When are Robust Contracts Linear?*

Daniel Walton ® Gabriel Carroll
Stanford University

May 18, 2019

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 show that if the possible responses to any given contract satisfy two properties — a *richness* and a *responsiveness* property — then a linear contract is optimal. This framework thus delineates a broad range of models in which linear contracts are optimally robust to uncertainty, including not only direct contracting with an agent, but also various models of hierarchical contracting and contracting with teams of agents. We also further apply the modeling apparatus to compare the principal's payoffs across different organizational structures.

1 Introduction

Suppose that a principal wishes to write an incentive contract to induce productive effort. The principal is poorly informed about the production technology, and therefore unsure of what actions might be taken in response to any given contract. How can she best write

*We thank Rohan Pitchford, Kieron Meagher, Andrés Carvajal, Ilya Segal, Idione Meneghel, Oleg Itskhoki, Ayça Kaya, Marina Halac, Stephen Morris, Matt Jackson, and Laura Doval for helpful comments and discussions, as well as audiences at ANU, BYU, UC Davis, Caltech, Johns Hopkins, and Texas A & M. ® denotes random author order (Ray ® Robson, 2018). This research was supported by a Sloan Foundation Fellowship. Parts of this work were done while the second author was visiting the Cowles Foundation at Yale and the Research School of Economics at ANU, and he gratefully acknowledges their hospitality.

a contract that is robust to this uncertainty? And how does the nature of this problem depend on the internal organization of the party on the other side of the contract?

Our starting point is the intuition that linear contracts — which simply pay some fixed fraction of the output produced — provide simple assurances by aligning the interests of the two contracting parties. Consider, for example, a principal and agent, both risk-neutral, who agree to a contract that pays the agent 1/4 of whatever output he produces. Suppose the agent is known to have some productive action he can take that assures him an expected payoff of at least 1500 under this contract. Then, even without knowing anything about what other actions the agent might have 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 the output against the agent's 1/4). This argument was developed in a previous paper by Carroll (2015), which formalized this idea of a guarantee via a worst-case criterion for the principal, and showed more generally that linear contracts are actually optimal under such a criterion. The intuitive explanation is as above: The principal cares about expected output, while the agent cares about expected payment. Linear contracts align these two objectives; and with enough uncertainty, there is no other way of better aligning them.

In reality, however, agency often takes place beyond simple bilateral relationships. For example, the principal may be a firm or government agency, 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. Does the robustness argument for linear contracts still hold up? Notice that this setting cannot simply be reduced to the principal-agent model above by treating the supplier as a single agent: even if the principal knows that some particular action is available to the worker within the supplier firm, it may not be implementable for the supplier due to the internal agency problem.

To express the question more concretely, extend the principal-agent model to a simple hierarchy: The principal contracts with a supervisor, specifying payment to the supervisor as a function of output; the supervisor then subcontracts with an agent, and the agent chooses the action that determines output. There are (at least) three natural ways to write this model, with different informational assumptions:

(i) As in the simpler principal-agent 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 technology.

(ii) The supervisor may know more actions than the principal does, but suspects that the agent may have 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).

As it turns out, 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 when the linearity argument will apply, and why.

Inspired by this contrast, the main contribution of our paper is to identify a broad class of models of contracting under uncertainty in which linear contracts give the best guarantees. This class will include the basic principal-agent model and variants (i) and (ii) of the hierarchical model above. We also give two other examples to illustrate the range of such models: one where a supervisor manages multiple agents in differentiated roles, and another that is a simplified version of the robust incentives for teams model of Dai and Toikka (2018). Delineating this class serves two purposes. First, it strips down the "linear is robust" argument to its essential features, helping to understand (for example) the relevant properties that are shared by models (i) and (ii) above but fail in (iii). And second, since linear contracts are relatively tractable, our approach can offer a guide to modelers looking to write convenient models for other organizational settings.

To achieve this generality, 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 correspondence $\Phi(w)$ specifying 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. We maintain risk-neutrality and the worst-case criterion throughout; we view these as natural "background" assumptions, since if nonlinearity were introduced into the model at these points we would not expect linearity in the conclusion. Instead, the focus is on the correspondence Φ.

We identify two properties of Φ that are jointly sufficient to ensure that linear contracts are optimal.[1] One is a *richness* property, requiring that the set of possible responses to any

[1]To be precise, in the models we will study, it may happen that the optimum is not attained, so that

given contract is diverse enough. The other is a *responsiveness* property, which essentially says that when the principal changes the contract, the set of possible responses changes in a way consistent with revealed preference, as if the counterparty were a single, risk-neutral and expected-utility-maximizing agent. This property is satisfied when the counterparty is run by a "leader" who can fully foresee the consequences of her own choices (as in model (i) above), but as we shall see, it can hold in other settings as well. It thus formalizes, in a broad way, the idea of the "counterparty's interest" being related to expected payment, an essential step in the argument that linear contracts align the counterparty's interest with the principal's.

After setting up this general framework in Section 3, we proceed in Section 4 to illustrate by formally detailing each of the specific models listed above, defining the resulting Φ, and checking that the richness and responsiveness properties are satisfied in each case. Version (iii) of the hierarchical model, which lies outside our framework, is covered in Section 5.

To substantiate our claim that the linearity result helps with tractability, in Section 6 we return to hierarchical model (i) and show how to complete the analysis of the optimal contract, i.e. how to identify the optimal slope. We do this by identifying the worst-case scenario for any given linear contract. This analysis also underscores the point that, although our models of different organizational structures share the prediction of linearity, they are not all equivalent to each other. In the simpler principal-agent model, the worst case is simply that the agent has one unknown action, that optimally exploits the given contract. By contrast, in the hierarchical model, the worst case involves a continuum of additional unknown actions, described by a differential equation. (We will also analyze model (ii), and show that the worst case is essentially also one with a single unknown action, though still different from the one critical unknown action in the principal-agent model.)

Since our investigations lead to a unified development of many different models of contracting under uncertainty, it is natural to place them side-by-side and compare. In Section 7, we illustrate by studying the question: Is the principal worse off in a hierarchy than she would be if she could contract with the agent directly? The answer is not entirely obvious: On one hand, more layers of the hierarchy lead to the "double marginalization of rents" problem that makes incentivizing effort more expensive (Tirole, 1994). On the other hand, with a supervisor, the principal may get to leverage the supervisor's

linear contracts can get arbitrarily close to — but not achieve — the supremum of the principal's payoff. We will address this more formally below, but for now will ignore the distinction.

better information about the agent's technology. We show that the first effect dominates. More specifically, we can compare versions (i) and (ii) of the hierarchical model against each other and also against the robust principal-agent model, holding fixed the principal's knowledge of the agent's technology. We find that the principal's maxmin guarantee is highest when contracting with the agent directly, followed by the fully-informed supervisor, and the partially-informed supervisor is worst.

Our work appears to be carving out a new niche in the theory of incentive contracts and organizations. There is a substantial existing literature seeking to explain the prevalence of relatively simple functional forms for contracts, including some approaches based on "richness" assumptions on the action space in a parallel spirit with ours (Holmstrom and Milgrom (1987); Innes (1990); Diamond (1998); Barron et al. (2019)). Among these, Barron et al. (2019) give an argument for linear contracts in their model that parallels the argument we give in Section 3. 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)). Yet there does not seem to be overlap between these two branches. But our work, particularly the analysis in Section 7, suggests that studying the former topic may be fruitful for the latter as well.

There is a separate strand of literature on contracting in hierarchies, such as Tirole (1986), that focuses on issues of collusion between the supervisor and agent in making reports to the principal. This is not relevant to our application to hierarchical contracting, since we do not allow communication back to the principal.

This work also contributes to the literature on explaining simple incentive structures as robust in unknown environments. Besides Carroll (2015), other examples studying various kinds of agency problems include Frankel (2014), Garrett (2014), and Carroll and Meng (2016). Closest in spirit to this work are the recent paper by Marku and Ocampo Diaz (2019), which applies a version of the robust contracting approach from Carroll (2015) to a common agency model; and that of Dai and Toikka (2018), which studies robust incentives for teams and which inspired one of the example models considered here.

2 Overview of Examples

Before setting up our general modeling framework, we first give brief verbal descriptions of the various examples for which we will apply it. This is meant to illustrate the range of situations that can be covered within our framework. The examples will be presented in formal detail in Section 4.

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 robust principal-agent model now applies to describe the relationship between the supervisor and the agent, implying 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.

This model 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 5.

Supervised team with differentiated roles. In this example, the supervisor oversees a team of two agents, who will both simultaneously take costly actions. Agent 1's action produces some intermediate good (in a stochastic amount), and agent 2's action determines how this intermediate good is (stochastically) mapped to final output. The intermediate good is not contractible, so instead the supervisor offers both agents contracts that specify payment as a function of final output. We assume the agents observe each other's contracts; the incentives provided by these contracts then determine a game between the agents, and they play a Nash equilibrium of this game.

As in hierarchical model (i), we assume the supervisor knows the technology fully. The principal does not: she knows some actions available to each agent, but each agent may have additional unknown actions, and the principal again uses a worst-case guarantee.

Unsupervised team. This example is based on the teams model of Dai and Toikka (2018). There is no supervisor in this model; the counterparty to the principal's contract consists directly of a team of several agents, who will simultaneously take actions. The principal knows of some actions (including their costs) available to each agent, but there may be other actions. For any action profile consisting of known actions, the principal knows what output distribution will result; but if at least one agent takes an unknown action, the distribution is potentially arbitrary. In Dai and Toikka's model, the principal can write contracts with each individual agent, specifying payment as a function of output. A key result is that such contracts cannot give the principal any positive guarantee unless the payments to different agents are linearly related to each other. Here, we bypass their argument (but draw inspiration from it) by assuming that the principal can only write a single contract with the team, and that payments to the team will automatically be split equally among the agents.

We acknowledge that one might quibble with embedded assumptions in some of these models. For example, the limited liability restriction, which we maintain throughout, may be less natural in the firm-to-firm contracting settings that we have used to motivate the

hierarchical models than in contracting with individuals. Nonetheless, it seems plausible in many settings, such as when the relevant firm division has a limited budget, or simply when large payments from the agent firm to the principal are ruled out by convention.[2]

In addition, one might view worst-case optimization as a strong assumption. As argued in Carroll (2015), in a principal-agent model, we can view this assumption not as a literal description of decision-making, but rather simply as a way of formalizing a robustness property of linear contracts. If linear contracts maximize the worst-case criterion, this identifies a sense in which they are special (as opposed to merely being one of many contractual forms that possess some robustness). By contrast, hierarchical models (ii) and (iii) assume that the supervisor uses a worst-case criterion; these models require a more positive interpretation of the criterion, since the principal must use it as a prediction of the supervisor's behavior. Nonetheless, our goal here is not to exhaustively explore the foundations, but just to offer a range of simple examples to illustrate the breadth of our framework. If a theorist wishes to propose alternative models for how the principal should expect the supervisor to make decisions, she could likewise check whether those alternative models satisfy the Richness and Responsiveness properties.

3 General Framework and Main Result

First, some notational conventions. We write $\Delta(X)$ for the space of Borel distributions on $X \subseteq \mathbb{R}$. We equip $\Delta(X)$ with the weak topology, represented by the Prohorov metric. 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 $C(X)$ 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

[2]Limited liability is indeed important. If we instead allowed payments to be arbitrarily negative, we would need to add a participation constraint; if we did so in the manner sketched in Subsection 3.3, 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.

incentives by promising payments to the counterparty. We abstract away from the internal structure of the counterparty (whether a single agent, a hierarchy, a team, etc.).

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 $w \in C^+(Y)$. Note that this definition incorporates the limited liability restriction: the contract must pay a nonnegative amount.[3]

We have a particular interest in linear contracts. A *linear contract* is one of the form

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

where $\alpha \geq 0$ is a constant.

We assume that we are given a nonempty-valued correspondence $\Phi : C^+(Y) \rightrightarrows \Delta(Y)$, the *outcome correspondence*, which summarizes the contracting situation. $\Phi(w)$ 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 there is some possible environment in which the counterparty would indeed generate F if offered contract w. For now we treat Φ as exogenously given; in each of the individual applications in Section 4, we will in turn define Φ from more primitive objects.

Any contract 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 to evaluate contract w is

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

Of course, we will need some conditions on Φ to obtain any results. We consider the following properties that Φ may have.

Property 1 (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

[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 principle one could replace continuity with other, weaker requirements. 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.

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.

Property 2 (Responsiveness). Suppose $w, w' \in C^+(Y)$, and $F \in \Delta(Y)$ such that $F \notin \Phi(w)$. If

$$\mathbb{E}_{F'}[w'(y)] - \mathbb{E}_F[w'(y)] \geq \mathbb{E}_{F'}[w(y)] - \mathbb{E}_F[w(y)] \qquad \text{for all } F' \in \Phi(w),$$

then $F \notin \Phi(w')$.

This is a "revealed preference" property, expressing how the possible outcomes respond to the incentives provided by expected payment. One way to understand it 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 two contracts w and w' satisfying the hypotheses, and let F be a distribution not in $\Phi(w)$. 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 $\mathbb{E}_{F'}[w(y)] - c(F') > \mathbb{E}_F[w(y)] - c(F)$, or equivalently $\mathbb{E}_{F'}[w(y)] - \mathbb{E}_F[w(y)] > c(F') - c(F)$. The hypothesis of Responsiveness implies the same holds with w' in place of w, so that 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) or in the differentiated-team model). But it also turns out to be satisfied in some other environments, such as hierarchy (ii) where the leader is not expected-value-maximizing, or the unsupervised team case where there is no leader.

For simplicity, we have not included a participation constraint. In Subsection 3.3 below, we describe how to accommodate such a constraint.

3.2 Linearity result

Now we come to our main result.

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 $V_P(w') \geq V_P(w)$.*

The proof comes in two steps. In Step 1, we observe that we can focus on concave contracts, since Richness implies that any non-concavities would potentially be exploited anyway. In Step 2, we show that a concave contract can in turn be improved to a linear contract, which offers the same expected payment but higher marginal incentives, in the worst case for the initial contract. In both steps, we rely on Responsiveness to show that changing the contract affects outcomes in the expected way.

Proof. Step 1. Let $w \in C^+(Y)$. Let $\hat{w}$ denote the pointwise concavification[4] of w. This is a concave function defined on $\text{co}(Y)$, the convex hull of Y. The restriction of $\hat{w}$ to Y is a contract; abusing notation, we will denote this contract also by $\hat{w}$. Note that $\hat{w}(y) \geq w(y)$ for all $y \in Y$. The goal of this step is to show that $V_P(\hat{w}) \geq V_P(w)$.

It is standard[5] that for each $y \in \text{co}(Y)$,

$$\hat{w}(y) = \sup \{\alpha w(y_1) + (1-\alpha) w(y_2) | \alpha \in [0,1], y_1, y_2 \in Y,\ \alpha y_1 + (1-\alpha) y_2 = y\}.$$

Since $Y \times Y \times [0,1]$ is compact and w continuous, this supremum is attained. For each $y \in \text{co}(Y)$, choose some y_1, y_2, α attaining the supremum, and define $F_y = \alpha \delta_{y_1} + (1-\alpha)\delta_{y_2}$. Note that $\hat{w}$ must coincide with w at points y_1 and y_2. (We may have $y_1 = y_2$.)

Now consider any $F \in \Phi(\hat{w})$, and let $\nu = \mathbb{E}_F[y]$. Take F_ν as above, and y_1, y_2, α as in the definition of F_ν. By Jensen's inequality,

$$\mathbb{E}_{F_\nu}[\hat{w}(y)] = \alpha \hat{w}(y_1) + (1-\alpha)\hat{w}(y_2) = \hat{w}(\nu) \geq \mathbb{E}_F[\hat{w}(y)].$$

Richness then implies that $F_\nu \in \Phi(\hat{w})$.

We use Responsiveness to show that $F_\nu \in \Phi(w)$ as well. Suppose, for contradiction, that $F_\nu \notin \Phi(w)$. Consider any $\tilde{F} \in \Phi(w)$. Since $\hat{w}$ coincides with w on the set $\{y_1, y_2\} = \text{supp}(F_\nu)$, we have $\mathbb{E}_{F_\nu}[w(y)] = \mathbb{E}_{F_\nu}[\hat{w}(y)]$. Since $w \leq \hat{w}$ everywhere, we also have $\mathbb{E}_{\tilde{F}}[w(y)] \leq \mathbb{E}_{\tilde{F}}[\hat{w}(y)]$. Consequently, $\mathbb{E}_{\tilde{F}}[\hat{w}(y)] - \mathbb{E}_{F_\nu}[\hat{w}(y)] \geq \mathbb{E}_{\tilde{F}}[w(y)] - \mathbb{E}_{F_\nu}[w(y)]$, which by Responsiveness implies $F_\nu \notin \Phi(\hat{w})$. This contradicts the previous paragraph. Therefore, we conclude $F_\nu \in \Phi(w)$.

[4]The *pointwise concavification* of the function w is defined pointwise as $\hat{w}(y) = \inf \{g(y) | g(y') \geq w(y')\ \forall y' \in Y,\ g \text{ concave on co}(Y)\}$ for $y \in \text{co}(Y)$.

[5]cf. Rockafellar (1970), Corollary 17.1.6.

Then, for each $F \in \Phi(\hat{w})$, we have found $F_\nu \in \Phi(w)$ such that

$$\mathbb{E}_F[y - \hat{w}(y)] \geq \mathbb{E}_{F_\nu}[y - \hat{w}(y)] = \mathbb{E}_{F_\nu}[y - w(y)]$$

and taking infima over $\Phi(\hat{w})$ and $\Phi(w)$ yields $V_P(\hat{w}) \geq V_P(w)$.

Step 2. We continue to refer to the function $\hat{w}$ from Step 1. We will construct a linear contract w' for which $V_P(w') \geq V_P(\hat{w})$.

Assume henceforth that $V_P(\hat{w}) > 0$, since otherwise we can just take the linear contract $w'(y) = 0$, and this step is already done.

Let $\mu^* = \inf_{F \in \Phi(\hat{w})} \mathbb{E}_F[y]$. If $\mu^* = 0$, i.e., arbitrarily low mean output may be produced, then evidently $V_P(\hat{w}) \leq 0$ and we are in the case above. So we can assume $\mu^* > 0$. Define $\lambda = \hat{w}(\mu^*)/\mu^*$. Define the function $w'(y) = \lambda y$ on $\text{co}(Y)$. Again, when restricted to Y, this is a (linear) contract, and we denote this contract also by w'.

Because $\hat{w} - w'$ is concave, is nonnegative at 0 and zero at μ^*, we have

$$y \geq \mu^* \implies w'(y) \geq \hat{w}(y), \tag{1}$$
$$y \leq \mu^* \implies w'(y) \leq \hat{w}(y). \tag{2}$$

Suppose $F \in \Delta(Y)$ is any distribution such that $\mathbb{E}_F[y] < \mu^*$. We will show that $F \notin \Phi(w')$. Put $\nu = \mathbb{E}_F[y]$. By definition of μ^*, $F \notin \Phi(\hat{w})$; and also, $F_\nu \notin \Phi(\hat{w})$.

Suppose by way of contradiction that $F \in \Phi(w')$. By linearity, $\mathbb{E}_{F_\nu}[w'(y)] = w'(\nu) = \mathbb{E}_F[w'(y)]$, so by Richness, $F_\nu \in \Phi(w')$ also. Now, let $\tilde{F}$ be any distribution in $\Phi(\hat{w})$, so $\mathbb{E}_{\tilde{F}}[y] \geq \mu^*$. Applying Jensen's inequality, (1), and (2), respectively, we obtain

$$\mathbb{E}_{\tilde{F}}[\hat{w}(y)] \leq \hat{w}\left(\mathbb{E}_{\tilde{F}}[y]\right) \leq w'\left(\mathbb{E}_{\tilde{F}}[y]\right) = \mathbb{E}_{\tilde{F}}[w'(y)];$$
$$\mathbb{E}_{F_\nu}[\hat{w}(y)] = \hat{w}(\nu) \geq w'(\nu) = \mathbb{E}_{F_\nu}[w'(y)].$$

These inequalities imply

$$\mathbb{E}_{\tilde{F}}[w'(y)] - \mathbb{E}_{F_\nu}[w'(y)] \geq \mathbb{E}_{\tilde{F}}[\hat{w}(y)] - \mathbb{E}_{F_\nu}[\hat{w}(y)],$$

and by Responsiveness, $F_\nu \notin \Phi(w')$, a contradiction.

