A General Framework
for Robust Contracting Models*

Daniel Walton

Uber Technologies

Gabriel Carroll

University of Toronto

February 22, 2022

Abstract

We study a class of models of moral hazard in which a principal contracts with a counterparty, which may have its own internal organizational structure. The principal has non-Bayesian uncertainty as to what actions might be taken in response to the contract, and wishes to maximize her worst-case payoff. We identify conditions on the counterparty's possible responses to any given contract that imply that a linear contract solves this maxmin problem. In conjunction with a *Richness* property motivated by much previous literature, we identify a *Responsiveness* property that is sufficient—and, in an appropriate sense, also necessary—to ensure that linear contracts are optimal. We illustrate by contrasting several possible models of contracting in hierarchies. The analysis demonstrates how one can distill key features of contracting models that allow their findings to be carried beyond the bilateral setting.

1 Introduction

Suppose that a principal wishes to write an incentive contract to induce productive effort. How should she structure the incentives so as to optimally ensure their effectiveness? A rich theoretical literature has explored this question, giving arguments in favor of one or another form of contract.

However, this literature has generally focused on models involving a single principal and a single agent. In reality, agency often takes place beyond simple bilateral relationships. For example, the principal may be a firm or government office, procuring a good of unpredictable quality from a supplier, and committing to a payment that depends on the realized quality; but the supplier has its own internal agency problem, since the representative who signs the contract with the principal may not be the same worker who produces the good. Can we abstract away from the specificity of bilateral contracting models to understand when and why their lessons carry over to models of more complex organizations?

In this paper, we focus on one such lesson that has arisen frequently: that linear contracts—which simply pay some fixed fraction of the output produced—perform well by aligning the expected payoffs of the two parties. The literature on this theme has generally drawn on the idea that there may be a large space of possible actions by the agent, a notion that we formalize subsequently under the name of "Richness." With a narrow space of possible actions, the structure of the optimal contract may be nonlinear and finely tuned to the known possibilities; but under richness, any such nonlinearities are vulnerable to strategic gaming by the agent. Incarnations of this idea have appeared in static moral hazard models (Diamond, 1998; Carroll, 2015; Barron et al., 2020; Antić, 2021), dynamic moral hazard (Holmström and Milgrom, 1987), and screening as well (Malenko and Tsoy, 2020).

We focus more specifically on the following version of the argument, based on robustness to uncertainty: Consider a principal ("she") and an agent ("he"), both risk-neutral, who agree to a contract that pays the agent 1/4 of whatever output he produces. Suppose that the principal does not know exactly what productive actions the agent is able to take, but she knows he has some action available that will give him an expected payoff of at least 1500 under this contract. Then, even without any information about what other actions are available, the principal can be sure the agent will get a payoff at least 1500, and therefore she gets at least 4500 for

herself (since she receives 3/4 of output against the agent's 1/4). This argument was developed previously by Carroll (2015), which formalized this idea of a guarantee for the principal via a worst-case criterion, and showed more generally that linear contracts are optimal under such a criterion.

To see the difficulty in generalizing this conclusion beyond the simple bilateral setting, consider now a three-player hierarchy: The principal contracts with a supervisor (also "she"); the supervisor then subcontracts with an agent, and the agent chooses the action that determines output. Payments in both contracts are functions of output only. There are (at least) three natural ways to write this model, with different informational assumptions:

(i) As in the bilateral model, the principal knows some actions available to the agent, but there may be other actions that she does not know about. The supervisor, however, fully knows the agent's production technology.

(ii) The supervisor may know more actions than the principal does, but she suspects that the agent has still more actions available. Thus, the supervisor maximizes a worst-case objective with respect to unknown actions the agent may have; the principal has uncertainty over both the agent's possible actions and the supervisor's knowledge.

(iii) The supervisor knows no more than the principal does; both of them face the same uncertainty about the agent's technology (and both maximize for the worst case).

We outline these three models more carefully in Section 2. All three of them allow a large space of possible actions by the agent, and indeed, all three satisfy our formal Richness condition. Yet, it turns out that linear contracts maximize the principal's worst-case criterion in models (i) and (ii), but not in model (iii) in general. (This will be shown in our later analysis.) Thus, the details of the model matter, and it may not be initially obvious what, beyond Richness, is needed.

Our paper aims to identify the additional condition, at a high level of generality—and, in the process, to obtain a better understanding of the essential ingredients behind the linearity argument. To do this, we abstract away from any particular organizational form. Instead, the principal contracts with a *counterparty* of unspecified structure. The principal's uncertainty about the environment is described by a

3

correspondence Φ, where $\Phi(w)$ specifies the distributions over output that she thinks may potentially arise when she offers contract w. The principal wants to choose w to maximize her expected net profit in the worst case. Throughout, we maintain the background assumptions that the principal is risk-neutral and uses the worst-case criterion; the focus is on understanding the properties of Φ that are central to the linearity argument.

We formalize the *Richness* property by requiring that, whenever some distribution over output is a possible response to a given contract w (i.e. lies in $\Phi(w)$), any other distribution with the same expected output but higher expected payment to the counterparty is also possible. That is, the counterparty has the flexibility to extract maximal payment for a given expected output. In this form, it is clear that Richness by itself cannot pick out linear contracts, since it says nothing about how Φ varies when the contract changes.

The needed additional property that we identify is *Responsiveness*, which expresses that the counterparty's behavior responds to the incentives provided by expected payment. Responsiveness requires that when one contract w is replaced by a new contract w', such that every distribution that might have been chosen in response to w earns a higher expected payment than before, while some other distribution not chosen under w earns a lower payment than before, this unchosen distribution remains unchosen.

We show that the Richness and Responsiveness properties together imply that linear contracts give the best guarantees for the principal.[1] This allows the lesson about the robustness of linear contracts to apply to a broad class of models of contracting with diverse organizational forms. We formally develop the general framework, define Richness and Responsiveness, and present this result in Section 3.

Having noted that, as a supplement to the Richness property, Responsiveness is sufficient to make linear contracts optimal, we next ask if it is also necessary. To address this question, in Section 4, we give a converse result. For this, we develop an auxiliary framework in which contracts specify payment as a function of a physical outcome, and Φ describes the counterparty's behavior in response to such contracts; the principal's value for each possible outcome is a separate parameter of the model.

---

[1]To be precise, in the models we study, it may happen that the optimum of the principal's objective is not attained, so that linear contracts only approach the supremum. For now, we will ignore this distinction.

(For example, one can think of a supplier that can produce different goods, and a contract specifies a payment for each. How the supplier reacts to any given contract is independent of how much the principal values each good.) The Responsiveness property and a strengthened version of Richness can be expressed in this setting. When they hold, our main linearity result immediately implies that the principal can always maximize her guarantee by offering a linear contract, meaning one that pays proportionally to the principal's value for the realized outcome. Our converse shows that, once we ignore certain contracts that can be ruled out a priori as never optimal, if Responsiveness is violated, then there exists a valuation for the principal under which linear contracts are not optimal. This shows that our Responsiveness condition captures, at a formal level, the specific cross-contract restriction on behavior needed for the linearity result.

After presenting the results above, in Section 5 we analyze the robust principal-agent model and hierarchical models (i)–(ii) sketched above in more detail to indicate how the Richness and Responsiveness properties can be verified. (The Supplemental Material, in Section S-1, presents two further applications to illustrate the breadth of our framework.) Section 6 examines hierarchical model (iii), where Responsiveness is violated and linear contracts can fail to be optimal.

A reader might think that our whole exercise is unnecessary because the models where linear contracts turn out to be optimal can be easily reduced to bilateral contracting models anyway. This reaction is misplaced. For example, one might try to reduce hierarchical model (i) to a robust principal-agent model by combining the supervisor and agent into a single entity, whose cost of producing any output distribution is defined as the cost for the supervisor to induce the agent to choose that distribution in the original hierarchical model. However, this reduction fails because it does not preserve the structure of uncertainty needed to apply the result from the robust principal-agent model. In Section 7, we explain this failure in further detail. We also describe what the worst-case environment in the hierarchical model actually looks like; it is quite different from the worst-case in the principal-agent model.

Although the substantive results in this paper concern linear contracts, our conceptual framework more generally offers a way to express and prove results for contracting models without relying on a particular organizational structure. Robustness arguments for linear contracts naturally call for such a framework, because, as Section

5

7 shows, there does not seem to be any easy reduction argument that would allow us to directly extend the results from bilateral environments to more complex ones. In principle, the same methodology is applicable in the analysis of other forms of contracts and properties of contracting models that favor them. To illustrate by example, we demonstrate, in Section S-3 of the Supplementary Material, a result on concave contracts analogous to our main theorem on linear contracts: under Responsiveness and a particular weakening of Richness, concave contracts are optimal.

Our work connects to several branches of literature. First, it naturally relates to the body of work on linear contracts and their robustness against large spaces of actions, mentioned previously. While many of the arguments in this literature are thematically related, which helps motivate our interest in focusing on linear contracts, we certainly do not claim that all of these previous findings are special cases of the results developed here. Also related is the literature explaining other kinds of simple incentive structures as robust in unknown environments, such as Frankel (2014), Garrett (2014), and Carroll and Meng (2016).

There is also considerable previous work on incentives in hierarchies and more complex structures, mostly focusing on comparison across organizational forms (surveyed in Mookherjee (2006), Mookherjee (2013)). (There is a separate and more distant strand of literature on hierarchies, such as Tirole (1986), that focuses on issues of collusion.) Yet there seems to be little work studying how organizational structure interacts with the optimal choice of contractual form.

Finally, closest in spirit to the present work are several recent papers that study robust moral hazard contracting in different organizational environments. In particular, there is the work of Dai and Toikka (2021), which studies robust incentives for a team (and which inspired one of the additional applications explored in the Supplemental Material). Others in this line are Marku and Ocampo Diaz (2019), which takes up a common agency model; and Kambhampati (2022), which considers two agents who produce independently but are known to share a common technology.

2 Overview of Examples

We begin with brief descriptions of our main example applications, meant to give context for the general framework introduced in Section 3. The examples will be presented in formal detail in Section 5.

Robust principal-agent model. In the basic application, the principal contracts directly with an agent, offering a contract that specifies payment as a function of output. Limited liability applies (in this example and throughout the paper): the contract can never pay less than zero. The agent can take any of various actions; an action is modeled as a pair, consisting of a probability distribution over output and a (nonnegative) effort cost incurred by the agent. The principal knows of some set of actions that are definitely available to the agent. But the principal does not know the true production technology, i.e. the set of actions actually available. For any contract she can offer, she evaluates it based on her guaranteed payoff; that is, her expected net profit (after paying the agent) in the worst case over all possible technologies consistent with her knowledge. The guarantee of a contract is typically strictly positive, because the principal knows that the agent is optimizing under the true production technology, so will not take a totally unproductive action if he is known to have a better action available. The analysis of Carroll (2015) showed that the best guarantee for the principal is attained by a linear contract, and we shall recover this result as one instance of our general framework.

Hierarchical model (i). In this model, the principal offers a contract to a supervisor, again specifying (nonnegative) payment as a function of output. The supervisor, after seeing this contract, in turn offers a contract to the agent, also specifying (nonnegative) payment as a function of output. The agent privately chooses his action, output is produced, and then both the supervisor and agent are paid according to their respective contracts.

In this model, we assume that the supervisor knows the agent's technology, so when she writes a contract, she is solving a standard, Bayesian version of a principal-agent problem, in which the "output" produced by the agent is not the output in the original model but rather the payment received by the supervisor. The principal, as before, knows only some actions available to the agent, but does not know the full technology, and evaluates contracts by the worst-case expected payoff over possible technologies.

Hierarchical model (ii). The hierarchical structure is as in the previous model, but now the supervisor's knowledge is different: she may know of actions that the principal does not, but is uncertain as to whether there are still more actions available, and writes her contract with the agent to maximize her own worst-case guarantee. Note that the relationship between the supervisor and the agent is now described

7

by the robust principal-agent model above; this implies that the supervisor has an optimal contract in which she offers the agent some fixed fraction of the payment she receives from the principal.

The principal does not know the full technology, nor how much of it is known by the supervisor, and again uses the worst-case criterion.

Hierarchical model (iii). In this version of the hierarchical model, the supervisor and the principal are symmetrically uninformed: the supervisor knows only as much about the technology as the principal does, and (as in model (ii)) maximizes a worst-case guarantee when contracting with the agent.

Model (iii) can be expressed in the language of our general framework below, but it does *not* satisfy the conditions for our linearity result (in particular, the Responsiveness property is violated), and indeed the result may fail, as we shall show in Section 6.

In Section S-1 of the Supplemental Material, we give two more examples to illustrate the breadth of potential applications for our framework. In the first example, a supervisor contracts with a team of two agents who play differentiated roles in producing output. The second example is a simplified version of the model of Dai and Toikka (2021); the principal contracts directly with a team of agents, and we simply assume that payments to the team are split equally among the agents.

One might argue that these models make some demanding assumptions. The limited liability restriction, which we maintain throughout, may be less natural in the kinds of firm-to-firm settings where hierarchies naturally arise than it would be in contracting with individuals.[2] In addition, hierarchical models (ii) and (iii) impose the worst-case criterion as a positive description of the supervisor's behavior, unlike in the principal-agent model, where worst-case maximization can be viewed simply as a language for expressing statements about robustness properties of contracts. Nonetheless, our goal here is not to write down the most defensible model of contracting in hierarchies, but simply to illustrate that there are various models one could consider, only some of which deliver linear contracts, and thereby to motivate the search for their common features.

---

[2]Limited liability is indeed important. If we instead allowed payments to be arbitrarily negative, we would need to add a participation constraint. If this were done following the approach indicated at the end of Subsection 3.1, one can show that it would always be optimal to use a "selling the firm" contract, giving the counterparty all the output minus some constant.

