Robustness and Linear Contracts

Gabriel Carroll, Stanford University
gdc@stanford.edu

July 10, 2014

Abstract

We consider a moral hazard problem where the principal is uncertain what the agent can and cannot do: She knows some actions available to the agent, but other, unknown actions may also exist. The principal demands robustness, evaluating possible contracts by their worst-case performance, over unknown actions the agent might potentially take. The model assumes risk-neutrality and limited liability, and no other functional form assumptions. Very generally, the optimal contract is linear. The model thus offers a new explanation for linear contracts in practice. It also introduces a flexible modeling approach for moral hazard under non-quantifiable uncertainty.

Keywords: limited liability, linear contracts, principal-agent problem, robustness, worst-case

Thanks to (in random order) Larry Blume, Daron Acemoglu, Luis Zermeño, Paul Milgrom, Drew Fudenberg, Iván Werning, Alex Wolitzky, Mike Riordan, Arun Chandrasekhar, Lucas Maestri, David Easley, Ben Golub, Alex Frankel, Sylvain Chassang, Petra Persson, Abhijit Banerjee, Bengt Holmström, and Tomasz Sadzik for helpful discussions and advice, as well as several referees, and other seminar participants at Zurich, UIUC, the Minneapolis Fed, NYU, WashU, UCLA, Toulouse, Caltech, Duke, Wisconsin, Columbia, and Harvard EconCS for their valuable comments.

1 Introduction

Imagine a principal who contracts with an agent, but who has only limited knowledge of what the agent can and cannot do. She wants to write a contract that is robust to this uncertainty. What should such a contract look like?

In our model, as in standard moral hazard models, the agent takes an unobserved costly action, which produces a stochastic output. The principal gives incentives by paying the agent based on observed output. She wishes to maximize the expected value of output minus the wage paid out. But unlike in most of the literature, our principal does not know exactly what actions the agent can take. She knows of some available actions; but other, unknown actions may also exist, and our principal does not even have a prior belief about these unknown actions. In this nonprobabilistic setting, we assume a simple (arguably the simplest) criterion to evaluate contracts: Any contract is judged by its worst possible performance, given the principal's knowledge. The principal and agent are financially risk-neutral, and payments are constrained by limited liability.

One way the principal can obtain a worst-case payoff guarantee is to use a linear contract — paying the agent a fixed share of output. For example, suppose the principal considers a contract that pays the agent one-third of output, keeping two-thirds for herself, and suppose she knows some action the agent can take that would give him an expected payoff of 400 under this contract. Then, any unknown action the agent might rationally choose would also give him at least 400. Since the principal's ex-post payoff is always at least twice the agent's, she thus is guaranteed at least 800.

Besides linear contracts, many other contracts can also provide positive guarantees. But the main finding of this paper is that the *best* such guarantee, out of all possible contracts, comes from a linear contract. This result holds without any assumptions on the structure of the set of known actions. It also persists — with suitable modifications — through a number of extensions and variations of the basic model.

Briefly, the intuition is as follows. When the principal proposes a contract, in the face of her uncertainty about the agent's technology, she knows very little about what will happen; but the one thing she does know is a lower bound on the agent's expected payoff (from the actions that are known to be available). The only effective way to turn this into a lower bound on her own expected payoff is via a linear relationship between the two, as in the example above. Even when a contract is nonlinear, whatever guarantee it gives is still driven by a linear relationship, which in general is an inequality. Linear contracts are the ones for which this relationship is tight, and this is why they are optimal.

The importance of our finding can be viewed in three different ways. First, it addresses a longstanding problem in contract theory: Why are linear contracts common in practice — while textbook models often predict more complicated functional forms? As Holmström and Milgrom write in their classic paper on linear contracts in dynamic environments [15,

p. 326]:

> It is probably the great robustness of linear rules based on aggregates that accounts for their popularity. That point is not made as effectively as we would like by our model; we suspect that it cannot be made effectively in any traditional Bayesian model. But issues of robustness lie at the heart of explaining any incentive scheme which is expected to work well in practical environments.

This paper aims to answer their implicit call with a forthrightly non-Bayesian model of robustness.

The second view of our contribution is that it provides concrete advice to people faced with the practical task of designing incentive contracts under non-quantifiable uncertainty. And, third, it adds to the arsenal of tools for analyzing agency problems, offering a new and flexible modeling framework that can be used to make more complex moral hazard problems tractable.

Mathematically, the main result of this paper is rather simple. This makes it all the more surprising that it did not appear much earlier in the agency theory literature. There have been results on optimality of linear contracts using other maxmin-type criteria, due to Hurwicz and Shapiro [16] and recently Chassang [4, Corollary 1]. Diamond [8] also gave a Bayesian model in which related intuitions apply. However, the present paper offers a relatively general class of environments, together with a mathematical argument for robustness based on the alignment between the principal's and agent's goals, that differentiate it from previous literature. (The connections with these previous works will be discussed in more detail in the concluding section.)

Section 2 of the paper formally presents the basic version of the model and result. The model is kept as simple and clean as possible here. Section 3 then shows how the logic of the result persists under various extensions that either remedy unrealistic features of the basic model or otherwise enrich it. This includes assuming some knowledge about the costs of various actions, replacing the single set of known actions with multiple possible minimal sets of actions, and allowing a participation constraint, as well as allowing the principal to screen agents by their potential actions. These extensions also illustrate how the method extends beyond the basic model.

This paper joins a recently growing literature exploring mechanism design with worst-case objectives. This includes the work of Hurwicz and Shapiro [16] mentioned above, Frankel [11], and Garrett [12], also on contracting with unknown agent preferences; the work initiated by Bergemann and Morris [1] and Chung and Ely [7] on mechanism design with unknown higher-order beliefs; and work such as Yamashita's [24] on maxmin expected welfare under weak assumptions on agent behavior (in this case, assuming only that agents play undominated strategies). A broader mechanism design literature provides nearly optimal worst-case performance in various settings, without optimizing exactly; recent examples include the work of Chassang [4], Segal [21], Chawla, Hartline, Malec, and Sivan [5], and Micali and Valiant [20]. There is also a less closely related strand of literature, such as Madarász and Prat [18], that looks at local robustness when the model of the environment is slightly misspecified.

This paper also adds to the literature on explanations for linear contracts — including the maxmin-optimality papers mentioned above as well as several others. Again, discussion of the relationship to that literature is deferred to the concluding section. The conclusion also gives some discussion of interpretation and how to connect the stark assumptions of the model to real-world contract design.

2 The basic model

We start with the basic version of the model. The model here is kept simple, at some costs of realism, which will be addressed later.

2.1 Notation

We write $\Delta(X)$ for the space of Borel distributions on $X \subseteq \mathbb{R}^k$, equipped with the weak topology. For $x \in X$, δ_x is the degenerate distribution putting probability 1 on x. $\mathbb{R}^+$ is the set of nonnegative reals.

2.2 Setup

A principal contracts with an agent, who is to take a costly action that leads to a stochastic output. The action is not observable to the principal; only the resulting output, y, is observable. Thus, payment to the agent can depend only on y, and this dependence is what provides incentives. Both parties are financially risk-neutral.

We write Y for the set of possible output values, and assume Y is a compact subset of $\mathbb{R}$. Y may be finite or infinite. We normalize $\min(Y) = 0$.

To model the agent's actions, we abstract away from their physical description and record only the features that affect behavior and payoffs: the cost of each action to the agent, and the resulting probability distribution over output. Thus, an *action* is a pair $(F, c) \in \Delta(Y) \times \mathbb{R}^+$. The interpretation is that the agent pays cost c, and output is drawn $y \sim F$. c may be interpreted literally as a monetary cost, or an additive disutility of effort. We give $\Delta(Y) \times \mathbb{R}^+$ the natural product topology.

A *technology* is a compact subset of $\Delta(Y) \times \mathbb{R}^+$, describing a possible set of actions available to the agent. The agent has a technology $\mathcal{A}$, which he knows but the principal does not. Instead, the principal knows only some set $\mathcal{A}_0$ of actions available to the agent, and she believes $\mathcal{A}$ may be any technology such that $\mathcal{A}_0 \subseteq \mathcal{A}$.

The exogenous $\mathcal{A}_0$ may be any technology, subject to the following *nontriviality* assumption: There exists $(F, c) \in \mathcal{A}_0$ such that $E_F[y] - c > 0$. This assumption ensures that the principal benefits from hiring the agent.

It is natural to assume that the agent can always exert no effort; this corresponds to assuming $(\delta_0, 0) \in \mathcal{A}_0$. However our results will not require this assumption. Also, we say $\mathcal{A}_0$ satisfies the *full-support condition* if, for all $(F, c) \in \mathcal{A}_0$ such that $(F, c) \neq (\delta_0, 0)$, F has full support on Y. Our main result becomes stronger when this condition holds.

Next we define the space of contracts. Any contract must specify how much the agent is paid for each level of output. We assume one-sided limited liability: the agent can never be paid less than zero. Thus, a *contract* is any continuous function $w : Y \to \mathbb{R}^+$.[1]

We can now summarize the timing of the game:

1. the principal offers a contract w;
2. the agent, knowing $\mathcal{A}$, chooses action $(F, c) \in \mathcal{A}$;
3. output $y \sim F$ is realized;
4. payoffs are received: $y - w(y)$ to the principal and $w(y) - c$ to the agent.

Describing the agent's behavior is simple, since he maximizes expected utility. Given contract w, and technology $\mathcal{A}$, the set of actions the agent is willing to choose is

$$A^*(w|\mathcal{A}) = \arg\max_{(F,c)\in\mathcal{A}} \left(E_F[w(y)] - c\right).$$

Continuity and compactness ensure this set is nonempty. It will also be useful to write

$$V_A(w|\mathcal{A}) = \max_{(F,c)\in\mathcal{A}} \left(E_F[w(y)] - c\right)$$

[1]Requiring continuity of w ensures the agent's optimization problem has a solution. If, say, Y is an arbitrarily fine discrete grid, then continuity is a vacuous assumption. Alternatively, we could relax continuity to upper semi-continuity, and all arguments would go through, with a few extra verifications.

for the agent's expected payoff. If the agent is indifferent among several actions, we assume he maximizes the principal's utility. Thus the principal's expected payoff under technology $\mathcal{A}$ is

$$V_P(w|\mathcal{A}) = \max_{(F,c)\in A^*(w|\mathcal{A})} E_F[y - w(y)].$$

Finally, we assume the principal evaluates contracts by their worst-case expected payoff, over all possible technologies $\mathcal{A}$:

$$V_P(w) = \inf_{\mathcal{A}\supseteq\mathcal{A}_0} V_P(w|\mathcal{A}).$$

Our focus is on the principal's problem, namely to maximize $V_P(w)$. In the next subsection, we will show that the maximum exists, and identify the contract that attains it.