We have shown that a distribution F cannot be in $\Phi(w')$ unless $\mathbb{E}_F[y] \geq \mu^*$.

Now take any $\varepsilon > 0$. Since $\hat{w}$ is continuous on $[0, \bar{y}]$, there exists $\delta > 0$ such that $|\hat{w}(y) - \hat{w}(\mu^*)| < \varepsilon$ whenever $|y - \mu^*| < \delta$. Hence, by taking $F \in \Phi(\hat{w})$ whose mean μ is sufficiently close to μ^*, we can ensure that $\mu \leq \mu^* + \varepsilon$ and $\hat{w}(\mu) \geq \hat{w}(\mu^*) - \varepsilon$. As we saw

in step 1, $F \in \Phi(\hat{w})$ implies $F_\mu \in \Phi(\hat{w})$ by Richness. Consequently,

$$V_P(\hat{w}) = \inf_{F \in \Phi(\hat{w})} \mathbb{E}_F[y - \hat{w}(y)] \leq \mathbb{E}_{F_\mu}[y - \hat{w}(y)] = \mu - \hat{w}(\mu) \leq \mu^* - \hat{w}(\mu^*) + 2\varepsilon,$$

and taking $\varepsilon \to 0$ gives

$$V_P(\hat{w}) \leq \mu^* - \hat{w}(\mu^*) = (1 - \lambda)\mu^*.$$

Our assumption $V_P(\hat{w}) > 0$ implies $\lambda < 1$. Now notice that by linearity of w', and the fact that $F \in \Phi(w')$ implies $\mathbb{E}_F[y] \geq \mu^*$ (established above), we have

$$\inf_{F \in \Phi(w')} \mathbb{E}_F[y - w'(y)] = \inf_{F \in \Phi(w')} \mathbb{E}_F[(1 - \lambda)y] \geq (1 - \lambda)\mu^* \geq V_P(\hat{w}).$$

This says exactly that $V_P(w') \geq V_P(\hat{w})$.

Combining Steps 1 and 2 completes the proof of the theorem. □

We comment briefly that the two properties as stated are actually much stronger than needed for the proof, since we use Richness only for certain distributions (with support size at most two), and likewise use Responsiveness only for specific pairs of contracts and distributions. However, we find the statements given here more succinct and interpretable than they would be if we tried to write the weakest possible versions.

Note also that both properties are indeed needed. 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 5 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.[6]

[6]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 tiebreaking provision (and show in detail that Responsiveness still holds) in several of our models in Section 4.

3.3 Participation constraint

Our general framework above does not include a participation constraint. One interpretation is that the outside option is low enough that any such constraint is non-binding: any contract satisfying limited liability would always be accepted by the counterparty.

However, we can also slightly extend the framework to model the possibility that the counterparty could reject the contract. We will briefly describe how to do so here, although for brevity we will not concern ourselves with the participation issue for the rest of the paper.

Let $Z \subseteq C^+(Y)$ be some (exogenously specified) set of contracts, which we interpret as the contracts that the counterparty would definitely accept. Let us assume the principal is interested in maximizing the worst-case guarantee V_P over contracts in Z. (She could then compare the resulting payoff to her outside option, which is the guarantee she would get by offering a contract outside Z.)

Suppose that Z satisfies the following property: If $w \in Z$, and $w' \in C^+(Y)$ such that $\mathbb{E}_F[w'(y)] \geq \mathbb{E}_F[w(y)]$ for all $F \in \Phi(w)$, then $w' \in Z$. This is essentially an analogue of Responsiveness for the participation decision. Notice that if $w \in Z$ is any contract satisfying $V_P(w) > 0$, then the contract w' constructed in the proof of Theorem 1 is also in Z. (More specifically: the property just stated directly implies that the $\hat{w}$ constructed in Step 1 is in Z; then to go from there to the w' constructed in step 2, notice that for every $F \in \Phi(\hat{w})$ we have $\mathbb{E}_F[w'(y)] = w'(\mathbb{E}_F[y]) \geq \hat{w}(\mathbb{E}_F[y]) \geq \mathbb{E}_F[\hat{w}(y)]$, where the first inequality is from (1) and the second is from concavity of $\hat{w}$.) Thus, when the principal is restricted to the set of contracts Z, she can still focus on linear contracts, as long as her outside option is nonnegative.

3.4 Existence of optimum

We have still been imprecise on one point: The verbal interpretation given to Theorem 1 is that it implies that a linear contract is optimal for the principal. Indeed, *if* an optimal contract exists, then there is one that is linear (just take w in Theorem 1 to be the optimal contract; then the w' in the theorem must also be optimal). However, the properties of Φ that we have stated do not assure existence of an optimum. If none exists, then 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.

Arguably, the existence question is a technical issue rather than an economic one. Nonetheless, it is 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 hemi-continuous, 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 some of the examples in Section 4, we use this result to show that an optimal contract exists.

4 Applications

We now proceed to detail the various applications of our framework previewed in Section 2, and show that the Richness and Responsiveness properties are satisfied. We also, in some cases, illustrate lower hemi-continuity in order to verify existence of an optimal contract. Hierarchical model (iii) does not satisfy Responsiveness, and its formal analysis is left to Section 5. In each of these applications, the main task is to define the outcome correspondence Φ corresponding to the model, and then show that Φ satisfies Richness and Responsiveness. To define the outcome correspondence, we must make assumptions about the organizational structure of the counterparty, behavior of agents within this structure, and which environmental details are uncertain from the perspective of the principal.

4.1 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 $(F, c) \in \mathcal{A}$. The principal is uncertain about what actions the agent can take, meaning we assume that the principal doesn't know $\mathcal{A}$. Instead, there is an exogenously given technology $\mathcal{A}^0$, representing all actions that are known by the principal to be available to the agent. The agent's actual technology is known to satisfy $\mathcal{A} \supseteq \mathcal{A}^0$. It is natural to assume that the agent can always choose to exert no effort and cause no output to be

produced with probability 1 (and that the principal knows this); this corresponds to the assumption that $(\delta_0, 0) \in \mathcal{A}^0$. However, we will not need to make this assumption.

Given this description of the model, we can define the outcome correspondence. Let $\Gamma_A(w, \mathcal{A}) = \{F \in \Delta(Y) | \exists c \geq 0, (F, c) \in \arg\max_{\mathcal{A}} V_A(\cdot, \cdot | w)\}$. In words, $\Gamma_A(w, \mathcal{A})$ is the set of distributions over output for which there exists a corresponding action $(F, c) \in \mathcal{A}$ such that (F, c) maximizes the agent's objective over $\mathcal{A}$ given w. Continuity of w and compactness of $\mathcal{A}$ ensure that $\Gamma_A(w, \mathcal{A})$ is nonempty. Finally, we assume that when there are multiple maximizers of the agent's objective, an action most beneficial to the principal is chosen. This assumption is a *tiebreaking condition*, as it says how we resolve the agent's indifference. We refer to the elements that satisfy the tiebreaking condition as *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

$$\Phi^{PA}(w) = \bigcup_{\text{technology } \mathcal{A} \supseteq \mathcal{A}^0} \Gamma_A^P(w, \mathcal{A}).$$

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. We verify that Richness and Responsiveness hold in this model. We will also verify that the restricted correspondence $\widetilde{\Phi}$ is lower hemi-continuous, 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 3. *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 tiebreaking technicalities.

Proof. (Richness) Let $w \in C^+(Y)$, $F \in \Phi^{PA}(w)$, so that there exists a technology $\mathcal{A} \supseteq \mathcal{A}^0$ 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. Take any technology $\mathcal{A} \supseteq \mathcal{A}^0$ containing (F, c) for some $c \geq 0$, and $(F', c') \in \Gamma_A^P(w, \mathcal{A})$ an action chosen by the agent under $\mathcal{A}$ and w, so that $F' \in \Phi^{PA}(w)$. Since $F \notin \Gamma_A^P(w, \mathcal{A})$, it must be that

$$V_A(F', c'|w) = \mathbb{E}_{F'}[w(y)] - c' > \mathbb{E}_F[w(y)] - c = V_A(F, c|w) \tag{3}$$

or that

$$V_A(F', c'|w) = V_A(F, c|w), \qquad \mathbb{E}_{F'}[y - w(y)] > \mathbb{E}_F[y - w(y)]. \tag{4}$$

Moreover, by hypothesis

$$\mathbb{E}_{F'}[w'(y)] - \mathbb{E}_F[w'(y)] \geq \mathbb{E}_{F'}[w(y)] - \mathbb{E}_F[w(y)]. \tag{5}$$

Then

$$\begin{aligned} V_A(F', c'|w') = \mathbb{E}_{F'}[w'(y)] - c' &\geq \mathbb{E}_F[w'(y)] + (\mathbb{E}_{F'}[w(y)] - c' - \mathbb{E}_F[w(y)]) \\ &\geq \mathbb{E}_F[w'(y)] - c \\ &= V_A(F, c|w') \end{aligned} \tag{6}$$

where the first inequality is by (5) and the second inequality is by (3) or (4). If (6) holds strictly, then $F \notin \Gamma_A(w', \mathcal{A})$ (not agent-optimal). Otherwise, we must be in case (4), and (5) holds as an equality. Equality in (6) means that if $F \in \Gamma_A(w', \mathcal{A})$, then $F' \in \Gamma_A(w', \mathcal{A})$ as well. Combining the second statement of (4) with the equality in (5) implies that $\mathbb{E}_{F'}[y - w'(y)] > \mathbb{E}_F[y - w'(y)]$, so that $F \notin \Gamma_A^P(w', \mathcal{A})$. Hence, no matter

whether (6) is strict or not, $F \notin \Gamma_A^P(w', \mathcal{A})$. Since $\mathcal{A}$ was arbitrary, $F \notin \Phi^{PA}(w')$.

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 $\epsilon > 0$. We want to show that there exists $\eta > 0$ such that $\alpha' \in \mathcal{B}_\eta(\alpha)$ implies that $\widetilde{\Phi}^{PA}(\alpha') \cap \mathcal{B}_\epsilon(F)$ is nonempty, where $\mathcal{B}_\eta(\alpha)$ is the Euclidean ball of radius η around α restricted to $[0,1]$, and $\mathcal{B}_\epsilon(F)$ is the ϵ-ball about F in the Prohorov metric.

If F has mean $\bar{y}$, then $F = \delta_{\bar{y}}$, and 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. So we can assume that $\mathbb{E}_F[y] < \bar{y}$.

Let $\mathcal{A}$ be a technology for which the agent produces distribution F. 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. Choose $F' \in \mathcal{B}_\epsilon(F)$ such that $\mathbb{E}_{F'}[y] > \mathbb{E}_F[y]$, which can be done by taking $F' = (1 - \beta)F + \beta\delta_{\bar{y}}$ and choosing $\beta > 0$ small.

Consider $\alpha > 0$. Since $f^*(\alpha) = \alpha\mathbb{E}_F[y] < \alpha\mathbb{E}_{F'}[y]$ and f^* is continuous, there exists some η such that $\alpha' \in \mathcal{B}_\eta(\alpha)$ implies $f^*(\alpha') < \alpha'\mathbb{E}_{F'}[y]$.

If $\alpha = 0$, we can still find $\eta > 0$ with $\alpha' \in \mathcal{B}_\eta(\alpha) \implies f^*(\alpha') < \alpha'\mathbb{E}_{F'}[y]$ for $\alpha' \neq 0$; otherwise, there exists some sequence $(F_n, c_n) \subseteq \mathcal{A}$, which we can assume (by taking a subsequence and using compactness) converges 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 $\Gamma_A(w_{\alpha'}, \mathcal{A}') = \Gamma_A^P(w_{\alpha'}, \mathcal{A}') = \{(F', 0)\}$, and $F' \in \widetilde{\Phi}^{PA}(\alpha') \cap \mathcal{B}_\epsilon(F)$. □

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 zero cost is not a substantive assumption; the logic can be carried over to more detailed models that explicitly restrict the plausible effort costs as a function of expected output. The equivalent step consists of adding an action to the technology that produces F at the lowest allowable cost. (See Carroll (2015), section II.A, for more details.)

4.2 Hierarchical Model (i)

In the three hierarchical models which we analyze, the hierarchical structure is the following. A 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 4.1. 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 $\mathcal{A}^0$, 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.

In hierarchical model (i), we assume that the supervisor is perfectly informed of $\mathcal{A}$. It must again include the actions known to the principal, that is, $\mathcal{A} \supseteq \mathcal{A}^0$. 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 S-A contracts available to the supervisor 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 so that the model is well-defined: it ensures that for each $w \in C^+(Y)$ and technology $\mathcal{A}$, there exists $w_A \in \mathcal{S}$ that maximizes the supervisor's objective function. (The necessity of some kind of restriction on S-A contracts here is demonstrated in Appendix B, where we show that otherwise the supervisor may fail to have a maximizing contract.)

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 for which the supervisor benefits the most, given that the agent is maximizing his objective. This again represents a tiebreaking condition, and we refer to elements of Γ_A^S as *supervisor-preferred*. The supervisor's objective in hierarchical model (i) is then

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

where we slightly abuse notation by writing $\mathbb{E}_{\Gamma_A^S(w,w_A,\mathcal{A})}$: the subscript is a set of distributions, not a single distribution, but the expectation is well-defined since it is independent of which distribution we choose in this set (and the set is nonempty, see below). The "i" in V_S^i stands for "informed." The supervisor maximizes V_S^i over $w_A \in \mathcal{S}$. In words, the supervisor maximizes the expected payment she receives from the principal minus the payment she makes to the agent, taking the agent's strategic action choice into account.

We impose one more tiebreaking condition in order to achieve lower hemicontinuity of the outcome correspondence. Define the *principal-preferred set* $\Gamma_A^{PS}(w, w_A, \mathcal{A}) = \arg\max_{F \in \Gamma_A^S(w,w_A,\mathcal{A})} \mathbb{E}_F[y - w(y)]$. This says that if there are multiple elements of $\Gamma_A^S(w, w_A, \mathcal{A})$, then the one maximizing the principal's payoff is chosen. Define

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

In words, for fixed P-S contract w and true technology $\mathcal{A}$, this is the set of output distributions that are possible, given that the supervisor is optimally choosing contract w_A and the agent is maximizing given w_A, along with the tiebreaking conditions.

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

$$\Phi^{PSA1}(w) = \bigcup_{\text{technology } \mathcal{A} \supseteq \mathcal{A}^0} \Gamma_S(w, \mathcal{A}).$$

The principal then evaluates contracts according to

$$V_P^{PSA1}(w) = \inf_{F \in \Phi^{PSA1}(w)} \mathbb{E}_F[y - w(y)].$$

This completes the description of the model. We should make sure Φ^{PSA1} is nonempty-valued: Recall that $\Gamma_A(w_A, \mathcal{A})$ is nonempty, and furthermore it is compact and upper hemicontinuous in w_A, by Berge's Theorem. Hence the set $\{(w_A, F) \in \mathcal{S} \times \Delta(Y) | F \in \Gamma_A(w_A, \mathcal{A})\}$ is compact. By continuity of the supervisor objective as a function of (w_A, F), there exists a maximizing pair (w_A, F), and the set of maximizers is compact. In turn, continuity of the principal's payoff ensures that $\Gamma_A^{PS}(w, w_A, \mathcal{A})$ is always nonempty. Hence, the set $\Gamma_S(w, \mathcal{A})$ is nonempty for each $w \in C^+(Y)$ and technology $\mathcal{A}$, which certainly ensures $\Phi^{PSA1}(w)$ nonempty.

To analyze the model, let us break down the definition of Φ^{PSA1}. For F to be in $\Phi^{PSA1}(w)$ means the following: there exists a technology $\mathcal{A} \supseteq \mathcal{A}^0$, cost $c \geq 0$ such that

$(F, c) \in \mathcal{A}$, and S-A contract $w_A \in \mathcal{S}$, satisfying

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

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

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

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

We argue that in checking whether criteria (a)–(d) can be satisfied for a given F, it suffices to consider technologies $\mathcal{A}$ containing $(F, 0)$ and $w_A \equiv 0$.

Lemma 4. *For any $w \in C^+(Y)$, $F \in \Phi^{PSA1}(w)$ if and only if $\exists$ a technology $\mathcal{A} \supseteq \mathcal{A}_0$ with $(F, 0) \in \mathcal{A}$, such that $\mathcal{A}$, action $(F, 0)$ and S-A contract $w_A \equiv 0$ satisfy (a)–(d).*

In showing this, we take a perspective on the supervisor's problem that will repeatedly prove useful in later applications as well: 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.

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 tiebreaking conditions; we leave the details to Appendix A.

We can use this lemma to show that the model falls under our general framework, and thus:

Proposition 5. *There exists a linear contract maximizing V_P^{PSA1}.*

We verify the Richness and Responsiveness properties, as well as the lower hemicontinuity property, by arguments very similar to those used in the robust principal-agent model. Along the way, Lemma 4 helps to simplify by reducing the space of possibilities to consider. In view of the similarity to the earlier arguments, we do not give the full proof of Proposition 5 in the text; it is left to Appendix A.

4.3 Hierarchical Model (ii)

Hierarchical model (ii) closely resembles the previously discussed hierarchical model. Here, the key difference is the assumption that the supervisor is not perfectly informed of $\mathcal{A}$. Instead, the supervisor is partially informed, at least as well as the principal is. Specifically, we assume that the principal knows about technology $\mathcal{A}^0$, the supervisor knows about technology $\mathcal{A}^1$, and the true technology is $\mathcal{A}$ such that $\mathcal{A}^0 \subseteq \mathcal{A}^1 \subseteq \mathcal{A}$. The principal is uncertain about both $\mathcal{A}$ and $\mathcal{A}^1$. Since the model continues to focus on the principal's problem, $\mathcal{A}^0$ is a primitive of the model, whereas $\mathcal{A}^1$ 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 contract in this model; 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 then

$$V_S^u(w_A | w, \mathcal{A}^1) = \inf_{\mathcal{A} \supseteq \mathcal{A}^1} V_S^i(w_A | w, \mathcal{A})$$

where V_S^i is the informed supervisor objective of hierarchical model (i), and $\mathcal{A}^1$ is the supervisor's knowledge of 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 $\mathcal{A}^1$.

For fixed P–S contract w, and technologies $\mathcal{A}, \mathcal{A}^1$, define

$$\Gamma_S(w, \mathcal{A}^1, \mathcal{A}) = \bigcup_{w_A \in \arg\max_{C^+(Y)} V_S^u(\cdot | w, \mathcal{A}^1)} \Gamma_A^{PS}(w, w_A, \mathcal{A}).$$

This is the set of output distributions such that the supervisor is choosing maximizing contract w_A according to $V_S^u(\cdot | w, \mathcal{A}^1)$, and the agent is maximizing (with supervisor-preferred and then principal-preferred tie-breaking) given w_A and $\mathcal{A}$. Note that the supervisor is maximizing according to her knowledge of technology $\mathcal{A}^1$, while the agent is maximizing according to the true technology $\mathcal{A}$.

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

$$\Phi^{PSA2}(w) = \bigcup_{\text{tech } \mathcal{A} \supseteq \mathcal{A}^1 \supseteq \mathcal{A}^0} \Gamma_S(w, \mathcal{A}^1, \mathcal{A}).$$

The principal thus evaluates contracts according to

$$V_P^{PSA2}(w) = \inf_{F \in \Phi^{PSA2}(w)} \mathbb{E}_F[y - w(y)].$$

As before, we should check nonemptiness. We know from the robust principal-agent model that the set of maximizers of the supervisor objective is nonempty, since there is a maximizing contract which is linear in the payment from the principal. (However, we have not restricted the supervisor to use such a contract; there may also be other optimal choices of w_A.) $\Gamma_A^{PS}(w, w_A, \mathcal{A})$ is nonempty as in hierarchical model (i). Hence the set $\Gamma_S(w, \mathcal{A}^1, \mathcal{A})$ is indeed nonempty for each $w \in C^+(Y)$, and technologies $\mathcal{A}^1$, $\mathcal{A}$, and consequently Φ^{PSA2} is nonempty-valued. Note that this is where we use the lower hemicontinuity result from the robust principal-agent model.

