Econometrica Supplementary Material

SUPPLEMENT TO "A GENERAL FRAMEWORK FOR ROBUST CONTRACTING MODELS"

(*Econometrica*, Vol. 90, No. 5, September 2022, 2129–2159)

DANIEL WALTON
Uber Technologies

GABRIEL CARROLL
Department of Economics, University of Toronto

THIS SUPPLEMENTARY MATERIAL for "A General Framework for Robust Contracting Models" gives additional results and applications. Sections, results, and equations in this Supplementary Material are numbered starting with "S"; numbers without "S" refer to the main paper.

To briefly outline, we begin in Section S-1 by discussing two applications of the framework beyond the ones described in the main paper: one that involves contracting with a supervisor who subcontracts with a team of agents, which we call the "supervised team" model, and one where the principal contracts directly with many agents, which we call the "unsupervised team" model. The latter is based on the model in Dai and Toikka (2022). In Section S-2, we show how to compare worst-case outcomes in the robust principal-agent model and hierarchical models (i) and (ii). Section S-3 shows how a variant on our Richness property, combined with Responsiveness, leads to concave (rather than linear) contracts. To finish, Section S-4 contains all remaining proofs, both those omitted from the main text as well as those for results within this Supplement.

S-1. TEAM APPLICATIONS

In the following two subsections, we present two models of contracting with a team that fall within the framework of Section 3, to further illustrate the breadth of the framework. We demonstrate that both models satisfy Richness and Responsiveness, so Theorem 1 applies. For brevity, we do not concern ourselves with existence of an optimal contract here; this means we can also dispense with the extra detail of principal-preferred tie-breaking.

S-1.1. *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.[1] Given $K \in C(Y_1, \Delta(Y))$, each $y_1 \in Y_1$ defines

Daniel Walton: dbw.econ@gmail.com
Gabriel Carroll: gabriel.carroll@utoronto.ca

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

 https://doi.org/10.3982/ECTA17386

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 $KG \in \Delta(Y)$ on Borel sets $A \subseteq Y$ as

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

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 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 in hierarchical model (i), we assume that the principal only knows $\mathcal{A}_1^P \subseteq \mathcal{A}_1$ and $\mathcal{A}_2^P \subseteq \mathcal{A}_2$, and the supervisor and agents 1 and 2 all know $\mathcal{A}_1$ and $\mathcal{A}_2$. Thus $\mathcal{A}_1^P$ and $\mathcal{A}_2^P$ 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,

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

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 Glicksberg's existence theorem (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 $\mathbb{E}_{H(\sigma)}[w(y) - w_{A1}(y) - w_{A2}(y)]$. 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 tie-breaking assumptions in the hierarchical models. We then write

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

for (any) $\sigma \in \mathcal{E}^S(w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2)$, and write $\Gamma_A^S(w, w_{A1}, w_{A2}, \mathcal{A}_1, \mathcal{A}_2)$ 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

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

In words, for fixed P-S 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-preferred Nash equilibrium given

w_{A1}, w_{A2}. We can verify nonempty-valuedness of $\Gamma^i_S(w, \mathcal{A}_1, \mathcal{A}_2)$ using compactness arguments as in the hierarchical models. The outcome correspondence is then defined to be

$$\Phi^{ST}(w) = \bigcup_{\text{technologies } \mathcal{A}_1 \supseteq \mathcal{A}^P_1, \mathcal{A}_2 \supseteq \mathcal{A}^P_2} \Gamma^i_S(w, \mathcal{A}_1, \mathcal{A}_2),$$

and the principal evaluates contracts according to the resulting objective $V^{ST}_P(w)$.

PROPOSITION S-1: *For any $w \in C^+(Y)$, there exists a linear $w' \in C^+(Y)$ such that $V^{ST}_P(w') \geq V^{ST}_P(w)$.*

The proof of Proposition S-1 is similar to those in the hierarchical 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. 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.