2.3 Analysis

In the above model, the principal considers the worst case over a very wide range of technologies. Faced with this huge uncertainty, can she even guarantee herself a positive expected payoff? Yes; one simple way to get such a guarantee is to use a linear contract — one of the form $w(y) = \alpha y$ for constant $\alpha \in [0, 1]$. The argument was sketched in the introduction, and now we write it out formally. (This same calculation appears also in Chassang [4, Theorem 1].) Suppose the principal offers such a contract, with $\alpha > 0$. Note that whatever technology $\mathcal{A} \supseteq \mathcal{A}_0$ the agent has, and whatever optimal action (F, c) he chooses, his expected payment satisfies

$$E_F[w(y)] \quad \geq \quad E_F[w(y)] - c \quad = \quad V_A(w|\mathcal{A}) \quad \geq \quad V_A(w|\mathcal{A}_0). \tag{2.1}$$

Here the second inequality holds because $\mathcal{A}$ contains $\mathcal{A}_0$, and having more actions available can only make the agent better off. Now, the principal receives a fraction $1 - \alpha$ of output while the agent receives fraction α, hence their ex-post payoffs are related via

$$y - w(y) \quad = \quad \frac{1-\alpha}{\alpha}\, w(y). \tag{2.2}$$

Combining with (2.1) gives a lower bound on the principal's expected payoff:

$$E_F[y - w(y)] \quad \geq \quad \frac{1-\alpha}{\alpha}\, E_F[w(y)] \quad \geq \quad \frac{1-\alpha}{\alpha}\, V_A(w|\mathcal{A}_0).$$

Since this holds regardless of the technology,

$$V_P(w) \quad \geq \quad \frac{1-\alpha}{\alpha} V_A(w|\mathcal{A}_0). \tag{2.3}$$

The nontriviality assumption implies that if α is close to 1 then $V_A(w|\mathcal{A}_0) > 0$, and so we have a positive lower bound on the principal's worst-case payoff.

This shows how to obtain a payoff guarantee from a linear contract. But is it possible that some other, subtler contract form would give a better guarantee? The answer is no.

Theorem 2.1. *There exists a linear contract that maximizes* V_P*.*

To show this, we start from any arbitrary contract w, and show that its guarantee is driven by a linear relation between the principal's and agent's payoffs — in the same way that equation (2.2) drives the guarantee from a linear contract above; but in general the driving relation will be an *inequality.* We then construct a contract w' that satisfies the same linear relation with equality. We show that as a result, w' gives the principal the same guarantee (or better). Thus, any contract can be improved on by a linear contract.

The mechanics of the argument are depicted in Figure 1. Consider any contract $w(y)$, which may be nonlinear and may not even be monotone in y, as shown by the thick curve in panel (a). For any action (F, c) the agent may potentially take, consider the point whose coordinates are the expected output and expected payment to the agent, $(E_F[y], E_F[w(y)])$. This point evidently lies in the convex hull of the curve w — the union of the two shaded regions in panel (a). Moreover, since the agent can certainly assure himself a payoff of at least $V_A(w|\mathcal{A}_0)$, he will only take actions that pay at least this much — those corresponding to the darker shaded region. From this we can identify the worst case for the principal, call it point Q: it is the point in the dark region where her expected profit, $E_F[y] - E_F[w(y)]$, is lowest. Aside from uninteresting cases, this is where the horizontal line $V_A(w|\mathcal{A}_0)$ hits the left boundary of the convex hull, as in the figure.

Now take the support line to the convex hull at this point Q, as shown in panel (b). This line exactly delineates the driving inequality: for any action to the right of the line, if it gives the agent at least $V_A(w|\mathcal{A}_0)$ (dashed line) then the principal's resulting payoff is no worse than at Q. But we can also regard the line as being itself a contract, call it w'. w' is basically a linear contract; more precisely it is an *affine* contract, one of the form $w'(y) = \alpha y + \beta$ for constant α, β. Because w' lies above w, it indeed assures the agent a payoff at least as high as the dashed line; and because w' still satisfies the driving inequality (as an equality), it follows that the principal does no worse than Q. Therefore, the principal's worst-case guarantee is at least as good under w' as under w.

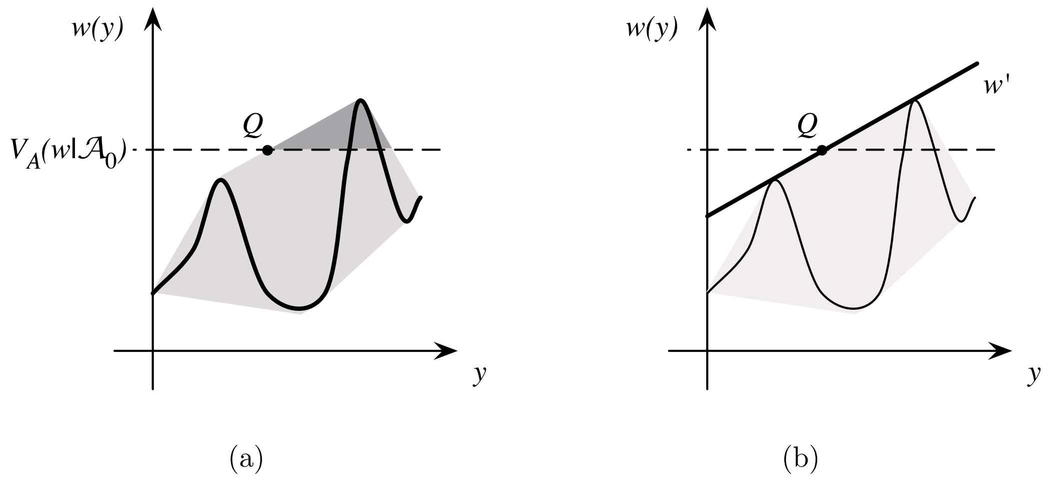

Figure 1: Sketch of the main proof: (a) Identifying the worst-case point for a given contract $w(y)$. (b) Improving w to an affine contract w'.

We now proceed to walk through each step of the argument in more detail, formally developing the argument as a series of lemmas. Some technical verifications will be left to Appendix A.

The first step is to exactly identify the guarantee $V_P(w)$ from any candidate contract w. This characterization is given by Lemma 2.2 below, which formalizes our verbal description of the worst-case point Q above: the point in the dark shaded region that gives the lowest expected payoff for the principal. However, our assumption of tie-breaking in favor of the principal introduces some technicalities.

One technicality is that we need to deal with the zero contract ($w(y) = 0$ for all y) separately. We abusively denote this contract by 0. Suppose there exists $(F, c) \in \mathcal{A}_0$ with $c = 0$; that is, the agent can definitely produce some output costlessly. Then the agent is willing to take any such action, so the principal's guarantee is simply the highest value of $E_F[y]$ over such F. That is, $V_P(0) = \max_{(F,0)\in\mathcal{A}_0} E_F[y]$. If there is no action $(F, 0) \in \mathcal{A}_0$, then the principal is not guaranteed any positive payoff: if it turns out the agent can either take actions in $\mathcal{A}_0$ or produce 0 output at cost 0, he will choose the latter; hence, $V_P(0) = 0$.

Another minor technicality is that the worst-case payoff $V_P(w)$ may be approached, but not actually attained for any technology. (This is why we defined it as an infimum over $\mathcal{A}$, and not a minimum.)

Also, to avoid some extra cases, we say a contract w is *eligible* if $V_P(w) \geq V_P(0)$

and $V_P(w) > 0$. From our observations so far, we know some eligible contract exists — indeed, an eligible linear contract exists: $w(y) = \alpha y$, for α close to 1, is eligible, unless $V_P(w) < V_P(0)$ in which case the zero contract is eligible. Therefore, in our search for optimality, we may restrict ourselves to eligible contracts.

All these points duly addressed, we state the characterization of the principal's guarantee:

Lemma 2.2. *Let w be any eligible contract, different from the zero contract. Then,*

$$V_P(w) = \min E_F[y - w(y)] \quad \text{over } F \in \Delta(Y) \text{ such that } E_F[w(y)] \geq V_A(w|\mathcal{A}_0). \tag{2.4}$$

Moreover, for any F attaining the minimum, the constraint holds with equality: $E_F[w(y)] = V_A(w|\mathcal{A}_0)$.

The proof is straightforward and left to Appendix A.

Note that the equality statement in Lemma 2.2 implies that (2.3), the guarantee of a linear contract, is actually an equality. We record this as a separate lemma:

Lemma 2.3. *For any $\alpha \in (0, 1]$, if the linear contract $w(y) = \alpha y$ is eligible, then*

$$V_P(w) = \frac{1-\alpha}{\alpha} V_A(w|\mathcal{A}_0) = \max_{(F,c)\in\mathcal{A}_0} \left((1-\alpha)E_F[y] - \frac{1-\alpha}{\alpha} c \right). \tag{2.5}$$

This is also valid for $\alpha = 0$, if we interpret the term $-\frac{1-\alpha}{\alpha}c$ as being 0 for $c = 0$ and $-\infty$ for $c > 0$.

The next step, and the heart of the argument, is to identify the linear inequality relating the principal's and agent's payoffs that drives the guarantee of contract w.