Like we did for the previous hierarchical model, let's break down the definition. For F to be in $\Phi^{PSA2}(w)$, this means the following: there exist technologies $\mathcal{A}$, $\mathcal{A}^1$ satisfying $\mathcal{A} \supseteq \mathcal{A}^1 \supseteq \mathcal{A}^0$, cost $c \geq 0$ such that $(F, c) \in \mathcal{A}$, and S-A contract $w_A \in C^+(Y)$, satisfying

(a) Supervisor maximization: the contract w_A maximizes $V_S^u(\cdot|w, \mathcal{A}^1)$ over $C^+(Y)$.

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

(c) Supervisor-preferred tiebreaking: given w, w_A, action (F, c) maximizes the supervisor's payoff $\mathbb{E}_F[w(y) - w_A(y)]$ over actions satisfying (b).

(d) Principal-preferred tiebreaking: given w, w_A, action (F, c) maximizes the principal's payoff $\mathbb{E}_F[y - w(y)]$ over actions satisfying (b)–(c).

Again, we argue that when checking if criteria (a)–(d) can be satisfied for a given F, it suffices to consider technologies $\mathcal{A} = \mathcal{A}^1$ containing $(F, 0)$ and $w_A \equiv 0$. The proof is similar to that for Lemma 4 and is in Appendix A. This fact is useful in showing that the model satisfies our two properties, and thereby obtaining our linearity result, Proposition 7.

Lemma 6. *For any $w \in C^+(Y)$, $F \in \Phi^{PSA2}(w)$ if and only if $\exists$ technologies $\mathcal{A}^1 = \mathcal{A}$ containing $(F, 0)$, such that these technologies, together with action $(F, 0)$ and S-A contract $w_A \equiv 0$, satisfy (a)–(d).*

Proposition 7. *There exists a linear contract maximizing V_P^{PSA2}.*

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

It might not be obvious that this model satisfies Responsiveness, since the supervisor is no longer an expected-utility maximizer. The key is Lemma 6 which shows, in effect, that we can reduce to a crucial subset of possible environments in which the supervisor does act like an expected-utility maximizer.

4.4 Supervised Team with Differentiated Roles

In this model, the supervisor oversees a team of two agents, who both simultaneously take costly actions. Agent 1's action produces some intermediate good in a compact set $Y_1 \subseteq \mathbb{R}^+$, and agent 2's action determines how the intermediate good is mapped to final output, which is some element of Y. Modeling this requires some more notation. Let $C(Y_1, \Delta(Y))$ denote the space of continuous functions from Y_1 to $\Delta(Y)$, endowed with the topology of uniform convergence.[7] Given $K \in C(Y_1, \Delta(Y))$, each $y_1 \in Y_1$ defines a probability measure $K(y_1) \in \Delta(Y)$. For $G \in \Delta(Y_1)$, and $K \in C(Y_1, \Delta(Y))$, define the probability measure $KG \in \Delta(Y)$ by

$$KG(A) = \int_{Y_1} [K(y_1)](A)\, G(dy_1),$$

for each subset A of Y.

The principal contracts with the supervisor through $w \in C^+(Y)$. The supervisor contracts with both agent 1 and agent 2 by choosing contracts w_{A1} and w_{A2}. The supervisor only observes final output, and must compensate both agents based only on this. We assume w_{A1} and w_{A2} are constrained to lie in $\mathcal{S}$, an exogenously specified, compact and convex subset of $C^+(Y)$ that contains all linear contracts with slopes $\alpha \in [0, 1]$. Agent 1 has access to an intermediate technology $\mathcal{A}_1$, a compact subset of $\Delta(Y_1) \times \mathbb{R}^+$. Agent 2 has access to an intermediate-to-final-output conversion technology, $\mathcal{A}_2$, which is a compact

[7] As in Chapter 19 of Aliprantis and Border (2006), this is the space of Markov transitions satisfying the *Feller property*.

subset of $C(Y_1, \Delta(Y)) \times \mathbb{R}^+$. When actions $(G, c_1) \in \mathcal{A}_1$ and $(K, c_2) \in \mathcal{A}_2$ are chosen by agents 1 and 2, respectively, final output is produced stochastically according to KG.

Like hierarchical model (i), we assume that the principal only knows $\mathcal{A}_1^0 \subseteq \mathcal{A}_1$ and $\mathcal{A}_2^0 \subseteq \mathcal{A}_2$, and the supervisor and agents 1 and 2 all know $\mathcal{A}_1$ and $\mathcal{A}_2$. Thus $\mathcal{A}_1^0$ and $\mathcal{A}_2^0$ are the primitives of the model. Given contracts w_{A1}, w_{A2} and actions $(G, c_1) \in \mathcal{A}_1$ and $(K, c_2) \in \mathcal{A}_2$, agent 1 and 2's payoffs are, respectively,

$$V_{A1}(G, c_1 | w_{A1}, K) = \mathbb{E}_{KG}[w_{A1}(y)] - c_1,$$
$$V_{A2}(K, c_2 | w_{A2}, G) = \mathbb{E}_{KG}[w_{A2}(y)] - c_2.$$

These payoffs (for fixed w_{A1}, w_{A2} and fixed technologies $\mathcal{A}_1$ and $\mathcal{A}_2$) define a simultaneous-move game between agents 1 and 2. Since the agent payoffs are continuous and action sets compact, there exists at least one mixed Nash equilibrium in this game, by an extension of Nash's existence theorem due to Glicksberg (1952). For any such equilibrium $\sigma = (\sigma_1, \sigma_2)$, we can write the resulting distribution over final output as $H(\sigma) = K(\sigma_2)G(\sigma_1)$, where $G(\sigma_1)$ is the weighted average over G generated by mixed strategy σ_1, and likewise $K(\sigma_2)$.

Let $\mathcal{E}(w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2)$ be the set of equilibria of the game, and let $\mathcal{E}^S(w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2) \subseteq \mathcal{E}(w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2)$ be the subset of equilibria that maximize the supervisor's payoff $\mathbb{E}_{H(\sigma)}[w(y) - w_{A1}(y) - w_{A2}(y)]$. These are the distributions induced from Nash equilibria that are most preferred by the supervisor. We thus assume that the supervisor can direct the agents as to which Nash equilibrium to play, given contracts w_{A1} and w_{A2}. This is similar to the supervisor-preferred tiebreaking assumptions in the previous models. We then write

$$V_S(w_{A1}, w_{A2} | w, \mathcal{A}_1, \mathcal{A}_2) = \mathbb{E}_{H(\sigma)}[w(y) - w_{A1}(y) - w_{A2}(y)]$$

for (any) $\sigma \in \mathcal{E}^S(w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2)$, and write $\Gamma_A^S(w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2)$ for the corresponding set of distributions $H(\sigma)$. Thus V_S is the supervisor's objective, and Γ_A^S is the set of distributions that may ensue. Now define

$$\Gamma_S(w, \mathcal{A}_1, \mathcal{A}_2) = \bigcup_{(w_{A1}, w_{A2}) \in \arg\max_{\mathcal{S} \times \mathcal{S}} V_S(\cdot, \cdot | w, \mathcal{A}_1, \mathcal{A}_2)} \Gamma_A^S(w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2).$$

In words, for fixed principal-supervisor contract w and true technologies $\mathcal{A}_1$ and $\mathcal{A}_2$, this is the set of final output distributions such that (a) the supervisor is choosing maximizing contracts w_{A1}, w_{A2} and (b) the agents are playing a supervisor-optimal Nash equilibrium

given w_{A1}, w_{A2}. Since Γ_A is nonempty and compact, by continuity of the supervisor's problem Γ_A^S is nonempty, and optimal w_{A1}, w_{A2} exist by compactness; hence $\Gamma_S(w, \mathcal{A}_1, \mathcal{A}_2)$ is nonempty. The outcome correspondence is then defined to be

$$\Phi^{ST}(w) = \bigcup_{\text{tech } \mathcal{A}_1 \supseteq \mathcal{A}_1^0, \mathcal{A}_2 \supseteq \mathcal{A}_2^0} \Gamma_S(w, \mathcal{A}_1, \mathcal{A}_2),$$

and the principal evaluates contracts according to

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

Here the "ST" stands for "supervised team."

Proposition 8. *For any $w \in C^+(Y)$, there exists a linear $w' \in C^+(Y)$ such that $V_P^{ST}(w') \geq V_P^{ST}(w)$.*

(For brevity, we do not concern ourselves with conditions for existence of the optimum in this model, or the next. Hence we simply check Richness and Responsiveness. This also means we do not bother with the extra principal-preferred tie-breaking as we did in earlier models.)

The proof of Proposition 8 is similar to those in the previous models. However, the argument for Richness requires a little more subtlety than before. In the earlier models with a single agent, the argument ran essentially as follows: take the technology under which the agent would produce the given distribution F, add to it the option to produce the new F' at cost 0, and check that distribution F' would indeed result. In the present model, the analogue is to add to agent 2's technology an extra action that always produces distribution F' (regardless of the value of y_1) at cost 0. When we do this, it is clear that the supervisor can induce F' by giving both agents the zero contract (analogously to the earlier hierarchical models), but it is not immediate that she would actually want to do so. The issue is that we cannot add F' without also making other new opportunities available to the supervisor, namely mixed Nash equilibria in which agent 2 mixes between $(F', 0)$ and one or more other actions (call this remaining part of 2's strategy σ_2). But with a little extra work, we can show that the supervisor cannot prefer to induce one of these other equilibria without contradicting the assumption that F was optimal originally.

4.5 Unsupervised Team

We now consider a different formulation with teams, one that is based on the main model in Dai and Toikka (2018), with some simplification. In the unsupervised teams model, the principal directly contracts with a team of $I \geq 2$ agents, indexed $i = 1, \ldots, I$. Agents simultaneously take unobservable costly actions, which jointly determine final output. The principal has uncertainty about both which costly actions the agents can take, and which distribution over output the unknown actions induce. Adopting the formalism from Dai and Toikka, a technology consists of a finite set $\mathcal{A} = \times_{i=1}^{I} \mathcal{A}_i$ (each $\mathcal{A}_i$ nonempty), and mappings $c_i : \mathcal{A}_i \to \mathbb{R}^+$ for each agent i, and $H : \mathcal{A} \to \Delta(Y)$. Agent i's action set is $\mathcal{A}_i$. Given a profile of mixed actions $\sigma = (\sigma_1, \ldots, \sigma_I)$ (so each σ_i is an element of $\Delta(\mathcal{A}_i)$), we define $H(\sigma) = \sum_{a \in \mathcal{A}} \sigma(a) H(a)$, where $\sigma(a) = \prod_i \sigma_i(a_i)$ is the probability of action profile a being played under σ. We also define for each player i the average cost of mixed action σ_i as $c_i(\sigma_i) = \sum_{a_i \in \mathcal{A}_i} \sigma_i(a_i) c_i(a_i)$.

The departure from Dai and Toikka's model comes through the contracts that the principal offers the agents. We assume that the principal can offer a contract $w \in C^+(Y)$, and the payment from w is equally split among agents, so that each agent receives $w(y)/I$ when y is the realized output. In contrast, Dai and Toikka assume that the principal can offer each agent a different contract, so that the principal chooses $(w_1, \ldots, w_I) \in C^+(Y)^I$. However, their analysis shows that the principal can only get a positive guarantee by offering the agents incentives that are affine transformations of each other. Our model thus adds just a slight further simplification by assuming that the payments offered to all agents are equal, thus allowing the model to fit within the single-contract framework developed in Section 3.

The payoff of agent i under pure strategy profile a and contract w is defined as

$$\mathbb{E}_{H(a)}\left[w(y)/I\right] - c_i(a_i).$$

We extend payoffs to mixed strategy profiles linearly, as usual. With these payoffs, a contract w and technology $(\mathcal{A}, H, c)$ defines a simultaneous-move normal form game. Denote $\mathcal{E}(w, \mathcal{A}, H, c)$ as the set of (mixed) Nash equilibria, which is nonempty since $\mathcal{A}$ is finite. In the case that there are many equilibria, we assume that an equilibrium σ maximizing the sum of agents' payoffs, $\mathbb{E}_{H(\sigma)}[w(y)] - \sum_i c_i(\sigma_i)$, is selected (henceforth, such an equilibrium is called "agents-optimal").[8] We denote the set of agents-optimal equilibria as

[8]Dai and Toikka (2018) instead assume the equilibrium that is best for the principal is played. This version of the model would require a bit more argumentation to fit with our framework, as Responsiveness

$\mathcal{E}^A(w, \mathcal{A}, H, c) \subseteq \mathcal{E}(w, \mathcal{A}, H, c)$. Let $\Gamma(w, \mathcal{A}, H, c) = \big\{H(\sigma) : \sigma \in \mathcal{E}^A(w, \mathcal{A}, H, c)\big\}$.

Consistent with all of the previous applications, we assume that the principal is poorly informed of the technology, so that the principal only knows $(\mathcal{A}^0, c_1^0, \ldots, c_I^0, H^0)$; these are the model primitives. It is assumed that the unknown $\mathcal{A} \supseteq \mathcal{A}^0$ is finite, $c_i : \mathcal{A}_i \to \mathbb{R}^+$ for all i, and $H : \mathcal{A} \to \Delta(Y)$, such that $c_i\big|_{\mathcal{A}_i^0} = c_i^0$ for all i, and $H\big|_{\mathcal{A}^0} = H^0$. We use notation $(\mathcal{A}, H, c) \supseteq (\mathcal{A}^0, H^0, c^0)$ to denote this relationship. Additionally, we assume that $c_i^0(a_i) > 0$ for all $a_i \in \mathcal{A}_i^0$, for all i. (This assumption simplifies the proofs, and will be discussed more later.) Hence the outcome correspondence is defined as

$$\Phi^{UT}(w) = \bigcup_{(\mathcal{A},H,c)\supseteq(\mathcal{A}^0,H^0,c^0)} \Gamma(w, \mathcal{A}, H, c).$$

The principal evaluates contracts according to

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

The "UT" stands for "unsupervised teams." We show that the outcome correspondence Φ^{UT} satisfies Richness and Responsiveness, hence linear contracts are optimal in this environment, as Dai and Toikka also show.

We begin with a useful characterization of Φ^{UT}. First, observe that for any fixed $(\mathcal{A}, H, c)$ and contract w a potential game is induced, with potential $P : \mathcal{A} \to \mathbb{R}$ defined by

$$P(a) = \mathbb{E}_{H(a)}[w(y)] - I \sum_{i=1}^{I} c_i(a_i).$$

(Precisely, the potential is $(1/I) \cdot P(a)$, but it will be more convenient for us to work with P.) Let a^0 denote a maximizer of P among action profiles in $\mathcal{A}^0$, and let $w^0 = P(a^0)$ be the corresponding maximum value.

Lemma 9. $\Phi^{UT}(w) = \{F \in \Delta(Y) : \mathbb{E}_F[w(y)] > w^0\}$.

The proof, which adapts techniques from Dai and Toikka (2018), is in Appendix A. The argument that every distribution that may be chosen does indeed satisfy $\mathbb{E}_F[w(y)] > w^0$ is essentially a direct application of the potential game structure, with additional use of the equilibrium selection criterion and the assumption $c_i^0(a_i) > 0$ to ensure the

can be violated for some (undesirable) contracts.

inequality holds strictly. For the converse, given a distribution F satisfying the inequality, we construct a new technology by adding a single zero-cost action to each agent's action set, so that when all agents play the new action, the resulting distribution is F. We carefully specify the distributions at all of the other new profiles (where some, but not all, agents play their new action) to make the new action dominant for each agent, thereby making F the unique equilibrium outcome.

With the characterization of Φ^{UT} in Lemma 9, it is straightforward to check Richness and Responsiveness, allowing us to apply Theorem 1.

Proposition 10. *For any $w \in C^+(Y)$, there exists a linear $w' \in C^+(Y)$ such that $V_P^{UT}(w') \geq V_P^{UT}(w)$.*

Again, the proof is in Appendix A.

In order to obtain this result without resorting to more intricate arguments, we made two simplifying assumptions: (a) that all agents receive the same share of w, $1/I$, and (b) the cost of every action in $\mathcal{A}^0$ is strictly positive. In fact, both of these assumptions may be relaxed: we can allow all agents to receive different shares of w (including a 0 share), as long as they sum to 1, and we can allow some actions in $\mathcal{A}^0$ to have zero cost. Under the relaxation of (a), our characterization of $\Phi^{UT}(w)$ remains valid but requires a significantly more careful argument.[9] Under the relaxation of (b), the characterization is almost true: $\Phi^{UT}(w)$ contains the actions identified in Lemma 9, but may also include the boundary $\{F \in \Delta(Y) : \mathbb{E}_F[w(y)] = w^0\}$. Richness still holds in this case, but Responsiveness may not hold. Instead, a weakened version of Responsiveness holds, sufficient to preserve the linearity result, with only minor changes to the proof. This property is stated below.

Property 2′ (Generalized Responsiveness). Suppose $F \in \Delta(Y)$ such that $F \notin \text{cl}(\Phi(w))$, where $\text{cl}(\cdot)$ denotes closure. If

$$\mathbb{E}_{F'}[w'(y)] - \mathbb{E}_F[w'(y)] \geq \mathbb{E}_{F'}[w(y)] - \mathbb{E}_F[w(y)] \qquad \text{for all } F' \in \Phi(w),$$

then $F \notin \Phi(w')$.

5 A Counterexample

In this section we describe version (iii) of the hierarchical model, where the supervisor shares the principal's ignorance about the technology, knowing only that it is a superset of

[9] The full argument is contained in Dai and Toikka (2018), Lemma A.6 of the Appendix.

the given $\mathcal{A}^0$, and 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, and identify an alternative contract that is optimal in the example. In the process, we also observe that this model satisfies Richness, so it must be a failure of Responsiveness that prevents Theorem 1 from applying.

The model is structured quite similarly to hierarchical models (i) and (ii). As in model (ii) (and unlike (i)), we will not restrict the supervisor to a compact set of contracts $\mathcal{S}$, since the restriction is not needed for existence of a best reply by the supervisor.

Here are the details. We take Y and $\mathcal{A}^0 \subseteq \Delta(Y) \times \mathbb{R}^+$ as in the first two versions of the model. For any contract w_A to the agent, and any true technology $\mathcal{A} \supseteq \mathcal{A}^0$, define $\Gamma_A^S(w_A, w, \mathcal{A})$ as before. Define $V_S^i(w_A|w, \mathcal{A})$ as in model (i), and define

$$V_S^u(w_A|w, \mathcal{A}^0) = \inf_{\mathcal{A} \supseteq \mathcal{A}^0} V_S^i(w_A|w, \mathcal{A})$$

as in model (ii). This is the supervisor's objective.

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 will *assume* for now that the supervisor uses an optimal contract of this form (but if there happen to be multiple choices of β that 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. At the end of this section, we will argue that removing the restriction will not change the main result.

Accordingly, define

$$\Gamma_S(w, \mathcal{A}^0, \mathcal{A}) = \bigcup_{\beta \in \arg\max_{\beta \in [0,1]} V_S^u(\beta w|w, \mathcal{A}^0)} \Gamma_A^S(\beta w, w, \mathcal{A}).$$

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 $(w, \mathcal{A}^0)$. The union arises due to the possibility of multiple optimal choices of β. Finally, the outcome correspondence is given by

$$\Phi^{PSA3}(w) = \bigcup_{\mathcal{A} \supseteq \mathcal{A}^0} \Gamma_S(w, \mathcal{A}^0, \mathcal{A}).$$

Accordingly, the principal's objective is

$$V_P^{PSA3}(w) = \inf_{F \in \Phi^{PSA3}(w)} \mathbb{E}_F[y - w(y)].$$

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

Let us first formally note, as promised previously:

Proposition 11. *The correspondence Φ^{PSA3} satisfies Richness.*

This is fairly immediate; the formal proof is in Appendix A.

Now, to proceed further, let us characterize $\Phi^{PSA3}(w)$ in more detail. If the principal offers 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 $(F, c) \in \mathcal{A}^0$ 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 $\mathcal{A}^0$, 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 Appendix A):

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

$$E_{F'}[w(y)] \geq E_F[w(y)] - \sqrt{c \cdot E_F[w(y)]} \tag{7}$$

where (F, c) is the targeted action leading to slope β.