3   Main Framework and Result

First, some notational conventions. We write $\Delta(X)$ for the space of Borel distributions on a metric space X. We equip $\Delta(X)$ with the weak topology. For $x \in X$, δ_x is the degenerate distribution putting probability 1 on x. We also write $\mathbb{R}^+$ for the set of nonnegative real numbers, and equip it with the usual topology. We write $\mathrm{co}(X)$ for the convex hull of X, when $X \subseteq \mathbb{R}$. We write $C(X)$ for the space of continuous functions from X to $\mathbb{R}$, equipped with the sup-norm, $||f|| = \sup_{x \in X} f(x)$. Recall that when X is compact, $C(X)$ is a Banach space. We write $C^+(X)$ for the subset of $C(X)$ consisting of functions whose values lie in $\mathbb{R}^+$.

3.1   The modeling framework

There is a principal, who contracts with a counterparty, which will subsequently produce (stochastic) output that accrues naturally to the principal. The principal can provide incentives by promising payments to the counterparty.

There is an exogenously given set $Y \subseteq \mathbb{R}$ of possible output values. We assume Y is nonempty and compact, and normalize $\min(Y) = 0$, and denote $\bar{y} = \max(Y)$. A *contract* is a function $w \in C^+(Y)$.[3] Note that this definition incorporates the limited liability restriction: the contract must pay a nonnegative amount.

We are particularly interested in *linear contracts*, which are of the form

$$w_\alpha(y) = \alpha y \qquad \text{for all } y,$$

where $\alpha \geq 0$ is a constant. The special case w_0 is called the *zero contract*.

We take as given a nonempty-valued correspondence $\Phi : C^+(Y) \rightrightarrows \Delta(Y)$, the *outcome correspondence*. $\Phi(w)$ describes the set of distributions over output y that the counterparty may generate in response to contract w, from the principal's point of view. The multiple-valuedness of $\Phi(\cdot)$ thus reflects the principal's uncertainty (about the production technology, or other aspects of the environment). Note that the interpretation of $F \in \Phi(w)$ is not simply that distribution F may be physically feasible, but rather that it might actually occur in response to contract w. For

---

[3]The continuity assumption on contracts is not actually needed for the linearity result; we impose it only to guarantee existence of best responses in the applications, so that everyone's behavior is well-defined. In any case, it has no bite when Y is an arbitrarily fine discrete grid, so we do not view it as a substantive restriction.

example, if the principal knows that the counterparty is able to produce output 0 with probability 1, but would never do so in response to w because some other distribution is better-incentivized, then we would have $\delta_0 \notin \Phi(w)$. For now we treat the correspondence Φ as exogenously given; in each of the individual applications in Section 5, we will in turn define Φ from more primitive objects.

Any contract w is then evaluated by its worst-case guarantee for the principal across environments. Since the principal's ex-post payoff equals the output she receives minus the payment made to the counterparty, the relevant criterion is

$$V_P(w) = \inf_{F \in \Phi(w)} \mathbb{E}_F[y - w(y)].$$

We now consider the following properties that Φ may have.

Richness. *Suppose $w \in C^+(Y)$, $F \in \Phi(w)$, and $F' \in \Delta(Y)$ is another distribution such that $\mathbb{E}_{F'}[y] = \mathbb{E}_F[y]$ and $\mathbb{E}_{F'}[w(y)] \geq \mathbb{E}_F[w(y)]$. Then, $F' \in \Phi(w)$.*

This property essentially says that the set of possible responses to a given contract is sufficiently broad: for any distribution that the counterparty might produce, any other distribution with the same expected output but higher average payment to the counterparty is also possible. Even more simply put, for any given expected output, the principal worries that the counterparty will extract the highest possible average payment.

Responsiveness. *Suppose $w, w' \in C^+(Y)$, and $F \in \Delta(Y)$ such that $F \notin \Phi(w)$. Suppose that $\mathbb{E}_{F'}[w'(y)] \geq \mathbb{E}_{F'}[w(y)]$ for all $F' \in \Phi(w)$, while $\mathbb{E}_F[w'(y)] \leq \mathbb{E}_F[w(y)]$. Then, $F \notin \Phi(w')$.*

This property expresses how the possible outcomes respond to the incentives provided by expected payment. If an "old" contract w is replaced by a "new" contract w', such that any distribution that the counterparty might have produced under the old contract now pays more (in expectation) than before, while some other distribution F pays less than before, the counterparty will not switch to choosing F.

One way to understand Responsiveness is to consider a standard principal-agent problem without uncertainty: The counterparty is a single agent, and there is some fixed, mutually known set of output distributions F that he can produce, each with an associated cost $c(F)$. When the principal offers contract w, the agent chooses F to maximize $\mathbb{E}_F[w(y)] - c(F)$. Thus $\Phi(w)$ is the set of maximizers F. This model satisfies Responsiveness: Consider any w, w', and $F \notin \Phi(w)$ for which the hypotheses of

Responsiveness are satisfied. If F is not even feasible, clearly $F \notin \Phi(w')$. Otherwise, F is feasible but not optimal under w. Then, let F' be an optimal choice. So $\mathbb{E}_{F'}[w(y)] - c(F') > \mathbb{E}_F[w(y)] - c(F)$. Since $\mathbb{E}_{F'}[w'(y)] \geq \mathbb{E}_{F'}[w(y)]$ while $\mathbb{E}_F[w'(y)] \leq \mathbb{E}_F[w(y)]$, we have $\mathbb{E}_{F'}[w'(y)] - c(F') > \mathbb{E}_F[w'(y)] - c(F)$; that is, F remains non-optimal under w'.

Intuitively, we would expect Responsiveness to be satisfied when the counterparty is a single agent who maximizes expected value as in the example above, or more generally, when the counterparty has a "leader" who understands the environment and maximizes expected value (such as the supervisor in hierarchical model (i)). But it also turns out to be satisfied in some other models, such as hierarchy (ii) where the leader is not expected-value-maximizing. We discuss further in Subsection 5.3.

For simplicity, we have not included a participation constraint. An earlier version of the paper (Walton and Carroll, 2019) describes how such a constraint can be accommodated, by restricting the principal's choice of contracts to a subset of $C^+(Y)$, interpreted as the set of contracts that the agent is sure to accept, and assuming an appropriate structure on this subset.

3.2   Linearity result

Now we come to the first main result: our conditions are sufficient for optimality of linear contracts.

Theorem 1. *Suppose the correspondence $\Phi(\cdot)$ has the Richness and Responsiveness properties. Then, for any contract w, there is a linear contract w' such that $V_P(w') \geq V_P(w)$.*

We prove Theorem 1 constructively, by using the worst-case scenario under w to determine the slope of w'. The argument is illustrated in Figure 1. First, for any given level of expected output, say μ, Richness ensures that the principal is worried about the highest expected payment ν; this highest expected payment is given by the concavification of w, call it $\hat{w}$. Concavity of $\hat{w}$ implies that the ratio $\hat{w}(\mu)/\mu$, the expected payment per dollar of expected output, is decreasing in μ. So, among all distributions in $\Phi(w)$, the one with the lowest expected output μ is also the one for which the fraction of output ceded to the agent is highest; hence, this must be the worst-case distribution. The resulting outcome is shown as point (μ^*, ν^*) in Figure 1. We then take w' to be the linear contract that passes through this same point. Under
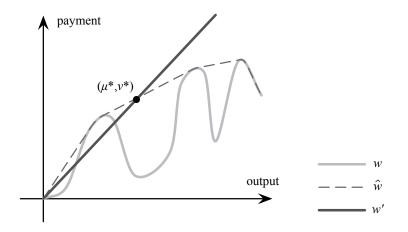
Figure 1: Constructing a linear contract w' with as good or better guarantee than some initial contract w.

w', the payment-per-output ratio is constant, so this new contract both pays more than the old one for higher expected output and pays less for lower expected output. By Responsiveness, w' can only motivate the counterparty to produce higher expected output than w—which in turn means higher expected profit for the principal, since linearity ensures that expected output and expected profit are aligned.

The above proof sketch is imprecise about the distinction between weak and strict inequalities, and also implicitly assumes that the inf in the definition of $V_P(w)$ is attained, which it may not be. The full proof below fills in these gaps. While we have used the concavification $\hat{w}$ for intuition, the formal proof does not need to refer to it.

Proof. We may assume that $V_P(w) > 0$, since otherwise we can just take w' to be the zero contract. Now, let $\lambda = \sup_{F \in \Phi(w)}(\mathbb{E}_F[w(y)]/\mathbb{E}_F[y])$. Note that this expression is well-defined, since every $F \in \Phi(w)$ satisfies $\mathbb{E}_F[y] \geq V_P(w) > 0$. In fact, for every $F \in \Phi(w)$, we have $\mathbb{E}_F[w(y)]/\mathbb{E}_F[y] = 1 - (\mathbb{E}_F[y - w(y)]/\mathbb{E}_F[y]) \leq 1 - V_P(w)/\bar{y}$, and thus λ is bounded above by $1 - V_P(w)/\bar{y} < 1$.

Now define the linear contract $w'(y) = \lambda y$. We will show that $V_P(w') \geq V_P(w)$.

Let $F_1, F_2, \ldots$ be a sequence of distributions in $\Phi(w)$ approaching the inf in the definition of $V_P(w)$: $\mathbb{E}_{F_k}[y - w(y)] \to V_P(w)$. By taking a subsequence, we can assume that F_k converges to some limiting distribution F^*. Put $\mu_k = \mathbb{E}_{F_k}[y]$, $\mu^* = \mathbb{E}_{F^*}[y]$ and $\nu^* = \mathbb{E}_{F^*}[w(y)]$. Since $\mathbb{E}_{F_k}[w(y)] \leq \lambda \mathbb{E}_{F_k}[y]$ for each k, we have $\nu^* \leq \lambda \mu^*$. We claim the equality must hold. If not, pick λ' with $\nu^*/\mu^* < \lambda' < \lambda$. By definition of λ,

there exists $F' \in \Phi(w)$ such that $\mathbb{E}_{F'}[w(y)] > \lambda' \mathbb{E}_{F'}[y]$. Since $\mathbb{E}_{F'}[y - w(y)] \geq V_P(w)$, we have

$$\mathbb{E}_{F'}[y] > \frac{\mathbb{E}_{F'}[y - w(y)]}{1 - \lambda'} \geq \frac{V_P(w)}{1 - \lambda'} = \frac{\mu^* - \nu^*}{1 - \lambda'} > \mu^*.$$

Hence, for sufficiently large k, we have $\mathbb{E}_{F'}[y] > \mu_k$ as well. Take $\tilde{F}$ to be an appropriate convex combination of δ_0 and F' so that $\mathbb{E}_{\tilde{F}}[y] = \mu_k = \mathbb{E}_{F_k}[y]$. Then we have $\mathbb{E}_{\tilde{F}}[w(y)] \geq \lambda' \mathbb{E}_{\tilde{F}}[y]$ (since the latter inequality holds both for the component F' and, trivially, for δ_0). Hence, either $\mathbb{E}_{F_k}[w(y)] \geq \lambda' \mu_k$, or else we can apply Richness to the distributions F_k and $\tilde{F}$ to conclude $\tilde{F} \in \Phi(w)$. Either way, $\Phi(w)$ contains a distribution with expected output μ_k and expected payment at least $\lambda' \mu_k$. This means that $V_P(w) \leq \mu_k - \lambda' \mu_k$. But taking $k \to \infty$ gives $V_P(w) \leq \mu^* - \lambda' \mu^* < \mu^* - \nu^* = V_P(w)$, a contradiction. Thus, we conclude $\nu^* = \lambda \mu^*$ as claimed.

This implies that $V_P(w) = \mu^* - \nu^* = (1 - \lambda)\mu^*$. We can complete the proof by showing that $V_P(w') \geq (1-\lambda)\mu^*$. Since every distribution F satisfies $\mathbb{E}_F[y - w'(y)] = (1 - \lambda)\mathbb{E}_F[y]$, it suffices to show that $\mathbb{E}_F[y] \geq \mu^*$ for every $F \in \Phi(w')$.

Consider any distribution $\tilde{F}$ such that $\mathbb{E}_{\tilde{F}}[y] < \mu^*$; we need to show $\tilde{F} \notin \Phi(w')$. Put $\mu = \mathbb{E}_{\tilde{F}}[y]$. Let F be a convex combination of δ_0 and F^* such that $\mathbb{E}_F[y] = \mu$. We have $\mathbb{E}_F[w(y)] \geq \lambda \mathbb{E}_F[y]$, since this inequality holds both for the component F^* and (trivially) for δ_0. Since $\mathbb{E}_F[y - w(y)] \leq (1 - \lambda)\mathbb{E}_F[y] < (1 - \lambda)\mu^* = V_P(w)$, we have $F \notin \Phi(w)$. Moreover, $\mathbb{E}_F[w(y)] \geq \lambda \mathbb{E}_F[y] = \mathbb{E}_F[w'(y)]$, while for every $F' \in \Phi(w)$, we have $\mathbb{E}_{F'}[w(y)] \leq \lambda \mathbb{E}_{F'}[y] = \mathbb{E}_{F'}[w'(y)]$; thus, Responsiveness implies $F \notin \Phi(w')$. Since F has the same expected output and the same expected payment under w' as $\tilde{F}$ does, Richness then implies $\tilde{F} \notin \Phi(w')$.

$\square$

This shows that Richness and Responsiveness are sufficient for linear contracts; but are they necessary? In Section 4 we will give a converse result that aims toward addressing this question.

For the moment, we simply note that neither property can be dropped entirely. Richness alone would not give us the result, since we clearly need some assumption on how $\Phi(w)$ varies with w. For a more concrete example, in Section 6 we will note that in hierarchical model (iii), Φ satisfies Richness, but the conclusion of Theorem 1 can fail.

To see that Responsiveness alone is not sufficient, just consider a standard principal-agent problem without uncertainty, as was used to illustrate Responsiveness above.