Lemma 2.4. *Let w be any eligible contract, different from the zero contract. Then there exist numbers κ, λ, with $\lambda > 0$, such that*

$$y - w(y) \geq \kappa + \lambda\, w(y) \qquad \text{for all } y \in Y; \tag{2.6}$$

$$V_P(w) = \kappa + \lambda\, V_A(w|\mathcal{A}_0). \tag{2.7}$$

To be clear about why we say (2.6) drives the guarantee of contract w, simply consider any action (F, c) the agent might take under any technology, and apply expectations:

$$E_F[y - w(y)] \geq \kappa + \lambda\, E_F[w(y)] \geq \kappa + \lambda\, V_A(w|\mathcal{A}_0) \tag{2.8}$$

where the second inequality holds as in (2.1). But by (2.7), the right side of (2.8) is exactly equal to $V_P(w)$. So essentially, starting from (2.6) and taking expectations is all it takes to show that the principal is indeed guaranteed at least $V_P(w)$.

Inequality (2.6) carves out a half-plane whose bounding line is the support line shown in Figure 1. In the formal proof of the lemma, we identify this support line by an application of the separating hyperplane theorem. The earlier characterization of the worst-case point (Lemma 2.2) plays an essential role by showing that this point lies on the boundary of the convex hull, so that the support line exists.

The proof of Lemma 2.4 requires a little extra work (checking that λ is well-defined and positive), so we leave it to Appendix A.

Now we are ready to define our improved contract w'. We rearrange (2.6) as

$$w(y) \leq \frac{1}{1+\lambda}\, y - \frac{\kappa}{1+\lambda}. \tag{2.9}$$

This leads us to define

$$w'(y) = \frac{1}{1+\lambda}\, y - \frac{\kappa}{1+\lambda}, \tag{2.10}$$

an affine contract. We have $w' \geq w$ pointwise. Notice that this immediately implies $w'(y) \geq 0$ for all y, so that w' is indeed a contract.

The next step shows how the driving inequality ensures a guarantee for the principal from w' that is at least as good as the guarantee from w.

Lemma 2.5. *Suppose that w satisfies (2.6) and (2.7). Then the contract w' defined by (2.10) satisfies $V_P(w') \geq V_P(w)$.*

Proof. Rearrange the definition (2.10) to see that w' satisfies the driving inequality (2.6) (as an equality). So repeating the argument from (2.8) above, we get

$$E_F[y - w'(y)] \quad \geq \quad \kappa + \lambda\, V_A(w'|\mathcal{A}_0) \tag{2.11}$$

for any action (F, c) the agent might potentially choose under any technology. Thus the principal is guaranteed at least this amount. But since $w' \geq w$ everywhere, the agent is certainly at least as well off under w' as under w: $V_A(w'|\mathcal{A}_0) \geq V_A(w|\mathcal{A}_0)$. Applying this to the right side of (2.11), we see that by using w', the principal is guaranteed at least $\kappa + \lambda\, V_A(w|\mathcal{A}_0)$. But this equals $V_P(w)$ by (2.7). □

At this point the proof of Theorem 2.1 is almost complete. We have just a couple clean-up steps left:

Lemma 2.6. *For any affine contract w', there is a linear contract w'' that does at least as well as w': $V_P(w'') \geq V_P(w')$, with strict inequality unless w' is already linear.*

Proof. Write $w'(y) = \alpha y + \beta$, and note $\beta = w'(0) \geq 0$. Put $w''(y) = \alpha y = w'(y) - \beta$. This increases the principal's payoff by β, since a constant shift does not affect the agent's incentives for choice of action. □

Lemma 2.7. *Within the class of linear contracts, there exists an optimal one for the principal.*

Proof. Recall the formula (2.5), which was a lower bound for the guarantee of the linear contract with share α, with equality whenever the contract is eligible (which must be true for some linear contract). Since (2.5) is continuous in the share $\alpha \in [0, 1]$, it achieves a maximum, and therefore this maximum is also the optimal guarantee over all linear contracts. □

And now we can put all the pieces together.

Proof of Theorem 2.1. By Lemma 2.7, among the linear contracts, there is an optimal one, call it w^*. Then in particular w^* is eligible. If w is any other, nonlinear contract that does better than w^*, then by Lemmas 2.4, 2.5 and 2.6, there is a linear contract that in turn does at least as well as w. But this contradicts the fact that w^* is the best possible linear contract. Thus, w^* is optimal among all contracts. □

The above argument shows that there is an optimal contract that takes the linear form. However, there may potentially be many optimal contracts. If the full-support condition is satisfied, however, then we get a stronger conclusion:

Corollary 2.8. *If $\mathcal{A}_0$ satisfies the full-support condition, then* every *contract that maximizes V_P is linear.*

This just requires a couple extra verifications at certain points in the proof of Theorem 2.1, which we leave to Appendix A.

We have now reached our main goal, of showing that an optimal contract is linear, $w(y) = \alpha y$. We may as well wrap up the analysis of the principal's problem by identifying exactly what the share α is. From Lemma 2.3, the optimal share is found by maximizing

$$(1 - \alpha) E_F[y] - \frac{1 - \alpha}{\alpha} c$$

jointly over $(F, c) \in \mathcal{A}_0$ and $\alpha \in [0, 1]$. So we can first find the optimal α for any given (F, c), then maximize over actions (F, c). When $E_F[y] < c$, the maximum over α is 0 (given by $\alpha = 1$). Otherwise, the optimal α is equal to $\sqrt{c/E_F[y]}$, and the objective reduces to

$$E_F[y] + c - 2\sqrt{c\, E_F[y]} = (\sqrt{E_F[y]} - \sqrt{c})^2. \tag{2.12}$$

Therefore, the optimal contract is chosen by taking $(F^*, c^*) \in \mathcal{A}_0$ to maximize $\sqrt{E_F[y]} - \sqrt{c}$, and then choosing $\alpha^* = \sqrt{c^*/E_{F^*}[y]}$ to be the share. If it happens that there are several actions in $\mathcal{A}_0$ attaining the maximum (a knife-edge case), then there can be several optimal linear contracts.

We remark that, aside from the support-line approach taken here, there is also another, directly constructive way to show that any contract is (weakly) outperformed by a linear contract. That alternative proof is a little faster, but generalizes less readily. See Appendix C for details and discussion.

2.4 Discussion of assumptions

Before moving on to the extensions, we should comment on some assumptions, their role in the mechanics of the model and their consequences for interpretation.

The uncertainty on the principal's part is clearly essential: If the principal knew for certain that $\mathcal{A} = \mathcal{A}_0$, then the optimal contract would in general not be linear (see e.g. Diamond [8]). For example, with Y finite and $\mathcal{A}$ containing only two actions, the optimal way to incentivize the costlier action would be to pay a positive amount only for the value of output having the highest likelihood ratio, and zero for all other realizations of output. Moreover, the space of actions that may potentially be available to the agent needs to be sufficiently rich. This is needed in Lemma 2.2, which shows that the worst case for the principal lies on the boundary of the convex hull of w. If we assumed a less rich space of potential actions, the worst case might lie inside the convex hull, and then the support line would not be defined.

The limited liability assumption is also crucial. If we removed this assumption, and instead constrained payments from below by imposing a participation constraint (say, the agent must be assured a nonnegative expected payoff), then the standard solution of "selling the firm to the agent" would apply: clearly the principal could not be guaranteed any higher payoff than the total surplus under $\mathcal{A}_0$, namely $s_0 = \max_{(F,c)\in\mathcal{A}_0}(E_F[y] - c)$, and she could achieve this payoff by setting $w(y) = y - s_0$. Thus our argument depends on having some exogenous minimum payment to the agent. However, our assumption that this minimum payment is 0 is simply a normalization. One could instead assume that the minimum payment is some (positive or negative) constant $\underline{w}$, and the straightforward analogue of Theorem 2.1 would say that an optimal contract has the form $w(y) = \alpha y + \underline{w}$.[2]

Likewise, the model assumes that any action must entail a nonnegative cost to the agent. This can be relaxed modestly to allow actions that have private benefits: If we instead assume that the minimum possible cost of an action is $\underline{c} < 0$, so that an action is defined as an element of $\Delta(Y) \times [\underline{c}, \infty)$, then the resulting model is equivalent to an instance of the original model with the cost of each action translated by $-\underline{c}$. Thus, the nontriviality assumption would now require some $(F, c) \in \mathcal{A}_0$ with $E_F[y] - c > -\underline{c}$, and as long as this is satisfied, a linear contract is optimal. (Subsection 3.1 offers another variation on this theme.)

If this version of the nontriviality assumption is not satisfied, then no contract will guarantee the principal any positive payoff. Thus, the model is not suitable for describing situations where the private benefits from undesirable actions could potentially be very large, e.g. where the agent might be able to steal all the output for personal consumption.

We have also made an assumption of favorable tie-breaking — that if the agent is indifferent among actions, he chooses the best one for the principal. This may seem contrary to the worst-case spirit of the model, but it can be read as a modeling shorthand for the standard notion of a contract as consisting of both a payment rule and a recommended action. (Here the recommended action would be contingent on the technology, or we could simply imagine the blanket recommendation "break ties in favor of the principal.") Other tie-breaking rules would lead to essentially the same results, but may introduce technical complications: e.g. in some instances the optimal contract may not exist, so that $\sup_w V_P(w)$ is approached, but not attained, by linear contracts.

One more, subtler, assumption is hidden in the maxmin expected utility formulation: There is non-quantifiable uncertainty about the set of possible actions, but for any *particular* action, the risk associated with the action is quantifiable (and moreover, the principal and the agent agree about how to quantify it). One way to make sense of this combination of non-quantifiable and quantifiable uncertainty is that the risk inherent in any given action depends on physical events occurring in the world, which might be relatively familiar concepts, whereas technologies are too abstract for the principal to be able to reason probabilistically about them. We could also try to appeal to decision-theoretic foundations to justify the maxmin expected utility formulation (see e.g. [23] for references

[2]Similarly, our assumption $\min(Y) = 0$ is simply an additive normalization of the principal's payoffs; without this assumption, an optimal contract would take the form $w(y) = \alpha(y - \min(Y))$.

to several such axiomatizations), although such an appeal by itself would not explain why technology appears in the nonprobabilistic parameter space rather than the probabilistic state space.

3 Extensions

In this section we consider several variations of the basic model. The purpose is twofold: to study how the result persists when the model is made more realistic, and to show how the analytical tools extend to more complex models.