Now let us henceforth focus on a particular, parametric specification of $\mathcal{A}^0$ under which we can identify optimal contracts. Assume that Y is finite, so that we can avoid worrying about continuity restrictions 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 $y_L/c_L > y_H/c_H > 1$. Let

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

That is, there are three known actions, all deterministic: the agent can produce a high output level at high cost, low output at low cost, or no output at no cost. For brevity, we will call these actions "action H," "action L," and "action 0."

Consider any contract w that the principal could offer. A degenerate contract cannot give a positive guarantee (since the supervisor could choose $\beta = 0$, and then action 0 is optimal for the agent under technology $\mathcal{A}^0$), so we may focus on non-degenerate contracts. These fall into two sets: those that target action L, and those that target H.

Let w be any contract targeting action L. Note that we can safely replace w by the contract w' given by $w'(y_L) = w(y_L)$ and $w'(y) = 0$ for all other y. Indeed, since w' lies pointwise below w, with equality at y_L, (7) holds for a weakly smaller set of distributions F' under w', i.e. $\Phi(w') \subseteq \Phi(w)$. (Note also that w' does not target any action other than L.) So the minimization problem defining $V_P(w')$ is over a smaller set than that defining $V_P(w)$, and the minimand is also weakly higher for w'; hence $V_P(w') \geq V_P(w)$.

This shows that in looking for an optimum among contracts targeting L, we can restrict attention to those satisfying $w(y_L) = \psi$ and $w(y) = 0$ for all other y, where ψ is a value greater than c_L. We can also restrict to $\psi < y_L$, since otherwise the principal gets no positive guarantee (the agent may produce δ_{y_L}). Moreover, for any such contract, Lemma 12 identifies the set of distributions $\Phi(w)$: it consists of all distributions that produce y_L with probability p such that $p \cdot \psi \geq \psi - \sqrt{c_L \cdot \psi}$, i.e. all distributions that produce y_L with probability at least $1 - \sqrt{c_L/\psi}$. Clearly, the worst case for the principal is that y_L is produced with exactly this probability, and otherwise 0 is produced. This leads to the worst-case payoff

$$V_P(w) = \left(1 - \sqrt{\frac{c_L}{\psi}}\right)(y_L - \psi). \tag{8}$$

We note in passing that there does not seem to be a convenient analytical expression for the optimal choice of ψ (which is the solution to a cubic equation).

By similar logic, among the contracts targeting H, we can restrict attention to those satisfying $w(y_H) = \psi$ and $w(y) = 0$ for all other y, where $c_H < \psi < y_H$. Such a contract does not target any other action, and by exactly the same reasoning as above, its worst-case payoff is

$$V_P(w) = \left(1 - \sqrt{\frac{c_H}{\psi}}\right)(y_H - \psi). \tag{9}$$

Overall, then, an optimal contract can be found by maximizing each of (8) and (9) over ψ, and choosing the larger of the two values.

What if the principal uses a linear contract, $w(y) = \alpha y$? 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

$$V_P(w_\alpha) = (1-\alpha)\left(y_L - \sqrt{\frac{c_L y_L}{\alpha}}\right). \tag{10}$$

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

$$V_P(w_\alpha) = (1-\alpha)\left(y_H - \sqrt{\frac{c_H y_H}{\alpha}}\right). \tag{11}$$

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 we can directly compare the payoff from maximizing over linear contracts to the payoffs obtainable by maximizing (8) and (9), and see that for some parameter values, the latter do strictly better. For example, take $(y_L, y_H, c_L, c_H) = (5, 15, 1, 6)$. We can directly calculate the maximum in (9), which occurs for $\psi = 9.71$ and has payoff guarantee 1.13. By contrast, the guarantee from a linear contract $V_P(w_\alpha)$ is graphed in Figure 1; the supremum occurs as $\alpha \to \alpha^* = 0.7841$ from above, and the corresponding payoff guarantee is $0.93 < 1.13$. (The supremum is not attained, since the value of the guarantee jumps downward at the discontinuity point.)

What is happening in this example? Notice that the formula for the payoff from a linear contract that targets action L is equal to formula (8), under the substitution

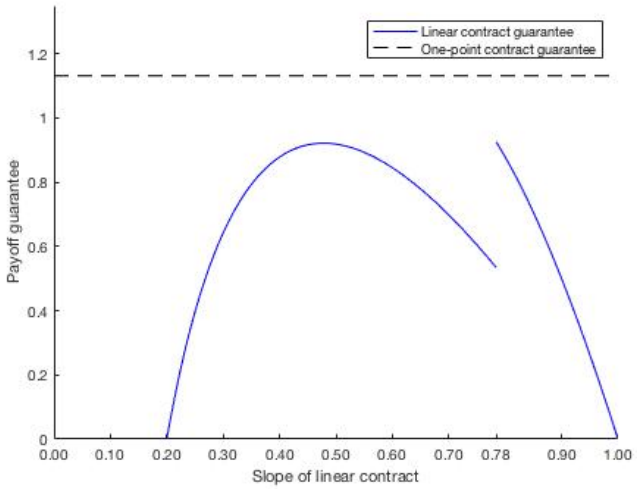

Figure 1: Guarantee from a linear contract, in example for hierarchical model (iii).

$\psi = \alpha y_L$. Similarly for action H and formula (9), with $\psi = \alpha y_H$. Thus, the principal is maximizing the same objective when choosing an optimal linear contract as an optimal nonlinear contract, but is maximizing over a restricted parameter range. Indeed, with the numbers above, the principal would like to target action H with $\psi = 9.71$, which corresponds under the above substitution to $\alpha = 0.6475$; but a linear contract with this slope α targets action L instead.

A rough intuition, then, is that under a linear contract, since the supervisor receives only a fraction of the output, she has less inclination than the principal does 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. Notice, also, that this argument would not apply in model (i) (or model (ii)) because there the supervisor may know of other actions besides H or L: in the worst case, the supervisor targets some new action that produces output y_H with some probability and 0 otherwise, and nonlinearity will not help the principal avoid this bad outcome.

Before leaving the discussion of model (iii), there is one loose end to tie up. We have assumed the supervisor always offers the agent a contract of the form $w_A = \beta w$, but there may also be other contracts that are optimal from the supervisor's point of view. We commented earlier that we could let the supervisor offer any such w_A, without changing the basic conclusion that linear contracts w can fail to be optimal. Now is the time to justify this comment. For brevity we keep the discussion less formal.

Keep the same Y and $\mathcal{A}^0$ as in the example above, but suppose that the supervisor is no longer restricted to contracts $w_A = \beta w$. In the example, the principal's optimal contract took the "one-point" form $w(y_H) = \psi > 0$, $w(y) = 0$ for all $y \neq y_H$, and this was strictly better than any linear contract. Dropping the restriction on the supervisor's behavior can only expand the set of F's that may ensue, and so can only lower the value of the worst-case guarantee V_P for any given contract. Let us now argue that for the one-point w, the supervisor has no optimal contracts other than those of the form βw. This will imply that $V_P(w)$ remains unchanged when we drop the restriction on w_A, and therefore this w remains strictly better than any linear contract.

Indeed, suppose the principal offers w as above. Suppose the supervisor offers some w_A that is not a multiple of w, i.e. $w_A(y) > 0$ for some $y \neq y_H$. Consider replacing w_A by the contract w'_A such that $w'_A(y_H) = w_A(y_H)$, and $w'_A(y) = 0$ for all other y. It is not hard to check that the worst case for the supervisor under w_A is strictly worse than under w'_A (briefly, under w_A the agent puts some mass on y with $w(y) = 0 < w_A(y)$, which is costly to the supervisor). Hence w_A is not optimal for the supervisor.

6 Optimal Contract Slope

Now that we have shown how several models fit into our linearity framework, we shift focus toward identifying the particular contract that is optimal in hierarchical models (i) and (ii) (and its guarantee). Part of our overall claim for the value of linearity results is that they aid in writing tractable models, and a natural test of tractability is whether we can actually characterize the optimal contract. Thus, this section aims to illustrate how these example models meet this test.

We have already shown that we can focus on linear contracts in these environments. To characterize the optimal slope analytically, 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 specific function. This makes it possible to apply the Lagrange conditions and solve for the optimal slope.

6.1 Hierarchical Model (i)

We begin with the analysis for hierarchical model (i). Assume the principal offers a particular linear contract w_α.

The first step is to show that, in defining $\Phi^{PSA1}(w_\alpha)$, 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, where this cost is a convex, nondecreasing function of the mean. We show that any distribution that could ever be produced could in fact arise for some technology of this form.

To be precise: first fix a number $\overline{c}$ such that $\overline{c} > \bar{y}$ and $\overline{c} > \max_{(F,c)\in\mathcal{A}^0} c$. Now, say that a function $\kappa : \text{co}(Y) \to \mathbb{R}^+$ is a *valid cost function* if it is (weakly) convex, continuous, nondecreasing, satisfies $\kappa(0) = 0$, and $\kappa(\mathbb{E}_F[y]) \leq c$ for every $(F, c) \in \mathcal{A}^0$. For any such κ, define the technology $\mathcal{A}_\kappa$ to consist of all actions $(F, c) \in \Delta(Y) \times \mathbb{R}^+$ such that $\kappa(\mathbb{E}_F[y]) \leq c \leq \overline{c}$. This is indeed a technology (the $\overline{c}$ bound ensures compactness), and the assumption on κ ensures that it contains $\mathcal{A}^0$.

The next step helps in characterizing the supervisor's behavior when the technology is of this form. In particular, we will find that the supervisor offers a linear contract in this case. One may note that the fact that both the principal and supervisor offer linear contracts totally aligns the way they rank outcomes: as long as $\alpha > 0$, both strictly prefer a higher mean outcome to a lower mean outcome, and are indifferent among outcomes with the same mean.

For any valid cost function κ, write $\kappa'(\mu)$ as the left-hand derivative of κ at μ, when $\mu > 0$. For $\mu = 0$, put $\kappa'(\mu) = 0$.

Proposition 13. *Let κ be a valid cost function, w a linear principal-supervisor contract, and suppose the true technology is $\mathcal{A}_\kappa$. Let $\mu \in [0, \overline{y}]$ and $F \in \Delta(Y)$ such that $\mathbb{E}_F[y] = \mu$. Then $F \in \Gamma_A(w_{\kappa'(\mu)}, \mathcal{A}_\kappa)$, where $w_{\kappa'(\mu)}$ is the linear contract with slope $\kappa'(\mu)$. For w_A such that $\mathbb{E}_F[w_A(y)] < \mu\kappa'(\mu)$, $F \notin \Gamma_A(w_A, \mathcal{A}_\kappa)$, i.e. the supervisor cannot induce a distribution with mean μ at a cost less than $\mu\kappa'(\mu)$.*

Proof. Suppose the supervisor offers a linear contract with slope β. Then the agent can choose any distribution with any mean μ, and get an expected payoff of $\beta\mu - \kappa(\mu)$. The first-order condition implies that any given choice of μ is optimal for the agent if $\beta = \kappa'(\mu)$ (and in this case, the agent is indifferent among all distributions with mean μ, since they are all equally costly and pay the same to the agent). This proves the first part of the proposition.

To see that a distribution with mean μ cannot be induced at cost less than $\mu\kappa'(\mu)$, let w_A be any contract that the supervisor offers to the agent, and suppose the agent chooses distribution F with mean μ. Assume $\mu > 0$ (otherwise the desired conclusion is obvious).

Consider any $\tilde{\mu} \in (0, \mu)$, and let $\tilde{F}$ be the distribution $\frac{\tilde{\mu}}{\mu}F + \left(1 - \frac{\tilde{\mu}}{\mu}\right)\delta_0$. By assumption, the agent can produce $\tilde{F}$ at cost $\kappa(\mathbb{E}_{\tilde{F}}[y]) = \kappa(\tilde{\mu})$. So the fact that he is willing to produce F rather than $\tilde{F}$ implies

$$\begin{aligned}
\mathbb{E}_F[w_A(y)] - \kappa(\mu) &\geq \mathbb{E}_{\tilde{F}}[w_A(y)] - \kappa(\tilde{\mu}) \\
&= \frac{\tilde{\mu}}{\mu}\mathbb{E}_F[w_A(y)] + \left(1 - \frac{\tilde{\mu}}{\mu}\right) w_A(0) - \kappa(\tilde{\mu}) \\
&\geq \frac{\tilde{\mu}}{\mu}\mathbb{E}_F[w_A(y)] - \kappa(\tilde{\mu}),
\end{aligned}$$

or by rearranging,

$$\mathbb{E}_F[w_A(y)] \geq \mu \left(\frac{\kappa(\mu) - \kappa(\tilde{\mu})}{\mu - \tilde{\mu}}\right).$$

Taking $\tilde{\mu} \to \mu$ from below, the right side converges to $\mu\kappa'(\mu)$, so this is a lower bound on the expected payment to the agent. □

This proposition pins down the cost to the supervisor to induce any given mean output. Consequently, the supervisor's maximization problem boils down to: $\max_{\mu \in [0, \bar{y}]}(\alpha - \kappa'(\mu))\mu$.[10]

Now, as indicated earlier, we can show that we can restrict our attention to technologies of the form $\mathcal{A}_\kappa$ for some κ.

Proposition 14. *Let w_α be a linear contract, $\alpha \in [0, 1]$, and let $F^* \in \Phi^{PSA1}(w_\alpha)$. Then there exists a valid cost function $\kappa : \text{co}(Y) \to \mathbb{R}^+$, such that $\kappa(\mathbb{E}_{F^*}[y]) = 0$, and $F^* \in \Gamma_S(w_\alpha, \mathcal{A}_\kappa)$.*

The proof is in the Appendix, but here is a summary: From Lemma 4, we know that there must be some technology $\mathcal{A}$ in which F is available at cost zero, and the supervisor induces F by offering the zero contract. We then enlarge $\mathcal{A}$ by specifying that whenever some distribution is available, all other distributions with weakly lower mean are available at the same cost; we also convexify the technology. We check that the new technology can be described in the form $\mathcal{A}_\kappa$, and that the supervisor would still prefer to induce F (at cost zero) rather than any other distribution. This latter fact draws on Proposition 13.

[10] Tiebreaking issues may arise when κ' is constant on some interval. Notice, however, that Proposition 13 implies that the max asserted here is an upper bound on the supervisor's payoff, and also that the supervisor can make the agent willing to choose an action that gives her this payoff, which means by supervisor-preferred tiebreaking that it is a lower bound as well.

The next step is to identify the lowest mean output that might be induced under a given linear contract w_α and known technology $\mathcal{A}^0$. We do this by further restricting the cost functions κ under consideration to take a specific form. Namely, suppose that under $\mathcal{A}_\kappa$, the supervisor induces the agent to take action with mean μ, as in Proposition 14. Our goal is to show that the same mean output μ can arise when we replace κ by a valid cost function $\tilde{\kappa}$ that satisfies

$$\tilde{\kappa}(\tilde{\mu}) = 0 \qquad \text{for } \tilde{\mu} \le \mu \tag{12}$$

$$\mu\alpha = \tilde{\mu}[\alpha - \tilde{\kappa}'(\tilde{\mu})] \qquad \text{for } \tilde{\mu} > \mu \tag{13}$$

The condition (13) says that the supervisor is indifferent over all mean output levels $\tilde{\mu} \in [\mu, \bar{y}]$ that she could induce, so that in particular, she is indeed willing to induce mean output μ. This condition defines a differential equation for $\tilde{\kappa}$ with boundary condition $\tilde{\kappa}(\mu) = 0$ (given by (12)). We can solve the differential characterization $\mu\alpha = \tilde{\mu}[\alpha - \tilde{\kappa}'(\tilde{\mu})]$ with $\tilde{\kappa}(\mu) = 0$ to find $\tilde{\kappa}$. For $\tilde{\mu} > \mu$, integration yields

$$\tilde{\kappa}(\tilde{\mu}) = \tilde{\kappa}(\tilde{\mu}) - \tilde{\kappa}(\mu) = \int_\mu^{\tilde{\mu}} \left(1 - \frac{\mu}{x}\right) \alpha \, dx = \alpha[(\tilde{\mu} - \mu) - \mu(\log \tilde{\mu} - \log \mu)]$$

and the full form of $\tilde{\kappa}$ is

$$\tilde{\kappa}(\tilde{\mu}) = \begin{cases} 0 & \tilde{\mu} \le \mu \\ \alpha[(\tilde{\mu} - \mu) - \mu \log(\tilde{\mu}/\mu)] & \tilde{\mu} > \mu. \end{cases} \tag{14}$$

(For the limiting case $\mu = 0$, we interpret $\mu \log \mu$ as 0, thus giving $\tilde{\kappa}(\tilde{\mu}) = \alpha\tilde{\mu}$.)

To check that this cost function $\tilde{\kappa}$ is valid, we need to verify that it still satisfies $\tilde{\kappa}(\mathbb{E}_F[y]) \le c$ for all known actions $(F, c) \in \mathcal{A}^0$. (The other conditions defining a valid cost function are easily checked.) Since the original κ was a valid cost function, it suffices to check that $\tilde{\kappa}(\tilde{\mu}) \le \kappa(\tilde{\mu})$ for all $\tilde{\mu} \in [0, \bar{y}]$. This is done in the proof of the following lemma, which is in Appendix A.

Lemma 15. *Suppose that, under linear contract w_α and technology $\mathcal{A}_\kappa$ (for some given valid cost function κ), the supervisor is willing to induce an action with mean cost μ. Then $\tilde{\kappa}$ given by (14) is also a valid cost function.*

Since the same mean output μ can be induced under the new cost function $\tilde{\kappa}$, we have proved the following proposition.

Proposition 16. *When evaluating the payoff guarantee V_P^{PSA1} for a linear contract w_α, we can restrict attention to technologies $\mathcal{A}_{\tilde{\kappa}}$, where $\tilde{\kappa}$ is a valid cost function of the form (14), for some $\mu \in [0, \bar{y}]$.*

To emphasize dependence on the point μ and contract slope α, we will henceforth write the function in (14) as $\kappa(\tilde{\mu}, \mu, \alpha)$. This function behaves well as we vary the parameters, as summarized in the following proposition, proven in Appendix A.

Proposition 17. *Define $\kappa(\tilde{\mu}, \mu, \alpha)$ as in (14). Then:*

(a) κ is nonincreasing in μ;

(b) κ is nondecreasing in α.

We can now characterize the worst-case technology for the principal, given a linear contract w_α. Since the contract is linear, the principal's payoff depends on the output distribution only through its mean. Consequently, our question is the same as asking what is the smallest mean output that can be induced over the class of valid cost functions (14), as μ varies.

Theorem 18. *For each linear contract between principal and supervisor, $w_\alpha(y) = \alpha y$, there exists a cost function of the form (14) that achieves the infimum in the principal's payoff guarantee. That is, there exists a $\mu \in [0, \bar{y}]$ such that $\kappa(\cdot, \mu, \alpha)$ is a valid cost function and*

$$V_P^{PSA1}(w_\alpha) = (1 - \alpha)\mu.$$

Proof. It suffices to show that the set

$$\chi(\alpha) = \left\{\mu : \kappa(\mathbb{E}_F[y], \mu, \alpha) \leq c \; \forall (F, c) \in \mathcal{A}^0\right\}$$

is compact and nonempty. Indeed, our work so far shows that mean output μ will be produced for some technology if and only if $\kappa(\cdot, \mu, \alpha)$ is a valid cost function, i.e. if and only if $\mu \in \chi(\alpha)$, and so we just need to show that $\chi(\alpha)$ has a minimum element.