As is well-known, usually a nonlinear contract is strictly optimal. For example, under standard specifications with a discrete output space and just two possible distributions F, an optimal contract pays only for the one realization of output that achieves the highest likelihood ratio, and pays zero for all other realizations.[4]

3.3 Existence of optimum

We have still been imprecise on one point: The verbal interpretation given to Theorem 1 is that a linear contract is optimal for the principal. Indeed, *if* an optimal contract exists, then there is one that is linear. However, it may happen that no optimal contract exists. In this case, under the conditions of Theorem 1, the supremum payoff $\sup_{w \in C^+(Y)} V_P(w)$ is approached, but not attained, by linear contracts.

It can be useful to have a handy way to check that existence is indeed satisfied in any given model. Define the correspondence $\widetilde{\Phi} : [0,1] \rightrightarrows \Delta(Y)$ by $\widetilde{\Phi}(\alpha) = \Phi(w_\alpha)$ (recall that w_α was the linear contract of slope α).

Proposition 2. *Suppose that Φ satisfies Richness and Responsiveness. If moreover $\widetilde{\Phi}$ is lower hemicontinuous, then there exists a contract maximizing V_P (and in fact, the maximum is attained by a linear contract).*

The proof (a straightforward limiting argument) is in Appendix A. In the examples in Section 5, we use this result to show that an optimal contract exists.

4 A Converse Result

We have argued that, given Richness, the addition of Responsiveness is sufficient for optimal contracts to be linear. To argue that we have really identified the right condition, we should show that Responsiveness is necessary as well. Of course, in the framework so far, this cannot be exactly right: many contracts are clearly far from optimal (for example, any contract that always pays more than the value of output), so violations of Responsiveness among such contracts are irrelevant. This suggests we should look for a more general framework, in which Responsiveness can be defined

---

[4]To be precise, in order for the optimal contract to exist, we should modify the model by specifying that whenever the agent is indifferent between multiple actions, he chooses the one preferred by the principal. For simplicity we have skipped over this here. We do make the analogous tie-breaking provision (and show in detail that Responsiveness still holds) for the applications in Section 5.

at the level of a class of models, so that Responsiveness becomes necessary to ensure that all instances within the class deliver linear contracts.

Specifically, we will now consider a framework in which output is not directly measured in payoff units. Instead, the counterparty produces "physical" outputs: for example, the counterparty may be a supplier that can produce different types of goods. Contracts specify payment as a function of the physical output. The principal, in turn, derives some monetary value from each possible physical output. The principal's valuation of outputs is now an additional parameter in the model; it matters for the principal's preferences but is irrelevant to the counterparty's behavior. Here, Responsiveness will be sufficient to ensure that, no matter what this valuation is, a linear contract (one that pays proportionally to the principal's value) is optimal, and we will argue that Responsiveness is essentially necessary for this conclusion as well.

Thus, for this section only, we consider a given nonempty set Z of *physical outputs*. We take Z to be finite (and endowed with the discrete metric), and we denote a typical element by z. A *contract* is now a function $w \in C^+(Z)$. We take as given a nonempty-valued *outcome correspondence* $\Phi : C^+(Z) \rightrightarrows \Delta(Z)$.

We reformulate the Richness property for this setting as follows:

Strong Richness.

(a) Suppose $w \in C^+(Z)$, $F \in \Phi(w)$, and $F' \in \Delta(Z)$ such that $\mathbb{E}_{F'}[w(z)] \geq \mathbb{E}_F[w(z)]$. Then, $F' \in \Phi(w)$.

(b) For every $w \in C^+(Z)$, the set $\Phi(w) \subseteq \Delta(Z)$ is closed.

Part (a) of this property is stronger than our original Richness property because it drops the restriction that F and F' should have the same expected output; this restriction cannot be formulated when output does not have a numeric value. We do not view this loss as a major sacrifice. In our applications in Section 5, this restriction matters only because it allows us to include tie-breaking assumptions on the counterparty's behavior (e.g. assuming that if the agent is indifferent between multiple distributions, he chooses the one that is better for the principal); without such assumptions, an optimal contract can sometimes fail to exist. Part (b) of Strong Richness is a technical condition that helps rule out some inconvenient boundary cases.

Together, parts (a) and (b) imply that, for each w, there exists a *threshold* $h(w) \in [\min_z w(z), \max_z w(z)]$ such that $\Phi(w) = \{F \in \Delta(Z) \mid \mathbb{E}_F[w(z)] \geq h(w)\}$.

We can formulate Responsiveness exactly as in our main framework, since it made no reference to the value of output:

Responsiveness. *Suppose $w, w' \in C^+(Z)$, and $F \in \Delta(Z)$ such that $F \notin \Phi(w)$. Suppose that $\mathbb{E}_{F'}[w'(z)] \geq \mathbb{E}_{F'}[w(z)]$ for all $F' \in \Phi(w)$, while $\mathbb{E}_F[w'(z)] \leq \mathbb{E}_F[w(z)]$. Then, $F \notin \Phi(w')$.*

We define a *valuation* to be a function $v : Z \to \mathbb{R}^+$ such that $\min_z v(z) = 0$. For any valuation and any contract, the principal's guarantee is then defined as

$$V_P(w|v) = \inf_{F \in \Phi(w)} \mathbb{E}_F[v(z) - w(z)].$$

Setting the minimum value of v to 0 is just a normalization that brings this setting into line with the assumption $\min(Y) = 0$ in our previous framework.

We can now say that a contract w is *linear given* v if there exists a constant $\alpha \geq 0$ such that $w(z) = \alpha v(z)$ for all z.

We will also say that w is *grounded* if $\min_z w(z) = 0$. We will focus our attention on grounded contracts (note that, given any v, any linear contract is grounded). This is justified if, for example, $\Phi(w) = \Phi(w')$ whenever w and w' are two contracts that differ by a constant: then, if a contract is not grounded, the principal can subtract a constant from it without changing incentives and so improve her own payoff. (All of our example applications satisfy this property of invariance to constant translations.)

Suppose that Φ satisfies Strong Richness and Responsiveness. Theorem 1 implies that, given any valuation v, for any contract w, there is a contract w' that is linear given v and satisfies $V_P(w'|v) \geq V_P(w|v)$.

A conjecture might be that this conclusion fails whenever Responsiveness is violated: that is, if Φ satisfies Strong Richness but not Responsiveness, there exists some choice of valuation v for which a nonlinear contract gives a strictly higher guarantee than any linear one. We will show a slightly weaker version of this statement: Given Φ, some contracts can be quickly ruled out as never optimal regardless of v, except in degenerate cases. We can think of these contracts as "irrelevant." We will show that if there is a violation of Responsiveness involving "relevant" (and grounded) contracts, then v can be chosen so that a nonlinear contract does better than linear.

Given Φ satisfying Strong Richness, let h be the threshold function defined above;

and, for $w \in C^+(Z)$ and $\beta \geq 0$, write βw for the contract obtained by scaling w pointwise by β.

Now say that a contract w is *scaling-dominated* if, for every v such that $V_P(w|v) > 0$, there exists some $\beta \geq 0$ such that $V_P(\beta w|v) > V_P(w|v)$. These are the contracts we regard as irrelevant. They can be identified as non-optimal using only a very small part of the correspondence Φ (namely, its values on the scalar multiples βw). Thus, we can identify whether w is scaling-dominated using only the values of h on its scalar multiples, and the next proposition characterizes explicitly when this happens.

Proposition 3. *Assume Strong Richness, and let w be a grounded contract. Then, w is scaling-dominated if and only if it satisfies at least one of the following five conditions:*

   (i) $h(w) = 0$.

  (ii) *There exists $0 < \beta < 1$ such that $h(\beta w) > \beta h(w)$.*

 (iii) *For every positive number K, there exists $0 < \beta < 1$ such that $h(w) - h(\beta w) > K(h(w) - h(\beta w)/\beta)$.*

 (iv) *For every number $K > 1$, there exists $\beta > 1$ such that $h(\beta w) > \beta h(w)$ and $h(\beta w) - h(w) < K(h(\beta w)/\beta - h(w))$.*

  (v) *There exist $0 < \underline{\beta} < 1$ and $\overline{\beta} > 1$ such that $h(\overline{\beta}w)/\overline{\beta} > h(w) > h(\underline{\beta}w)/\underline{\beta}$ and*

$$\frac{h(w) - h(\underline{\beta}w)}{h(w) - h(\underline{\beta}w)/\underline{\beta}} > \frac{h(\overline{\beta}w) - h(w)}{h(\overline{\beta}w)/\overline{\beta} - h(w)}.$$

Moreover, if none of these conditions is satisfied, we can choose v such that w is linear given v, $V_P(w|v) > 0$, and $V_P(w|v) \geq V_P(\beta w|v)$ for all $\beta \geq 0$.

With our focus on contracts that are not scaling-dominated, we can now give our statement on the necessity of Responsiveness.

Theorem 4. *Assume Φ satisfies Strong Richness but fails to satisfy Responsiveness between some two contracts w, w', where w' is grounded and not scaling-dominated. Then, the valuation v can be chosen so that w gives a strictly higher guarantee than any linear contract.*

The proofs of Proposition 3 and Theorem 4 are left to Appendix A; we just give sketches here.

For the "if" direction of Proposition 3, we go through the conditions one by one. In each case, we use the statement of the condition to identify the relevant rescaling βw that does better than w. (In conditions (iii) and (iv), where β depends on K, we must first choose K appropriately depending on the valuation v.) For the "only if" direction, it suffices to prove the last statement of the proposition; thus we wish to choose v to be proportional to the given contract w (thus making w linear) and pick the constant of proportionality so that w is in fact optimal among linear contracts. Writing down the conditions needed for this to happen, it turns out that we run into difficulty precisely when one of (i)–(v) holds.

For Theorem 4, the key observation is that when Responsiveness is violated between contracts w, w' and w' is linear, then w gives a strictly better guarantee than w' does. (Refer back to Figure 1: a violation of Responsiveness means that some low-output distributions are possible under w' but not under w.) Thus, to prove Theorem 4, we can invoke the last statement of Proposition 3 to choose v such that the given w' is optimal among linear contracts. The foregoing observation then implies that w does strictly better than w', and thus better than any linear contract.

As a final remark, we note that the lengthy conditions in Proposition 3 can be somewhat simplified if we assume that Φ satisfies the following additional property:

Scaling Monotonicity. *For any $w \in C^+(Z)$ and any $\beta > 1$, $\Phi(\beta w) \subseteq \Phi(w)$.*

Given Strong Richness, this property is equivalent to the statement that $h(\beta w)/\beta$ is weakly increasing in β. In words, it says that the counterparty responds positively to rescaling of incentives (but does not say anything about responses to changes in the shape of incentives). It is satisfied in the robust principal-agent model and all three versions of the hierarchical model.

If Scaling Monotonicity is imposed, then condition (ii) of the proposition can never arise, and conditions (iii)–(v) can be rewritten respectively as (iii) $\underline{R}(w) = \infty$, (iv) $\overline{R}(w) = 1$, (v) $\underline{R}(w) > \overline{R}(w)$, where

$$\underline{R}(w) = \sup_{0 < \underline{\beta} < 1} \frac{h(w) - h(\underline{\beta}w)}{h(w) - h(\underline{\beta}w)/\underline{\beta}}, \qquad \overline{R}(w) = \inf_{\overline{\beta} > 1} \frac{h(\overline{\beta}w) - h(w)}{h(\overline{\beta}w)/\overline{\beta} - h(w)}.$$

5 Applications

We now return to our main framework where output $y \in Y$ is directly measured in payoff units. We proceed to detail the various applications that were previewed in Section 2, showing how they are instances of the general framework. We indicate how to verify Richness and Responsiveness, as well as lower hemicontinuity. (Some of the formalities are deferred to the Supplemental Material.) Hierarchical model (iii) does not satisfy Responsiveness and so is left to the next section.

5.1 Robust Principal-Agent Model

In this model, the counterparty consists of a single agent. An *action* the agent may take is modeled as a pair $(F, c) \in \Delta(Y) \times \mathbb{R}^+$. If action (F, c) is taken, output is drawn according to the distribution F and the agent incurs an effort cost of c. We define a *technology* to be a nonempty, compact subset of $\Delta(Y) \times \mathbb{R}^+$, interpreted as the set of actions available to the agent. Given a contract $w \in C^+(Y)$ and technology $\mathcal{A}$, the agent maximizes objective

$$V_A(F, c|w) = \mathbb{E}_F[w(y)] - c$$

over $(F, c) \in \mathcal{A}$. We assume that the principal doesn't know $\mathcal{A}$. Instead, there is an exogenously given technology $\mathcal{A}^P$, representing all actions that are known by the principal to be available to the agent. The agent's actual technology is known only to satisfy $\mathcal{A} \supseteq \mathcal{A}^P$.

Given this description of the model, we can define the outcome correspondence. Let $\Gamma_A(w, \mathcal{A}) = \{F \in \Delta(Y) | \exists c \geq 0, (F, c) \in \arg\max_{\mathcal{A}} V_A(\cdot, \cdot|w)\}$; this is the set of distributions that can result when the agent maximizes V_A over $\mathcal{A}$ given w. Continuity of w and compactness of $\mathcal{A}$ ensure that $\Gamma_A(w, \mathcal{A})$ is nonempty. Finally, as a tie-breaking condition, we assume that when the agent is indifferent among actions, he chooses the one best for the principal; we call such actions *principal-preferred*. Formally, this set is denoted as $\Gamma_A^P(w, \mathcal{A}) = \arg\max_{F \in \Gamma_A(w, \mathcal{A})} \mathbb{E}_F[y - w(y)]$. This assumption helps to ensure that an optimal contract w exists (discussed more momentarily). Now the

outcome correspondence is defined as

$$\Phi^{PA}(w) = \bigcup_{\text{technology } \mathcal{A} \supseteq \mathcal{A}^P} \Gamma_A^P(w, \mathcal{A}).$$

The principal then evaluates contracts according to $V_P^{PA}(w) = \inf_{F \in \Phi^{PA}(w)} \mathbb{E}_F[y - w(y)]$.