S-1.2. *Unsupervised Team*

We now consider a different formulation with a team, one that is based on the model in Dai and Toikka (2022), with some simplification. In the unsupervised team model, the principal directly contracts with a team of $I \geq 2$ agents, indexed $i = 1, \dots, I$. 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 specifying action costs, $c_i : \mathcal{A}_i \to \mathbb{R}^+$ for each agent i, and the output distribution produced by any action profile, $H : \mathcal{A} \to \Delta(Y)$. Agent i's action set is $\mathcal{A}_i$. Given a profile of mixed actions $\sigma = (\sigma_1, \dots, \sigma_I)$ (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 $c_i(\sigma_i) = \sum_{a_i \in \mathcal{A}_i} \sigma_i(a_i) c_i(a_i)$.

Consistent with all of the previous applications, we assume that the principal is poorly informed of the technology, so that the principal only knows some partial action sets and corresponding functions, $(\mathcal{A}^P, c^P_1, \dots, c^P_I, H^P)$; these are the model primitives. It is assumed that the unknown $\mathcal{A} \supseteq \mathcal{A}^P$ is finite, $c_i : \mathcal{A}_i \to \mathbb{R}^+$ for all i, and $H : \mathcal{A} \to \Delta(Y)$, such that $c_i(a_i) = c^P_i(a_i)$ for all $i, a_i \in \mathcal{A}^P_i$, and $H(a) = H^P(a)$ for all $a \in \mathcal{A}^P$. We use notation $(\mathcal{A}, H, c) \supseteq (\mathcal{A}^P, H^P, c^P)$ to denote this relationship. Additionally, we assume that $c^P_i(a_i) > 0$ for all $a_i \in \mathcal{A}^P_i$, for all i. (This assumption simplifies the proofs, though it diverges from Dai and Toikka.)

Our substantive 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 $(w_1, \dots, w_I) \in C^+(Y)^I$. 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. Whereas this affine equivalence is a result in their model, we take it as a starting point, and we add 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 thus

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

We extend payoffs to mixed strategy profiles linearly, as usual. With these payoffs, a contract w and technology $(\mathcal{A}, H, c)$ define 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").[2] We denote the set of agents-optimal equilibria as $\mathcal{E}^A(w, \mathcal{A}, H, c) \subseteq \mathcal{E}(w, \mathcal{A}, H, c)$. Let $\Gamma(w, \mathcal{A}, H, c) = \{H(\sigma) : \sigma \in \mathcal{E}^A(w, \mathcal{A}, H, c)\}$. Hence, the outcome correspondence is defined as

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

The principal evaluates contracts according to the resulting objective $V_P^{UT}(w)$.

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 among the agents, with potential $P : \mathcal{A} \to \mathbb{R}$ defined by

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

(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 $\mathcal{A}^P$, and let $w^0 = P(a^0)$ be the corresponding maximum value.

LEMMA S-2: $\Phi^{UT}(w) = \{F \in \Delta(Y) : \mathbb{E}_F[w(y)] > w^0\}$.

The proof, which adapts techniques from Dai and Toikka (2022), is in Section S-4. The argument that every distribution that may be chosen does indeed satisfy $\mathbb{E}_F[w(y)] > w^0$

[2]Dai and Toikka (2022) 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 can be violated for some (undesirable) contracts.

is essentially a direct application of the potential game structure, with additional use of the equilibrium selection criterion and the assumption $c_i^P(a_i) > 0$ to ensure the 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 S-2, it is straightforward to check Richness and Responsiveness, allowing us to apply Theorem 1.

PROPOSITION S-3: *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 Section S-4.

S-2. MODEL COMPARISONS

As a further application of the hierarchical models considered in Section 5, we investigate the possibility of comparing outcomes across the robust principal-agent model, hierarchical model (i), and hierarchical model (ii), holding fixed the known technology $\mathcal{A}^P$. How does the principal's guarantee change as we move from one organizational structure to another? 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 the principal can benefit by delegating contract-writing to the supervisor, if the supervisor can write a cheaper contract that incentivizes the agent to produce more. So the comparison of the models is not so obvious.

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.)