Specifically, we consider: refining the principal's knowledge by adding a lower bound on the cost of producing any given output distribution, or by otherwise changing the set of possible technologies; adding a participation constraint; and allowing the principal to screen by offering different contracts depending on the agent's technology $\mathcal{A}$. Note that these extensions are independent of each other; we do not pursue the task of writing a single model that is as general as possible.

3.1 Lower bounds on cost

An immediate criticism of the basic model is that it unrealistically allows the agent to produce large amounts of output for free. Indeed, as the proof of Lemma 2.2 (in Appendix A) shows, the worst-case action for any contract is one that produces an undesirable distribution F at cost 0. We might wish to change the model to rule this out. One way is to suppose instead that the principal knows a lower bound on the cost of producing any given level of expected output.

To model this, suppose there is given a convex function $b : \mathbb{R} \to \mathbb{R}^+$, and amend the definition of a technology $\mathcal{A}$ to require that every $(F, c) \in \mathcal{A}$ should satisfy $c \geq b(E_F[y])$. We suppose that the known technology $\mathcal{A}_0$ also satisfies this condition. We again define $V_P(w)$ as the infimum of $V_P(w|\mathcal{A})$ over all possible technologies $\mathcal{A} \supseteq \mathcal{A}_0$. Everything else is as in the original model. Then, it turns out that a linear contract is still optimal.

In fact, a significant generalization holds too. We can allow the known lower bound on cost, b, to depend not only on the expected value of output but also on other moments. (For example, it may be that producing a high level of output for certain is known to be expensive, but producing the same mean output with high variance might be less costly.) Following Holmström [14], we can also allow there to be other observable variables, besides output, that are informative about the bound on cost. The general result is that the optimal contract is an affine function of output and whatever other relevant variables are observed.

To give the general formulation, we allow for a vector of observables $z = (z_1, \dots, z_k)$, taking values in a compact set $Z \subseteq \mathbb{R}^k$. We assume output is included as one component of z, say $y = z_1$, and thus assume $\min\{z_1 \mid z \in Z\} = 0$. An action now consists of a distribution on Z and an associated cost. In our model, the principal knows a lower bound on the cost of any distribution that depends on the expected values of all the z_i. Thus, we assume given a convex function $b : \mathbb{R}^k \to \mathbb{R}^+$, and define an *action* to be a pair (F, c) with $F \in \Delta(Z)$ and $c \geq b(E_F[z])$. A *technology* is a compact set of actions. We assume given a technology $\mathcal{A}_0$, the set of known actions, and the true $\mathcal{A}$ may be any technology $\mathcal{A}_0$. We make the same nontriviality assumption as before.

A *contract* is now a continuous function $w : Z \to \mathbb{R}^+$. The timing of the game and payoffs are as before: Given contract w and technology $\mathcal{A}$, the agent's utility and his choice set are

$$V_A(w|\mathcal{A}) = \max_{(F,c)\in\mathcal{A}} (E_F[w(z)] - c), \qquad A^*(w|\mathcal{A}) = \arg\max_{(F,c)\in\mathcal{A}} (E_F[w(z)] - c);$$

the principal's expected payoff is

$$V_P(w|\mathcal{A}) = \max_{(F,c)\in A^*(w|\mathcal{A})} E_F[z_1 - w(z)].$$

The principal's objective $V_P(w)$ is defined to be the infimum of $V_P(w|\mathcal{A})$ over all technologies $\mathcal{A} \supseteq \mathcal{A}_0$.

The main result for this model is:

Theorem 3.1. *There exists a contract that maximizes V_P and is affine — that is*

$$w(z) = \alpha_1 z_1 + \cdots + \alpha_k z_k + \beta$$

for some real numbers $\alpha_1, \dots, \alpha_k$ and β.

The argument here is an extension of the ideas used for the basic model. However, the analogue of Lemma 2.4, finding the driving inequality for any candidate contract w, is now a bit subtler: we apply the separating hyperplane theorem to separate two sets, one given by the convex hull of w as before and the other determined by the shape of the function b. In addition, identifying the worst-case action for a given contract involves separately addressing a boundary case that previously applied only for the zero contract, but now can occur more widely and so requires more careful treatment. The details of the proof are deferred to Appendix B. We also give an example there to illustrate how this generalized model also serves to describe a situation where only output is observed but the cost bound depends on higher moments.

3.2 Alternative sets of technologies

The basic model assumes that the true technology might be much, much larger than the set of actions known to the principal, since any $\mathcal{A} \supseteq \mathcal{A}_0$ is considered possible. However, all of the same results hold if the principal considers a much smaller set of possible technologies $\mathcal{A}$: either $\mathcal{A}_0$ itself, or $\mathcal{A}_0$ with just one more action (F, c) added. To see this, just check that when $V_P(w)$ is redefined as the infimum of $V_P(w|\mathcal{A})$ over this restricted set of technologies, its value does not change.

In fact, we do not even need to assume that there is a single minimal technology $\mathcal{A}_0$. Here is a more general formulation that allows for multiple minimal technologies, and also encompasses the simplification in the previous paragraph. Suppose simply that there is some nonempty collection $\mathcal{T}$ of possible technologies, and the principal's value from any contract w is defined as $V_P(w) = \inf_{\mathcal{A} \in \mathcal{T}} V_P(w|\mathcal{A})$. Suppose that $\mathcal{T}$ has the following property: For any $\mathcal{A} \in \mathcal{T}$, and any arbitrary action (F, c), then there exists some $\mathcal{A}' \subseteq \mathcal{A}$ such that $\mathcal{A}' \cup \{(F, c)\} \in \mathcal{T}$. Then, Theorem 2.1 continues to hold.

The proof is essentially the same as before, using the following generalization of Lemma 2.2: If w is a nonzero, eligible contract (eligibility defined as before), then

$$V_P(w) = \min E_F[y - w(y)] \quad \text{over } F \in \Delta(Y) \text{ such that } E_F[w(y)] \geq \inf_{\mathcal{A} \in \mathcal{T}} V_A(w|\mathcal{A});$$

and for any F attaining the minimum, $E_F[w(y)] = \inf_{\mathcal{A}} V_A(w|\mathcal{A})$. (The proof that the optimal contract exists is slightly more work than before, but one can derive an analogue to (2.5) and check that it is upper semi-continuous in α, which is enough for existence of the optimum.)

One can also show that linear contracts are uniquely optimal under an appropriate version of the full-support condition.

This discussion stresses that the model does not depend on allowing the agent's technology to be outrageously large. However, we do need that, *as the technology varies*, the range of actions that may potentially be chosen should be sufficiently rich, as discussed in Subsection 2.4 above. We could not (for example) restrict attention to technologies that contain only actions "close" to those in the known technology $\mathcal{A}_0$ and expect the same results to hold.

3.3 Participation constraint

In the basic model, the only constraint that imposed a lower bound on payments to the agent was limited liability. We could instead imagine that there is also a participation constraint, so that the principal is required to guarantee the agent an expected payoff of at least $\overline{U}_A$. This operates differently than limited liability, since it applies to the agent's payoff net of cost. Such a constraint could be modeled by restricting the principal to only propose contracts w satisfying $E_F[w(y)] - c \geq \overline{U}_A$ for some $(F, c) \in \mathcal{A}_0$. Let us assume some eligible w satisfies this constraint.

In this case, the same argument as before shows that every contract is weakly outperformed by an affine contract. Indeed, since the contract w' constructed in Lemma 2.5 satisfies $w' \geq w$ everywhere, if w satisfies the participation constraint, so does w'. However, the one step of the original argument that does not go through is Lemma 2.6, going from the affine contract $\alpha y + \beta$ to the linear αy: the latter may not satisfy the participation constraint.

Actually, a little more work shows that the optimal contract is still linear. Intuitively, if the optimal contract were affine, $\alpha y + \beta$ with $\beta > 0$, then the value of β would be determined by the participation constraint binding. But as long as the participation constraint is binding, the principal would rather increase α slightly, better aligning the agent's incentives with her own, and decrease β so that the participation constraint still binds. Thus it would be an improvement to increase the slope α up to the point where limited liability binds instead ($\beta = 0$).

Theorem 3.2. *In the model with a participation constraint added, there is still an optimal contract that is linear.*

The details are in Appendix A.

3.4 Screening on technology

We proved Theorem 2.1 by improving any contract to a linear contract. One might instead try a a more direct method of proof: find the contract with the best possible guarantee for the principal, $\max_w V_P(w)$, and identify an adversarial technology $\mathcal{A}$ that prevents the principal from doing any better. It turns out, however, that this proof approach would not work, because in general such an adversarial technology does not exist.

Proposition 3.3. *Let $w^*(y) = \alpha^* y$ be the linear contract that maximizes V_P, and suppose that $\alpha^* > 0$. Then there exists $\overline{V}_P > V_P(w^*)$ such that, for every technology $\mathcal{A} \supseteq \mathcal{A}_0$, there is some contract w with $V_P(w|\mathcal{A}) \geq \overline{V}_P$.*

The proof is in Appendix A.

For another interpretation of this proposition, imagine that the principal could somehow make the contract she offers be a function of the agent's technology. The proposition says that she could achieve a strictly higher guarantee — in worst case over all possible agent types — than in the basic model where she is constrained to only offer a single contract.

This result naturally brings to mind the question the possibility of screening: What if the principal could offer multiple contracts, inducing different agent types to self-select into different contracts? Such screening would be less flexible than the technology-dependent contract choice above, because it has to be incentive-compatible. It turns out this overturns the result of the proposition: Screening with a menu does not give a better worst-case guarantee than using a single contract.

To formalize this, we imagine that the principal offers a menu of contracts $\mathcal{W} = (w_{\mathcal{A}})$, one for each possible technology $\mathcal{A}$ that the agent could have, such that the agent with any technology $\mathcal{A}$ chooses the corresponding contract (this is without loss of generality by the revelation principle). Thus, we require

$$V_A(w_{\mathcal{A}}|\mathcal{A}) \geq V_A(w_{\mathcal{A}'}|\mathcal{A}) \qquad \text{for all} \qquad \mathcal{A}, \mathcal{A}' \supseteq \mathcal{A}_0. \tag{3.1}$$

We write the principal's worst-case payoff from the menu as

$$V_P(\mathcal{W}) = \inf_{\mathcal{A} \supseteq \mathcal{A}_0} V_P(w_{\mathcal{A}}|\mathcal{A}).$$

Theorem 3.4. *The principal cannot do any better, in terms of worst-case guarantee, with a menu of contracts than she can with a single contract. That is, for any menu $\mathcal{W}$,*

$$V_P(\mathcal{W}) \leq \max_w V_P(w).$$

Again, the proof is in Appendix A.