This is the same model as the one considered in Carroll (2015). In our framework, we reproduce the main result of that paper, by verifying that Richness and Responsiveness hold in this model. We also verify that the restricted correspondence $\widetilde{\Phi}$ is lower hemicontinuous, so a maximizing linear contract exists; this existence is needed later when we embed this model in a principal-supervisor-agent hierarchy, as it ensures that the supervisor's behavior is well-defined.

Proposition 5. *There exists a linear contract maximizing V_P^{PA}.*

Essentially, Richness holds because, for any F that might be chosen for some technology, the more-remunerative F' might then also be chosen if it turned out to also be available. Responsiveness holds by the same argument as in the principal-agent model without uncertainty sketched in Subsection 3.1, repeated for each possible technology. The formal proof ends up a bit lengthy because of tie-breaking technicalities.

Proof. (Richness) Let $w \in C^+(Y)$, $F \in \Phi^{PA}(w)$, so that there exists a technology $\mathcal{A} \supseteq \mathcal{A}^P$ containing action (F, c) such that $\mathbb{E}_F[w(y)] - c \geq \mathbb{E}_{\tilde{F}}[w(y)] - \tilde{c}$ for all $(\tilde{F}, \tilde{c}) \in \mathcal{A}$. Let $F' \in \Delta(Y)$ such that $\mathbb{E}_F[y] = \mathbb{E}_{F'}[y]$ and $\mathbb{E}_F[w(y)] \leq \mathbb{E}_{F'}[w(y)]$. Consider an alternative technology $\mathcal{A}' = \mathcal{A} \cup \{(F', 0)\}$. Then

$$\mathbb{E}_{F'}[w(y)] - 0 \geq \mathbb{E}_F[w(y)] - c \geq \mathbb{E}_{\tilde{F}}[w(y)] - \tilde{c}$$

for all $(\tilde{F}, \tilde{c}) \in \mathcal{A} = \mathcal{A}' \setminus \{(F', 0)\}$, so $F' \in \Gamma_A(w, \mathcal{A}')$. If $\mathbb{E}_F[w(y)] = \mathbb{E}_{F'}[w(y)]$, then F being principal-preferred implies that F' is principal-preferred in $\mathcal{A}'$. Otherwise, $\mathbb{E}_F[w(y)] < \mathbb{E}_{F'}[w(y)]$, and hence the agent strictly prefers taking action $(F', 0)$ to all other actions in $\mathcal{A}'$, so F' is principal-preferred since it is the only element of $\Gamma_A(w, \mathcal{A}')$. So $F' \in \Gamma_A^P(w, \mathcal{A}') \subseteq \Phi^{PA}(w)$.

(Responsiveness) Let $w, w' \in C^+(Y)$ and $F \notin \Phi^{PA}(w)$ satisfy the conditions of the Responsiveness property. If we can show for any technology $\mathcal{A} \supseteq \mathcal{A}^P$ that $F \notin \Gamma_A^P(w', \mathcal{A})$, then $F \notin \Phi^{PA}(w')$. Take any technology $\mathcal{A} \supseteq \mathcal{A}^P$ containing (F, c)

for some $c \geq 0$, and take $(F', c') \in \Gamma_A^P(w, \mathcal{A})$. By the hypothesis of Responsiveness and agent optimality,

$$\mathbb{E}_{F'}[w'(y)] - c' \geq \mathbb{E}_{F'}[w(y)] - c' \geq \mathbb{E}_F[w(y)] - c \geq \mathbb{E}_F[w'(y)] - c. \tag{1}$$

If any of these equalities are strict, $F \notin \Gamma_A(w', \mathcal{A})$. Otherwise, they are all equalities. In this case, it is possible for $F \in \Gamma_A(w', \mathcal{A})$ only if $F \in \Gamma_A(w, \mathcal{A})$ and $F' \in \Gamma_A(w', \mathcal{A})$ as well. If this holds, principal tie-breaking under w and the hypothesis of Responsiveness along with the equalities in (1) imply

$$\mathbb{E}_{F'}[y - w'(y)] = \mathbb{E}_{F'}[y - w(y)] > \mathbb{E}_F[y - w(y)] = \mathbb{E}_F[y - w'(y)]$$

in which case F is not principal-preferred under w', so $F \notin \Gamma_A^P(w', \mathcal{A})$.

Now we have verified Richness and Responsiveness, so by Theorem 1, we can restrict to linear contracts when maximizing V_P^{PA}. It remains to verify that $\widetilde{\Phi}^{PA}$ is lower hemicontinuous.

(Lower Hemicontinuity) Let $\alpha \in [0, 1]$, $F \in \widetilde{\Phi}^{PA}(\alpha)$, and let $\mathcal{O}$ be an open neighborhood of F in $\Delta(Y)$. We want to show that there exists $\eta > 0$ such that $\alpha' \in \mathcal{B}_\eta(\alpha)$ implies that $\widetilde{\Phi}^{PA}(\alpha') \cap \mathcal{O}$ is nonempty, where $\mathcal{B}_\eta(\alpha)$ is the Euclidean ball of radius η around α restricted to $[0, 1]$.

If $F = \delta_{\bar{y}}$, then any technology $\mathcal{A}$ containing $(F, 0)$ has $F \in \Gamma_A(w_{\alpha'}, \mathcal{A})$ for any $\alpha' \in [0, 1]$, and F is principal-preferred, giving $F \in \widetilde{\Phi}^{PA}(\alpha')$ already. So we can assume that $F \neq \delta_{\bar{y}}$. Then, choose $F' \in \mathcal{O}$ such that $\mathbb{E}_{F'}[y] > \mathbb{E}_F[y]$.

Let $\mathcal{A}$ be a technology for which the agent produces distribution F. We can assume without loss that $(F, 0) \in \mathcal{A}$. By Berge's Theorem, $f^* : [0, 1] \to \mathbb{R}$ defined by $f^*(\alpha) = \max_{(\tilde{F}, \tilde{c}) \in \mathcal{A}} (\alpha \cdot \mathbb{E}_{\tilde{F}}[y] - \tilde{c})$ is continuous.

We claim that there exists some η such that $\alpha' \in \mathcal{B}_\eta(\alpha) \setminus \{0\}$ implies $f^*(\alpha') < \alpha' \mathbb{E}_{F'}[y]$. If $\alpha > 0$, then this claim follows from the fact that $f^*(\alpha) = \alpha \mathbb{E}_F[y] < \alpha \mathbb{E}_{F'}[y]$ together with continuity of f^*. So focus on $\alpha = 0$, and suppose no such η exists. Then there is some sequence of $\alpha'_n \to 0$ and corresponding optimal actions $(F_n, c_n) \in \mathcal{A}$, which we can assume (by taking a subsequence) converge to $(F^*, 0) \in \mathcal{A}$, where $\mathbb{E}_{F^*}[y] \geq \mathbb{E}_{F'}[y] > \mathbb{E}_F[y]$. This contradicts that F has largest mean among zero-cost actions in $\mathcal{A}$ (which follows from principal-preferred tie-breaking).

Hence, for any $\alpha' \in \mathcal{B}_\eta(\alpha) \setminus \{0\}$, constructing new technology $\mathcal{A}' = \mathcal{A} \cup \{(F', 0)\}$ yields $(F', 0)$ as the unique maximizer of $V_A(\cdot, \cdot | w_{\alpha'})$ over $\mathcal{A}'$, so $\Gamma_A(w_{\alpha'}, \mathcal{A}') = \Gamma_A^P(w_{\alpha'}, \mathcal{A}')$

$= \{F'\}$, and $F' \in \widetilde{\Phi}^{PA}(\alpha') \cap \mathcal{O}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

One comment on interpretation: The above proof relies (as do many others later) on adding an arbitrary action of the form $(F, 0)$ to the technology. It may seem unrealistic to allow the agent to produce large amounts of output at zero cost. However, the fact that the unknown actions are totally unrestricted is not crucial; the logic can be carried over to more detailed models that incorporate lower bounds on plausible effort costs. (See Carroll (2015), section II.A, for more details.)

5.2 Hierarchical Model (i)

The three hierarchical models which we analyze have the following structure. The principal contracts with a supervisor, who, after observing this contract, writes a contract with an agent. We assume that, for reasons outside the model, the principal cannot contract directly with the agent. We assume that the supervisor does not directly affect production in any way; the only role the supervisor plays is as an intermediary between the principal and the agent. Technology for the agent is the same as in Subsection 5.1. The contract from the principal to the supervisor is the w of our general framework; the contract from the supervisor to the agent is denoted w_A, and we assume both contracts depend solely on output, so that $w, w_A \in C^+(Y)$.

The agent's objective is the same as in the robust principal-agent model, but now the agent receives payment from the supervisor, not directly from the principal. Thus, given contract w_A and technology $\mathcal{A}$, the agent maximizes objective $V_A(F, c|w_A)$ over $\mathcal{A}$.

In all versions of the hierarchical model, we assume that the principal doesn't know $\mathcal{A}$. Like the robust principal-agent model, there is an exogenously given technology $\mathcal{A}^P$, representing all actions known by the principal to be available to the agent. Let $\Gamma_A(w_A, \mathcal{A})$ be defined as before, noting that w_A refers to the contract between supervisor and agent (henceforth "S-A contract").

In hierarchical model (i), we assume that the supervisor is perfectly informed of $\mathcal{A}$, which must include the actions known to the principal; that is, $\mathcal{A} \supseteq \mathcal{A}^P$. The supervisor wants to maximize the expected difference between payments from the principal and payments to the agent. In addition, we restrict the set of permitted S-A contracts to some exogenously given compact set $\mathcal{S} \subseteq C^+(Y)$, which is assumed to contain all linear contracts with slope in the interval $[0, 1]$. This assumption is

necessary to ensure the supervisor always has a best reply. (Without some such restriction, one can find situations, similar to Mirrlees (1999), in which no optimal contract for the supervisor exists. See the earlier version, Walton and Carroll (2019), for an example and more discussion.)

To formally specify the supervisor's behavior, first, for any w, w_A and $\mathcal{A}$, define $\Gamma_A^S(w, w_A, \mathcal{A}) = \arg\max_{F \in \Gamma_A(w_A, \mathcal{A})} \mathbb{E}_F[w(y) - w_A(y)]$. Thus Γ_A^S is the set of distributions that are best for the supervisor among the agent's optimal choices. This again represents a tie-breaking condition, and we refer to elements of Γ_A^S as *supervisor-preferred*. The supervisor's objective in hierarchical model (i) is then

$$V_S^i(w_A|w, \mathcal{A}) = \mathbb{E}_{\Gamma_A^S(w, w_A, \mathcal{A})}[w(y) - w_A(y)].$$

The subscript in $\mathbb{E}_{\Gamma_A^S(w, w_A, \mathcal{A})}$ is a slight abuse of notation: this is a set of distributions, not a single distribution, but the expectation is well-defined since it is the same for all distributions in this set (and the set is nonempty; see the proof of Lemma 6 below). The "i" in V_S^i stands for "informed."

We impose one more layer of tie-breaking—favoring *principal-preferred* actions—in order to achieve lower hemicontinuity of the outcome correspondence. Explicitly, define $\Gamma_A^{PS}(w, w_A, \mathcal{A}) = \arg\max_{F \in \Gamma_A^S(w, w_A, \mathcal{A})} \mathbb{E}_F[y - w(y)]$. Now define

$$\Gamma_S^i(w, \mathcal{A}) = \bigcup_{w_A \in \arg\max_S V_S^i(\cdot|w, \mathcal{A})} \Gamma_A^{PS}(w, w_A, \mathcal{A}).$$

In words, for the fixed principal-supervisor ("P-S") contract w and true technology $\mathcal{A}$, this is the set of output distributions that are possible, given that the supervisor is choosing contract $w_A \in \mathcal{S}$ to optimize V_S^i, and the agent is maximizing given w_A, along with the tie-breaking conditions.

Finally, the outcome correspondence in hierarchical model (i) is defined as

$$\Phi^{PSA(i)}(w) = \bigcup_{\text{technology } \mathcal{A} \supseteq \mathcal{A}^P} \Gamma_S^i(w, \mathcal{A})$$

and the principal's corresponding objective is denoted $V_P^{PSA(i)}(w)$.

This completes the description of the model. We should make sure $\Phi^{PSA(i)}$ is nonempty-valued; this is done in the lemma below:

Lemma 6. *For each w, $\Phi^{PSA(i)}(w)$ is nonempty.*

The proof of this lemma, and remaining proofs in this section, are deferred to the Supplemental Material, as they are either technical or similar to previous proofs.

To analyze the model, it helps to restate the definition of $\Phi^{PSA(i)}$. The distribution F lies in $\Phi^{PSA(i)}(w)$ if and only if there exists a triple $(\mathcal{A}, c, w_A)$, where $\mathcal{A} \supseteq \mathcal{A}^P$ is a technology, $c \geq 0$, and $w_A \in \mathcal{S}$, that satisfies the following conditions:

(a) Supervisor maximization: the contract w_A maximizes $V_S^i(\cdot | w, \mathcal{A})$ over $\mathcal{S}$;

(b) Agent maximization: action (F, c) lies in $\mathcal{A}$ and, given contract w_A, this action maximizes the agent's payoff over $\mathcal{A}$;

(c) Supervisor-preferred tie-breaking: given w, w_A, action (F, c) maximizes the supervisor's payoff over actions satisfying (b);

(d) Principal-preferred tie-breaking: given w, w_A, action (F, c) maximizes the principal's payoff over actions satisfying (b)–(c).

A triple $(\mathcal{A}, c, w_A)$ satisfying these conditions will be called a *PSA(i)-certificate for F under w*.

