1990, 'Reform of the Core Public Sector: New Zealand Experience,' *Governance*, 3, 138-167.

[30] Segal, Ilya and Michael Whinston, 2003, 'Robust Predictions for Bilateral Contracting with Externalities,' *Econometrica*, 71, 757-791.

[31] Strausz, Roland, 2006, 'Deterministic versus stochastic mechanisms in principal-agent models,' *Journal of Economic Theory*, 128, 306-314.

[32] Topkis, Donald, 1978, 'Minimizing a Submodular Function on a Lattice,' *Operations Research*, 26, 305-321.

[33] Wolak, Frank A., 1994, 'An Econometric Analysis of the Asymmetric Information, Regulator-Utility Interaction,' *Annales d'Economie et de Statistique*, 34, 13-69.

A Appendix: Proofs of all results

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

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

where $e(\beta) = \beta - C(\beta)$ for all β, and where VG is given by (3).

Proof. Let the agent of type β 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 ψ is assumed Lipschitz continuous, $\psi(\beta - C)$ is equi-Lipschitz continuous in β across $C \in \mathcal{C}$, with the Lipschitz constant the same as for ψ. This ensures the satisfaction of Assumption A3 of Carbajal and Ely. Note that satisfaction of their Conditions A1-A2 is immediate.[14]

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 ψ. 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

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

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

is nonempty. $S(\beta)$ is single-valued in case the above limits are equal at $(C(\beta), \beta)$, and a closed interval of positive length otherwise. By convexity of ψ, $\bar{d}v(-C, \beta)$ and $\underline{d}v(-C, \beta)$ are non-increasing in (C, β); 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 β 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 $\beta - C\left(\hat{\beta}\right)$ 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 $V(\beta) = y(\beta) - \psi(\beta - C(\beta)) = \max_{\hat{\beta} \in [\underline{\beta}, \bar{\beta}]} \left\{ y\left(\hat{\beta}\right) - \psi\left(\beta - C\left(\hat{\beta}\right)\right) \right\}$. Since A1-A3 and M of Carbajal and Ely are satisfied, Theorem 1 of their paper applies. Hence, for any $\beta \in [\underline{\beta}, \bar{\beta}]$,

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

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 β (i.e., equal to the upper bound for S), and by setting $V\left(\bar{\beta}\right) = 0$ (since individual rationality requires $V\left(\bar{\beta}\right) \geq 0$). In order for an agent of type β 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

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

as desired. Q.E.D.

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

By the Inada condition, for each $\beta \in \left[\underline{\beta}, \bar{\beta}\right]$, there exists $u > 0$ such that $VG(e, \beta) < 0$ for all $e < 0$ and all $e > u$. Note that, because ψ is convex, the left derivative of ψ, i.e. $\partial_{-}\psi$, is left-continuous and non-decreasing. Hence $VG(\cdot, \beta)$ is upper semi-continuous for all β. This means that the maximizers $E^*(\beta) \equiv \arg\max [VG(e, \beta)]$ are non-empty and closed for each β. Since $F(\beta)/f(\beta)$ is increasing, standard monotone comparative statics arguments (see Topkis, 1978) imply that $E^*(\beta)$ is non-increasing in the strong set order. We can then consider monotone (non-increasing) 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 which 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 β (hence $C^*(\cdot)$ is non-decreasing). 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 $U\left(\beta, \hat{\beta}\right)$ be the payoff obtained by type β when reporting $\hat{\beta}$ and choosing effort to meet the cost target. We have

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

The third equality follows using 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 non-decreasing functions. Given that the agent finds it optimal to meet the cost target $C\left(\hat{\beta}\right)$ 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 one.*