PROPOSITION S-4: *Given $\mathcal{A}^P$ and a grounded contract $w \in C^+(Y)$, we have $\Phi^{PA}(w) \subseteq \Phi^{PSA(i)}(w)$, and if $\mathcal{S}$ contains all contracts $w_A = \beta w$ with $\beta \in [0, 1]$, then $\Phi^{PSA(i)}(w) \subseteq \Phi^{PSA(ii)}(w)$. These facts imply that*

$$\max_{w \in C^+(Y)} V_P^{PA}(w) \geq \max_{w \in C^+(Y)} V_P^{PSA(i)}(w) \geq \max_{w \in C^+(Y)} V_P^{PSA(ii)}(w).$$

For some technologies $\mathcal{A}^P$, the optimal robust guarantee is the same in all three models, so the bounds in Proposition S-4 are tight. For instance, this happens under any technology $\mathcal{A}^P$ in which the highest-mean-output action actually has cost 0. To obtain more precise comparisons across models for specific $\mathcal{A}^P$, we must apply Theorem 1 and take advantage of the analysis described in Section 7 to solve for optimal robust guarantees.

S-3. CONCAVE CONTRACTS

In this section, we give an example to illustrate how the same methodology we have used to identify organization-free conditions for linear contracts can also be used for other kinds of contracts. This example is inspired by Barron, Georgiadis, and Swinkels (2020). They consider a Bayesian principal-agent model in which the agent can, after privately observing output, costlessly add mean-zero random noise to produce the "final" output, and only final output is contractible. This means that if the agent can produce a distribution F, he can also produce any mean-preserving spread of F at the same effort cost. They show that (weakly) concave contracts are optimal in such a model.

In our general framework, with the worst-case criterion, an assumption in the same spirit is the following weakening of Richness for Φ.

SPREAD-RICHNESS: *Suppose $w \in C^+(Y)$, $F \in \Phi(w)$, and $F' \in \Delta(Y)$ is another distribution such that F' is a mean-preserving spread of F and $\mathbb{E}_{F'}[w(y)] \geq \mathbb{E}_F[w(y)]$. Then $F' \in \Phi(w)$.*

This is weaker than Richness, since the latter requires $F' \in \Phi(w)$ when F' has the same mean output as F (and higher expected payment) even if F' is not a mean-preserving spread of F.

We have not assumed Y is an interval, so let us say that a contract is *concave* if it can be obtained as the restriction to Y of some concave function defined on $\text{co}(Y)$. We then have the following result (whose proof is in Section S-4).

PROPOSITION S-5: *Suppose the correspondence $\Phi(\cdot)$ has the Spread-Richness and Responsiveness properties. Then, for any contract w, there is a concave contract w' such that $V_P(w') \geq V_P(w)$.*

S-4. REMAINING OMITTED PROOFS

Proofs from Section 5.2: Hierarchical Model (i)

PROOF OF LEMMA 6: For any w_A and $\mathcal{A}$, $\Gamma_A(w_A, \mathcal{A})$ is nonempty and, furthermore, it is compact and upper hemicontinuous in w_A, by Berge's theorem. Hence, the set of pairs $\{(w_A, F) \in \mathcal{S} \times \Delta(Y) | F \in \Gamma_A(w_A, \mathcal{A})\}$ 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 $\Gamma_S^i(w, \mathcal{A})$ is nonempty for each $w \in C^+(Y)$ and technology $\mathcal{A}$, which certainly ensures $\Phi^{PSA(i)}(w)$ nonempty. *Q.E.D.*

PROOF OF LEMMA 7: Sufficiency is immediate. For necessity, let $(\mathcal{A}, c, w_A)$ be a PSA(i)-certificate for F under w. Create a new technology $\mathcal{A}' = \mathcal{A} \cup \{(F, 0)\}$ and put $w'_A = w_0$. We check that $(\mathcal{A}', 0, w'_A)$ is also a PSA(i)-certificate. Technology $\mathcal{A}'$ 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, conditions (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, that is, $\mathcal{A} = \mathcal{A}'$, and the contract w_A must have paid 0 for F, therefore also for F' (otherwise

(b) would have been violated for the original certificate). 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. *Q.E.D.*

PROOF OF PROPOSITION 8: *(Richness)* Suppose $w \in C^+(Y)$, $F \in \Phi^{PSA(i)}(w)$, $F' \in \Delta(Y)$ such that $\mathbb{E}_F[y] = \mathbb{E}_{F'}[y]$, $\mathbb{E}_F[w(y)] \leq \mathbb{E}_{F'}[w(y)]$. By Lemma 7, there exists a PSA(i)-certificate for F under w of the form $(\mathcal{A}, 0, w_0)$.