The following lemma shows that in searching for such a certificate, we can focus on cases where $c = 0$ and $w_A = w_0$ (recall that w_0 denotes the zero contract).

Lemma 7. *For any w and F, we have $F \in \Phi^{PSA(i)}(w)$ if and only if there exists a PSA(i)-certificate for F under w of the form $(\mathcal{A}, 0, w_0)$.*

To see why this is, take the following perspective on the supervisor's problem: Any choice of contract w_A will induce the agent to produce some distribution F. Rather than view the supervisor as choosing w_A, we can view her as directly choosing what F to induce, and then inducing it in the least costly way. Now, if there is some technology $\mathcal{A}$ under which the supervisor would choose to induce F, then under $\mathcal{A}' = \mathcal{A} \cup \{(F, 0)\}$, the supervisor is all the more inclined to induce F, since she can do so costlessly by offering the agent the zero contract. The lemma follows from this observation, together with careful verification of the tie-breaking conditions.

We can now show that the model falls under our general framework, and thus:

Proposition 8. *There exists a linear contract maximizing $V_P^{PSA(i)}$.*

We verify the Richness and Responsiveness properties, as well as the lower hemi-continuity property, by arguments very similar to those used in the robust principal-agent model. Along the way, Lemma 7 helps to simplify by reducing the space of possibilities to consider.

5.3 Hierarchical Model (ii)

Hierarchical model (ii) closely resembles the previous hierarchical model. The key difference is that the supervisor is not perfectly informed of $\mathcal{A}$. Instead, the supervisor is now uncertain about $\mathcal{A}$; but we assume she is at least as well informed as the principal is. Specifically, the principal knows about technology $\mathcal{A}^P$, the supervisor knows about technology $\mathcal{A}^S$, and the true technology is $\mathcal{A}$, with $\mathcal{A} \supseteq \mathcal{A}^S \supseteq \mathcal{A}^P$. The principal, for her part, is uncertain about both $\mathcal{A}$ and $\mathcal{A}^S$. Since the model continues to focus on the principal's problem, $\mathcal{A}^P$ is a primitive of the model, whereas $\mathcal{A}^S$ and $\mathcal{A}$ are free variables. We also no longer restrict the supervisor to contracts in $\mathcal{S}$, since such a restriction will not be needed for existence of an optimal S-A contract; thus, the supervisor may offer any contract $w_A \in C^+(Y)$.

In this model, ignoring the principal for a moment, the relationship between supervisor and agent looks much like the robust principal-agent model. We now formally describe the supervisor's behavior. Define $\Gamma_A^S(w, w_A, \mathcal{A}) = \arg\max_{F \in \Gamma_A(w_A, \mathcal{A})} \mathbb{E}_F[w(y) - w_A(y)]$ and $\Gamma_A^{PS}(w, w_A, \mathcal{A}) = \arg\max_{F \in \Gamma_A^S(w, w_A, \mathcal{A})} \mathbb{E}_F[y - w(y)]$ as in hierarchical model (i). The supervisor's objective in hierarchical model (ii) is

$$V_S^u(w_A | w, \mathcal{A}^S) = \inf_{\text{technology } \mathcal{A} \supseteq \mathcal{A}^S} V_S^i(w_A | w, \mathcal{A})$$

where V_S^i is the informed supervisor objective of hierarchical model (i), and $\mathcal{A}^S$ is the supervisor's knowledge of the technology. We write "u" to denote "uninformed." In words, the supervisor maximizes expected money received minus money paid, given the agent's strategic response, in the worst case over all possible technologies containing $\mathcal{A}^S$.

For fixed P-S contract w, and technologies $\mathcal{A}, \mathcal{A}^S$, define

$$\Gamma_S^u(w, \mathcal{A}^S, \mathcal{A}) = \bigcup_{w_A \in \arg\max_{C^+(Y)} V_S^u(\cdot | w, \mathcal{A}^S)} \Gamma_A^{PS}(w, w_A, \mathcal{A}).$$

This is the analogue of the set $\Gamma^i_S(w, \mathcal{A})$ from hierarchical model (i). It now depends also on $\mathcal{A}^S$ since this determines the supervisor's behavior.

Now, we define the outcome correspondence for hierarchical model (ii) to be

$$\Phi^{PSA(ii)}(w) = \bigcup_{\text{technologies } \mathcal{A}, \mathcal{A}^S \text{ s.t. } \mathcal{A} \supseteq \mathcal{A}^S \supseteq \mathcal{A}^P} \Gamma^u_S(w, \mathcal{A}^S, \mathcal{A})$$

and the principal's objective is denoted by $V_P^{PSA(ii)}(w)$.

As before, we should check nonemptiness:

Lemma 9. *For each w, $\Phi^{PSA(ii)}(w)$ is nonempty.*

We comment that this is where we use the lower hemicontinuity result from the robust principal-agent model: it ensures that the supervisor has an optimal choice of w_A—in fact, one in which w_A is proportional to w—so that the union in the definition of Γ^u_S is not empty.

To analyze the model, we proceed in a fashion similar to the previous hierarchical model. Distribution F is in $\Phi^{PSA(ii)}(w)$ if there exists a situation in which the agent would choose it—where now such a situation is described by a quadruple $(\mathcal{A}, \mathcal{A}^S, c, w_A)$, meeting conditions analogous to (a)–(d) in hierarchical model (i). We refer to such a quadruple as a *PSA(ii)-certificate for F under w*. We can formulate an analogue of Lemma 7 for this model (see Lemma S-6 in the Supplemental Material): F lies in $\Phi^{PSA(ii)}(w)$ if and only if there exists such a PSA(ii)-certificate in which $c = 0$, $w_A = w_0$, and moreover $\mathcal{A} = \mathcal{A}^S$.

This fact is useful in showing that the model satisfies our two properties, and thereby obtaining our linearity result:

Proposition 10. *There exists a linear contract maximizing $V_P^{PSA(ii)}$.*

Again, the proof follows the same basic argument as for the principal-agent model (and as for hierarchical model (i)).

It might not be obvious that this model satisfies Responsiveness, since the supervisor is no longer modeled as an expected-utility maximizer. However, the key is that Responsiveness is a property on the correspondence Φ, which gathers the counterparty's behavior across all possible environments (in this case, all possible $\mathcal{A}^S$ and $\mathcal{A}$); it does not have to hold in each environment individually. In this model, Lemma S-6 shows, in effect, that we can restrict attention to a crucial subset of possible

environments: those in which the supervisor knows that she can induce distribution F for free by offering the zero contract, and she chooses to do so. In this subset of environments, the supervisor does act like an expected-utility maximizer, and this suffices to imply Responsiveness.

6   An Example Where Responsiveness Fails

In this section we describe version (iii) of the hierarchical model, where the principal is no longer uncertain as to what the supervisor does and doesn't know. Instead, the supervisor, like the principal, knows only that the technology $\mathcal{A}$ is a superset of the given $\mathcal{A}^P$, and she contracts with the agent so as to maximize her own worst-case payoff. We give an example to show that linear contracts can fail to be optimal. In the process, we also observe that this model satisfies Richness, but not Responsiveness in general. In fact, this example is an illustration of Theorem 4, as will be discussed briefly later.

Here are the details. We take $\mathcal{A}^P \subseteq \Delta(Y) \times \mathbb{R}^+$ as in the first two versions of the model. For any contract[5] $w_A \in C^+(Y)$ to the agent, and any true technology $\mathcal{A} \supseteq \mathcal{A}^P$, define $\Gamma_A^S(w, w_A, \mathcal{A})$ as before. Define $V_S^i(w_A|w, \mathcal{A})$ as in model (i), and define $V_S^u(w_A|w, \mathcal{A}^P) = \inf_{\mathcal{A} \supseteq \mathcal{A}^P} V_S^i(w_A|w, \mathcal{A})$ as in model (ii). This is the supervisor's objective.

Assume for the moment that $\min_y w(y) = 0$ (in the language of Section 4, w is grounded). As in model (ii), we can apply the robust principal-agent analysis to the supervisor-agent relationship here to conclude that the supervisor has an optimal contract that takes the form $w_A(y) = \beta w(y)$ for some constant $\beta \in [0, 1]$; however, there may also exist optimal choices of w_A that are not of this form. We *assume* that the supervisor uses an optimal contract of this form (but if multiple choices of β are optimal, we remain agnostic about which one is chosen). This restriction on the supervisor's behavior will simplify the analysis, but it is not in keeping with model (ii) where no such restriction was made. With a little more work, one can verify that the lessons of this section are unchanged if the restriction is removed; details are included in an earlier version of this paper (Walton and Carroll, 2019).

---

[5]Note that, as in model (ii), the supervisor is not restricted to a compact set of contracts.

Accordingly, define

$$\widetilde{\Gamma}^u_S(w, \mathcal{A}^P, \mathcal{A}) = \bigcup_{\beta \in \arg\max_{\tilde{\beta} \in [0,1]} V^u_S(\tilde{\beta}w | w, \mathcal{A}^P)} \Gamma^S_A(w, \beta w, \mathcal{A}).$$

This is the set of distributions that may be chosen when the agent's technology is $\mathcal{A}$, and the supervisor has presented him with a contract that is optimal and linear (from the supervisor's point of view) given $(w, \mathcal{A}^P)$. The union arises due to the possibility of multiple optimal choices of β. And now, the outcome correspondence is given by

$$\Phi^{PSA(iii)}(w) = \bigcup_{\mathcal{A} \supseteq \mathcal{A}^P} \widetilde{\Gamma}^u_S(w, \mathcal{A}^P, \mathcal{A}).$$

(For simplicity, we do not bother with principal-preferred tie-breaking; adding such a tie-break would not change the substantive conclusions.)

For completeness, we should also define $\Phi(w)$ when w is not grounded, although such contracts will play no role in the subsequent analysis. For simplicity, for any such w, let w' be the grounded contract obtained by subtracting the constant $\min_y w(y)$ from w, and just put $\Phi^{PSA(iii)}(w) = \Phi^{PSA(iii)}(w')$.

Now $\Phi^{PSA(iii)}$ is fully defined, and accordingly, the principal's objective $V^{PSA(iii)}_P(w)$ is defined as usual.

Now let us analyze the model. We first formally note, as promised previously:

Proposition 11. *The correspondence $\Phi^{PSA(iii)}$ satisfies Richness.*

This is fairly immediate; the formal proof is in the Supplemental Material.

Next, if the principal offers a (grounded) contract w, what fraction β will the supervisor share with the agent? The analysis of the robust principal-agent problem (see Carroll (2015)) gives the answer: the supervisor will identify action $(F, c) \in \mathcal{A}^P$ for which $\sqrt{\mathbb{E}_F[w(y)]} - \sqrt{c}$ is maximal, and as long as this quantity is positive, the supervisor will set the corresponding value $\beta = \sqrt{c / \mathbb{E}_F[w(y)]}$. (If there happens to be more than one optimal (F, c), then all corresponding β's are optimal for the supervisor. Also, if there is no known action with $\sqrt{\mathbb{E}_F[w(y)]} - \sqrt{c} > 0$, then the supervisor cannot obtain a positive guarantee, so every $\beta \in [0, 1]$ is optimal — they all give the supervisor a guarantee of zero.) Accordingly, let us say that the contract w *targets* the action (F, c) if this action maximizes $\sqrt{\mathbb{E}_F[w(y)]} - \sqrt{c}$ over $\mathcal{A}^P$, and the

contract is *non-degenerate* if the corresponding value of $\sqrt{\mathbb{E}_F[w(y)]} - \sqrt{c}$ is strictly positive.

We can further use the principal-agent analysis to explicitly characterize the possible responses by the agent (proof in the Supplemental Material):

Lemma 12. *Suppose w is grounded and non-degenerate, and $\beta > 0$ is an optimal choice for the supervisor. A distribution F' lies in $\Gamma_A^S(w, \beta w, \mathcal{A})$ for some $\mathcal{A}$ if and only if*

$$E_{F'}[w(y)] \geq E_F[w(y)] - \sqrt{c \cdot E_F[w(y)]} \tag{2}$$

where $(F, c) \in \mathcal{A}^P$ is the targeted action leading to slope β.

Therefore, if w is grounded and non-degenerate, $\Phi^{PSA(iii)}(w)$ consists of all distributions F' that satisfy (2) for some targeted action (F, c).

Henceforth, for concreteness, let us focus on a particular, parametric specification of $\mathcal{A}^P$. Assume that Y is finite, so that we can ignore the continuity restriction on contracts, and let y_L, y_H be elements of Y with $y_H > y_L > 0$. Also let c_L, c_H be positive numbers with $c_L/y_L < c_H/y_H < 1$. Let

$$\mathcal{A}^P = \{(\delta_{y_H}, c_H), (\delta_{y_L}, c_L), (\delta_0, 0)\}.$$

Thus, there are three known actions, all deterministic. For brevity, we will call these actions "action H," "action L," and "action 0."

Suppose first that the principal uses a linear contract, $w_\alpha(y) = \alpha y$. We can see which action is targeted depending on the value of α:

- for $\alpha < c_L/y_L$, the contract targets action 0;

- for $c_L/y_L < \alpha < \alpha^*$, the contract targets action L;

- for $\alpha^* < \alpha$, the contract targets action H,

where $\alpha^* = \left(\frac{\sqrt{c_H} - \sqrt{c_L}}{\sqrt{y_H} - \sqrt{y_L}}\right)^2$. (At boundary cases, two actions are targeted.)