Proof of Lemma A.2. Let $e^{\min}\left(\underline{\beta}\right)$ be the minimal element of $E^*\left(\underline{\beta}\right)$. Note that $[\partial_-\psi]\left(e^{\min}\left(\underline{\beta}\right)\right) \leq 1$; if $[\partial_-\psi]\left(e^{\min}\left(\underline{\beta}\right)\right) > 1$, effort can be reduced from $e^{\min}\left(\underline{\beta}\right)$ while increasing surplus, contradicting the definition of $e^{\min}\left(\underline{\beta}\right)$. In addition, $[\partial_-\psi](e) < 1$ for all $e < e^{\min}\left(\underline{\beta}\right)$. Given the first claim and convexity of ψ, the only way this can fail to be true is if $[\partial_-\psi]\left(e^{\min}\left(\underline{\beta}\right)\right) = [\partial_-\psi](e) = 1$ for some $e < e^{\min}\left(\underline{\beta}\right)$. However, in this case, ψ is linear on $\left[e, e^{\min}\left(\underline{\beta}\right)\right]$ with gradient equal to one, contradicting that $e^{\min}\left(\underline{\beta}\right)$ is the minimum of the efficient effort choices.

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

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

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

We next determine further properties of optimal effort policies.

Lemma A.3. *Optimal effort $e^{*}\left(\cdot\right)$ is essentially unique and essentially non-increasing.*

Proof of Lemma A.3. First, consider why any selection from optimal effort policies E^{*} must be non-increasing (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 $e^{*}\left(\beta'\right) < e^{*}\left(\beta''\right)$. From the previous lemma, $[\partial_{-}\psi]\left(e^{*}\left(\beta''\right)\right) < 1$, and hence, since ψ is convex, we conclude that $e^{*}\left(\beta''\right) - \psi\left(e^{*}\left(\beta''\right)\right) > e^{*}\left(\beta'\right) - \psi\left(e^{*}\left(\beta'\right)\right)$. Hence, if $[\partial_{-}\psi]\left(e^{*}\left(\beta''\right)\right) = [\partial_{-}\psi]\left(e^{*}\left(\beta'\right)\right)$, $e^{*}\left(\beta'\right)$ does not maximize the virtual gains $VG\left(e,\beta'\right)$. Suppose then that $[\partial_{-}\psi]\left(e^{*}\left(\beta''\right)\right) > [\partial_{-}\psi]\left(e^{*}\left(\beta'\right)\right)$, and note

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

because $e^*(\beta')$ maximizes virtual gains $VG(e,\beta')$. 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 $VG(e,\beta'')$. 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 $\mathbb{R}$. Q.E.D.

This completes the proof of Proposition 3.1. Q.E.D.

Proof of Proposition 4.1. Step 1: $\{P(\gamma) : \gamma \in \Gamma\} =co\{P(\gamma_x) : \ x \in [\underline{\beta}, \bar{\beta}]\}$. 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 γ_x is in Γ, this point must reside in $\{P(\gamma) : \gamma \in \Gamma\}$; i.e.,

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

Conversely, any point $P(\gamma)$, $\gamma \in \Gamma$, can be approximated arbitrarily closely by points $P(\gamma^k)$, with γ^k being right-continuous step functions and hence convex combinations of the step functions γ_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 $\{P(\gamma) : \gamma \in \Gamma\} \subset \text{co}\{P(\gamma_x) : \ x \in [\underline{\beta}, \bar{\beta}]\}$, which implies the result.

Step 2: Z^* strictly increasing and convex. That Z^* is strictly increasing and convex follows immediately from observing that $(R, Z^*(R))$, 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 γ^* described by the cut-offs β_l and β_u follows because points on the lower boundary of $\text{co}\left\{P\left(\gamma_x\right):\ x \in \left[\underline{\beta}, \bar{\beta}\right]\right\}$ can be written as convex combinations of $P\left(\gamma_x\right)$ 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 $\left(R, Z^*\left(R\right)\right)$. This point belongs to the intersection of $\text{co}\left\{P\left(\gamma_x\right):\ x \in \left[\underline{\beta}, \bar{\beta}\right]\right\}$ 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 which can be written as a convex combination of the step functions γ_x for two values of x. Q.E.D.