Nonemptiness of $\chi(\alpha)$ follows from the fact that $\bar{y} \in \chi(\alpha)$, as the corresponding κ is identically zero. Since $\chi(\alpha)$ is contained in $[0, \bar{y}]$, compactness is equivalent to it being closed. For this, note that for each individual $(F, c) \in \mathcal{A}^0$, the set $\{\mu : \kappa(\mathbb{E}_F[y], \mu, \alpha) \leq c\}$ is closed, since $\kappa(\tilde{\mu}, \mu, \alpha)$ is continuous in μ. Then $\chi(\alpha)$ is an intersection of closed sets and so is closed. □

Define $\chi(\alpha)$ as in the proof of Theorem 18. Define

$$\mu^*(\alpha) = \min \chi(\alpha)$$

for each $\alpha \in [0,1]$. Then the worst-case technology for the principal, given contract $w_\alpha(y) = \alpha y$, is given by the valid cost function $\kappa(\cdot, \mu^*(\alpha), \alpha)$, under which the supervisor offers the agent the zero contract and the agent chooses action $(\delta_{\mu^*(\alpha)}, 0)$ (or any other zero cost action with mean output $\mu^*(\alpha)$). Thus we arrive at the expression for principal payoff guarantee $V_P^{PSA1}(w_\alpha) = \mu^*(\alpha)(1-\alpha)$. As we know from Proposition 5, there exists a solution $\alpha^* \in [0,1]$ that maximizes this objective.

Finally, let us give an alternative, slightly more explicit characterization of the guarantee from a given linear contract, and of the optimal contract.

For a given share $\alpha \in [0,1]$, write $\mu_\alpha(\tilde{\mu}, c)$ for the smallest value μ such that $\kappa(\tilde{\mu}, \mu, \alpha) \leq c$. Note that Proposition 17(a), together with $\kappa(\tilde{\mu}, \mu, \alpha) = 0$ for $\mu \geq \tilde{\mu}$ and $\to \alpha\tilde{\mu}$ as $\mu \to 0$, imply that this value is the unique solution of $\kappa(\tilde{\mu}, \mu, \alpha) = c$ if $0 < c < \alpha\tilde{\mu}$; if $c = 0$ then it is $\tilde{\mu}$, and if $c \geq \alpha\tilde{\mu}$ then it is 0.

We claim that $\mu^*(\alpha) = \max_{(F,c)\in\mathcal{A}^0} \mu_\alpha(\mathbb{E}_F[y], c)$. Indeed: this follows from

$$\begin{aligned} \mu \geq \mu^*(\alpha) &\Leftrightarrow \kappa(\mathbb{E}_F[y], \mu, \alpha) \leq c \text{ for all } (F,c) \in \mathcal{A}^0 \\ &\Leftrightarrow \mu \geq \mu_\alpha(\mathbb{E}_F[y], c) \text{ for all } (F,c) \in \mathcal{A}^0. \end{aligned}$$

The principal's guarantee from the optimal contract is then equal to

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

where

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

We thus have a moderately explicit description of the optimal contract, and the principal's guarantee: it can be determined by identifying the action $(F,c) \in \mathcal{A}^0$ that maximizes the function g; then the guarantee is simply the value of g, and the share is given by the α that attains the max. Unfortunately, these objects are defined implicitly and we cannot give a closed-form solution.

Let us consider the simplest example, with $\mathcal{A}^0 = \{(0,0), (\delta_{y^*}, c^*)\}$ for some particular values $y^*, c^* > 0$ with $c^* < y^*$. We can implicitly characterize the optimal guarantee by calculating $g(\mathbb{E}_F[y], c)$ for each of the two actions in $\mathcal{A}^0$. For $(0,0)$ it is simply zero. For (δ_{y^*}, c^*), the first-order condition for the maximization over μ is

$$c^*(y^* - \mu) = ((y^* - \mu) - \mu \log(y^*/\mu))^2. \tag{15}$$

Note that if we take square roots of both sides of (15) then the left side becomes a concave function of μ, the right side a convex function, and they are equal at $\mu = y^*$ (where both sides are zero), so there is at most one other point in the interval $(0, y^*)$ where this first-order condition is satisfied. Note also that such an interior maximum must indeed exist, since otherwise the maximum possible guarantee would be zero, and this is not the case (a linear contract with a slope sufficiently close to 1 gives a positive guarantee). Finally, with this value of μ identified, the share of the corresponding contract is $\alpha(\mu) = c^*/((y^* - \mu) - \mu \log(y^*/\mu))$, and its guarantee is $(1 - \alpha(\mu))\mu$.

6.2 Hierarchical Model (ii)

Now we characterize the worst-case technology for hierarchical model (ii). Note that in model (i) we had to show that the supervisor optimally offered a linear contract to the agent. For hierarchical model (ii), we automatically have this result, since the relationship between the supervisor and agent in hierarchical model (ii) is isomorphic to the robust principal-agent model.

More generally, for any given linear contract w_α (with slope $\alpha \in (0,1]$) that the principal may offer, we can exploit the analysis of the robust principal-agent model in Carroll (2015) to characterize optimal behavior for the supervisor under each $\mathcal{A}^1$, and the possible responses by the agent. Our task is further simplified by Lemma 6, telling us that F is a possible response by the agent if and only if it can occur under a technology $\mathcal{A}^1$ that contains $(F, 0)$ and incentivizes the supervisor to offer the zero contract. So we just need to identify the distributions F for which some such $\mathcal{A}^1$ exists.

This analysis leads to the following lemma, proven in Appendix A. Given any technology $\mathcal{A}$, let $\mathcal{A}|_\alpha$ denote the subset $\{(F,c) \in \mathcal{A} : \alpha \mathbb{E}_F[y] \geq c\}$.

Lemma 19. *Suppose the principal offers a linear contract w_α, with $\alpha > 0$. Then*

$$\Phi^{PSA2}(w_\alpha) = \{F \in \Delta(Y) : \alpha \mathbb{E}_F[y] \geq \left(\sqrt{\alpha \mathbb{E}_{F'}[y]} - \sqrt{c'}\right)^2 \text{ for all } (F', c') \in \mathcal{A}^0|_\alpha\}.$$

This identifies the minimum expected output that the agent could potentially produce when the principal offers contract w_α: namely, $\mu^*(\alpha) = \max_{(F,c)\in\mathcal{A}^0|_\alpha} \mu_\alpha(\mathbb{E}_F[y], c)$, where we define

$$\mu_\alpha(\tilde{\mu}, c) = \frac{1}{\alpha}\left(\sqrt{\alpha\tilde{\mu}} - \sqrt{c}\right)^2$$

(and define $\mu^*(\alpha) = 0$ if $\mathcal{A}^0|_\alpha$ is empty).

The case $\alpha = 0$ requires separate treatment, but the same formula applies: If the principal offers the zero contract, then it is uniquely optimal for the supervisor to offer the agent the zero contract as well. Then for each $\mathcal{A}$, $\Gamma_A^S(0, 0, \mathcal{A}) = \Gamma_A(0, \mathcal{A}) = \{F \in \Delta(Y) : (F, 0) \in \mathcal{A}\}$, so principal-preferred tiebreaking selects a distribution with the largest mean among those with cost 0. This implies that $\Phi^{PSA2}(w_0) = \{F \in \Delta(Y) : \mathbb{E}_F[y] \geq \mathbb{E}_{F'}[y] \text{ for all } (F', 0) \in \mathcal{A}^0|_0\}$. We can naturally define $\mu_0(\tilde{\mu}, c) = \tilde{\mu}$ if $c = 0$ and $-\infty$ if $c > 0$, and use this to define $\mu^*(\alpha)$ for $\alpha = 0$ by the same formula as for $\alpha > 0$, and we see that the minimum expected output over $\Phi^{PSA2}(w_0)$ is $\mu^*(0)$.

Now, we proceed as in hierarchical model (i) to explicitly write down the guarantee from the optimal contract: it is equal to

$$\begin{aligned}\max_{\alpha\in[0,1]} \mu^*(\alpha)(1-\alpha) &= \max_{\alpha\in[0,1]} \max_{(F,c)\in\mathcal{A}^0|_\alpha} (1-\alpha)\mu_\alpha(\mathbb{E}_F[y], c) \\ &= \max_{(F,c)\in\mathcal{A}^0|_1} g(\mathbb{E}_F[y], c)\end{aligned}$$

where

$$g(\tilde{\mu}, c) = \max_{\alpha\in\left[\frac{c}{\tilde{\mu}}, 1\right]} (1-\alpha)\mu_\alpha(\tilde{\mu}, c) = \begin{cases} \tilde{\mu} & \text{if } c = 0, \\ \max_{\mu\in[0,(\sqrt{\tilde{\mu}}-\sqrt{c})^2]} \left(1 - \frac{c}{(\sqrt{\tilde{\mu}}-\sqrt{\mu})^2}\right)\mu & \text{if } c > 0. \end{cases}$$

We can actually continue to make this more explicit. For the case of $c > 0$, we can solve the maximization over μ by taking a first-order condition with respect to μ, and after some algebra we obtain a cubic equation in μ with one real solution,

$$\mu = c^{2/3}\tilde{\mu}^{1/3} + \tilde{\mu} - 2c^{1/3}\tilde{\mu}^{2/3} = \tilde{\mu}^{1/3}(\tilde{\mu}^{1/3} - c^{1/3})^2.$$

One may verify by a second order condition that the function is strictly concave, and this is the unique maximum. Inserting this formula into g, and simplifying, we obtain

$$g(\tilde{\mu}, c) = (\tilde{\mu}^{1/3} - c^{1/3})^3.$$

This was obtained assuming $c > 0$, but note that in fact it holds in the $c = 0$ case also. Thus we can explicitly compare all actions $(F, c) \in \mathcal{A}^0|_1$, and the principal's guarantee can be identified by finding the action $(F, c) \in \mathcal{A}^0|_1$ that maximizes the function g.

At this point, we comment on a comparison of the worst-case scenarios between models. We mentioned that just adding a single point of the form $(F, 0)$ to $\mathcal{A}^0$ is all that is needed to obtain the worst-case technology in hierarchical model (ii); under this new technology, the supervisor finds it optimal to offer the zero contract to the agent, and the agent in the worst case will choose action $(F, 0)$. (In the principal-agent model, the worst case also involved adding a single point.) For hierarchical model (i), however, we needed to add an entire continuum of actions parameterized by the cost function κ to ensure that the worst-case $(F, 0)$ would be induced. This difference arises because in model (i), the supervisor knows the technology available to the agent, and if only the one action $(F, 0)$ were added, she could still induce higher-output actions cheaply. The intermediate actions need to be included in order to make it more costly for the supervisor to induce high-output actions, by providing tempting deviations for the agent that the supervisor needs to deter. In model (ii), the supervisor does not know whether these kinds of intermediate actions are available, which foils any attempt on her part to induce a high mean action at a relatively low cost. Hence the zero-cost action is sustained without the need to explicitly include intermediate actions. Finally, from the above analysis we note that the solutions to both hierarchical models are quite different from the robust principal-agent model, thus it seems very unlikely that one could give an alternate proof of linearity in these models by somehow reducing them to the robust principal-agent model.

7 Model Comparisons

As a further application of the hierarchical models considered in Section 4, we investigate the extent to which one can compare outcomes across the robust principal-agent model, hierarchical model (i), and hierarchical model (ii). Does the principal's optimal contract in the hierarchical models produce a better or worse payoff guarantee than the optimal contract in the robust principal-agent model? In the hierarchical models, is it better for the principal if the supervisor has full or partial information about the agent's technology? In the hierarchical models, the supervisor does not produce anything and takes some portion of the payoff, leading one to believe that the principal would be better off directly contracting with the agent. On the other hand, the supervisor has better information than the principal about the technology accessible to the agent, so perhaps by delegat-

ing contract-writing to the supervisor, the supervisor can write a cheaper contract that incentivizes the agent to produce more, benefiting both the principal and the supervisor. So the comparison of the models is not so obvious.

In the traditional setting, where all parties know the true technology, there is a simple proof that the principal does better without the supervisor: For any action (F, c), the expected amount that she has to pay the supervisor to induce that action is at least as high as she would have to pay the agent directly to incentivize the action, since the principal has to at least cover the supervisor's cost of incentivizing the action. Since every action becomes more expensive with the supervisor present, the principal's payoff can only become lower. For the robust version of the model, one might try to adapt this argument as follows: consider the worst-case technology in the principal-agent model; then apply the argument above to show that the principal gets an even lower payoff in the principal-supervisor-agent model (under this same technology) than the principal-agent model. However, this adaptation does not work, because typically there is no one "worst-case technology" in the principal-agent model, without reference to a particular contract. More precisely, if we denote by $\bar{u}_P$ the value of the principal's guarantee in the optimal robust contract, there does not exist any single technology $\mathcal{A}$ that prevents the principal from achieving a payoff higher than $\bar{u}_P$; i.e. the principal's maxmin problem does not have a saddle point. (This is Proposition 1 in Carroll (2015), section II.D.)

Nonetheless, we can make a clean comparison between organizational structures. The main result of this section is that the payoff guarantee to the principal can be weakly ordered from highest to lowest as follows: first the robust principal-agent model, then hierarchical model (i), then hierarchical model (ii). In fact, the comparison across models holds for any fixed contract: we are able to show that the set of possible outcomes from the outcome correspondence grows as we move from model to model, which immediately implies that the worst-case outcome becomes weakly worse. (To compare hierarchical models (i) and (ii), we require an additional technical assumption, because model (i) had the added restriction that w_A had to lie in the exogenous set $\mathcal{S}$. The technical assumption is not binding for linear w and hence for optimal w.)

Theorem 20. *Given $\mathcal{A}^0$ and $w \in C^+(Y)$, we have $\Phi^{PA}(w) \subseteq \Phi^{PSA1}(w)$, and if $\mathcal{S}$ contains all contracts $w_A = \beta w$ with $\beta \in [0, 1]$, then $\Phi^{PSA1}(w) \subseteq \Phi^{PSA2}(w)$. These facts imply that*

$$\max_{w \in C^+(Y)} V_P^{PA}(w) \geq \max_{w \in C^+(Y)} V_P^{PSA1}(w) \geq \max_{w \in C^+(Y)} V_P^{PSA2}(w).$$

The proof is in Appendix A.

For some technologies $\mathcal{A}^0$, the optimal robust guarantee is the same in all three models, so the bounds in Theorem 20 are tight. For instance, this happens under any technology $\mathcal{A}^0$ in which the highest-mean-output action actually has cost 0. To obtain more precise comparisons across models for specific $\mathcal{A}^0$, we must apply Theorem 1 and take advantage of the analysis in Section 6 to solve for optimal robust guarantees.

It is possible to go beyond Theorem 20 and make more detailed comparisons. For example, one can show that the difference in the principal's optimal guarantee between hierarchical models (i) and (ii) is no larger than the difference between the principal-agent model and hierarchical model (i). This can be interpreted roughly as saying that being able or unable to contract directly with the agent is more important than the particular kind of supervisor who intervenes. The statement can be proven using the worst-case analyses from Section 6; we omit the details.

8 Conclusion

The idea that linear contracts provide robustness in situations of great uncertainty, by aligning the parties' interests without being sensitive to details of the environment, has intuitive appeal. Yet the argument is not automatic, and its formal validity depends on the detailed specification of the model, as the contrast between hierarchical models (i), (ii) and (iii) shows. This observation motivated us to identify a broad class of models of contracting with uncertainty in which linear contracts can be microfounded. We took a black-box modeling approach that avoids explicit description of the organizational environment. We identified two properties of the contracting environment, Richness and Responsiveness, that together are sufficient to ensure that linear contracts are indeed optimally robust. The first of these properties expresses the requirement of sufficient uncertainty about the possible outcomes; the second requires that when contracts change, the possible responses vary as if maximizing expected payment, so that the idea of linear contracts "aligning interests" applies. These sufficient conditions can cover a wide variety of organizational structures. Moreover, as a detailed worst-case analysis of some of these structures confirms, even though diverse models lead to the same form for optimal contracts, they are not equivalent to each other.

The contribution of our analysis serves two goals. On one hand, the argument for robustness as a way of understanding why linear contracts are so prevalent in the world is bolstered to the extent that this argument holds up across many models. After all,

contracting often does not place in simple bilateral relationships, but is embedded in a variety of more complex environments, and an effective theory should be able to accommodate this variety. On the other hand, our approach also provides a modeling tool, by suggesting a way to write down tractable models of more complex organizations, that can then be used to study more applied questions. More broadly, both these points suggest that the question of how the form of optimal contracts does or does not vary with the organizational environment — so far a relatively neglected area of contract theory — may deserve more careful study.

A Additional proofs

Here are proofs omitted from the main paper.

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 hemi-continuity 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 $V_P(w_{\alpha_k}) \to \overline{V}_P$. □

Proof of Lemma 4. Sufficiency is immediate. For necessity, suppose there exists a technology $\mathcal{A}$, cost c and S-A contract w_A satisfying criteria (a)–(d). Create a new technology $\mathcal{A}' = \mathcal{A} \cup \{(F, 0)\}$ and put $w'_A = 0$. This allows the supervisor to induce F at cost 0 to herself, which is clearly cheaper than any other way to induce F, and is also weakly more profitable to her than inducing any other action in $\mathcal{A}$ (since inducing (F, c) via w_A was optimal under $\mathcal{A}$). Thus, (a)–(c) are satisfied. It remains to check (d). Note that if another action $(F', c') \in \mathcal{A}'$ also passes (a)–(c) under w'_A, we must have $c' = 0$, and $\mathbb{E}_{F'}[w(y)] = \mathbb{E}_F[w(y)]$. So, for inducing F to have been optimal for the supervisor under $\mathcal{A}$, it must have already been available at cost 0, i.e. $\mathcal{A} = \mathcal{A}'$, and the contract w_A must have paid 0 for F, therefore also for F' (otherwise (b) would have been violated). Thus both F and F' survived (b)–(c) under w_A. Since F further survived (d) under w_A, we conclude $\mathbb{E}_{F'}[y] \leq \mathbb{E}_F[y]$, hence F survives (d) under w'_A as needed. □

Proof of Proposition 5. (Richness) Suppose $w \in C^+(Y)$, $F \in \Phi^{PSA1}(w)$, $F' \in \Delta(Y)$ such that $\mathbb{E}_F[y] = \mathbb{E}_{F'}[y]$, $\mathbb{E}_F[w(y)] \leq \mathbb{E}_{F'}[w(y)]$. Let $\mathcal{A}$ be a technology for which F is chosen. By Lemma 4, we may assume that $(F, 0) \in \mathcal{A}$ and the supervisor induces F using $w_A \equiv 0$.

Create a new technology $\mathcal{A}' = \mathcal{A} \cup \{(F', 0)\}$. In the new technology, the supervisor can also induce F' using the zero contract, and this is at least as good for her as inducing F was under $\mathcal{A}$, so (a) is satisfied with $\mathcal{A}'$ and $w'_A \equiv 0$. The agent is willing to take any zero-cost action, so (b) is satisfied for $(F', 0)$. The preceding observation also implies that (c) is satisfied. Finally, if (d) is violated, there is some other $(F'', c'') \in \mathcal{A}$ that also satisfies (a)–(c) with w'_A and is strictly better for the principal; but this means $c'' = 0$, and then

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

Here the first inequality is by the principal's strict preference; the second is because $\mathbb{E}_{F'}[y] = \mathbb{E}_F[y]$ by assumption but the supervisor was willing to induce F''; the third is because the supervisor was willing to induce $(F, 0)$ rather than $(F'', 0)$ in $\mathcal{A}$. We conclude that the principal strictly prefers F'' over F under w, which means that F would have also violated (d) under $\mathcal{A}$ and $w_A \equiv 0$, contrary to assumption.

Hence $F' \in \Gamma_A^{PS}(w, w'_A, \mathcal{A}') \subseteq \Gamma_S(w, \mathcal{A}')$, so $F' \in \Phi^{PSA1}(w)$, so Richness holds.

(Responsiveness) Let $w, w' \in C^+(Y)$, $F \notin \Phi^{PSA1}(w)$ satisfy the hypotheses of Responsiveness. Suppose for contradiction that $F \in \Phi^{PSA1}(w')$. By Lemma 4, there exists some technology $\mathcal{A}$ that contains $(F, 0)$ such that $(\mathcal{A}, (F, 0), w'_A \equiv 0)$ satisfy (a)–(d) under w'. Also, let $((\tilde{F}, \tilde{c}), \tilde{w}_A)$ satisfy (a)–(d) under $\mathcal{A}$ and w, so that $\tilde{F} \in \Phi^{PSA1}(w)$.