Create a new technology $\mathcal{A}' = \mathcal{A} \cup \{(F', 0)\}$. We argue that $(\mathcal{A}', 0, w_0)$ is a PSA(i)-certificate for F'. In the new technology, the supervisor can induce F' using the zero contract, and this is at least as good for her as inducing F was under $\mathcal{A}$, so condition (a) is satisfied. The agent is willing to take any zero-cost action, so (b) is satisfied. 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 $w_A = w_0$ and is strictly better for the principal; but this means $c'' = 0$, and then

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

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 $w_A = w_0$, contrary to assumption.

Hence, we have a PSA(i)-certificate for F', so $F' \in \Phi^{PSA(i)}(w)$, and Richness holds.

(Responsiveness) Let $w, w' \in C^+(Y)$, $F \notin \Phi^{PSA(i)}(w)$ satisfy the hypotheses of Responsiveness, and suppose, toward a contradiction, that $F \in \Phi^{PSA(i)}(w')$. By Lemma 7, we can find a PSA(i)-certificate for F under w' of the form $(\mathcal{A}, 0, w_0)$. Also, let $\tilde{F}$ be a distribution that could be chosen under $\mathcal{A}$ and w, and let $(\mathcal{A}, \tilde{c}, \tilde{w}_A)$ be a PSA(i)-certificate for $\tilde{F}$ accordingly.

Under w', the supervisor can still induce $(\tilde{F}, \tilde{c})$ via contract $\tilde{w}_A$, and applying the hypotheses of Responsiveness and optimality of inducing $(F, 0)$ under w' and $(\tilde{F}, \tilde{c})$ under w we have

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

Then all the above inequalities are equalities. In particular, $\mathbb{E}_F[w'(y)] = \mathbb{E}_F[w(y)]$ and $\mathbb{E}_{\tilde{F}}[w'(y)] = \mathbb{E}_{\tilde{F}}[w(y)]$. Then, under w, it remains optimal for the supervisor to induce F via the zero contract, and $F \in \Gamma^S_A(w, w_0, \mathcal{A})$. However, $F \notin \Gamma^{PS}_A(w, w_0, \mathcal{A})$, so it must be the case that there is another $F' \in \Gamma^{PS}_A(w, w_0, \mathcal{A})$, which gives both the supervisor and agent the same payoffs as F under (w, w_0), but gives the principal strictly higher payoff. Then, since $F' \in \Phi^{PSA(i)}(w)$, we have $\mathbb{E}_{F'}[w'(y)] \geq \mathbb{E}_{F'}[w(y)]$ by the hypothesis of Responsiveness. This must be an equality, otherwise $(F, 0)$ would not survive supervisor-preferred tie-breaking under w'. Therefore, $F' \in \Gamma^S_A(w', w_0, \mathcal{A})$, and

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

contradicting the assumption that F survived principal-preferred tie-breaking under w'. This checks Responsiveness.

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

(Lower Hemicontinuity) Let $\alpha \in [0,1]$, $F \in \widetilde{\Phi}^{PSA(i)}(\alpha)$. Let $(\mathcal{A}, 0, w_0)$ be a PSA(i)-certificate for F under w_α (using Lemma 7). Since F is principal-preferred, F has the highest mean among zero-cost actions in $\mathcal{A}$. If $F = \delta_{\bar{y}}$, then for any α, the supervisor cannot earn a higher amount than $\alpha\bar{y}$, so $F \in \widetilde{\Phi}^{PSA(i)}(\alpha)$ for all α. So we can assume $F \neq \delta_{\bar{y}}$.

Given a neighborhood $\mathcal{O}$ of F, choose F' as we did in the lower hemicontinuity argument in the robust P-A model, so that $F' \in \mathcal{O}$. 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, $f^*(\alpha') = \max_{w_A \in \mathcal{S}} V_S^i(w_A|w_{\alpha'}, \mathcal{A})$ is also $\bar{y}$-Lipschitz in α', since the max of a collection of k-Lipschitz functions is again k-Lipschitz. In particular, f^* is continuous.