We close this section with an additional observation: The fact that the principal can do strictly better than $\max_w V_P(w)$ for any given technology suggests that she should be able to improve her worst-case guarantee by deliberately randomizing over contracts. This naturally raises the question of what the worst-case-optimal randomized contract looks like. This seems a harder problem. Similarly, it is not clear whether the result of Theorem 3.4 — that screening does not help — persists when the principal can offer menus of randomized contracts. We leave these questions for future work.

4 Discussion

We have presented here a simple principal-agent model that illustrates the robustness value of linear contracts. In the face of uncertainty about the technology available to the agent, linearity is the only tool the principal can use to turn her assurance about the agent's expected payoff into a guarantee for herself, and so optimal contracts are linear.

We now return to discuss this model's potential to help explain the popularity of linear contracts in the real world. Many previous scholars have noticed that, while theoretical models of agency relations often predict complicated incentive schemes that are sensitive to the details of the model, when one does see explicit incentives in practice they often take simple forms, and linear contracts are one common form (see [2, pp. 763-4] and [6, fn. 3] for many references).

One way to try to explain this via our model would be to take the model literally, imagining that contract writers explicitly maximize a worst-case objective, are risk-neutral, and so on. A fuzzier story, but perhaps closer to the truth, is as follows: Just as economists work with stylized models for tractability, so, too, real-life decision-makers may not be able to write down (or solve) their decision problems in full precision. They may therefore be content to adopt a solution that is guaranteed to perform reasonably well in a class of approximate models (similarly to Simon's "satisficing" [22]). This paper begins by pointing out how such a guarantee can be obtained from linear contracts, with only slight reliance on knowledge of the environment. Our main result then shows that, while many other contracts can also offer some such guarantee, linear contracts play a distinguished role in this story, by providing the best possible guarantee.

How does our model relate to other explanations for linear contracts in previous literature? The paper of Holmström and Milgrom [15] quoted above was one early explanation, also invoking robustness. In their model, the principal and agent have CARA utility, and the agent controls the drift of a (possibly multidimensional) Brownian motion in continuous time. Although the principal can make payments depend on the entire path of motion, the optimal contract is simply a linear function of the endpoint. Holmström and Milgrom present this model as capturing robustness, in view of the agent's large strategy space. However, it is really the stationary structure of the model that underlies the conclusion:

the CARA utility implies that at each point in time, the optimal incentives going forward are independent of the previous history, and this leads to linearity.

Diamond [8] gives an argument for linear contracts that is close to the intuition of this paper. Diamond's Section 5 considers a model in which the agent can either choose no effort, producing a low expected output, or high effort, producing a higher expected output. For a given level of effort, the agent can choose among all distributions over output that have the same mean, and all such distributions are equally costly. A linear contract is then optimal. The argument rests on the same intuition as here — with such freedom to choose the distribution, only a linear relation can tie the principal's expected profit to the agent's expected compensation. However, the assumptions that there are exactly two effort levels, and that all distributions with a given mean are equally costly, are restrictive. Furthermore, there are actually many optimal contracts in Diamond's model. In our model, uncertainty about which distributions are actually possible can make the linear contract *uniquely* optimal.

Several other papers consider models where the contractible outcome variable combines effort with mean-zero additive noise, leading naturally to linear contracts (or, more precisely, what we have called affine contracts, but let's ignore that distinction). For example, a version of the model of Edmans and Gabaix [9] with linear utility and additive noise gives this result. However, their model focuses mainly on implementing a particular action, rather than maximizing the principal's payoff. Earlier works by Laffont and Tirole [17] and McAfee and McMillan [19] consider problems that combine moral hazard and adverse selection: a principal uses a menu of contracts to screen agents on ability. In both of their models, there is again an optimal menu in which payment is linear in output within each contract. Again, however, there may also be other optimal menus. In any case, the assumption of additive noise is quite specific.

Chassang [4] considers a dynamic model and gives the same lower bound as ours (2.3) on the performance of a linear contract, by the same calculation. Chassang also gives a worst-case optimality result for linear contracts in a certain class of environments (his Corollary 1). In that class, first-best total surplus may be arbitrarily small, so the objective used is the *ratio* of the principal's profit to first-best surplus. As in our proof here, Chassang argues by finding a bad environment for any given contract. However, there is no analogue to our argument of improving a nonlinear contract to a linear one (instead, Chassang directly calculates that the objective value for any contract is at most the lower bound for the best linear contract); nor any analogue to our driving inequality (2.6), which expresses the intuition that *any* contract's guarantee stems from its ability to align the principal's and agent's payoffs.

Finally, Hurwicz and Shapiro [16] also consider a maxmin contracting problem whose objective involves the ratio of principal's profits to first-best total surplus. They focus on a particular class of environments involving quadratic effort costs. Their paper does not discuss economic intuition behind the optimality argument, which involves a differential inequality; it seems quite different from the argument here.

Against this backdrop, then, the contribution of the current paper is a combination of features: The model allows many degrees of freedom (the set of *known* actions the agent has can be arbitrary, with no functional form assumptions); the concern for robust performance is modeled explicitly through the worst-case payoff objective; and we give a mathematical argument for optimality based on the simple intuition that any contract's guarantee is driven by its alignment between the agent's interests and the principal's. Also, in contrast to previous maxmin results, the simple expected-profit objective here might also be considered more natural than the ratio objective, although this distinction is a matter of taste.

The model's mathematical tractability is also a virtue: as discussed in the introduction, one main purpose of the model is to present a methodology that can be adopted to study more complicated contracting problems. The various extensions in Section 3 give a sampling of such possibilities. And one can certainly come up with others beyond those discussed in detail here: for one more example, suppose that, instead of being risk-neutral over money, the principal and agent maximize the expected value of some (known) utility functions; then a straightforward variation of our main argument here shows that the maxmin-optimal contract is now affine in utility space.[3] Another, more sophisticated, application of our machinery is in the companion paper [3], which takes the same modeling approach to the principal-expert problem of Zermeño [25, 26] to study worst-case-optimal incentives for information acquisition.

Relatedly, the modeling approach here may prove useful to future economic theorists, who need a flexible model of moral hazard that outputs a simple solution, to serve as just one of many moving parts in some larger model. However, this suggestion should be supplemented with a note of caution: It is common in applied theory models to assume full knowledge of the environment, but then exogenously impose a restriction to linear contracts for tractability (e.g. [10, 13]). The model here *cannot* be invoked as a justification for this practice, since the contract that is best among all linear contracts when the technology is known to be $\mathcal{A}_0$ is generally different from the maxmin-optimal contract

[3] Details of this model are available from the author.

studied here.

A Proofs omitted from main text

Proof of Lemma 2.2. First, consider any technology $\mathcal{A} \supseteq \mathcal{A}_0$. The agent's payoff is at least $V_A(w|\mathcal{A}_0)$. That is, his chosen action (F, c) satifies

$$E_F[w(y)] \geq E_F[w(y)] - c \geq V_A(w|\mathcal{A}_0).$$

Hence the principal's payoff, $V_P(w|\mathcal{A}) = E_F[y - w(y)]$, is at least the minimum given by (2.4). Thus, the principal's worst-case payoff $V_P(w)$ is no lower than given by (2.4).

To see this is tight, let F be a distribution attaining the minimum in (2.4). First suppose that F does not place full support on values of y for which w attains its maximum. Then let F' be a mixture of F with weight $1-\epsilon$, and a mass point δ_{y^*} with weight ϵ, where y^* is some point where w attains its maximum. Then $E_{F'}[w(y)] > E_F[w(y)] \geq V_A(w|\mathcal{A}_0)$. The strict inequality means that if $\mathcal{A} = \mathcal{A}_0 \cup \{(F', 0)\}$, then the agent's unique optimal action in $\mathcal{A}$ is $(F', 0)$, leading to expected payoff $(1 - \epsilon)E_F[y - w(y)] + \epsilon(y^* - w(y^*))$ for the principal. As $\epsilon \to 0$ this converges to the minimum in (2.4), so the principal cannot be guaranteed any higher expected payoff.

Now suppose F does place full support on values of y at which w attains its maximum. If $E_F[w(y)] > V_A(w|\mathcal{A}_0)$, then we can again proceed as above with $\mathcal{A} = \mathcal{A}_0 \cup \{(F, 0)\}$. This leaves only the case of equality — $V_A(w|\mathcal{A}_0) = \max_y w(y)$ — which is only satisfied when $\mathcal{A}_0$ contains some action of the form $(F, 0)$ with F supported at output levels for which w attains its maximum. Then, under technology $\mathcal{A}_0$ the agent will choose such an action. But then the agent would have been willing to choose the same action under the zero contract (and any $\mathcal{A} \supseteq \mathcal{A}_0$), which costs less to the principal (strictly, since w is not the zero contract). Thus $V_P(w) < V_P(0)$, contradicting eligibility.

This shows (2.4). Now let $F \in \Delta(Y)$ attain the minimum in (2.4). We have $E_F[y - w(y)] = V_P(w) > 0$ by eligibility. On the other hand, $y - w(y) \leq 0$ when $y = 0$. Now if we have $E_F[w(y)] > V_A(w|\mathcal{A}_0)$ strictly, then replace F by a mixture of F with weight $1 - \epsilon$ and δ_0 with weight ϵ, for small ϵ, to see that minimality is contradicted. Hence we have equality, $E_F[w(y)] = V_A(w|\mathcal{A}_0)$, as claimed. □

Proof of Lemma 2.4. Although the separation argument in Figure 1 is illustrated in outcome space, the proof (and result) of this lemma are most cleanly written in payoff space.

Thus, let $S \subseteq \mathbb{R}^2$ be the convex hull of all points $(w(y), y - w(y))$ for $y \in Y$. Let T be the set of all pairs $(u, v) \in \mathbb{R}^2$ such that $u > V_A(w|\mathcal{A}_0)$ and $v < V_P(w)$. The conclusion (2.4) of Lemma 2.2 implies that S and T are disjoint.