Under w', the supervisor can still induce $(\tilde{F}, \tilde{c})$ via contract $\tilde{w}_A$, thereby getting a payoff of

$$\begin{aligned}\mathbb{E}_{\tilde{F}}[w'(y) - \tilde{w}_A(y)] &\geq \mathbb{E}_{\tilde{F}}[w(y) - \tilde{w}_A(y)] + \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)] \\ &\geq \mathbb{E}_F[w(y)] + \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)] \\ &= \mathbb{E}_F[w'(y)] \\ &\geq \mathbb{E}_{\tilde{F}}[w'(y) - \tilde{w}_A(y)]\end{aligned}$$

where the first inequality is by the hypothesis of Responsiveness (and the fact that $\tilde{F} \in \Phi^{PSA1}(w)$); the second is by assumption that under w, the supervisor weakly prefers to induce $\tilde{F}$ using $\tilde{w}_A$ than F using the zero contract; and the third is by the assumption that under w', the supervisor weakly prefers to induce F using the zero contract rather than $\tilde{F}$ using $\tilde{w}_A$. So all the inequalities in this cycle must be equalities. This means that it is also optimal for the supervisor to induce $(F, 0)$ via $w'_A \equiv 0$ under w. That is, $(\mathcal{A}, (F, 0), w'_A)$ satisfy (a)–(c) under w.

Since $F \notin \Phi^{PSA1}(w)$, it must be (d) that is violated: some other $(F'', c'') \in \mathcal{A}$, which gives the agent and supervisor the same payoffs as $(F, 0)$ (under w, w'_A), is strictly better for the principal, and so $F'' \in \Gamma_A^{PS}(w, w'_A, \mathcal{A})$. But this means $c'' = 0$ and $\mathbb{E}_{F''}[w(y)] = \mathbb{E}_F[w(y)]$. Therefore, the hypothesis of Responsiveness implies $\mathbb{E}_{F''}[w'(y)] \geq \mathbb{E}_F[w'(y)]$. This must be an equality, otherwise $(F, 0)$ would not survive supervisor tiebreaking under w'.

However, the principal's strict preference for F'' under w then implies $\mathbb{E}_{F''}[y] > \mathbb{E}_F[y]$, which means principal tie-breaking under $\mathcal{A}$, w', w'_A also strictly favors F'' over F. This contradicts condition (d) for $(\mathcal{A}, (F, 0), w'_A)$ under w'. Responsiveness is proven.

So Theorem 1 applies, and we can restrict to linear contracts when maximizing V_P^{PSA1}. It remains to show that $\widetilde{\Phi}^{PSA1}$ is lower hemicontinuous, to ensure that the optimum is attained.

(Lower Hemicontinuity) Let $\epsilon > 0$, $\alpha \in [0, 1]$, $F \in \widetilde{\Phi}^{PSA1}(\alpha)$. Let $\mathcal{A} \supseteq \mathcal{A}^0$ be a technology such that $F \in \Gamma_S(w_\alpha, \mathcal{A})$, and we can assume that $(F, 0) \in \mathcal{A}$ and $w_A(y) \equiv 0$ is the contract offered by the supervisor in this setting, and since F is principal-preferred, F has the highest mean among zero-cost actions in $\mathcal{A}$. If F has mean $\bar{y}$, then $F = \delta_{\bar{y}}$, then for any α, the supervisor cannot earn a higher amount than $\alpha\bar{y}$, so $F \in \widetilde{\Phi}^{PSA1}(\alpha)$ for all α. So we can assume $\mathbb{E}_F[y] < \bar{y}$.

Assume $\alpha \in [0, 1]$, and choose β and F' as we did in the lower hemicontinuity argument in the robust P-A model, so that $F' \in \mathcal{B}_\epsilon(F)$. Note that $|V_S^i(w_A | w_{\alpha'}, \mathcal{A}) -$

$V_S^i(w_A|w_{\alpha''}, \mathcal{A})| \leq |\alpha' - \alpha''| \cdot \bar{y}$ for any α', α'' and w_A; consequently, $f^*(\alpha') = \max_{w_A \in \mathcal{S}} V_S^i(w_A|w_{\alpha'}, \mathcal{A})$ must be a continuous function, since when α' is moved by a small amount η, the max cannot fall by more than $\eta \cdot \bar{y}$.

Now, we can find $\eta > 0$ such that $\alpha' \in \mathcal{B}_\eta(\alpha) \setminus \{0\} \implies f^*(\alpha') < \alpha' \mathbb{E}_{F'}[y]$, via the same justifications as in the robust P-A argument. Hence constructing $\mathcal{A}' = \mathcal{A} \cup \{(F', 0)\}$ yields $\Gamma_S(w_{\alpha'}, \mathcal{A}') = \{F'\}$, hence $F' \in \widetilde{\Phi}^{PSA1}(\alpha') \cap \mathcal{B}_\epsilon(F)$. □

Proof of Lemma 6. Sufficiency is immediate. For necessity, suppose for F there exists technologies $\mathcal{A} \supseteq \mathcal{A}^1$, cost c and S-A contract w_A satisfying criteria (a)–(d). Create new technologies $\mathcal{A}' = \mathcal{A}^{1'} = \mathcal{A}^1 \cup \{(F, 0)\}$ and put $w'_A \equiv 0$. This ensures that the supervisor obtains at worst $V_S^u(w'_A|w, \mathcal{A}^{1'}) \geq \mathbb{E}_F[w(y)]$, which is at least as good as the worst case from any other contract (since the latter worst case either could also have happened under some superset of $\mathcal{A}^1$ or uses action $(F, 0)$); therefore (a) is satisfied. The agent is inclined to take any zero-cost action, satisfying (b). Inducing F at cost 0 is at least as good for the supervisor as any other zero cost action in $\mathcal{A}'$, since otherwise the supervisor could have been assured strictly better under $\mathcal{A}^1$ using $w_A = 0$, and (c) would have been violated under $\mathcal{A}^1$, $\mathcal{A}$ and w_A. Therefore (c) holds. To check (d), note that if a different action $(F', c') \in \mathcal{A}'$ passes (a)–(c) under w'_A, it must be that $c' = 0$, and $\mathbb{E}_{F'}[w(y)] = \mathbb{E}_F[w(y)]$. However, then $(F', 0) \in \mathcal{A}^1$, so under $\mathcal{A}^1$ the supervisor could already induce F' with the zero contract. Then, for F to have been induced under $\mathcal{A}$ and w_A, it must be that $(F, 0) \in \mathcal{A}$ and w_A paid 0 for F, therefore also for F' (otherwise (b) would have been violated). Thus both F and F' survived (b)–(c) under w_A and technologies $\mathcal{A}^1, \mathcal{A}$. Since F further survived (d) under w_A, we conclude that $\mathbb{E}_{F'}[y] \leq \mathbb{E}_F[y]$, so F survives (d) under $(w'_A, \mathcal{A}^{1'}, \mathcal{A}')$ as needed. □

Proof of Proposition 7. (Richness) Suppose $w \in C^+(Y)$, $F \in \Phi^{PSA2}(w)$, $F' \in \Delta(Y)$ such that $\mathbb{E}_F[y] = \mathbb{E}_{F'}[y]$ and $\mathbb{E}_F[w(y)] \leq \mathbb{E}_{F'}[w(y)]$. Let $\mathcal{A}$, $\mathcal{A}^1$ be the technologies associated to F. By Lemma 6, we may assume that $(F, 0) \in \mathcal{A}^1 = \mathcal{A}$ and the supervisor induces F using $w_A \equiv 0$.

Create new technologies $\mathcal{A}' = \mathcal{A}^{1'}$ by taking union with $\{(F', 0)\}$. Set $w'_A \equiv 0$. In the new technology, the supervisor can obtain $V_S^u(w'_A|w, \mathcal{A}^{1'}) \geq \mathbb{E}_{F'}[w(y)]$, and no other contract $\tilde{w}_A$ can guarantee better (just consider the worst-case technology $\widetilde{\mathcal{A}} \supseteq \mathcal{A}^1$; then $\widetilde{\mathcal{A}} \cup \{(F', 0)\}$ is a possible technology containing $\mathcal{A}^{1'}$, and the agent either takes the same action as under $\tilde{w}_A, \widetilde{\mathcal{A}}$ or takes action $(F', 0)$). So, (a) is satisfied with $\mathcal{A}^{1'}$ and $w'_A \equiv 0$. The agent is willing to take any zero-cost action, so (b) is satisfied for $(\mathcal{A}', (F', 0), w'_A)$. Since $\mathbb{E}_F[w(y)] \leq \mathbb{E}_{F'}[w(y)]$ and (c) held for $(F, 0)$, (c) is satisfied with $(\mathcal{A}', (F', 0), w'_A)$.

Finally, if (d) is violated, there is some other $(F'', c'') \in \mathcal{A}$ that also satisfies (a)–(c) with w'_A and is strictly better for the principal; but this means $c'' = 0$, and then

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

Here, the first inequality is by the principal's strict preference; the second is because $\mathbb{E}_{F'}[y] = \mathbb{E}_F[y]$ by assumption but supervisor-tiebreaking is sastisfied for F'' under $\mathcal{A}'$; the third is because supervisor-tiebreaking held for F, and $(F'', 0)$ was in $\mathcal{A}$. We conclude that the principal strictly prefers F'' over F under w, which means F would also have violated (d) under $\mathcal{A}$ and $w_A \equiv 0$, contrary to assumption.

Hence $F' \in \Gamma_A^{PS}(w, w'_A, \mathcal{A}') \subseteq \Gamma_S(w, \mathcal{A}^{1\prime}, \mathcal{A}')$.

(Responsiveness) Let $w, w' \in C^+(Y)$, and $F \notin \Phi^{PSA2}(w)$ satisfy the hypotheses of Responsiveness. Suppose for contradiction that $F \in \Phi^{PSA2}(w')$. By Lemma 6, there exist technologies $\mathcal{A}^1 = \mathcal{A}$ both containing $(F, 0)$ such that $(\mathcal{A}^1, \mathcal{A}, (F, 0), w'_A \equiv 0)$ satisfy (a)–(d) under w'. Also, let S-A contract $\tilde{w}_A$ (along with some action) satisfy (a)–(d) under $\mathcal{A}^1$, $\mathcal{A}$, and w.

Consider any possible technology $\widehat{\mathcal{A}} \supseteq \mathcal{A}^1$, and any resulting actions $\widehat{F} \in \Gamma_A^{PS}(w, \tilde{w}_A, \widehat{\mathcal{A}})$ and $\widehat{F}' \in \Gamma_A^{PS}(w', \tilde{w}_A, \widehat{\mathcal{A}})$. We have

$$\mathbb{E}_{\widehat{F}'}[w'(y) - \tilde{w}_A(y)] \geq \mathbb{E}_{\widehat{F}}[w'(y) - \tilde{w}_A(y)] \geq \mathbb{E}_{\widehat{F}}[w(y) - \tilde{w}_A(y)] + \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)],$$

where the first inequality occurs because the agent is indifferent between $\widehat{F}$ and $\widehat{F}'$ but breaks ties to favor the supervisor, and the second inequality comes from the hypothesis of Responsiveness since $\widehat{F} \in \Phi^{PSA2}(w)$. Taking infimum over technologies $\widehat{\mathcal{A}}$ gives

$$V_S^u(\tilde{w}_A | w', \mathcal{A}^1) \geq V_S^u(\tilde{w}_A | w, \mathcal{A}^1) + \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)].$$

Proceeding,

$$\begin{aligned}
V_S^u(\tilde{w}_A | w', \mathcal{A}^1) &\geq V_S^u(\tilde{w}_A | w, \mathcal{A}^1) + \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)] \\
&\geq V_S^u(w'_A | w, \mathcal{A}^1) + \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)] \\
&\geq \mathbb{E}_F[w(y)] + \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)] \\
&= \mathbb{E}_F[w'(y)] \\
&\geq V_S^u(w'_A | w', \mathcal{A}^1) \\
&\geq V_S^u(\tilde{w}_A | w', \mathcal{A}^1).
\end{aligned}$$

The first inequality was argued above; the second is by assumption that under w and $\mathcal{A}^1$, $\tilde{w}_A$ satisfies (a); the third is by the fact that the agent is willing to take action $(F, 0)$ when given w'_A; the fourth and fifth are because w'_A, F satisfy (a)–(c) with $\mathcal{A}$ under w'. So all the inequalities must be equalities. This means that it is also optimal for the supervisor to give contract $w'_A \equiv 0$ under w, so (a) holds with $(\mathcal{A}^1, w'_A)$ under w. Under w'_A, the agent is willing to take any zero-cost action, so (b) holds with $(\mathcal{A}, (F, 0), w'_A)$ under w. And the equalities also show that $V_S^u(w'_A|w, \mathcal{A}^1) = \mathbb{E}_F[w(y)]$, which means that (c) holds for $(\mathcal{A}, (F, 0), w'_A)$ under w, since $\mathcal{A} = \mathcal{A}^1$ achieves the infimum in $V_S^u(w'_A|w, \mathcal{A}^1)$.

Since $F \notin \Phi^{PSA2}(w)$, it must then be that (d) is violated; this implies that condition (d) is violated for $(\mathcal{A}, (F, 0), w'_A)$ under w', contradicting $F \in \Phi^{PSA2}(w')$. The argument is identical to the one given in the proof of Responsiveness in Proposition 5.

So Theorem 1 applies, and we can restrict to linear contracts when maximizing V_P^{PSA2}. It remains to show that $\widetilde{\Phi}^{PSA2}$ is lower hemicontinuous, to ensure that the optimum is attained.

(Lower Hemicontinuity) Let $\epsilon > 0$, $\alpha \in [0, 1]$, $F \in \widetilde{\Phi}^{PSA2}(\alpha)$. Let $\mathcal{A} \supseteq \mathcal{A}^1 \supseteq \mathcal{A}^0$ be technologies such that $(F, 0) \in \mathcal{A}^1$, $F \in \Gamma_A^{PS}(w_\alpha, 0, \mathcal{A})$, and $w_A(y) = 0$ is optimal for the supervisor in this setting. As in hierarchical model (i), if F has mean $\bar{y}$, then the supervisor can do no better than earning $w_\alpha(\bar{y})$, so for any neighborhood of α, the supervisor is at the very least indifferent between inducing $(F, 0)$ and any other action, so $F \in \Gamma_S(w_{\alpha'}, \mathcal{A}^1, \mathcal{A})$ for any α' in this neighborhood. So we can assume $\mathbb{E}_F[y] < \bar{y}$.

For any α', define $f^*(\alpha') = \max_{w_A \in C^+(Y)} V_S^u(w_A|w_{\alpha'}, \mathcal{A}^1)$, and let $w_A^*(\alpha') \in \arg\max_{w_A \in C^+(Y)} V_S^u(w_A|w_{\alpha'}, \mathcal{A}^1)$. Since $V_S^u(w_A|w_{\alpha'}, \mathcal{A}^1) \leq V_S^i(w_A|w_{\alpha'}, \mathcal{A}^1)$ for all w_A and $w_{\alpha'}$, $f^*(\alpha') \leq V_S^i(w_A^*(\alpha')|w_{\alpha'}, \mathcal{A}^1)$. We also know (as in the proof of Proposition 5) that there is $\eta > 0$ such that, when $\alpha' \in \mathcal{B}_\eta(\alpha) \setminus \{0\}$, $V_S^i(w_A^*(\alpha')|w_{\alpha'}, \mathcal{A}^1) < \alpha' \mathbb{E}_{F'}[y]$. Combining these steps, then, whenever α' is in this neighborhood of α (and is not zero), $f^*(\alpha') < \alpha' \mathbb{E}_{F'}[y]$, where F' was the distribution constructed in the proof of lower-hemicontinuity in Proposition 3. Hence constructing $\mathcal{A}' = \mathcal{A} \cup \{(F', 0)\}$ and $\mathcal{A}^{1\prime} = \mathcal{A}^1 \cup \{(F', 0)\}$ yields $\Gamma_S(w_{\alpha'}, \mathcal{A}^{1\prime}, \mathcal{A}') = \{F'\}$, hence $F' \in \widetilde{\Phi}^{PSA2}(\alpha') \cap \mathcal{B}_\epsilon(F)$.

□

Proof of Proposition 8. (Richness) Let $w \in C^+(Y)$, $F \in \Phi^{ST}(w)$, and $F' \in \Delta(Y)$ such that $\mathbb{E}_F[y] = \mathbb{E}_{F'}[y]$, $\mathbb{E}_F[w(y)] \leq \mathbb{E}_{F'}[w(y)]$. Then there exist technologies $\mathcal{A}_1 \supseteq \mathcal{A}_1^0$ and $\mathcal{A}_2 \supseteq \mathcal{A}_2^0$, and contracts w_{A1}, w_{A2} that maximize $V_S(\cdot, \cdot|w, \mathcal{A}_1, \mathcal{A}_2)$, such that F is induced in the supervisor-optimal Nash equilibrium. Consider a new technology for agent 2 defined as $\mathcal{A}_2' = \mathcal{A}_2 \cup \{(K', 0)\}$, where $K'(y_1) = F'$ for all $y_1 \in Y_1$. Note that under technologies

$\mathcal{A}_1, \mathcal{A}_2'$, the supervisor can induce F' as the outcome of a Nash equilibrium by offering contracts $w_{A1}(y) = w_{A2}(y) \equiv 0$ (and having agent 2 choose $(K', 0)$). We will show that it is optimal for the supervisor to do so, which will imply $F' \in \Phi^{ST}(w)$.

Suppose not; then there exist contracts w'_{A1}, w'_{A2} and a (mixed) Nash equilibrium (σ'_1, σ'_2) of the game between the agents, such that the supervisor's resulting payoff $\mathbb{E}_{H(\sigma'_1,\sigma'_2)}[w(y) - w'_{A1}(y) - w'_{A2}(y)]$ strictly exceeds $\mathbb{E}_{F'}[w(y)]$. Let π be the probability that σ'_2 places on action $(K', 0)$; thus we can write $\sigma'_2 = \pi \cdot (K', 0) + (1 - \pi) \cdot \sigma''_2$, where $\sigma''_2 \in \Delta(\mathcal{A}_2)$. If $\pi = 1$, then $H(\sigma'_1, \sigma'_2) = F'$, contradicting the assumption that the supervisor's payoff exceeds $\mathbb{E}_{F'}[w(y)]$. Hence $\pi < 1$.

We claim that under technologies $(\mathcal{A}_1, \mathcal{A}_2)$, if the supervisor instead offers contract $(1-\pi)w'_{A1}$ to agent 1 and w'_{A2} to agent 2, then (σ'_1, σ''_2) is a mixed-strategy equilibrium for the agents, and the supervisor's payoff is strictly higher than $\mathbb{E}_{F'}[y]$. (Note that $(1-\pi)w'_{A1}$ is in the allowed set of contracts $\mathcal{S}$, by convexity.) For the first part of the claim, note that because σ''_2 was part of a best reply by agent 2 against σ'_1 when agent 2 had technology $\mathcal{A}'_2$ and was offered w'_{A2}, it remains a best reply under $\mathcal{A}_2$ and w'_{A2}. As for agent 1, when he is offered $(1 - \pi)w'_{A1}$ and agent 2 plays σ''_2, his best-reply problem consists of choosing $(G, c_1) \in \mathcal{A}_1$ to maximize $\mathbb{E}_{K(\sigma''_2)G}[(1 - \pi)w'_{A1}(y)] - c_1$. Whereas when he was offered contract w'_{A1} and 2 played σ'_2, agent 1's objective was

$$\begin{aligned}\mathbb{E}_{K(\sigma'_2)G}[w'_{A1}(y)] - c_1 &= \mathbb{E}_{\pi F' + (1-\pi)K(\sigma''_2)G}[w'_{A1}(y)] - c_1 \\ &= \pi\mathbb{E}_{F'}[w'_{A1}(y)] + (1 - \pi)\mathbb{E}_{K(\sigma''_2)G}[w'_{A1}(y)] - c_1.\end{aligned}$$

So the two maximization problems differ only by a constant, so σ'_1 must remain a best reply for agent 1 under $(1 - \pi)w'_{A1}$ and σ''_2.