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 $\Gamma_S^i(w_{\alpha'}, \mathcal{A}') = \{F'\}$, and hence $F' \in \widetilde{\Phi}^{PSA(i)}(\alpha') \cap \mathcal{O}$. *Q.E.D.*

Proofs from Section 5.3: Hierarchical Model (ii)

PROOF OF LEMMA 9: We know from the robust principal-agent model that the set of w_A maximizing the supervisor objective is nonempty, since there is a maximizing contract which is linear in the payment from the principal. (More precisely, it is proportional to $w(y) - \min(w)$, since the principal-agent model assumed that minimum output was normalized to zero. There may also exist other optimal w_A not of this form.) $\Gamma_A^{PS}(w, w_A, \mathcal{A})$ is nonempty as in the proof of Lemma 6. Hence, the set $\Gamma_S^u(w, \mathcal{A}^S, \mathcal{A})$ is nonempty for each $w \in C^+(Y)$, and technologies $\mathcal{A}^S$, $\mathcal{A}$, and consequently $\Phi^{PSA(ii)}(w)$ is nonempty. *Q.E.D.*

We next define PSA(ii)-certificates, as described in the main paper. Fix a contract $w \in C^+(Y)$. Let $F \in \Delta(Y)$. Consider a quadruple $(\mathcal{A}, \mathcal{A}^S, c, w_A)$, consisting of technologies $\mathcal{A}, \mathcal{A}^S$ with $\mathcal{A} \supseteq \mathcal{A}^S \supseteq \mathcal{A}^P$, cost $c \geq 0$, and $w_A \in C^+(Y)$. Say that such a quadruple is a *PSA(ii)-certificate for F under w* if it satisfies:

(a) Supervisor maximization: the contract w_A maximizes $V_S^u(\cdot|w, \mathcal{A}^S)$ over $C^+(Y)$;
(b) Agent maximization: given contract w_A, action (F, c) maximizes the agent's payoff over $\mathcal{A}$;
(c) Supervisor-preferred tie-breaking: given w, w_A, action (F, c) maximizes the supervisor's payoff over actions satisfying (b);
(d) Principal-preferred tie-breaking: given w, w_A, action (F, c) maximizes the principal's payoff over actions satisfying (b)–(c).

We have the following analogue of Lemma 7.

LEMMA S-6: *For any w and F, $F \in \Phi^{PSA(ii)}(w)$ if and only if there exists a PSA(ii)-certificate for F under w of the form $(\mathcal{A}, \mathcal{A}^S, 0, w_0)$, with $\mathcal{A}^S = \mathcal{A}$.*

PROOF: Sufficiency is immediate. For necessity, let $(\mathcal{A}, \mathcal{A}^S, c, w_A)$ be a PSA(ii)-certificate for F under w. Create new technologies $\mathcal{A}' = \mathcal{A}^{S'} = \mathcal{A}^S \cup \{(F, 0)\}$; we show that $(\mathcal{A}', \mathcal{A}^{S'}, 0, w_0)$ is again a PSA(ii)-certificate. Contract $w'_A = w_0$ ensures that the supervisor obtains at worst $V_S^u(w'_A|w, \mathcal{A}^{S'}) \geq \mathbb{E}_F[w(y)]$, 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 $\mathcal{A}^S$ or uses action $(F, 0)$); therefore condition (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 $\mathcal{A}^S$ using the zero contract, and (c) would have been violated under $\mathcal{A}$ and $\mathcal{A}^S$. Therefore (c) holds. To check (d), note that if a different action $(F', c') \in \mathcal{A}'$ passes (a)–(c), it must be that $c' = 0$, and $\mathbb{E}_{F'}[w(y)] = \mathbb{E}_F[w(y)]$. However, then $(F', 0) \in \mathcal{A}^S$, so under $\mathcal{A}^S$ 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 the original certificate. Since F further survived (d), we conclude that $\mathbb{E}_{F'}[y] \leq \mathbb{E}_F[y]$, so F survives (d) under the new (claimed) certificate, as needed. *Q.E.D.*

PROOF OF PROPOSITION 10: *(Richness)* Suppose $w \in C^+(Y)$, $F \in \Phi^{PSA(ii)}(w)$, $F' \in \Delta(Y)$ such that $\mathbb{E}_F[y] = \mathbb{E}_{F'}[y]$ and $\mathbb{E}_F[w(y)] \leq \mathbb{E}_{F'}[w(y)]$. By Lemma S-6, we have a PSA(ii)-certificate for F under w of the form $(\mathcal{A}, \mathcal{A}^S, 0, w_0)$ with $\mathcal{A}^S = \mathcal{A}$.