For a contract targeting 0, no positive guarantee is possible. For a contract targeting L, Lemma 12 shows that the possible distributions are the ones for which $\mathbb{E}_F[\alpha y] \geq \alpha y_L - \sqrt{c_L \cdot \alpha y_L}$, or equivalently $\mathbb{E}_F[y] \geq y_L - \sqrt{c_L y_L/\alpha}$. Since the principal's payoff is $(1 - \alpha)\mathbb{E}_F[y]$, the payoff guarantee from the contract with slope α

is

$$V_P(w_\alpha) = (1 - \alpha) \left(y_L - \sqrt{\frac{c_L y_L}{\alpha}} \right). \tag{3}$$

By identical reasoning, for a linear contract targeting H, the payoff guarantee is

$$V_P(w_\alpha) = (1 - \alpha) \left(y_H - \sqrt{\frac{c_H y_H}{\alpha}} \right). \tag{4}$$

So overall, the principal's guarantee is

$$V_P(w_\alpha) = \begin{cases} 0 & \text{if } \alpha < c_L/y_L, \\ (1 - \alpha) \left(y_L - \sqrt{\frac{c_L y_L}{\alpha}} \right) & \text{if } c_L/y_L < \alpha < \alpha^*, \\ (1 - \alpha) \left(y_H - \sqrt{\frac{c_H y_H}{\alpha}} \right) & \text{if } \alpha^* < \alpha. \end{cases}$$

(And at the boundary values of α, the guarantee is given by the lower of the two neighboring formulas.)

Now consider a value of α such that the linear contract w_α targets L. Consider instead the nonlinear contract $\hat{w}_\alpha$ given by $\hat{w}_\alpha(y_H) = \alpha y_H$, and $\hat{w}_\alpha(y) = 0$ for all other y. Evidently, this contract cannot target L. As long as $\alpha > c_H/y_H$, it targets H (rather than 0). And in this case, Lemma 12 tells us that the distributions the agent may produce are the ones for which $\mathbb{E}_F[\hat{w}_\alpha(y)] \geq \alpha y_H - \sqrt{c_H \cdot \alpha y_H}$, or equivalently, those that produce y_H with probability at least $1 - \sqrt{c_H/\alpha y_H}$. Since the principal receives $(1 - \alpha)y_H$ if output y_H realizes, and receives at least 0 for any other output, the principal's guarantee from $\hat{w}_\alpha$ is at least

$$(1 - \alpha) y_H \left(1 - \sqrt{\frac{c_H}{\alpha y_H}} \right). \tag{5}$$

This is exactly the same as in (4), but note that it applies for a wider range of values of α, since α only needs to be higher than c_H/y_H, which is less than α^*.

This suggests that we can choose numeric values for the parameters such that the value of (5), for some appropriate $\alpha \in (c_H/y_H, \alpha^*)$, is higher than the maximum value of $V_P(w_\alpha)$. For example, take $(y_L, y_H, c_L, c_H) = (75, 105, 32, 50)$. It can be checked that the maximum value of $V_P(w_\alpha)$ occurs at $\alpha = 2/3$ and is equal to 5. However, this value of α lies in the range $(c_H/y_H, \alpha^*)$, and the corresponding value of expression (5) is equal to $35 - 5\sqrt{35} \approx 5.42$. Thus, the guarantee from $\hat{w}_\alpha$ at $\alpha = 2/3$ is higher than the guarantee from any linear contract. This comparison is shown in Figure 2.
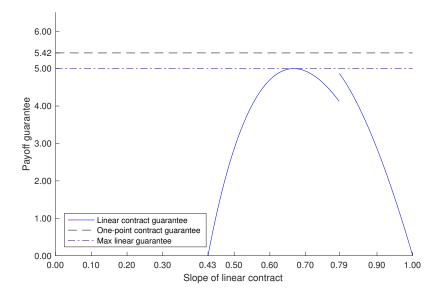
Figure 2: Guarantee from a linear contract, in example for hierarchical model (iii).

What is happening in this example? An intuition is that if the principal were restricted to linear contracts, she would like to leave the supervisor just a modest fraction of the output, but then the supervisor is less inclined than the principal is to offer the agent incentives targeted at action H. The principal can steer the supervisor back toward incentivizing H instead of L by not paying for the output of L. This gets the supervisor to offer stronger incentives to the agent, which in turn guards against the agent taking relatively unproductive actions. Notice, also, that this use of nonlinear contracts to get the supervisor to target H rather than L would not have worked in hierarchical model (i) (or model (ii)), because there the supervisor may know of other actions besides H or L. In the worst case there, the supervisor targets some new action that produces output y_H with some probability and 0 otherwise, and nonlinear w will not help the principal avoid this bad outcome.

We can also relate the failure of linearity in model (iii) back to the Responsiveness property. Notice that the property fails between $\hat{w}_\alpha$ and w_α (with the specific value $\alpha = 2/3$), since w_α pays weakly more than $\hat{w}_\alpha$ for *any* distribution, and pays equally much when the distribution is supported on $\{0, y_H\}$, yet some such distributions are in $\Phi(w_\alpha)$ but not in $\Phi(\hat{w}_\alpha)$. Given that Responsiveness is violated, we should expect from the results of Section 4 that the numerical parameters can be set in

such a way that linear contracts are non-optimal. In fact, our example can be cast in the framework of that section, by viewing "output H, output L, output 0" as physical outputs whose numerical values are initially left unspecified, and considering a contract w that pays 70, 50, 0 for these outputs respectively (and $\hat{w}$ pays 70, 0, 0). Applying Theorem 4 to this violation of Responsiveness leads to the valuation that values these outputs at $105, 75, 0$, thus recovering precisely this example.

7 Analysis of Hierarchical Models (i) and (ii)

Now that we have shown how several models fit into our linearity framework, we return to study hierarchical models (i) and (ii) in more detail. This analysis serves two purposes. First, we demonstrate by example why hierarchical models do not straightforwardly rewrite as special cases of the original robust principal-agent model. This is important, since if the lessons from principal-agent models already effortlessly extended to other organizational environments, our main linearity result would be redundant. And second, we sketch how one may numerically compute the optimal contract slopes and guarantees in hierarchical models (i) and (ii), thus completing the process of solving for the optimal contract in these models.

A further illustration of what one can do using a detailed analysis of multiple different contracting models appears in Section S-2 of the Supplemental Material. That section takes the principal-agent model, together with hierarchical models (i) and (ii), and places them side-by-side to compare the principal's payoffs.

7.1 Nonequivalence of Hierarchical Model (i) to the Robust Principal-Agent Model

One may be tempted to try to reduce hierarchical model (i) to a special case of the robust principal-agent model, as follows: collapse the supervisor and agent into a single "modified agent," whose cost of producing any distribution F is simply the cost that the supervisor would have to pay to incentivize F in the original hierarchical model. We show here why this reduction fails.

Formally, given a known technology $\mathcal{A}^P$ in hierarchical model (i), define $\tilde{\mathcal{A}}^P$ by the following procedure: for each $(F, c) \in \mathcal{A}^P$, let $\bar{w}_A \in \arg\min_{w_A \in \mathcal{S}} \mathbb{E}_F[w_A(y)]$ subject to $(F, c) \in \arg\max_{(\tilde{F}, \tilde{c}) \in \mathcal{A}^P} \mathbb{E}_{\tilde{F}}[w_A(y)] - \tilde{c}$, and let $\bar{c}(F)$ be the corresponding value of the

min (if there is no w_A satisfying the constraint then put $\bar{c}(F) = \infty$). Then define $\tilde{\mathcal{A}}^P = \{(F, \bar{c}(F)) : \exists c \geq 0 \text{ s.t. } (F, c) \in \mathcal{A}^P \text{ and } \bar{c}(F) < \infty\}$. Thus, we keep the possible output distributions in $\mathcal{A}^P$ the same, but adjust the cost of the action to mimic the expected amount the supervisor would have to pay to induce the action under $\mathcal{A}^P$. We then apply the robust principal-agent framework with known technology $\tilde{\mathcal{A}}^P$. Is $\Phi^{PA}(w)$, given $\tilde{\mathcal{A}}^P$, equal to $\Phi^{PSA(i)}(w)$ given $\mathcal{A}^P$?

To illustrate, consider $\mathcal{A}^P = \{(\delta_0, 0), (\delta_{y^*}, c)\}$ where $c > 0$ and $y^* = 5c$. With this simple known technology, $\tilde{\mathcal{A}}^P = \mathcal{A}^P$, since $\bar{w}_A(y) = 0$ makes $(\delta_0, 0)$ a maximizer of $\mathbb{E}_F[\bar{w}_A(y)] - c$ over $\mathcal{A}^P$, and $\bar{w}_A(y) = \frac{1}{5}y$ makes (δ_{y^*}, c) a maximizer of the same objective over $\mathcal{A}^P$. Now consider a linear contract $w(y) = \frac{1}{2}y$. Suppose the true technology in the hierarchical model (i) case includes just one additional action $(F', 0)$ where $F' = \frac{3}{5}\delta_0 + \frac{2}{5}\delta_{y^*}$. A straightforward computation shows that the supervisor chooses to induce $(F', 0)$, offering $\bar{w}_A(y) = 0$ (and using tie-breaking) to get a payoff of $\frac{1}{2} \cdot \frac{2}{5}y^* = c$, rather than inducing the agent to choose (δ_{y^*}, c) (which requires $\bar{w}_A(y^*) \geq \frac{5}{3}c$) for a payoff of $\frac{5}{6}c$, and hence $F' \in \Phi^{PSA(i)}(w)$.

Now we consider $\mathcal{A} = \tilde{\mathcal{A}}^P \cup \{(F', 0)\}$ in the robust principal-agent framework, with the same linear contract $w(y) = \frac{1}{2}y$. The agent's optimal choice is (δ_{y^*}, c), since $\frac{1}{2}y^* - c = \frac{3}{2}c > c = \frac{1}{2} \cdot \frac{2}{5}y^*$. Indeed, across all technologies $\mathcal{A}$, the lowest-mean (worst-case) action that can occur is one that has zero cost and mean $3c$, so $(F', 0)$ with mean $2c$ is not possible, so $F' \notin \Phi^{PA}(w)$ given $\tilde{\mathcal{A}}^P$.

Thus, one cannot reduce hierarchical model (i) to the robust principal-agent model as proposed. (Likewise, this reduction would not work for model (ii) either.) A simple intuition is that adding an extra, unknown action to the agent's technology $\mathcal{A}$ in the hierarchical model can potentially increase the supervisor's cost of incentivizing the original known actions, whereas in the principal-agent model, the costs of the known actions remain unchanged. Thus, the uncertainty about $\mathcal{A}$ has different effects in the two models. This is further underscored in the next subsection where we identify the worst-case technology in the hierarchical model (i), which differs in structure from the worst case in the robust principal-agent model.

Admittedly, the counterexample above does not rule out the possibility that some other, subtler argument might be available to reduce the hierarchical model to a special case of the robust principal-agent model. However, it at least suggests (to us) that any such argument would be non-obvious enough so that the hierarchical model is most naturally viewed as a separate model, as we have presented it.

33

7.2 Optimal Contract Slope

We have argued in Section 5 that optimal contracts in both hierarchical models (i) and (ii) are linear. Here, we characterize the optimal slope of the linear contract. The main task is to identify the worst-case technology for any given linear contract, which allows us to replace the infimum in the principal's objective with a more explicit function. We sketch the steps here; the method is fully laid out in an earlier version of this paper (Walton and Carroll, 2019).

We begin with the analysis for hierarchical model (i).[6] Assume the principal offers a particular linear contract w_α. The first step is to show that, in defining $\Phi^{PSA(i)}(w_\alpha)$, rather than taking the union over all possible technologies $\mathcal{A}$, we can consider a much smaller class of technologies. In particular, we can focus on technologies where the agent can produce any output distribution, and his cost of doing so depends only on the mean of the distribution; and moreover, this cost is a convex, nondecreasing function of the mean. Intuitively, once we have focused on linear contracts, only the mean output should matter for all parties involved. Thus a technology may be identified with a cost function $\kappa : \text{co}(Y) \to \mathbb{R}^+$, where the agent can produce any distribution F at cost $\kappa(\mathbb{E}_F[y])$. Note that such a choice of κ is consistent with the requirement $\mathcal{A}^P \subseteq \mathcal{A}$ if and only if $\kappa(\mathbb{E}_F[y]) \leq c$ for every $(F, c) \in \mathcal{A}^P$.

The next step is to identify the lowest mean output that might be induced under a given linear contract w_α and known technology $\mathcal{A}^P$. To do this, we show that for any κ as above, the supervisor's cost of inducing any given mean output μ is $\mu\kappa'(\mu)$; therefore, the supervisor chooses μ to maximize $(\alpha - \kappa'(\mu))\mu$. Thus, for any value of μ, the supervisor will induce mean output μ only if $(\alpha - \kappa'(\mu))\mu \geq (\alpha - \kappa'(\tilde{\mu}))\tilde{\mu}$ for all $\tilde{\mu}$. This is a family of inequality constraints on κ, and in the worst case, these inequalities will be binding. In particular, the worst-case mean output μ is the lowest value for which κ can be chosen to satisfy the equality $(\alpha - \kappa'(\mu))\mu = (\alpha - \kappa'(\tilde{\mu}))\tilde{\mu}$ for all $\tilde{\mu} > \mu$, while still respecting the constraint that $\kappa(\mathbb{E}_F[y]) \leq c$ for all $(F, c) \in \mathcal{A}^P$. We notate this worst-case mean output as $\mu^*(\alpha)$. The differential equation leads to a one-parameter family of solutions for κ, and we simply pick the worst one that satisfies the $\mathcal{A}^P$ constraint. Finally, $V^{PSA(i)}(w_\alpha) = \mu^*(\alpha)(1 - \alpha)$. In the earlier version of the paper, we fill in these steps with calculations, showing that the maximum value of

---

[6]Garrett et al. (2020) independently study a problem of optimal technology design in a principal-agent setting that shares some features with this analysis.