Finally, for the last part of the claim: we have assumed

$$\begin{aligned}\mathbb{E}_{F'}[w(y)] &< \mathbb{E}_{H(\sigma'_1,\sigma'_2)}[w(y) - w'_{A1}(y) - w'_{A2}(y)] \\ &= \pi\mathbb{E}_{F'}[w(y) - w'_{A1}(y) - w'_{A2}(y)] + (1 - \pi)\mathbb{E}_{H(\sigma'_1,\sigma''_2)}[w(y) - w'_{A1}(y) - w'_{A2}(y)].\end{aligned}$$

Since the first term on the right evidently is at most $\pi\mathbb{E}_{F'}[w(y)]$, we must have

$$\begin{aligned}\mathbb{E}_{F'}[w(y)] &< \mathbb{E}_{H(\sigma'_1,\sigma''_2)}[w(y) - w'_{A1}(y) - w'_{A2}(y)] \\ &\leq \mathbb{E}_{H(\sigma'_1,\sigma''_2)}[w(y) - (1 - \pi)w'_{A1}(y) - w'_{A2}(y)],\end{aligned}$$

which completes the proof of the claim.

But this shows that under $(\mathcal{A}_1, \mathcal{A}_2)$, the supervisor could have earned a payoff above $\mathbb{E}_{F'}[w(y)] \geq \mathbb{E}_F[w(y)]$, so that inducing F was not optimal, contradicting $F \in \Phi^{ST}(w)$. This contradiction completes the proof of Richness.

(Responsiveness) Let $w, w' \in C^+(Y)$ and $F \notin \Phi^{ST}(w)$ satisfy the hypotheses of Responsiveness. Let $\mathcal{A}_1 \supseteq \mathcal{A}_1^0$ and $\mathcal{A}_2 \supseteq \mathcal{A}_2^0$ be technologies, and let w_{A1}, w_{A2} be optimal contracts between the supervisor and agents under $w, \mathcal{A}_1, \mathcal{A}_2$, and let $F' \in \Gamma_A^S(w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2)$. Since $F \notin \Phi^{ST}(w)$, under $\mathcal{A}_1$ and $\mathcal{A}_2$, either (a) there does not exist $\tilde{w}_{A1}, \tilde{w}_{A2}$ such that F is induced in a Nash equilibrium (supervisor-preferred or otherwise), or (b) there do exist $\tilde{w}_{A1}, \tilde{w}_{A2}$ that induce F in a Nash equilibrium, but any such $\tilde{w}_{A1}, \tilde{w}_{A2}$ satisfy $\mathbb{E}_F[w(y) - \tilde{w}_{A1}(y) - \tilde{w}_{A2}(y)] < \mathbb{E}_{F'}[w(y) - w_{A1}(y) - w_{A2}(y)]$. Since changing w to w' does not affect the set $\mathcal{S}$ of contracts the supervisor can offer, if (a) holds, then it still holds under w', and therefore $F \notin \Gamma_S(w', \mathcal{A}_1, \mathcal{A}_2)$. Suppose (b) holds under w. Swapping w for w', observe that

$$\begin{aligned}
\mathbb{E}_{F'}[w'(y) - w_{A1}(y) - w_{A2}(y)] &\geq \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)] + \mathbb{E}_{F'}[w(y)] - \mathbb{E}_{F'}[w_{A1}(y) + w_{A2}(y)] \\
&> \mathbb{E}_F[w'(y)] - \mathbb{E}_F[w(y)] + \mathbb{E}_F[w(y)] - \mathbb{E}_F\left[\tilde{w}_{A1}(y) + \tilde{w}_{A2}(y)\right] \\
&= \mathbb{E}_F[w'(y) - \tilde{w}_{A1}(y) - \tilde{w}_{A2}(y)]
\end{aligned}$$

where the first inequality is by the hypothesis of Responsiveness, and the second inequality is by (b). Then F' is strictly preferred by the supervisor to F, and hence $F \notin \Gamma_S(w', \mathcal{A}_1, \mathcal{A}_2)$. Hence $F \notin \cup_{\mathcal{A}_1, \mathcal{A}_2} \Gamma_S(w', \mathcal{A}_1, \mathcal{A}_2) = \Phi^{ST}(w')$, and Responsiveness holds. □

Proof of Lemma 9. Let $\widehat{\Phi}(w)$ denote the set named in the lemma statement, so we wish to show $\Phi^{UT}(w) = \widehat{\Phi}(w)$.

($\Phi^{UT}(w) \subseteq \widehat{\Phi}(w)$). Let $F \in \Phi^{UT}(w)$, and let $(\mathcal{A}, H, c)$ be some valid technology, and σ an agents-optimal equilibrium under this technology with $F = H(\sigma)$. Let a be a potential-maximizing pure action profile, so that it is also an equilibrium (in pure strategies). Assume moreover that if a^0 remains potential-maximizing under the technology $(\mathcal{A}, H, c)$, then we have taken $a = a^0$.

Since σ is agents-optimal, and a is also an equilibrium,

$$\begin{aligned}\mathbb{E}_F[w(y)] &\geq \mathbb{E}_{H(\sigma)}[w(y)] - \sum_i c_i(\sigma_i) \\ &\geq \mathbb{E}_{H(a)}[w(y)] - \sum_i c_i(a_i) \\ &\geq \mathbb{E}_{H(a)}[w(y)] - I\sum_i c_i(a_i) \\ &\geq \mathbb{E}_{H(a^0)}[w(y)] - I\sum_i c_i(a_i^0) = w^0.\end{aligned}$$

Moreover, one of the inequalities is strict: either the potential is strictly higher under a than a^0 so that the fourth inequality is strict, or else (by assumption) $a = a^0$ and then the third inequality is strict since $\sum c_i(a_i) > 0$ and $I > 1$. Hence $\mathbb{E}_F[w(y)] > w^0$.

($\widehat{\Phi}(w) \subseteq \Phi^{UT}(w)$). First suppose w is nonconstant.

Let $F \in \widehat{\Phi}(w)$. Construct technology $(\mathcal{A}, H, c) \supseteq (\mathcal{A}^0, H^0, c^0)$ as follows: add a single action to each agent's original action set, a_i' at cost 0, and $H(a') = F$. Also write $\bar{w} = \max_{y\in Y} w(y)$ and $\underline{w} = \min_{y\in Y} w(y)$; since w is nonconstant, $\bar{w} > \underline{w}$. To define H at profiles where some but not all agents are playing the new action, we proceed as follows.

For any profile $a = (a_J', a_{-J})$ where a nonempty subset $J \subseteq I$ of agents are playing the new action, and all other agents (if any) are playing some profile $a_{-J} \in \mathcal{A}^0_{-J}$, let $p^J(a_{-J}) = \max_{a_J\in\mathcal{A}^0_J}\left\{\mathbb{E}_{H(a_J,a_{-J})}[w(y)] - I\sum_{j\in J} c_j(a_j)\right\}$. Note that $p^J(a_{-J}) < \bar{w}$, since the second term inside the max operator is strictly positive for all $a_J \in \mathcal{A}^0_J$. We also observe the recursive relationship $p^{J+i}(a_{-(J+i)}) = \max_{a_i\in\mathcal{A}^0_i}\left[p^J(a_i, a_{-(J+i)}) - Ic_i(a_i)\right]$ where $J + i$ is shorthand for $J \cup \{i\}$. And when $J = I$ (so a_{-J} is the empty profile), $p^J(a_{-J}) = w^0$.

Next, fix constants $\varepsilon_0 < \varepsilon_1 < \cdots < \varepsilon_{|I|}$ such that

- $\varepsilon_0 = 0$;
- all ε's are small enough so that $\varepsilon_{|J|} \leq \bar{w} - \max\{\underline{w}, p^J(a_{-J})\}$ for all nonempty J;
- if $w^0 \geq \underline{w}$, then $\varepsilon_{|I|} = \mathbb{E}_F[w(y)] - w^0$ (observe that this is consistent with the previous requirement);
- if $w^0 < \underline{w}$, then $\varepsilon_{|I|} < Ic_i(a_i)$ for all i and all $a_i \in \mathcal{A}_i^0$, and also $\varepsilon_{|I|} < \underline{w} - w^0$.

Now, whenever J is neither empty nor all of I, for any $a_{-J} \in \mathcal{A}^0_{-J}$, define $H(a_J', a_{-J})$ to be any distribution such that

$$\mathbb{E}_{H(a_J',a_{-J})}[w(y)] = \max\{\underline{w}, p^J(a_{-J})\} + \varepsilon_{|J|}. \tag{16}$$

The assumptions on the ε's ensure that the right side of (16) always lies in the interval $[\underline{w}, \bar{w}]$, so that the desired distribution indeed exists. Notice also that (16) holds for $J = I$ as well if $w^0 \geq \underline{w}$.

We claim that the profile a' is the unique equilibrium under this technology and contract w. In fact we will show that a'_i is a strictly dominant action for all i. Fix agent i, and fix profile $a_{-i} \in \mathcal{A}_{-i}$ and $a_i \in \mathcal{A}_i^0$. If $a \in \mathcal{A}^0$, we have

$$\begin{aligned}\mathbb{E}_{H(a'_i,a_{-i})}[w(y)/I] - c_i(a'_i) &= \max\{\underline{w}/I, \max_{\tilde{a}_i \in \mathcal{A}_i^0}\{\mathbb{E}_{H(\tilde{a}_i,a_{-i})}[w(y)/I] - c_i(\tilde{a}_i)\}\} + \varepsilon_1/I \\ &> \mathbb{E}_{H(a_i,a_{-i})}[w(y)/I] - c_i(a_i),\end{aligned}$$

so action a'_i is strictly preferred to a_i.

Now assume that at least 1 player $j \neq i$ is playing a'_j. If $\mathbb{E}_{H(a'_i,a_{-i})}[w(y)] = \bar{w}$, $\bar{w} \geq \mathbb{E}_{H(a_i,a_{-i})}[w(y)]$, so $P(a'_i, a_{-i}) - P(a_i, a_{-i}) > 0$. (The inequality is strict, since $c_i(a_i) > 0 = c_i(a'_i)$.) Otherwise, let J be the set of players different from i who are playing the new action in a_{-i}. As long as J is not all of $I \setminus \{i\}$, we have

$$\begin{aligned}P(a'_i, a_{-i}) - P(a_i, a_{-i}) &= \max[p^{J+i}(a_{-(J+i)}), \underline{w}] + \varepsilon_{|J|+1} \\ &\quad - \max[p^J(a_i, a_{-(J+i)}) - Ic_i(a_i), \underline{w} - Ic_i(a_i)] - \varepsilon_{|J|} \\ &\geq \max[p^{J+i}(a_{-(J+i)}), \underline{w}] + \varepsilon_{|J|+1} \\ &\quad - \max[p^{J+i}(a_{-(J+i)}), \underline{w} - Ic_i(a_i)] - \varepsilon_{|J|}\end{aligned}$$

where the inequality is by the recursive relationship of p^J. Clearly the first max term is weakly greater than the second max term, and $\varepsilon_{|J|+1} > \varepsilon_{|J|}$, so $P(a'_i, a_{-i}) - P(a_i, a_{-i}) > 0$.

If J is all of $I \setminus \{i\}$, but $w^0 \geq \underline{w}$ so that (16) still holds for I, then the same reasoning applies. The only remaining case is when $J = I \setminus \{i\}$ but $\underline{w} > w^0$. In this case, we have

$$\begin{aligned}P(a'_i, a_{-i}) - P(a_i, a_{-i}) &= \mathbb{E}_F[w(y)] - \max[p^J(a_i, a_{-(J+i)}) - Ic_i(a_i), \underline{w} - Ic_i(a_i)] - \varepsilon_{|J|} \\ &\geq \underline{w} - \max[p^{J+i}(a_{-(J+i)}), \underline{w} - Ic_i(a_i)] - \varepsilon_{|J|} \\ &= \min[\underline{w} - w^0, Ic_i(a_i)] - \varepsilon_{|J|} \\ &> 0\end{aligned}$$

where the last line uses the final assumption on the choice of ε's.

This analysis shows that a'_i is a strictly dominant strategy for each agent i. So, the unique equilibrium is the action profile a', and so $F \in \Phi^{UT}(w)$.

Finally, suppose w is constant, in which case $\widehat{\Phi}(w)$ is all of $\Delta(Y)$. Take any $F \in \Delta(Y)$. Construct technology $(\mathcal{A}, H, c) \supseteq (\mathcal{A}^0, H^0, c^0)$ by adding a single new action a_1' for agent 1, at cost $c_1(a_1') = 0$, and set $H(a_1', a_{-1}) = F$ for all $a_{-1} \in \mathcal{A}_{-1}$. Since w is constant, any profile where all agents are playing a minimum cost action is an agents-optimal equilibrium. Any such profile involves agent 1 playing a_1', which results in $F \in \Phi^{UT}(w)$. □

Proof of Proposition 10. (Richness) Follows directly from Lemma 9.

(Responsiveness) Suppose $F \notin \Phi^{UT}(w)$, and

$$\mathbb{E}_{F'}[w'(y)] - \mathbb{E}_F[w'(y)] \geq \mathbb{E}_{F'}[w(y)] - \mathbb{E}_F[w(y)] \qquad \text{for all } F' \in \Phi^{UT}(w).$$

By Lemma 9, $\mathbb{E}_F[w(y)] \leq w^0$. Let a^0 be the maximizer of P (the potential when contract w is given) over $\mathcal{A}^0$, and let $F' = H^0(a^0)$, and let $C = I \sum_i c_i^0(a_i^0)$, so $\mathbb{E}_F[w(y)] \leq P(a^0) = \mathbb{E}_{F'}[w(y)] - C$. Furthermore, since $\mathbb{E}_{F'}[w(y)] > w^0$, $F' \in \Phi^{UT}(w)$, again by Lemma 9. Let P' the potential under contract w'. Then

$$\begin{aligned}
\mathbb{E}_F[w'(y)] &\leq \mathbb{E}_{F'}[w'(y)] - \mathbb{E}_{F'}[w(y)] + \mathbb{E}_F[w(y)] \\
&\leq \mathbb{E}_{F'}[w'(y)] - \mathbb{E}_{F'}[w(y)] + \mathbb{E}_{F'}[w(y)] - C \\
&= P'(a^0) \leq w'^0.
\end{aligned}$$

Again applying Lemma 9, $F \notin \Phi^{UT}(w')$.

With Richness and Responsiveness proven, applying Theorem 1 yields the result. □

Proof of Proposition 11. Suppose w is a contract and $F \in \Phi^{PSA3}(w)$, so there is some $\beta \in [0, 1]$ for which βw is optimal for the supervisor, and some $\mathcal{A}$ such that $F \in \Gamma_A^S(\beta w, w, \mathcal{A})$. Let F' be any other distribution with $\mathbb{E}_{F'}[w(y)] \geq \mathbb{E}_F[w(y)]$. Then F' also leads to (weakly) higher expected values than w for both the agent's payment $\beta w(y)$ and the supervisor's payoff $w(y) - \beta w(y)$, so defining $\mathcal{A}' = \mathcal{A} \cup \{(F', 0)\}$, we have $F' \in \Gamma_A^S(\beta w, w, \mathcal{A}')$. Consequently, $F' \in \Phi^{PSA3}(w)$ as needed. □

The proof of Lemma 12 makes use of the following fact:

Lemma 21. *Suppose $x, y, \tilde{x}, \tilde{y}$ are nonnegative numbers with $\sqrt{x} - \sqrt{y} \geq \sqrt{\tilde{x}} - \sqrt{\tilde{y}}$ and $\sqrt{x} - \sqrt{y} > 0$. Put $\beta = \sqrt{y/x}$. Then, $\beta x - y \geq \beta \tilde{x} - \tilde{y}$.*

Proof. Put $u = \sqrt{x}, v = \sqrt{y}, \tilde{u} = \sqrt{\tilde{x}}, \tilde{v} = \sqrt{\tilde{y}}$. So $u - v \geq \tilde{u} - \tilde{v}$. Note that the function $f(t) = \frac{v}{u}(u - v + t)^2 - t^2$ is a negative quadratic in t, maximized when $t = v$. Hence,

$$\frac{v}{u}\tilde{u}^2 - \tilde{v}^2 \leq \frac{v}{u}(u - v + \tilde{v})^2 - \tilde{v}^2 = f(\tilde{v}) \leq f(v) = uv - v^2.$$

Writing in terms of x's and y's gives the inequality stated in the lemma. □

Proof of Lemma 12. First, note that since $\beta > 0$, the supervisor's payoff $w - \beta w$ is a scalar multiple of βw. This implies that the agent's choice is not affected by tie-breaking to favor the supervisor: $\Gamma_A^S(\beta w, w, \mathcal{A}) = \Gamma_A(\beta w, \mathcal{A})$.

Now to check the characterization in (7). If the agent chooses (F', c') under technology $\mathcal{A}$, then

$$E_{F'}[\beta w(y)] \geq E_{F'}[\beta w(y)] - c' \geq E_F[\beta w(y)] - c,$$

and dividing through by β yields (7). Conversely, suppose (7) is satisfied. Note that the targeted action (F, c) is indeed optimal for the agent among actions in $\mathcal{A}^0$, since $(\tilde{F}, \tilde{c}) \in \mathcal{A}^0$ implies $\sqrt{\mathbb{E}_F[w(y)]} - \sqrt{c} \geq \sqrt{\mathbb{E}_{\tilde{F}}[w(y)]} - \sqrt{\tilde{c}}$ by targeting, hence $\beta\mathbb{E}_F[w(y)] - c \geq \beta\mathbb{E}_{\tilde{F}}[w(y)] - \tilde{c}$ by Lemma 21. In turn, for any F' that satisfies (7), action $(F', 0)$ (if it is available) is at least as good for the agent as (F, c), and so under technology $\mathcal{A} = \mathcal{A}^0 \cup \{(F', 0)\}$, this action becomes optimal for the agent, i.e. $F' \in \Gamma_A^S(\beta w, w, \mathcal{A})$. □

Proof of Proposition 14. Begin with a technology $\mathcal{A}$ under which F^* arises; by Lemma 4, we may assume $(F^*, 0) \in \mathcal{A}$.

First, we add all distributions at cost $\bar{c}$ to $\mathcal{A}$, to form $\mathcal{A}'$. So $\mathcal{A}' = \mathcal{A} \cup (\Delta(Y) \times \{\bar{c}\})$. $\mathcal{A}'$ is still a valid technology, since it is the union of two compact sets. For any S-A contract w_A, for any $F \in \Delta(Y)$, $(F, \bar{c})$ being the optimal action under w_A implies $\mathbb{E}_F[w_A(y)] - \bar{c} \geq 0$, so $\mathbb{E}_F[w_A(y)] \geq \bar{c}$. But if the supervisor offers w_A, then her payoff is $\mathbb{E}_F[\alpha y - w_A(y)] < \alpha\mathbb{E}_F[y] - \bar{c} \leq (\alpha - 1)\bar{y} \leq 0$, so w_A is less preferred than the zero contract for the supervisor, so the agent will never take any action $(F, \bar{c})$ when the supervisor behaves optimally, hence $\Gamma_S(w_\alpha, \mathcal{A}) = \Gamma_S(w_\alpha, \mathcal{A}')$.

Next, for each distribution available in $\mathcal{A}'$, we add in all distributions with lower means at the same cost. That is, define $\mathcal{A}'' = \mathcal{A}' \cup \bigcup_{(F,c)\in\mathcal{A}'} \left\{(\tilde{F}, c) : \mathbb{E}_{\tilde{F}}[y] \leq \mathbb{E}_F[y]\right\}$. That $\mathcal{A}''$ is indeed a valid technology follows from the fact that it is a closed subset of a compact set $\Delta(Y) \times [0, \bar{c}]$: Let $(F_n, c_n) \to (F, c)$, where $(F_n, c_n) \in \mathcal{A}''$ for each n. Then there is a sequence $\{(G_n, d_n)\} \subseteq \mathcal{A}'$ such that $\mathbb{E}_{F_n}[y] \leq \mathbb{E}_{G_n}[y]$ and $d_n = c_n$ for each n. Since $\mathcal{A}'$ is compact, (G_n, d_n) has a convergent subsequence, which converges to, say, $(G, d) \in \mathcal{A}'$. Then $c = d$ and $\mathbb{E}_F[y] \leq \mathbb{E}_G[y]$, so $(F, c) \in \mathcal{A}''$ and $\mathcal{A}''$ is closed.