Put $\mathcal{A}' = \mathcal{A}^{S'} = \mathcal{A} \cup \{(F', 0)\}$, and $w'_A = w_0$. We show that $(\mathcal{A}', \mathcal{A}^{S'}, 0, w'_A)$ is a PSA(ii)-certificate for F'. In the new technology, the supervisor can obtain $V^u_S(w'_A | w, \mathcal{A}^{S'}) \geq \mathbb{E}_{F'}[w(y)]$, and no other contract $\tilde{w}_A$ can guarantee better (just consider the worst-case technology $\widetilde{\mathcal{A}} \supseteq \mathcal{A}^S$ under $\tilde{w}_A$; then $\widetilde{\mathcal{A}} \cup \{(F', 0)\}$ is a possible technology containing $\mathcal{A}^{S'}$, and the agent either takes the same action as under $\widetilde{\mathcal{A}}$ or takes action $(F', 0)$). So, condition (a) is satisfied. The agent is willing to take any zero-cost action, so (b) is satisfied. Since $\mathbb{E}_F[w(y)] \leq \mathbb{E}_{F'}[w(y)]$ and (c) held for the original certificate for F, (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

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

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 tie-breaking is satisfied for F'' under $\mathcal{A}'$; the third is because supervisor tie-breaking 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 the original certificate, contrary to assumption.

Hence, we have a PSA(ii)-certificate for F', and $F' \in \Phi^{PSA(ii)}(w)$.

(Responsiveness) Let $w, w' \in C^+(Y)$ and $F \notin \Phi^{PSA(ii)}(w)$ satisfy the hypotheses of Responsiveness, and suppose to the contrary that $F \in \Phi^{PSA(ii)}(w')$. By Lemma S-6, there exists a PSA(ii)-certificate for F under w' of the form $(\mathcal{A}, \mathcal{A}^S, 0, w_0)$ with $\mathcal{A}^S = \mathcal{A}$. Also, let $\tilde{w}_A$ be an S-A contract that maximizes $V^u_S(\cdot | w, \mathcal{A}^S)$.

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

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

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 $\widehat{F} \in \Phi^{PSA(ii)}(w)$. Taking infimum over technologies $\widehat{\mathcal{A}}$,

$$V^u_S(\tilde{w}_A | w', \mathcal{A}^S) \geq V^u_S(\tilde{w}_A | w, \mathcal{A}^S). \tag{S-1}$$

Now the following string of inequalities holds:

$$V^u_S(w_0 | w', \mathcal{A}^S) \geq V^u_S(\tilde{w}_A | w', \mathcal{A}^S) \geq V^u_S(\tilde{w}_A | w, \mathcal{A}^S) \geq V^u_S(w_0 | w, \mathcal{A}^S) \geq V^u_S(w_0 | w', \mathcal{A}^S)$$

and is a cycle, so all of the inequalities are equalities. The first inequality is the optimality of $w'_A = w_0$ under w', the second is from (S-1), the third is optimality of $\tilde{w}_A$ under w, and the fourth is from the hypothesis of Responsiveness $\mathbb{E}_F[w(y)] \geq \mathbb{E}_F[w'(y)]$ and $F \in \Gamma^S_A(w', w_0, \mathcal{A}^S)$, which together ensure the supervisor is assured at least as high a payoff by incentivizing $(F, 0)$ using $w'_A = w_0$ under w as she would be under w'.

This cycle of equalities means that it is also optimal for the supervisor to give contract $w'_A = w_0$ under w and $\mathcal{A}^S$, and $F \in \Gamma^S_A(w, w_0, \mathcal{A})$. By assumption, then F is defeated in principal-preferred tie-breaking, so there is $\tilde{F} \in \Gamma^{PS}_A(w, w_0, \mathcal{A})$, $\tilde{F} \neq F$, so that $\tilde{F} \in \Phi^{PSA(ii)}(w)$. The hypothesis of Responsiveness again says that $\mathbb{E}_{\tilde{F}}[w'(y)] \geq \mathbb{E}_{\tilde{F}}[w(y)]$ and $\mathbb{E}_F[w(y)] \geq \mathbb{E}_F[w'(y)]$, but it must be that $\mathbb{E}_F[w(y)] = \mathbb{E}_{\tilde{F}}[w(y)]$. Combining gives $\mathbb{E}_{\tilde{F}}[w'(y)] \geq \mathbb{E}_F[w'(y)]$, but we must have equality since otherwise $F \notin \Gamma^S_A(w', w_0, \mathcal{A})$. Then $\tilde{F} \in \Gamma^S_A(w', w_0, \mathcal{A})$, and

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

contradicting the assumption $F \in \Gamma^{PS}_A(w', w_0, \mathcal{A})$.