$V^{PSA(i)}$ over α can be more explicitly characterized as

$$\max_{\alpha \in [0,1]} \mu^*(\alpha)(1-\alpha) = \max_{(F,c) \in \mathcal{A}^P} g(\mathbb{E}_F[y], c)$$

where

$$g(\tilde{\mu}, c) = \begin{cases} \tilde{\mu} & \text{if } c = 0, \\ \max_{\mu \in [0,\tilde{\mu}]} \left(1 - \frac{c}{(\tilde{\mu}-\mu)-\mu\log(\tilde{\mu}/\mu)}\right)\mu & \text{if } c > 0, \end{cases}$$

and that the corresponding optimal value of α is equal to $c/((\mathbb{E}_F(y)-\mu)-\mu\log(\mathbb{E}_F[y]/\mu))$ for the choices of (F,c) and μ that attain the max. We thus have a fairly explicit description of the optimal contract, and the principal's guarantee, though there is no fully closed-form solution.

We now briefly discuss hierarchical model (ii); again, the details are in the earlier version of the paper. For any linear contract w_α, $\alpha \in (0, 1]$ that the principal offers, we can exploit the analysis of the robust principal-agent model to characterize optimal behavior for the supervisor under each possible $\mathcal{A}^S$. The task is further simplified by Lemma S-6, telling us that F is a possible response by the agent if and only if it can occur under a technology $\mathcal{A}^S$ such that $(F, 0) \in \mathcal{A}^S$ and such that the supervisor offers the zero contract under $\mathcal{A}^S$. So we just need to identify the distributions F for which some such $\mathcal{A}^S$ exists. This leads to the characterization

$$\Phi^{PSA(ii)}(w_\alpha) = \{F \in \Delta(Y) : \alpha\mathbb{E}_F[y] \geq \left(\sqrt{\alpha\mathbb{E}_{F'}[y]} - \sqrt{c'}\right)^2 \text{ for all } (F', c') \in \mathcal{A}^P|_\alpha\},$$

where $\mathcal{A}^P|_\alpha$ denotes the subset $\{(F, c) \in \mathcal{A}^P : \alpha\mathbb{E}_F[y] \geq c\}$.

This identifies the minimum expected output that the agent could potentially produce when the principal offers contract w_α. We can use this characterization to compute more explicitly the guarantee from any given linear contract, and then optimize over α. In this case, the computation turns out to give a closed-form solution:

$$\max_{\alpha \in [0,1]} V^{PSA(ii)}(w_\alpha) = \max_{(F,c) \in \mathcal{A}^P} (\mathbb{E}_F[y]^{1/3} - c^{1/3})^3,$$

and the optimal value of α equals $(c/\mathbb{E}_F[y])^{1/3}$ for the (F, c) that solves the maximization on the right-hand side.

We comment also that, in hierarchical model (ii) (unlike (i) but similar to the robust principal-agent model), the worst case is attained under a technology $\mathcal{A}^S$ that

consists of $\mathcal{A}^P$ plus just one additional action. This suggests a seeming "discontinuity" between models (ii) and (iii): if the supervisor is allowed to know just one additional action that the principal doesn't know, the result looks very different than if their knowledge definitely coincides exactly. Of course, this difference depends on the fact that the one additional action can be anything, so that there is already great uncertainty on the principal's part about how the supervisor will behave.

8   Conclusion

Economic analysis uses stylized, simplified models to develop concepts. In particular, agency theory commonly works with bilateral principal-agent models. But actual agency relations may be embedded in much more complex organizations. An effective theory should not fall silent when confronted with this variety.

In this paper, we have examined arguments concerning linear contracts, and their ability to provide robustness to uncertainty by aligning the payoffs of the two parties without being sensitive to details of the environment. To avoid assuming any specific organizational structure, we have proposed a "black-box" framework for reasoning about contracts, working in terms of the outcome correspondence Φ mapping contracts to possible responses. Past literature has pointed to the broad space of possible actions as key to robustness of linear contracts. We have formalized this in our framework as a *Richness* property on the correspondence Φ. But Richness alone is not enough to allow comparisons across contracts, let alone to identify optimal contracts. We identified a further *Responsiveness* property, requiring that when contracts change, the possible responses vary in a way that resembles maximizing expected payment. We showed that, as a supplement to Richness, Responsiveness is sufficient—and, in an appropriate sense, essentially necessary—to ensure that linear contracts are optimally robust, in the sense of solving a maxmin problem. We illustrated this in more detail by describing several specific ways to write down a model of hierarchical contracting, all of which satisfy Richness, and showing how Responsiveness distinguishes versions that lead to linear contracts from a version that does not. Our more detailed analysis of the individual models also showed that, even though multiple models lead to linear contracts being optimal, these models are not equivalent to each other.

Our focus has been on understanding when and why linear contracts are robust to

uncertainty, without relying on the bilateral principal-agent structure. But a similar methodology could potentially be applied to many other classes of contracts. We illustrate this in the Supplemental Material, Section S-3, by showing how alternative conditions on Φ would lead to concave rather than linear contracts. Another potential future application of the approach would be foundations for debt contracts: Antić (2021) shows how changing the space of uncertainty in a robust agency model can lead to debt contracts, rather than linear contracts, as optimal (and, a bit farther afield, Malenko and Tsoy (2020) do the same in a screening model); our work naturally suggests the question of whether some alternative properties on the correspondence Φ would reproduce this finding in an organization-free way. More generally, we hope that our work will spur the development of an appropriate methodology to separate the analysis of formal incentives from assumptions on the organizational environment in which they operate.

A Additional proofs

Here are proofs omitted from Sections 3 and 4 of the main paper. Remaining proofs are in the Supplemental Material.

Proof of Proposition 2. As noted in the text, Theorem 1 ensures that $\sup_w V_P(w)$ is approached within the set of linear contracts. Moreover, for any contract w_α whose slope α is greater than 1, $V_P(w_\alpha) \leq 0 = V_P(w_1)$, so it is sufficient to restrict attention to $\alpha \in [0, 1]$. Thus we need only verify that, on the restricted domain $\{w_\alpha \mid \alpha \in [0, 1]\}$, V_P has a maximum.

Define $\overline{V}_P = \sup_{\alpha \in [0,1]} V_P(w_\alpha)$, and let $\alpha_1, \alpha_2, \ldots$ be a sequence of values such that $V_P(w_{\alpha_k}) \to \overline{V}_P$. By compactness we may assume α_k has a limit α^*. Assume for contradiction that $V_P(w_{\alpha^*}) \neq \overline{V}_P$. Put $\varepsilon = \overline{V}_P - V_P(w_{\alpha^*}) > 0$. The definition of V_P means there exists $F \in \Phi(w_{\alpha^*})$ such that

$$\mathbb{E}_F[y - \alpha^* y] < V_P(w_{\alpha^*}) + \frac{\varepsilon}{2} = \overline{V}_P - \frac{\varepsilon}{2}.$$

Now, lower hemicontinuity means that for k large enough, there exists $F_k \in \Phi(w_{\alpha_k})$

such that $\mathbb{E}_{F_k}[y] \leq \mathbb{E}_F[y] + \frac{\varepsilon}{2}$. Then,

$$
\begin{aligned}
\limsup_{k \to \infty} V_P(w_{\alpha_k}) \ &\leq \ \limsup_{k \to \infty} \mathbb{E}_{F_k}[y - \alpha_k y] \\
&= \ (1 - \alpha^*) \limsup_{k \to \infty} \mathbb{E}_{F_k}[y] \\
&\leq \ (1 - \alpha^*) \left(\mathbb{E}_F[y] + \frac{\varepsilon}{2} \right) \\
&\leq \ (1 - \alpha^*) \mathbb{E}_F[y] + \frac{\varepsilon}{2} \qquad < \overline{V}_P.
\end{aligned}
$$

(Here the first inequality follows from the definition of V_P, and the other steps are straightforward.) This contradicts the assumption $V_P(w_{\alpha_k}) \to \overline{V}_P$. $\qquad\square$

Proof of Proposition 3. ("If" direction) We proceed through the conditions one by one, showing that each implies that w is scaling-dominated. Let v be a valuation; we need to show that either $V_P(w|v) \leq 0$ or $V_P(\beta w|v) > V_P(w|v)$ for some β. For brevity, we write $V_P(w)$ rather than $V_P(w|v)$ (and likewise $V_P(\beta w)$, etc.).

If (i) holds, then $\Phi(w)$ is all of $\Delta(Z)$, so $V_P(w) = \min_{z \in Z}(v(z) - w(z)) \leq \min_z v(z) = 0$.

If (i) is not satisfied but (ii) is for some β, then $\Phi(\beta w) \subseteq \Phi(w)$. So,

$$
V_P(w) = \min_{F \in \Phi(w)} \mathbb{E}_F[v(z) - w(z)] < \min_{F \in \Phi(w)} \mathbb{E}_F[v(z) - \beta w(z)] \leq \min_{F \in \Phi(\beta w)} \mathbb{E}_F[v(z) - \beta w(z)] = V_P(\beta w),
$$

where the strict inequality comes from $\mathbb{E}_F[w(z)] > 0$ since $h(w) > 0$.

Next, suppose (i)–(ii) are not satisfied but (iii) is. Let $\hat{z}$ be the output at which w attains its minimum value subject to $w(z) \geq h(w)$. Let z' be the output for which w attains its highest value less than $w(\hat{z})$ (this exists since w is grounded and (i) is assumed not to hold, so that $\min_z w(z) = 0 < h(w) \leq w(\hat{z})$). Let $\lambda = 1/(w(\hat{z}) - w(z'))$. Notice that, given any distribution F for which $\mathbb{E}_F[w(z)] \leq h(w)$, if we gradually move probability mass from output levels where $w(z) < h(w)$ to those with $w(z) \geq h(w)$, the expected value of w increases by at least $1/\lambda$ times the amount of mass moved; therefore, by moving at most $\lambda(h(w) - \mathbb{E}_F[w(z)])$ mass, we can reach a distribution F' with $\mathbb{E}_{F'}[w(z)] = h(w)$.

Let $\kappa = \max_z v(z)$. Take $K = \max\{\kappa\lambda, 1\}$ in condition (iii), and consider the value of β given by that condition.

Let $\underline{F}$ be the worst-case distribution for contract βw. Thus, $V_P(\beta w) = \mathbb{E}_{\underline{F}}[v(z)] - \beta\mathbb{E}_{\underline{F}}[w(z)]$. We also know that $\mathbb{E}_{\underline{F}}[\beta w(z)] \geq h(\beta w)$, so $\mathbb{E}_{\underline{F}}[w(z)] \geq h(\beta w)/\beta$.

If $\mathbb{E}_{\underline{F}}[w(z)] \geq h(w)$, then $\underline{F} \in \Phi(w)$, and we already have $V_P(w) \leq \mathbb{E}_{\underline{F}}[v(z) - w(z)] < \mathbb{E}_{\underline{F}}[v(z) - \beta w(z)] = V_P(\beta w)$. So assume $\mathbb{E}_{\underline{F}}[w(z)] < h(w)$. As noted above, we can move at most $\lambda(h(w) - \mathbb{E}_{\underline{F}}[w(z)])$ probability mass from $\underline{F}$ to obtain a distribution F' with $\mathbb{E}_{F'}[w(z)] = h(w)$, and

$$\mathbb{E}_{F'}[v(z)] \leq \mathbb{E}_{\underline{F}}[v(z)] + K(h(w) - \mathbb{E}_{\underline{F}}[w(z)])$$

(recalling $K \geq \kappa\lambda$). Therefore,

$$
\begin{aligned}
V_P(\beta w) &= \mathbb{E}_{\underline{F}}[v(z) - \beta w(z)] \\
&\geq \mathbb{E}_{F'}[v(z)] - K(h(w) - \mathbb{E}_{\underline{F}}[w(z)]) - \mathbb{E}_{\underline{F}}[\beta w(z)] \\
&= \mathbb{E}_{F'}[v(z)] + (K - \beta)\mathbb{E}_{\underline{F}}[w(z)] - Kh(w) \\
&\geq \mathbb{E}_{F'}[v(z)] + (K - \beta)h(\beta w)/\beta - Kh(w) \\
&= \mathbb{E}_{F'}[v(z)] - h(\beta w) - K(h(w) - h(\beta w)/\beta).
\end{aligned}
$$

(The second inequality in the above chain uses $K - \beta \geq 0$.) Finally, applying condition (iii), the last right-hand side is

$$> \mathbb{E}_{F'}[v(z)] - h(\beta w) - (h(w) - h(\beta w)) = \mathbb{E}_{F'}[v(z) - w(z)] \geq V_P(w).$$

Next, suppose conditions (i)–(iii) do not hold but (iv) does. Consider the problem of minimizing $\mathbb{E}_F[v(z) - w(z)]$ subject to the constraint $\mathbb{E}_F[w(z)] = h(w)$; let $\underline{F}$ attain the minimum, and let r be the corresponding objective value. Note that if the constraint were replaced by $\mathbb{E}_F[w(z)] \geq h(w)$, this new minimization problem would have objective value equal to $V_P(w)$, by definition of V_P. We observe that in this latter minimization problem, any solution must satisfy the constraint with equality: otherwise the constraint would not be binding, so we could remove it, but then the value of the problem is simply $\min_z(v(z) - w(z)) \leq \min_z v(z) = 0$, i.e. $V_P(w) \leq 0$, and we are done. This observation implies that $r = V_P(w)$, and also that for every z such that $w(z) > h(w)$, we must have $v(z) - w(z) > V_P(w)$ strictly.