Now, let $\mathcal{A}''' = \overline{\text{co}(\mathcal{A}'')}$, the closed convex hull of $\mathcal{A}''$. The set $\mathcal{A}'''$ is compact, and hence a valid technology.

Finally, let $\kappa(\mu) = \inf\{c : (F, c) \in \mathcal{A}''' \text{ for some } F,\ \mathbb{E}_F[y] = \mu\}$. Clearly, $\kappa(\mu) = 0$ for any $\mu \leq \mathbb{E}_{F^*}[y]$, since $(F^*, 0) \in \mathcal{A}$ and we added all distributions with smaller mean at cost 0. Because $\mathcal{A}'''$ is compact, κ is continuous. The function κ is convex since $\mathcal{A}'''$ is a convex set, and the lower envelope of a convex set is convex. Finally, κ is nondecreasing since we know if $(\tilde{F}, \tilde{c}) \in \mathcal{A}'''$, then $(\tilde{F}', \tilde{c}) \in \mathcal{A}'''$ whenever $\mathbb{E}_{\tilde{F}'}[y] \leq \mathbb{E}_{\tilde{F}}[y]$. And $\kappa(\mathbb{E}_F[y]) \leq c$ for all $(F, c) \in \mathcal{A}^0$ because $\mathcal{A}^0 \subseteq \mathcal{A}'''$. Thus κ is a valid cost function, and notice that $\mathcal{A}''' = \mathcal{A}_\kappa$.

It remains to show that $F^* \in \Gamma_S(w_\alpha, \mathcal{A}''')$. Note that if the S-A contract is the zero contract, then $F^* \in \Gamma_A^{PS}(w_\alpha, 0, \mathcal{A}''')$, since $F^* \in \Gamma_A^S(w_\alpha, 0, \mathcal{A})$, and no higher-mean output distributions were added at zero cost in the previous steps. By Proposition 13, the supervisor can do no better than offering linear contracts, so if we can show that the supervisor payoff to a linear contract is no better than the payoff under the zero contract, then $F^* \in \Gamma_S(w_\alpha, \mathcal{A}''')$. We already saw that inducing F^* via the zero contract was optimal for the supervisor under $\mathcal{A}'$. Consider any linear S-A contract $w_A(y) = ay$, under technology $\mathcal{A}'''$. Let $\tilde{F} \in \Gamma_A^{PS}(w_\alpha, w_A, \mathcal{A}')$. Then $\tilde{F} \in \Gamma_A(w_\alpha, w_A, \mathcal{A}'')$, since any $(\hat{F}, \hat{c}) \in \mathcal{A}''$ satisfies $\mu_{\hat{F}} \leq \mu_F, \hat{c} = c$ for some $(F, c) \in \mathcal{A}'$, thus showing $a\mu_{\tilde{F}} - \tilde{c} \geq a\mu_F - c \geq a\mu_{\hat{F}} - \hat{c}$ (using agent-optimality in $\mathcal{A}'$), so the action remains agent-optimal in $\mathcal{A}''$. Moreover, $\tilde{F}$ survives supervisor- and principal-preferred tiebreaking in $\mathcal{A}''$ since no greater mean actions are optimal for the agent, thus $\tilde{F} \in \Gamma_A^{PS}(w_\alpha, w_A, \mathcal{A}'')$. Since the agent's objective is continuous and linear on $\Delta(Y) \times [0, \bar{c}]$ under $w_A(y) = ay$, all maximizers of the agent's objective with w_A and $\mathcal{A}''$ are also maximizers with w_A and $\mathcal{A}'''$, and any new maximizers under $\mathcal{A}'''$ are (the limit of) convex combinations of maximizers under $\mathcal{A}''$. Thus, the mean output from any linear w_A doesn't increase from $\mathcal{A}''$ to $\mathcal{A}'''$, which shows that $\tilde{F} \in \Gamma_A^{PS}(w_\alpha, w_A, \mathcal{A}''')$. Thus, $w_A = 0$ is still an optimal choice for the supervisor, so $F^* \in \Gamma_S(w_\alpha, \mathcal{A}''')$. □

Proof of Lemma 15. As argued preceding the lemma statement, it suffices to check that $\tilde{\kappa}(\tilde{\mu}) \leq \kappa(\tilde{\mu})$ for each $\tilde{\mu}$. For $\tilde{\mu} \leq \mu$, $\tilde{\kappa}(\tilde{\mu}) = 0 \leq \kappa(\tilde{\mu})$ since costs are nonnegative. For $\tilde{\mu} > \mu$, from optimality of the supervisor's choice under κ, we have

$$\mu[\alpha - \kappa'(\mu)] \geq \tilde{\mu}[\alpha - \kappa'(\tilde{\mu})].$$

By rearranging this inequality, we obtain

$$\kappa'(\tilde{\mu}) \geq \left(1 - \frac{\mu}{\tilde{\mu}}\right) \alpha + \frac{\mu}{\tilde{\mu}} \kappa'(\mu).$$

Furthermore,

$$\tilde{\kappa}'(\tilde{\mu}) = \left(1 - \frac{\mu}{\tilde{\mu}}\right) \alpha,$$

and κ nondecreasing implies that $\kappa'(\mu) \geq 0$, hence

$$\kappa'(\tilde{\mu}) \geq \left(1 - \frac{\mu}{\tilde{\mu}}\right) \alpha = \tilde{\kappa}'(\tilde{\mu}).$$

Integrating this inequality from μ to $\tilde{\mu}$ gives $\kappa(\tilde{\mu}) - \kappa(\mu) \geq \tilde{\kappa}(\tilde{\mu}) - \tilde{\kappa}(\mu)$, and given that $\kappa(\mu) \geq \tilde{\kappa}(\mu)$, we conclude that $\kappa(\tilde{\mu}) \geq \tilde{\kappa}(\tilde{\mu})$. Hence $\tilde{\kappa} \leq \kappa$ everywhere. □

Proof of Proposition 17. (a) Fix α and $\tilde{\mu}$ and let $\mu' > \mu$. If $\tilde{\mu} \leq \mu'$, then $\kappa(\tilde{\mu}, \mu', \alpha) = 0 \leq \kappa(\tilde{\mu}, \mu, \alpha)$. And if $\tilde{\mu} > \mu'$, then the function $\kappa(\tilde{\mu}, \cdot, \alpha)$ is given by the second branch of (14) throughout the interval from μ to μ', so it suffices to check that the derivative of this formula with respect to the second argument of κ is negative over the relevant range:

$$\frac{\partial}{\partial \mu} \left(\alpha[(\tilde{\mu} - \mu) - \mu \log(\tilde{\mu}/\mu)]\right) = -\alpha \log(\tilde{\mu}/\mu) < 0.$$

(b) This is immediate from the definition, and the fact that $\kappa(\tilde{\mu}, \mu, \alpha) \geq 0$. □

Proof of Lemma 19. When the principal offers a linear contract w_α with slope $\alpha \in (0, 1]$ to the supervisor, for a fixed technology $\mathcal{A}^1$ known to the supervisor, her objective is

$$V_S^u(w_A | w_\alpha, \mathcal{A}^1) = \inf_{\mathcal{A} \supseteq \mathcal{A}^1} \mathbb{E}_{\Gamma_A^S(w_\alpha, w_A, \mathcal{A})}[\alpha y - w_A(y)].$$

Define $\Psi(w_A) = \bigcup_{\mathcal{A} \supseteq \mathcal{A}^1} \Gamma_A^S(w_\alpha, w_A, \mathcal{A})$; and further define $\tilde{y}$ to be αy, define $\tilde{F}$ as the distribution of $\tilde{y}$ when $y \sim F$, and define $\tilde{w}_A(\tilde{y}) = w_A(\tilde{y}/\alpha)$. Then note that the supervisor's objective is written as

$$\inf_{F \in \Psi(w_A)} \mathbb{E}_{\tilde{F}}[\tilde{y} - \tilde{w}_A(\tilde{y})].$$

Thus, the supervisor-agent relationship is the robust principal-agent model of Carroll

(2015) (and the subject of Section 4.1). The analysis of that paper shows that, under any $\mathcal{A}^1$, the supervisor's solution is given by identifying the action in $\mathcal{A}^1$ that maximizes the quantity

$$\sqrt{\mathbb{E}_{\tilde{F}}[\tilde{y}]} - \sqrt{c}, \tag{17}$$

and if this quantity is positive, then offering the agent a linear contract $\tilde{w}_A(\tilde{y}) = a\tilde{y}$ where $a = \sqrt{c/\mathbb{E}_{\tilde{F}}[\tilde{y}]}$. (If there are multiple actions maximizing (17), then any corresponding value of a is optimal for the supervisor. And if no action in $\mathcal{A}^1$ yields a positive value for expression (17), then the supervisor's best guarantee is zero, and it is optimal for her to offer the agent the zero contract, though other optimal contracts also exist.)

By Lemma 6, if distribution F ever arises for some technologies $\mathcal{A}^1$ and $\mathcal{A}$, then in particular it can arise for technologies such that $(F, 0) \in \mathcal{A}^1 \subseteq \mathcal{A}$ and the supervisor offers a contract of slope zero. This can only happen if this action maximizes (17) over $\mathcal{A}^1$; in particular, this requires $\alpha\mathbb{E}_F[y] = \left(\sqrt{\mathbb{E}_F[\alpha y]} - \sqrt{0}\right)^2 \geq \left(\sqrt{\alpha\mathbb{E}_{F'}[y]} - \sqrt{c'}\right)^2$ for all $(F', c') \in \mathcal{A}^0|_\alpha$, as in the lemma statement. (To be precise, the supervisor could also choose the zero contract if (17) is maximized by some other action also of the form $(F', 0) \in \mathcal{A}^1$, but then favorable tie-breaking would imply the agent would choose F' instead of F, a contradiction.)

Conversely, if F is a distribution such that $\alpha\mathbb{E}_F[y] \geq \left(\sqrt{\alpha\mathbb{E}_{F'}[y]} - \sqrt{c'}\right)^2$ for all $(F', c') \in \mathcal{A}^0|_\alpha$, then by taking $\mathcal{A} = \mathcal{A}^1 = \mathcal{A}^0 \cup \{(F, 0)\}$, the supervisor would indeed find it optimal to offer the zero contract and have the agent take action $(F, 0)$ (and this action choice is consistent with the tie-breaking conditions). □

Proof of Theorem 20. ($\Phi^{PA}(w) \subseteq \Phi^{PSA1}(w)$) Consider $F \in \Phi^{PA}(w)$. There exists some technology $\mathcal{A} \supseteq \mathcal{A}^0$ such that $\mathbb{E}_F[w(y)] \geq \mathbb{E}_{\tilde{F}}[w(y)] - \tilde{c}$ for all $(\tilde{F}, \tilde{c}) \in \mathcal{A}$, with $(F, 0) \in \mathcal{A}$. Consider hierarchical model (i). We will show that $F \in \Gamma_S(w, \mathcal{A})$. When the supervisor offers $w_A(y) = 0$, $F \in \Gamma_A(w_A, \mathcal{A})$, and she can obtain payoff $\mathbb{E}_F[w(y)]$. Consider any other action $(\tilde{F}, \tilde{c}) \in \mathcal{A}$. In order to get $\tilde{F} \in \Gamma_A(w_A, \mathcal{A})$ for some $w_A \in \mathcal{S}$, it must be that $\mathbb{E}_{\tilde{F}}[w_A(y)] - \tilde{c} \geq \mathbb{E}_F[w_A(y)] \geq 0$, so $\mathbb{E}_{\tilde{F}}[w_A(y)] \geq \tilde{c}$. Then the supervisor's payoff from offering any other contract and inducing $\tilde{F}$ (if this is even possible) must be

$$\mathbb{E}_{\tilde{F}}[w(y) - w_A(y)] \leq \mathbb{E}_{\tilde{F}}[w(y)] - \tilde{c} \leq \mathbb{E}_F[w(y)]$$

and therefore $w_A(y) = 0$ is a maximizer of the supervisor's payoff, and $F \in \Gamma_A^S(w, 0, \mathcal{A})$. If $(\tilde{F}, \tilde{c}) \in \mathcal{A}$ is *not* a maximizer of the agent's objective in the P-A model, the argument above shows that $\tilde{F}$ is indeed not a member of $\Gamma_A^S(w, 0, \mathcal{A})$. Then $F \in \Gamma_A^{PS}(w, 0, \mathcal{A})$ follows

from F being principal-preferred in the P-A model, so $F \in \Phi^{PSA1}(w)$.

($\Phi^{PSA1}(w) \subseteq \Phi^{PSA2}(w)$) Consider $F \in \Phi^{PSA1}(w)$. From Lemma 4, there exists $\mathcal{A} \supseteq \mathcal{A}^0$ such that $(F,0) \in \mathcal{A}$, $F \in \Gamma_A^S(w,0,\mathcal{A})$ and $w_A(y) = 0$ is a maximizer of $V_S^i(\cdot|w,\mathcal{A})$. Let $\mathcal{A}^1 = \mathcal{A}$. These hypotheses imply that $\mathbb{E}_F[w(y)] \geq \mathbb{E}_{\Gamma_A^S(w,w_A,\mathcal{A})}[w(y) - w_A(y)]$ for every $w_A \in \mathcal{S}$. We want to show that $F \in \Gamma_S(w,\mathcal{A}^1,\mathcal{A})$ for hierarchical model (ii). We know

$$\inf_{\tilde{\mathcal{A}} \supseteq \mathcal{A}^1} \mathbb{E}_{\Gamma_A^S(w,0,\tilde{\mathcal{A}})}[w(y)] = \mathbb{E}_F[w(y)].$$

It is without loss of generality to consider only contracts w_A of the form βw for $\beta \in [0,1]$, since Proposition 3 applied to the supervisor-agent relationship shows that there is an optimal contract of this form. By assumption, these contracts are contained in $\mathcal{S}$. Hence for any such contract $w_A = \beta w \in \mathcal{S}$,

$$\begin{aligned} V_S^u(0|w,\mathcal{A}^1) = \mathbb{E}_F[w(y)] &\geq \mathbb{E}_{\Gamma_A^S(w,w_A,\mathcal{A})}[w(y) - w_A(y)] \\ &\geq \inf_{\tilde{\mathcal{A}} \supseteq \mathcal{A}^1} \mathbb{E}_{\Gamma_A^S(w,w_A,\tilde{\mathcal{A}})}[w(y) - w_A(y)] = V_S^u(w_A|w,\mathcal{A}^1) \end{aligned}$$

where the first inequality is by $F \in \Gamma_S(w,\mathcal{A})$ and the second by definition of infimum. Hence $w_A(y) = 0$ is a maximizer of $V_S^u(\cdot|w,\mathcal{A}^1)$, and $F \in \Gamma_A^S(w,0,\mathcal{A})$. Moreover, since $\Gamma_A^S(w,0,\mathcal{A})$ is the same in both model (i) and (ii), if F survives principal-preferred tiebreaking within this set in model (i) then it also survives the tiebreaking in model (ii). Thus $F \in \Phi^{PSA2}(w)$.

For the last statement of the theorem: Whenever w is linear, our requirement on $\mathcal{S}$ is satisfied, and so we have shown that $\Phi^{PA}(w) \subseteq \Phi^{PSA1}(w) \subseteq \Phi^{PSA2}(w)$. So, taking the infima over the respective sets, $V_P^{PA}(w) \geq V_P^{PSA1}(w) \geq V_P^{PSA2}(w)$. Taking maxima over w, and noting that each maximum is attained for a linear w by our earlier results, completes the proof.

□

B Compactness of Supervisor's Contract Space

This section of the appendix discusses the assumption, made in hierarchical model (i), that the contracts offered by the supervisor are constrained to a compact set $\mathcal{S} \subseteq C^+(Y)$. (For simplicity we focus here on the hierarchical model, but note that the same assumption was also made in the supervised teams model, and very similar comments apply there.)

Indeed, some kind of restriction on the set of contracts allowed to the supervisor is

needed in order to ensure that the supervisor's maximization problem has a solution. Otherwise, for a given technology $\mathcal{A} \supseteq \mathcal{A}^0$ and fixed contract w between principal and supervisor, it is possible that $\Gamma_S(w, \mathcal{A})$ is empty. In such a case, the supervisor's behavior has not been defined. (It also would not make sense to try to patch the model by simply ruling out a priori the $\mathcal{A}$'s for which Γ_S is empty, since this set of $\mathcal{A}$'s depends on the P-S contract w.)

We emphasize this, because the concern for possible nonexistence of a solution to the supervisor's problem is not simply due to pathologies. Indeed, even when the supervisor receives a linear contract from the principal, and the true technology contains just two actions with continuous densities, it is possible that there exists no optimal contract for the supervisor in the space $C^+(Y)$. We give an example similar to that of Mirrlees (1999), but adapted to the context of hierarchical model (i).

Take $w(y) = y$, so the principal gives all output to the supervisor. Let $Y = [0, 1]$ and take technology $\mathcal{A} = \{(F, 0), (G, c)\}$, where F and G have densities

$$f(y) = -2y + 2$$
$$g(y) = 2y$$

respectively. Then $\mathbb{E}_F[y] = 1/3$ and $\mathbb{E}_G[y] = 2/3$. Suppose $0 < c < 1/3$, so the agent taking action (G, c) generates more total surplus than $(F, 0)$. To induce the agent to take action (G, c), the supervisor must pay the agent at least c in expectation under G. If the supervisor could pay just this amount to incentivize the agent to take action (G, c) over $(F, 0)$, then the supervisor could capture the entire expected surplus, $\mathbb{E}_G[y] - c$. In fact, the supervisor can induce action (G, c) by paying (in expectation) arbitrarily close to c, for example, by paying $c/(2 - 4\varepsilon)\varepsilon$ for realizations $y > 1 - \varepsilon$ and 0 for other realizations. (To be precise, this payment function is disallowed because it is not continuous, but it can be arbitrarily approximated by continuous functions.) However, the supervisor cannot pay exactly cost c, since F has full support and so any (nonzero) contract would leave the agent a positive rent under $(F, 0)$, hence must pay strictly more than c in order to induce the agent to take action (G, c). So the supremum payoff for the supervisor is not attained.

An alternative approach to ensuring existence of an optimal contract for the supervisor would be to modify the outcome space Y. The assumption that Y is finite is enough to establish existence of a payoff-maximizing contract to the supervisor, as demonstrated in Grossman and Hart (1983). More generally, we could take Y to be any compact subset of $\mathbb{R}$, but specify a finite subset $\tilde{Y}$ of Y, and restrict contracts w_A to be piecewise-linear with

kink points in $\tilde{Y}$. This restriction, together with any upper bound on the possible values of w_A, carves out a compact set of possible contracts. Thus finiteness of the outcome space is conceptually close to our approach of assuming a compact $\mathcal{S}$.

Finally, it may be possible to do away with the need for existence of maximizing contracts for the supervisor by making alternative assumptions about the supervisor's behavior, such as assuming the supervisor picks some ε-optimal contract rather than an optimal contract. We do not explore this route further here.

References

Aliprantis, Charalambos D. and Kim C. Border (2006) *Infinite Dimensional Analysis: A Hitchhiker's Guide*: Springer, 3rd edition.

Barron, Daniel, George Georgiadis, and Jeroen Swinkels (2019) "Optimal Contracts with a Risk-Taking Agent," February, Unpublished manuscript.

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 (2018) "Robust Incentives for Teams," April, Unpublished manuscript.

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.

Glicksberg, Irving L. (1952) "A Further Generalization of the Kakutani Fixed Point Theorem, with Application to Nash Equilibrium Points," *Proceedings of the American Mathematical Society*, **3** (1), 170–174.

Grossman, Sanford and Oliver Hart (1983) "An Analysis of the Principal-Agent Problem," *Econometrica*, **51** (1), 7–45.

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

Innes, Robert D. (1990) "Limited Liability and Incentive Contracting with Ex-ante Action Choices," *Journal of Economic Theory*, **52**, 45–67.

Marku, Keler and Sergio Ocampo Diaz (2019) "Robust Contracts in Common Agency," March, 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**, 367–390.

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

Ray, Debraj ® Arthur Robson (2018) "Certified Random: A New Order for Coauthorship," *American Economic Review*, **108** (2), 489–520.

Rockafellar, R. Tyrrell (1970) *Convex Analysis*: 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.

——— (1994) *The Theory of Industrial Organization*: The MIT Press.