So by the separating hyperplane theorem, there exist constants κ, λ, μ such that

$$\kappa + \lambda u - \mu v \leq 0 \qquad \text{for all} \qquad (u, v) \in S, \tag{A.1}$$

$$\kappa + \lambda u - \mu v \geq 0 \qquad \text{for all} \qquad (u, v) \in T, \tag{A.2}$$

and $(\lambda, \mu) \neq (0, 0)$. In addition, if we let F^* be the distribution attaining the minimum in (2.4), the pair $(E_{F^*}[w(y)], E_{F^*}[y - w(y)])$ lies in the closures of both S and T, hence

$$\kappa + \lambda\, E_{F^*}[w(y)] - \mu\, E_{F^*}[y - w(y)] = 0. \tag{A.3}$$

We will show that $\lambda, \mu > 0$. Condition (A.2) implies that $\lambda, \mu \geq 0$, so we just need to show that both inequalities are strict:

- If $\mu = 0$, then $\lambda > 0$, and (A.1) and (A.2) imply $\max_{y \in Y} w(y) \leq -\kappa/\lambda \leq V_A(w|\mathcal{A}_0)$. This can only happen if the agent has some action in $\mathcal{A}_0$ that guarantees him a payment of $\max_y w(y)$ and costs zero. As in the proof of Lemma 2.2, this implies $V_P(w) < V_P(0)$, contradicting eligibility.

- If $\lambda = 0$, then $\mu > 0$, and (A.1) and (A.2) imply $\min_{y \in Y}(y - w(y)) \geq \kappa/\mu \geq V_P(w)$. But $\min_{y \in Y}(y - w(y)) \leq 0 - w(0) \leq 0$, so $V_P(w) \leq 0$, again contrary to assumption.

Now we can rescale κ, λ, μ so as to assume $\mu = 1$. Then, λ remains positive, (A.1) implies (2.6), and (A.3) implies (2.7).

□

Proof of Corollary 2.8. Suppose w is an optimal contract. Note that the proof of Lemma 2.5 actually shows that the contract w' defined by (2.10) satisfies

$$V_P(w') \geq \kappa + \lambda\, V_A(w'|\mathcal{A}_0) = V_P(w) + \lambda(V_A(w'|\mathcal{A}_0) - V_A(w|\mathcal{A}_0)). \tag{A.4}$$

But under the full-support assumption, if w' is not identical to w then the difference $V_A(w'|\mathcal{A}_0) - V_A(w|\mathcal{A}_0)$ is strictly positive. (This follows because the action taken under w and technology $\mathcal{A}_0$ has full support, so gives the agent strictly higher payoff under w' than w.) Then (A.4) implies that $V_P(w') > V_P(w)$, contradicting optimality of w.

Therefore, $w' = w$. So w is an affine contract. In fact, w must be linear: otherwise the improvement given by Lemma 2.6 is strict, again contradicting optimality. Thus, every optimal contract is linear. □

Proof of Theorem 3.2. The same steps used for Theorem 2.1 show that there is an optimal contract that is affine, $w(y) = \alpha y + \beta$. Moreover, for any given α, the optimal choice of β is to be as small as possible subject to the nonnegativity and participation constraints:

$$\beta^*(\alpha) = \max \left\{ 0, \overline{U}_A - \max_{(F,c) \in \mathcal{A}_0} (\alpha E_F[y] - c) \right\}. \tag{A.5}$$

Evidently, the first case of the max holds when α is greater than some threshold $\overline{\alpha}$, and the second case holds for $\alpha \leq \overline{\alpha}$ (to be precise, $\overline{\alpha} = \min_{(F,c) \in \mathcal{A}_0} (\overline{U}_A + c)/E_F[y])$. So we have an analogue of Lemma 2.3: for any α, the guarantee of the best affine contract with slope α is at least

$$\max_{(F,c) \in \mathcal{A}_0} \left((1 - \alpha) E_F[y] - \frac{1 - \alpha}{\alpha} c \right) - \beta^*(\alpha), \tag{A.6}$$

with equality for eligible contracts. Thus finding the optimal contract reduces to maximizing (A.6) over α.

But when $\alpha \leq \overline{\alpha}$, (A.6) simplifies to

$$\max_{(F,c) \in \mathcal{A}_0} \left(E_F[y] - \frac{1}{\alpha} c \right) - \overline{U}_A, \tag{A.7}$$

which is increasing in α (or constant, in the case $c = 0$). And so we conclude that the maximum is attained at some $\alpha \geq \overline{\alpha}$, where $\beta^*(\alpha) = 0$. □

Proof of Proposition 3.3. Let $(F^*, c^*) \in \mathcal{A}_0$ be the action that maximizes (2.12). So we have $w^*(y) = \alpha^* y$, with $\alpha^* = \sqrt{c^*/E_{F^*}[y]}$, and the principal's guarantee is $V_P(w^*) = (\sqrt{E_{F^*}[y]} - \sqrt{c^*})^2$. Consider any technology $\mathcal{A}$, and let (F, c) be the agent's action under w^* and $\mathcal{A}$. Thus

$$\alpha^* E_F[y] - c \geq \alpha^* E_{F^*}[y] - c^*. \tag{A.8}$$

We consider two cases.

- If $c \geq c^*/2$, then the principal's payoff from contract w^* is

$$\begin{aligned}(1-\alpha^*)E_F[y] &\geq \frac{1-\alpha^*}{\alpha^*}\left(\alpha^* E_{F^*}[y] - c^* + c\right) \\ &= E_{F^*}[y] - 2\sqrt{c^* E_{F^*}[y]} + c^* + \frac{1-\alpha^*}{\alpha^*}c \\ &\geq V_P(w^*) + \frac{1-\alpha^*}{\alpha^*}\frac{c^*}{2}.\end{aligned}$$

- Now suppose $c \leq c^*/2$. We know that if the principal learns $\mathcal{A}$ before contracting, then by choosing an appropriate contract she can earn at least $(\sqrt{E_F[y]} - \sqrt{c})^2$ (since in fact this is her worst-case guarantee with $\mathcal{A}$ in place of $\mathcal{A}_0$ — note the condition $E_F[y] > c$ is met). We show that this expression is bounded strictly above $V_P(w^*)$. Define
$$g(x) = \sqrt{\frac{x^2 + (\alpha^* E_{F^*}[y] - c^*)}{\alpha^*}} - x \tag{A.9}$$
for $x \geq 0$. Then g is convex, and we check that the minimum is given by the first-order condition; this condition is satisfied (uniquely) by $x = \sqrt{c^*}$, with value $g(\sqrt{c^*}) = \sqrt{E_{F^*}[y]} - \sqrt{c^*}$. Now, holding c fixed, treat $E_F[y]$ as a variable, constrained by (A.8) and $E_F[y] > c$. Then $(\sqrt{E_F[y]} - \sqrt{c})^2$ is minimized by taking (A.8) to hold with equality, and in this case $\sqrt{E_F[y]} - \sqrt{c} = g(\sqrt{c})$. Thus we see that the principal can make a payoff of at least
$$\left(\sqrt{E_F[y]} - \sqrt{c}\right)^2 \geq \left(g(\sqrt{c})\right)^2 \geq \left(g(\sqrt{c^*/2})\right)^2.$$
Now observe that $(g(\sqrt{c^*/2}))^2 > (g(\sqrt{c^*}))^2 = V_P(w^*)$.

So in both cases, we have a lower bound for the principal's payoff when she knows $\mathcal{A}$ that is strictly above $V_P(w^*)$. □

Proof of Theorem 3.4. Consider any menu $\mathcal{W}$. Let $w_0 = w_{\mathcal{A}_0}$, the contract that the agent would choose when the technology is just $\mathcal{A}_0$. We claim that $V_P(w_0) \geq V_P(\mathcal{W})$, which will prove the theorem.

Suppose not. Then, there is some technology $\mathcal{A}_1$ under which, facing contract w_0, the agent chooses an action (F_1, c_1) that gives the principal payoff less than $V_P(\mathcal{W})$. We may assume that $\mathcal{A}_1 = \mathcal{A}_0 \cup \{(F_1, c_1)\}$. Note also that $(F_1, c_1) \notin \mathcal{A}_0$, since otherwise $\mathcal{A}_1 = \mathcal{A}_0$ and so $V_P(w_0|\mathcal{A}_0) < V_P(\mathcal{W})$ which is a contradiction. It must be that, under w_0, the agent earns strictly higher payoff from (F_1, c_1) than he does from any action in $\mathcal{A}_0$: otherwise he would be willing to take the same action under $\mathcal{A}_1$ as he does under $\mathcal{A}_0$, thereby giving the principal $V_P(w_0|\mathcal{A}_0) \geq V_P(\mathcal{W})$.

Now let $w_1 = w_{\mathcal{A}_1}$, the contract chosen from the menu when the technology is $\mathcal{A}_1$. Under w_1 and $\mathcal{A}_1$, the agent must choose action (F_1, c_1). Proof: If he chooses any action in $\mathcal{A}_0$, then his payoff is at most $V_A(w_0|\mathcal{A}_0)$ (by revealed preference (3.1)). On the other hand, his payoff under w_1 and $\mathcal{A}_1$ must be at least as high as his payoff from (F_1, c_1) under w_0 (by revealed preference again, since w_1 was chosen under $\mathcal{A}_1$), which is higher than $V_A(w_0|\mathcal{A}_0)$ by the previous paragraph.

Hence, (F_1, c_1) is the agent's uniquely chosen action under w_1, and

$$E_{F_1}[w_1(y)] - c_1 \geq E_{F_1}[w_0(y)] - c_1$$

again by (3.1). Then, the principal's payoff when the technology is $\mathcal{A}_1$ is

$$\begin{aligned} E_{F_1}[y - w_1(y)] &= E_{F_1}[y] - c_1 - (E_{F_1}[w_1(y)] - c_1) \\ &\leq E_{F_1}[y] - c_1 - (E_{F_1}[w_0(y)] - c_1) \\ &= E_{F_1}[y - w_0(y)] \\ &< V_P(\mathcal{W}) \end{aligned}$$

where the last line is by definition of (F_1, c_1). Since the principal should get at least $V_P(\mathcal{W})$ under every possible technology, we have a contradiction. □

B General lower bounds on cost

First, as promised in Subsection 3.1, we illustrate by example how the multiple-observables model allows us to describe situations where where the cost bound depends on higher moments of output. Suppose, for example, that only output y is observed, and the principal knows that any distribution F costs at least $h(E_F[y]) - \kappa \cdot Var_F[y]$, where h is some given convex function. Then, we would capture this by putting

$$Z = \{(y, y^2) \mid y \in Y\}$$

and

$$b(z_1, z_2) = \max\{0, h(z_1) - \kappa(z_2 - z_1^2)\}.$$

Theorem 3.1 would apply, and tell us that an optimal contract is quadratic in y.