Choose λ such that

$$0 < \lambda < \min_{z:w(z)>h(w)} \frac{v(z) - w(z) - V_P(w)}{w(z) - h(w)}$$

(interpreting the min as ∞ if no such z exists). The latter inequality implies $v(z) -

$(1 + \lambda)w(z) > V_P(w) - \lambda h(w)$ for all z such that $w(z) > h(w)$. Now, if we consider the problem of minimizing $\mathbb{E}_F[v(z) - (1 + \lambda)w(z)]$ subject to $\mathbb{E}_F[w(z)] \geq h(w)$, again the constraint must hold with equality at the optimum: otherwise, the constraint is not binding, and the minimum is attained when F is degenerate on some z with $w(z) > h(w)$, but we know that the objective value for any such F is higher than $V_P(w) - \lambda h(w)$ and the latter value is attained by $\underline{F}$. Therefore, this minimization problem is again solved by $\underline{F}$. We conclude that

$$\mathbb{E}_F[w(z)] \geq h(w) \qquad \text{implies} \qquad \mathbb{E}_F[v(z) - (1 + \lambda)w(z)] \geq V_P(w) - \lambda h(w). \qquad (6)$$

Now put $K = 1 + \lambda$, and let β be as given by condition (iv). Note that we must have $K > \beta$, since (iv) rearranges to $(K/\beta - 1)h(\beta w) > (K - 1)h(w) > 0$. Put $\kappa = (K/\beta - 1)h(\beta w) - (K - 1)h(w) > 0$. Then, for any distribution F such that $\mathbb{E}_F[\beta w(z)] \geq h(\beta w)$, we have

$$(\lambda + 1 - \beta)\mathbb{E}_F[w(z)] - \lambda h(w) \geq \kappa,$$

and therefore

$$
\begin{aligned}
\mathbb{E}_F[v(z) - \beta w(z)] &= \mathbb{E}_F[v(z) - w(z)] + (1 - \beta)\mathbb{E}_F[w(z)] \\
&\geq V_P(w) - \lambda h(w) + \lambda \mathbb{E}_F[w(z)] + (1 - \beta)\mathbb{E}_F[w(z)] \\
&= V_P(w) + (\lambda + 1 - \beta)\mathbb{E}_F[w(z)] - \lambda h(w) \\
&\geq V_P(w) + \kappa.
\end{aligned}
$$

(Here the first inequality comes from rearranging (6), which applies since $\mathbb{E}_F[w(z)] \geq h(\beta w)/\beta > h(w)$.)

We thus have $V_P(\beta w) \geq V_P(w) + \kappa > V_P(w)$ as needed.

Finally, suppose that conditions (i)–(iv) do not hold but (v) does. Let $\underline{\beta}, \overline{\beta}$ be as in that condition. We aim to show that one of $V_P(\underline{\beta}w) > V_P(w)$, $V_P(\overline{\beta}w) > V_P(w)$, or $V_P(w) \leq 0$ holds. So, assume that none of these holds, and seek a contradiction.

Take $\underline{F} \in \Delta(Z)$ to minimize $\mathbb{E}_F[v(z) - \underline{\beta}w(z)]$ subject to $\mathbb{E}_F[\underline{\beta}w(z)] \geq h(\underline{\beta}w)$, and similarly, $\overline{F}$ to minimize $\mathbb{E}_F[v(z) - \overline{\beta}w(z)]$ subject to $\mathbb{E}_F[\overline{\beta}w(z)] \geq h(\overline{\beta}w)$. Note that $h(\underline{\beta}w) > 0$, since otherwise the left side of the comparison (v) equals 1 but the right side is more than 1, impossible. Consequently, we have $\mathbb{E}_{\underline{F}}[w(z)] < h(w)$, since

otherwise

$$V_P(w) \leq \mathbb{E}_{\underline{F}}[v(z) - w(z)] < \mathbb{E}_{\underline{F}}[v(z) - \underline{\beta}w(z)] = V_P(\underline{\beta}w),$$

contrary to assumption. (The strict inequality step is where we use $h(\underline{\beta}w) > 0$.) Also, $\mathbb{E}_{\overline{F}}[w(z)] \geq h(\overline{\beta}w)/\overline{\beta} > h(w)$.

Write $\underline{\nu} = \mathbb{E}_{\underline{F}}[w(z)]$, $\nu = h(w)$, and $\overline{\nu} = \mathbb{E}_{\overline{F}}[w(z)]$, so we know that $\underline{\nu} < \nu < \overline{\nu}$. Put

$$F = \frac{\overline{\beta}\overline{\nu} - \nu}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}}\underline{F} + \frac{\nu - \underline{\beta}\underline{\nu}}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}}\overline{F}.$$

We claim $\mathbb{E}_F[w(z)] > h(w)$. Indeed, since $\overline{\beta}\overline{\nu} = \mathbb{E}_{\overline{F}}[\overline{\beta}w(z)] \geq h(\overline{\beta}w)$ and likewise $\underline{\beta}\underline{\nu} \geq h(\underline{\beta}w)$, we have

$$\frac{\nu - \underline{\beta}\underline{\nu}}{\nu - \underline{\nu}} \geq \frac{\nu - h(\underline{\beta}w)}{\nu - h(\underline{\beta}w)/\underline{\beta}} > \frac{h(\overline{\beta}w) - \nu}{h(\overline{\beta}w)/\overline{\beta} - \nu} \geq \frac{\overline{\beta}\overline{\nu} - \nu}{\overline{\nu} - \nu}.$$

Here the first inequality is because $(\nu - \underline{\beta}x)/(\nu - x)$ is increasing in x for $x < \nu$, the third inequality is similar, and the middle strict inequality comes from condition (v). Cross-multiplying,

$$(\nu - \underline{\beta}\underline{\nu})(\overline{\nu} - \nu) > (\overline{\beta}\overline{\nu} - \nu)(\nu - \underline{\nu}). \tag{7}$$

Then

$$\mathbb{E}_F[w(z)] = \frac{(\overline{\beta}\overline{\nu} - \nu)\underline{\nu} + (\nu - \underline{\beta}\underline{\nu})\overline{\nu}}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}} > \nu$$

where the inequality follows from rearranging (7).

Moreover, since $\mathbb{E}_F[w(z)] \geq h(w)$, we have $\mathbb{E}_F[v(z) - w(z)] \geq V_P(w)$. Combining this with the assumptions $V_P(w) \geq V_P(\underline{\beta}w) = \mathbb{E}_{\underline{F}}[v(z) - \underline{\beta}w(z)]$ and $V_P(w) \geq$

$V_P(\overline{\beta}w) = \mathbb{E}_{\overline{F}}[v(z) - \overline{\beta}w(z)]$, we have

$$
\begin{aligned}
\mathbb{E}_F[v(z) - w(z)] \;\geq\;& V_P(w) \\
\geq\;& \frac{\overline{\beta}\overline{\nu} - \nu}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}}\mathbb{E}_{\underline{F}}[v(z) - \underline{\beta}w(z)] + \frac{\nu - \underline{\beta}\underline{\nu}}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}}\mathbb{E}_{\overline{F}}[v(z) - \overline{\beta}w(z)] \\
=\;& \left(\frac{\overline{\beta}\overline{\nu} - \nu}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}}\mathbb{E}_{\underline{F}}[v(z)] + \frac{\nu - \underline{\beta}\underline{\nu}}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}}\mathbb{E}_{\overline{F}}[v(z)]\right) - \left(\frac{\overline{\beta}\overline{\nu} - \nu}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}}\underline{\beta}\underline{\nu} + \frac{\nu - \underline{\beta}\underline{\nu}}{\overline{\beta}\overline{\nu} - \underline{\beta}\underline{\nu}}\overline{\beta}\overline{\nu}\right) \\
=\;& \mathbb{E}_F[v(z)] - \nu \\
>\;& \mathbb{E}_F[v(z) - w(z)].
\end{aligned}
$$

This is our desired contradiction.

("Only if" direction) Evidently it suffices to prove the last statement of the proposition, since the conclusion of that sentence implies w is not scaling-dominated.

Assume none of (i)–(v) holds for w. Let $B = \{\beta > 1 | h(\beta w) > \beta h(w)\}$. Set $\underline{K} = \sup_{0 < \underline{\beta} < 1} \frac{h(w) - h(\underline{\beta}w)}{h(w) - h(\underline{\beta}w)/\underline{\beta}}$ and $\overline{K} = \inf_{\overline{\beta} \in B} \frac{h(\overline{\beta}w) - h(w)}{h(\overline{\beta}w)/\overline{\beta} - h(w)}$ (which we define as ∞ if $B = \emptyset$). For $0 < \underline{\beta} < 1$, it must be that $h(w) \geq h(\underline{\beta}w)/\underline{\beta}$, otherwise (ii) would hold. Furthermore, this inequality is strict, otherwise (iii) would hold (unless $h(w) = h(\underline{\beta}w) = 0$ but then (i) would hold). For any such $\underline{\beta}$, then, $\frac{h(w) - h(\underline{\beta}w)}{h(w) - h(\underline{\beta}w)/\underline{\beta}}$ is well-defined and is ≥ 1, hence $\underline{K} \geq 1$. We are ensured $\underline{K} < \infty$, again by the negation of (iii). We also have that $\underline{K} \leq \overline{K}$: either $B = \emptyset$ and $\overline{K} = \infty$, or if B is nonempty and $\underline{K} > \overline{K}$, property (v) holds. Finally, we know $\overline{K} > 1$, otherwise (iv) would hold.

Choose any finite K such that $K \in [\underline{K}, \overline{K}]$ and $K > 1$; the previous paragraph ensures this is possible. Set the valuation $v(z) = Kw(z)$. This is a valuation (i.e. its minimum value is 0) because w is grounded. Thus w is linear given v, with slope $1/K \in (0, 1)$. For any $\beta > 0$, under contract βw, the worst-case expected profit is $\left(\frac{K}{\beta} - 1\right)h(\beta w)$ (or lower, if $K < \beta$). So, to establish that $V_P(w|v) \geq V_P(\beta w|v)$, it suffices to show

$$(K - 1)h(w) \geq \left(\frac{K}{\beta} - 1\right)h(\beta w) \quad \text{for all } \beta > 0. \tag{8}$$

For $0 < \beta < 1$, this holds since $K \geq \underline{K}$ and $h(w) - h(\beta w)/\beta$ is positive. For $\beta \in B$, this holds since $K \leq \overline{K}$ and $h(\beta w)/\beta - h(w)$ is positive. If $\beta > 1, \beta \notin B$, then $\left(\frac{K}{\beta} - 1\right)h(\beta w) \leq (K - \beta)h(w) \leq (K - 1)h(w)$, unless $\beta > K$, in which case $\left(\frac{K}{\beta} - 1\right)h(\beta w) \leq 0 \leq (K - 1)h(w)$. Thus, (8) holds. Finally, for $\beta = 0$, $\Phi(\beta w) = $

$\Delta(Z)$, and we have $V_P(w|v) = (K-1)h(w) > 0 = V_P(\beta w|v)$ for this case as well.

$\square$

Proof of Theorem 4. Let v be the valuation obtained by applying the last statement of Proposition 3 to w'. Thus, $w'(z) = \alpha v(z)$ for some $\alpha \in (0, 1)$.

We hold v fixed. We will show that $V_P(w|v) > V_P(w'|v)$, which is sufficient, since we already know that w' has the highest guarantee among linear contracts.

Let F be the distribution for which Responsiveness fails. Since $F \in \Phi(w')$, it suffices to show that

$$\inf_{F' \in \Phi(w)} \mathbb{E}_{F'}[v(z) - w(z)] > \mathbb{E}_F[v(z) - w'(z)]. \tag{9}$$

By the hypothesis of Responsiveness, the left-hand side of (9) equals

$$\inf_{F' \in \Phi(w)} \mathbb{E}_{F'}[w'(z)/\alpha - w(z)] \geq \inf_{F' \in \Phi(w)} \mathbb{E}_{F'}\left[\left(\frac{1}{\alpha} - 1\right)w(z)\right] = \left(\frac{1}{\alpha} - 1\right)h(w),$$

the last inequality by definition of $h(w)$. Applying the other part of the hypothesis of Responsiveness, and using $F \notin \Phi(w)$, we have

$$\left(\frac{1}{\alpha} - 1\right)h(w) > \left(\frac{1}{\alpha} - 1\right)\mathbb{E}_F[w(z)] \geq \left(\frac{1}{\alpha} - 1\right)\mathbb{E}_F[w'(z)] = \mathbb{E}_F[v(z) - w'(z)]$$

which shows (9).

$\square$

References

Antić, Nemanja (2021) "Contracting with Unknown Technologies," Unpublished manuscript.

Barron, Daniel, George Georgiadis, and Jeroen Swinkels (2020) "Optimal Contracts with a Risk-Taking Agent," *Theoretical Economics*, 15, 715–761.

Carroll, Gabriel (2015) "Robustness and Linear Contracts," *American Economic Review*, 105 (2), 536–563.

Carroll, Gabriel and Delong Meng (2016) "Robust Contracting with Additive Noise," *Journal of Economic Theory*, 166, 586–604.

Dai, Tianjiao and Juuso Toikka (2021) "Robust Incentives for Teams," *Econometrica*, Forthcoming.

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

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

Garrett, Daniel F. (2014) "Robustness of Simple Menus of Contracts in Cost-Based Procurement," *Games and Economic Behavior*, 87, 631–641.

Garrett, Daniel F., George Georgiadis, Alex Smolin, and Balazs Szentes (2020) "Optimal Technology Design," Unpublished manuscript.

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

Kambhampati, Ashwin (2022) "Robust Performance Evaluation," Unpublished manuscript.

Malenko, Andrey and Anton Tsoy (2020) "Asymmetric Information and Security Design Under Knightian Uncertainty," Unpublished manuscript.

Marku, Keler and Sergio Ocampo Diaz (2019) "Robust Contracts in Common Agency," Unpublished manuscript.

Mirrlees, James A. (1999) "The Theory of Moral Hazard and Unobservable Behavior: Part I," *Review of Economic Studies*, 66 (1), 3–21.

Mookherjee, Dilip (2006) "Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective," *Journal of Economic Literature*, 44 (2), 367–390.

——— (2013) "Incentives in Hierarchies," in Gibbons, Robert and John Roberts eds. *Handbook of Organizational Economics*: Princeton University Press.

Tirole, Jean (1986) "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, & Organization*, 2 (2), 181–214.

Walton, Daniel and Gabriel Carroll (2019) "When are Robust Contracts Linear?", Unpublished manuscript.