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

We consider here the case where the cutoffs introduced in Proposition 4.1 satisfy $\underline{\beta} < \beta_l < \beta_u \leq \bar{\beta}$. Hence, there is an interval on which $\gamma^*\left(\beta\right) = 1$, an interval on which $\gamma^*\left(\beta\right) = \gamma^{\text{mid}}$ for $\gamma^{\text{mid}} \in (0,1)$, and possibly an interval on which $\gamma^*\left(\beta\right) = 0$. The remaining cases are where $\beta_l = \beta_u$ (so there is no interval on which $\gamma^*\left(\beta\right) = \gamma^{\text{mid}}$) and where $\beta_l = \underline{\beta}$ (so there is no interval on which $\gamma^*\left(\beta\right) = 1$), and these are treated in the main text. (All possible cases are given by thresholds β_l and β_u satisfying $\underline{\beta} \leq \beta_l \leq \beta_u \leq \bar{\beta}$, with either $\underline{\beta} = \beta_l$ or $\beta_l = \beta_u$, but not both.)

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

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

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

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

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

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

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

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

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

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 ψ_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

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

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

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

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

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, in order to approach the boundary of $\mathcal{U}$). Let $a, \varepsilon > 0$, with ε less than the aforementioned right derivative. Then consider the disutility function $\bar{\psi}(e; a, \varepsilon)$ as defined in Equation (5). Given this disutility function, the principal's virtual gains for innate cost β 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 ε is small enough, optimal effort for the disutility $\bar{\psi}(e; a, \varepsilon)$ is at least a for all β. Also, if the agent with innate cost β takes effort $\check{e} > 0$ in the optimal policy for disutility function ψ, 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., $[\partial_{-}\psi](\check{e}) = \left[\partial_{-}\bar{\psi}(\cdot; a, \varepsilon)\right](\check{e} + a)$). The (left) marginal disutility of effort for disutility function $\bar{\psi}(e; a, \varepsilon)$ is ε whenever the agent takes effort a. Also, the measure of β 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 ε is taken to zero. The agent's expected rents are either the same as under ψ (for instance, if the agent takes positive effort with probability one under ψ), or approach the value under ψ from above as $\varepsilon \to 0$. Q.E.D.

Proof of Corollary 5.1. First consider Part 1, and hence suppose $\frac{F(\beta)}{f(\beta)}$ is concave and $\mathbb{E}\left[\tilde{\beta}\right] \geq \frac{\underline{\beta}+\bar{\beta}}{2}$. Then, as observed in the main text, a solution to Problem I is $\gamma^*(\beta) = \frac{\underline{R}}{R}$ for all $\beta \in \left[\underline{\beta}, \bar{\beta}\right]$. 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 $\mathbb{E}\left[\tilde{\beta}\right] > \frac{\underline{\beta}+\bar{\beta}}{2}$.

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

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

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 $\mathbb{E}\left[\tilde{\beta}\right] \geq \frac{\underline{\beta}+\bar{\beta}}{2}$, and the inequality is strict if $\mathbb{E}\left[\tilde{\beta}\right] > \frac{\underline{\beta}+\bar{\beta}}{2}$. Also, $\frac{F(\beta)}{f(\beta)}$ and $\frac{\beta-\underline{\beta}}{f(\bar{\beta})(\bar{\beta}-\underline{\beta})}$ are functions taking 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 1 of the corollary therefore follows.

Now consider Part 2, and hence suppose $\frac{F(\beta)}{f(\beta)}$ is convex and $\mathbb{E}\left[\tilde{\beta}|\tilde{\beta} \leq \beta\right] \leq \frac{\underline{\beta}+\beta}{2}$ for all $\beta \in (\underline{\beta}, \bar{\beta}]$. For a given value $R \in (0, \bar{R})$, there is a solution γ^* 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, $\mathbb{E}_{\bar{F}}\left[\tilde{\beta}\right] \leq \frac{\underline{\beta}+\beta^*}{2}$. Hence, considering the expression in Equation (9) evaluated for the distribution $\bar{F}$, with upper limit of the support β^*, 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 $\mathbb{E}_{\bar{F}}\left[\tilde{\beta}\right] < \frac{\underline{\beta}+\beta^*}{2}$. This establishes the result. Q.E.D.