We now embark on the proof of the theorem, which follows the same outline as in Subsection 2.3. We first characterize the payoff guarantee of any given contract w. The situation is a bit more complex than before, because the tie-breaking assumption requires careful treatment of the boundary case in which the agent's best action under any possible technology is already available in $\mathcal{A}_0$.

For $F \in \Delta(Z)$ and a given contract w, define $h(F|w) = E_F[w(z)] - b(E_F[z])$, the highest expected payoff the agent could possibly get from producing distribution F. Since b is convex, h is concave in F.

Lemma B.1. *Let w be any contract. Then one of the following two cases occurs:*

(i) $V_P(w) = \min E_F[z_1 - w(z)]$ *over* $F \in \Delta(Z)$ *such that* $h(F|w) \geq V_A(w|\mathcal{A}_0)$.

(ii) $\max_{F\in\Delta(Z)} h(F|w) = V_A(w|\mathcal{A}_0)$.

Proof. Let F_0 be a distribution attaining the minimum in (i). (The constraint set is nonempty since it is satisfied by the action chosen under $\mathcal{A}_0$.) Suppose that F_0 does not also maximize $h(F|w)$ over all $F \in \Delta(Z)$. Then, choose F_1 yielding a higher value of h, and put $F' = (1-\epsilon)F_0 + \epsilon F_1$ for small ϵ. By concavity, $h(F'|w) \geq (1-\epsilon)h(F_0|w) + \epsilon h(F_1|w) > h(F_0|w)$. So if $\mathcal{A} = \mathcal{A}_0 \cup \{(F', b(E_{F'}[z]))\}$, then the agent's unique optimal action in $\mathcal{A}$ is $(F', b(E_{F'}[z]))$. As $\epsilon \to 0$ the principal's resulting payoff tends to $E_{F_0}[z_1 - w(z)]$. Thus the principal cannot be guaranteed more than the value in (i). On the other hand the principal is guaranteed at least this much, just as in the proof of Lemma 2.2.

Also, if $h(F_0|w) > V_A(w|\mathcal{A}_0)$ strictly, then let $\mathcal{A} = \mathcal{A}_0 \cup \{(F_0, b(E_{F_0}[z]))\}$. With this technology, the agent's unique optimal action is $(F_0, b(E_{F_0}[z]))$, and again the principal cannot be guaranteed more than the value in (i). Thus in either of these situations $V_P(w)$ is as specified by conclusion (i).

We are left with the situation in which F_0 maximizes $h(F|w)$ over all $F \in \Delta(Z)$ and $h(F_0|w) = V_A(w|\mathcal{A}_0)$. In this case, we have conclusion (ii). □

Now we prove Theorem 3.1 by the same process as before: given a non-affine contract w, use a separation argument to replace it by an affine contract w' that is pointwise above it and gives a weakly greater guarantee to the principal. We will perform the separation in outcome space, not in payoff space as in our proof of Lemma 2.4. In addition, we use two different versions of the argument, depending which case of Lemma B.1 applies.

Proof of Theorem 3.1. We may assume that the convex hull of Z is a full-dimensional set in $\mathbb{R}^k$. (This can be accomplished by a linear change of coordinates to embed Z

in a smaller-dimensional space if necessary, unless $Y = \{0\}$ but the latter situation is uninteresting.)

Consider any non-affine contract w. As usual, we may restrict attention to eligible contracts, since nontriviality ensures such a contract exists. One of the two cases of Lemma B.1 holds, and we deal with the two separately.

Case (i). We define

$$t(z) = \max\{b(z) + V_A(w|\mathcal{A}_0), z_1 - V_P(w)\}$$

and observe that t is a convex function. Now, we define two sets in $\mathbb{R}^{k+1} = \mathbb{R}^k \times \mathbb{R}$. Let S be the convex hull of all pairs $(z, w(z))$, for $z \in Z$. Let T be the set of all pairs (z, c) such that z lies in the convex hull of Z, and $c > t(z)$.

Both of these sets are convex. We claim they are disjoint. If not, there exists some $F \in \Delta(Z)$ such that $E_F[w(z)] > t(E_F[z])$. In particular,

$$E_F[w(z)] > b(E_F[z]) + V_A(w|\mathcal{A}_0)$$

implying

$$h(F|w) > V_A(w|\mathcal{A}_0),$$

and also

$$E_F[w(z)] > E_F[z_1] - V_P(w)$$

implying

$$E_F[z_1 - w(z)] < V_P(w).$$

This is a direct contradiction to our statement (i).

So by the separating hyperplane theorem, there are constants $\lambda_1, \ldots, \lambda_k, \mu, \nu$ such that

$$\sum_i \lambda_i z_i + \mu c \le \nu \qquad \text{for all} \qquad (z, c) \in S, \tag{B.1}$$

$$\sum_i \lambda_i z_i + \mu c \ge \nu \qquad \text{for all} \qquad (z, c) \in T, \tag{B.2}$$

and some λ_i or μ is nonzero. Inequality (B.2) implies $\mu \ge 0$. In fact, $\mu > 0$. Proof: Suppose $\mu = 0$. Since the projection of either S or T onto the first k coordinates contains Z, (B.1) gives $\sum_i \lambda_i z_i \le \nu$ for all $z \in Z$, while (B.2) gives $\sum_i \lambda_i z_i \ge \nu$ for all $z \in Z$. Hence, $\sum_i \lambda_i z_i = \nu$ for all $z \in Z$. Since not all λ_i are zero, this contradicts the full-dimensionality of Z.

Now we can rewrite (B.1) as

$$w(z) \le \frac{\nu - \sum_i \lambda_i z_i}{\mu} \qquad \text{for all } z \in Z.$$

This motivates us to define

$$w'(z) = \frac{\nu - \sum_i \lambda_i z_i}{\mu}, \tag{B.3}$$

an affine contract satisfying $w' \ge w$ pointwise.

Now we are ready to check that $V_P(w') \ge V_P(w)$. Let (F_0, c_0) be the action that the agent chooses under w and technology $\mathcal{A}_0$.

Consider any technology $\mathcal{A} \supseteq \mathcal{A}_0$. As in the original proof of Theorem 2.1, we certainly have have $V_A(w'|\mathcal{A}) \ge V_A(w'|\mathcal{A}_0) \ge V_A(w|\mathcal{A}_0)$. Let (F, c) be the action chosen under w' and $\mathcal{A}$. Then (B.2) implies

$$\begin{aligned} t(E_F[z]) &\ge \frac{\nu - \sum_i \lambda_i E_F[z_i]}{\mu} \\ &= E_F[w'(z)] \\ &= V_A(w'|\mathcal{A}) + c \\ &\ge V_A(w|\mathcal{A}_0) + c \\ &\ge b(E_F[z]) + V_A(w|\mathcal{A}_0). \end{aligned}$$

If the inequality is strict, then $t(E_F[z]) = E_F[z_1] - V_P(w)$, and so we have

$$V_P(w'|\mathcal{A}) = E_F[z_1 - w'(z)] = t(E_F[z]) + V_P(w) - E_F[w'(z)] \ge V_P(w).$$

Otherwise, $t(E_F[z]) = b(E_F[z]) + V_A(w|\mathcal{A}_0)$ and so all the inequalities in the stacked chain above are equalities. In particular, the second inequality is an equality, implying $V_A(w'|\mathcal{A}) = V_A(w'|\mathcal{A}_0) = V_A(w|\mathcal{A}_0)$. Since the agent does at least as well as $V_A(w|\mathcal{A}_0)$ by taking action (F_0, c_0), this action is in his choice set under w' and $\mathcal{A}$, and so the principal gets at least the corresponding payoff: $V_P(w'|\mathcal{A}) \ge E_{F_0}[z_1 - w'(z)]$. This is equal to $E_{F_0}[z_1 - w(z)]$, since otherwise $V_A(w'|\mathcal{A}_0) > V_A(w|\mathcal{A}_0)$. But $E_{F_0}[z_1 - w(z)] = V_P(w|\mathcal{A}_0) \ge V_P(w)$.

Thus in either case, $V_P(w'|\mathcal{A}) \ge V_P(w)$. This holds for all $\mathcal{A}$, so $V_P(w') \ge V_P(w)$.

Case (ii). In this case, define S to be the convex hull of all pairs $(z, w(z))$, and T to be the set of all (z, c) with z in the convex hull of Z and $c > b(z) + V_A(w|\mathcal{A}_0)$. These are convex, and disjoint: otherwise, there exists F such that

$$E_F[w(z)] > b(E_F[z]) + V_A(w|\mathcal{A}_0)$$

which reduces to

$$h(F|w) > V_A(w|\mathcal{A}_0),$$

in contradiction to the statement of (ii). Using the same arguments as in case (i), we find $\lambda_1, \ldots, \lambda_k, \mu, \nu$ such that (B.1) and (B.2) hold, and we show that $\mu > 0$. Again, we define an affine contract w' by (B.3); from (B.1) we know that $w' \geq w$ pointwise.

Consider the agent's behavior under contract w'. For any action (F, c) chosen by the agent under any possible technology, we have

$$E_F[w'(z)] - c \leq E_F[w'(z)] - b(E_F[z]) = w'(E_F[z]) - b(E_F[z]) \leq V_A(w|\mathcal{A}_0)$$

where the second inequality follows from (B.2). That is, the agent can never earn a higher expected payoff than $V_A(w|\mathcal{A}_0)$. On the other hand, the agent can always earn at least this much, since $V_A(w'|\mathcal{A}) \geq V_A(w'|\mathcal{A}_0) \geq V_A(w|\mathcal{A}_0)$ as usual. So we have equality. From here the argument finishes just as at the end of case (i), and we have $V_P(w') \geq V_P(w)$.

Existence of an optimum. We have shown that any contract w with $V_P(w) > 0$ can be (weakly) improved to an affine contract. So it now suffices to show existence of an optimum within the class of affine contracts, analogous to Lemma 2.7, and this contract will then be optimal among all contracts.

Put $\overline{b} = \max_{z \in Z} b(z)$ and $\overline{y} = \max(Y)$. Note that for any contract w satisfying $\max_{z \in Z} w(z) - \overline{b} \geq \overline{y}$, the agent can potentially attain a payoff greater than $\overline{y}$, which means that the principal cannot be guaranteed a positive payoff. Hence we can restrict attention to contracts with $w(z) \in [0, \overline{y} + \overline{b}]$ for all z. By full-dimensionality, this implies a compact range of possible values for the parameters α and β defining the affine contract. We will show below that $V_P(w)$ is upper semi-continuous with respect to w, under the sup-norm topology on the space of contracts. (It may not be fully continuous.) Since the affine contract w in turn varies continuously in α, β under this topology, it will then follow that $V_P(w)$ is upper semi-continuous in α, β, so that the maximum is attained.

Let $w_1, w_2, \ldots$ be any contracts that converge to some contract w_∞ in the sup norm. We wish to show that $V_P(w_\infty) \geq \limsup_k V_P(w_k)$. We can replace the sequence (w_k) with a subsequence along which $V_P(w_k)$ converges to its lim sup on the original sequence; thus, we assume henceforth that $V_P(w_k)$ converges. Now consider any technology $\mathcal{A}$, and let

(F_k, c_k) be the agent's chosen action under $\mathcal{A}$ and contract w_k. We may again pass to a subsequence and assume that (F_k, c_k) has some limit $(F_\infty, c_\infty) \in \mathcal{A}$. Then straightforward continuity arguments show that (F_∞, c_∞) is an optimal action (perhaps not the only one) for the agent under w_∞, and its payoff to the principal is the limit of the corresponding payoffs of (F_k, c_k) under w_k. Hence,

$$V_P(w_\infty|\mathcal{A}) \geq E_{F_\infty}[z_1 - w_\infty(z)] = \lim_k E_{F_k}[z_1 - w_k(z)] = \lim_k V_P(w_k|\mathcal{A}) \geq \lim_k V_P(w_k),$$

and so $V_P(w_\infty) \geq \lim_k V_P(w_k)$ as needed. □

C An alternative approach

We give here another, more direct approach to the main argument of Theorem 2.1: that for any contract w, there is a linear contract w' that guarantees at least as much for the principal. (The argument here was suggested by Lucas Maestri.)

Consider any eligible w, and let (F_0, c_0) be the action that the agent would choose under technology $\mathcal{A}_0$. Put $\alpha = E_{F_0}[w(y)]/E_{F_0}[y]$. (The denominator must be positive, since the principal is guaranteed a positive payoff.) Put $w'(y) = \alpha y$. Notice that under this contract, the agent can again take action (F_0, c_0) to earn a payoff of

$$E_{F_0}[\alpha y] - c_0 = E_{F_0}[w(y)] - c_0 = V_A(w|\mathcal{A}_0),$$

and the principal then earns

$$E_{F_0}[(1-\alpha)y] = E_{F_0}[y - w(y)] = V_P(w|\mathcal{A}_0) \geq V_P(w).$$

We will show that the principal does at least as well under w' as under w. Consider an arbitrary technology $\mathcal{A}$, and let (F, c) be the action the agent would take under contract w'; we need to show that the principal's resulting payoff, $V_P(w'|\mathcal{A})$, is at least $V_P(w)$. If $E_F[y] \geq E_{F_0}[y]$, then the principal gets

$$(1-\alpha)E_F[y] \geq (1-\alpha)E_{F_0}[y] = V_P(w|\mathcal{A}_0) \geq V_P(w).$$

Also, we have $E_F[w'(y)] - c \geq V_A(w'|\mathcal{A}_0) \geq V_A(w|\mathcal{A}_0)$ by optimality for the agent; and if equality holds throughout, then the agent would also be willing to choose (F_0, c_0), again giving the principal at least $V_P(w)$; thus $V_P(w'|\mathcal{A}) \geq V_P(w)$ in this case too. So we can focus on the case when $E_F[y] < E_{F_0}[y]$ and $E_F[w'(y)] - c > V_A(w|\mathcal{A}_0)$.

Put $\lambda = E_F[y]/E_{F_0}[y]$, and let F' be the mixture $\lambda F_0 + (1-\lambda)\delta_0$. Then, consider contract w when the technology is $\mathcal{A}_0 \cup \{(F', c)\}$. The agent's payoff from (F', c) is

$$\begin{aligned}
E_{F'}[w(y)] - c &= \lambda E_{F_0}[w(y)] + (1-\lambda)w(0) - c \\
&\geq \lambda E_{F_0}[w(y)] - c \\
&= \lambda \alpha E_{F_0}[y] - c \\
&= \alpha E_F[y] - c \\
&= E_F[w'(y)] - c \\
&> V_A(w|\mathcal{A}_0)
\end{aligned}$$

which means that the agent would strictly prefer to take action (F', c) over any other action. This leaves the principal with a payoff of

$$\begin{aligned}
E_{F'}[y - w(y)] &= \lambda E_{F_0}[y - w(y)] - (1-\lambda)w(0) \\
&\leq \lambda E_{F_0}[y - w(y)] \\
&= (1-\alpha)E_F[y] \\
&= E_F[y - w'(y)] \\
&= V_P(w'|\mathcal{A}).
\end{aligned}$$

Thus we have $V_P(w) \leq V_P(w'|\mathcal{A})$. So we have shown this inequality holds for all $\mathcal{A}$, implying $V_P(w) \leq V_P(w')$.

We comment that, while this proof is quicker and more direct than the support-line approach in the main text, we have focused on that approach for two reasons. One is that it generalizes readily, in particular to the multiple-observables extension of Appendix B and to the principal-expert problem in [3]. The approach above depends on taking a convex combination of an arbitrary distribution with δ_0 to attain a specific expected output; it is not clear how to extend it when the space of observable outcomes is not one-dimensional. The second reason is that Corollary 2.8 — only linear contracts are optimal with full support — is almost immediate with the support-line approach; with the argument here it seems to require more work.

References

[1] Dirk Bergemann and Stephen Morris (2005), "Robust Mechanism Design," *Econometrica* 73 (6), 1771–1813.

[2] Sugato Bhattacharyya and Francine Lafontaine (1995), "Double-Sided Moral Hazard and the Nature of Share Contracts," *RAND Journal of Economics* 26 (4), 761–781.

[3] Gabriel Carroll (2012), "Robust Incentives for Information Acquisition," working paper, Stanford University.

[4] Sylvain Chassang (2013), "Calibrated Incentive Contracts," *Econometrica* 81 (5), 1935–1971.

[5] Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan (2010), "Multi-Parameter Mechanism Design and Sequential Posted Pricing," in *Proceedings of the 42nd ACM Symposium on the Theory of Computing* (STOC-10), 311–320.

[6] Leon Yang Chu and David E. M. Sappington (2007), "Simple Cost-Sharing Contracts," *American Economic Review* 97 (1), 419–428.

[7] Kim-Sau Chung and J. C. Ely (2007), "Foundations of Dominant-Strategy Mechanisms," *Review of Economic Studies* 74 (2), 447–476.

[8] Peter Diamond (1998), "Managerial Incentives: On the Near Linearity of Optimal Compensation," *Journal of Political Economy* 106 (5), 931–957.

[9] Alex Edmans and Xavier Gabaix (2011), "Tractability in Incentive Contracting," *Review of Financial Studies* 24 (9), 2865–2894.

[10] Gerald A. Feltham and Jim Xie (1994), "Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations," *Accounting Review* 69 (3), 429–453.

[11] Alexander Frankel (2014), "Aligned Delegation," *American Economic Review* 104 (1), 66–83.

[12] Daniel Garrett (forthcoming), "Robustness of Simple Menus of Contracts in Cost-Based Procurement," *Games and Economic Behavior*.

[13] Robert Gibbons and Kevin J. Murphy (1992), "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy* 100 (3), 468–505.

[14] Bengt Holmström (1979), "Moral Hazard and Observability," *Bell Journal of Economics* 10 (1), 74–91.

[15] Bengt Holmström and Paul Milgrom (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica* 55 (2), 303–328.

[16] Leonid Hurwicz and Leonard Shapiro (1978), "Incentive Structures Maximizing Residual Gain under Incomplete Information," *Bell Journal of Economics* 9 (1), 180–191.

[17] Jean-Jacques Laffont and Jean Tirole (1986), "Using Cost Observation to Regulate Firms," *Journal of Political Economy* 94 (3), 614–641.

[18] Kristof Madarász and Andrea Prat (2014), "Sellers with Misspecified Models," working paper, London School of Economics.

[19] R. Preston McAfee and John McMillan (1987), "Competition for Agency Contracts," *RAND Journal of Economics* 18 (2), 296–307.

[20] Silvio Micali and Paul Valiant (2008), "Resilient Mechanisms for Truly Combinatorial Auctions," working paper, Massachusetts Institute of Technology.

[21] Ilya Segal (2003), "Optimal Pricing Mechanisms with Unknown Demand," *American Economic Review* 93 (3), 509–529.

[22] Herbert A. Simon (1956), "Rational Choice and the Structure of the Environment," *Psychological Review* 63 (2), 129–138.

[23] Jörg Stoye (2012), "New Perspectives on Statistical Decisions under Ambiguity," *Annual Review of Economics* 4, 257–282.

[24] Takuro Yamashita (2012), "Robust Trading Mechanisms to Strategic Uncertainty," working paper, Toulouse School of Economics.

[25] Luis Zermeño (2011), "A Principal-Expert Model and the Value of Menus," working paper, Massachusetts Institute of Technology.

[26] Luis Zermeño (2012), "The Role of Authority in a General Principal-Expert Model," working paper, Massachusetts Institute of Technology.