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

(Lower Hemicontinuity) Let $\alpha \in [0, 1]$, $F \in \widetilde{\Phi}^{PSA(ii)}(\alpha)$. Using Lemma S-6, let $(\mathcal{A}, \mathcal{A}^S = \mathcal{A}, 0, w_0)$ be a PSA(ii)-certificate for F under w_α. As in hierarchical model (i), if $F = \delta_{\bar{y}}$, 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 $F \in \Gamma^u_S(w_{\alpha'}, \mathcal{A}^S, \mathcal{A})$ for any α' in this neighborhood. So we can assume $F \neq \delta_{\bar{y}}$.

Given a neighborhood $\mathcal{O}$ of F, let $F' \in \mathcal{O}$ be the distribution constructed in the proof of lower hemicontinuity in Proposition 5. For any α', define $f^*(\alpha') = \max_{w_A \in C^+(Y)} V^u_S(w_A | w_{\alpha'}, \mathcal{A}^S) = \max_{w_A \in C^+(Y)}(\inf_{\mathcal{A}' \supseteq \mathcal{A}^S} V^i_S(w_A | w_{\alpha'}, \mathcal{A}'))$. As noted in the proof of Proposition 8, V^i_S is $\bar{y}$-Lipschitz continuous with respect to α'; hence, $\inf_{\mathcal{A}' \supseteq \mathcal{A}^S} V^i_S(w_A | w_{\alpha'}, \mathcal{A}')$ is $\bar{y}$-Lipschitz with respect to α', since the supremum or infimum of an arbitrary collection of k-Lipschitz functions is again k-Lipschitz for any positive k. Thus $f^*(\alpha')$ is continuous as well. Now steps analogous to the proof of Proposition 5 show that there is $\eta > 0$ such that, when $\alpha' \in \mathcal{B}_\eta(\alpha) \setminus \{0\}$, $f^*(\alpha') < \alpha' \mathbb{E}_{F'}[y]$. Hence, constructing $\mathcal{A}^{S'} = \mathcal{A}' = \mathcal{A} \cup \{(F', 0)\}$ yields $\Gamma^u_S(w_{\alpha'}, \mathcal{A}^{S'}, \mathcal{A}') = \{F'\}$, and hence $F' \in \widetilde{\Phi}^{PSA(ii)}(\alpha') \cap \mathcal{O}$. *Q.E.D.*

Proofs from Section 6: An Example Where Responsiveness Fails

PROOF OF PROPOSITION 11: We can assume the contract w is grounded. Suppose $F \in \Phi^{PSA(iii)}(w)$, so there is some $\beta \in [0, 1]$ for which βw is optimal for the supervisor, and some $\mathcal{A}$ such that $F \in \Gamma^S_A(w, \beta w, \mathcal{A})$. 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 $F' \in \Gamma^S_A(w, \beta w, \mathcal{A}')$. Consequently, $F' \in \Phi^{PSA(iii)}(w)$ as needed. *Q.E.D.*

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

LEMMA S-7: *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. *Q.E.D.*

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: $\Gamma_A^S(w, \beta w, \mathcal{A}) = \Gamma_A(\beta w, \mathcal{A})$.

Now to check the characterization in (2). If the agent chooses (F', c') under technology $\mathcal{A}$, then

$$\mathbb{E}_{F'}\big[\beta w(y)\big] \geq \mathbb{E}_{F'}\big[\beta w(y)\big] - c' \geq \mathbb{E}_F\big[\beta w(y)\big] - c,$$

and dividing through by β yields (2). Conversely, suppose (2) is satisfied. Note that the targeted action (F, c) is indeed optimal for the agent among actions in $\mathcal{A}^P$, since $(\tilde{F}, \tilde{c}) \in \mathcal{A}^P$ 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 S-7. In turn, for any F' that satisfies (2), action $(F', 0)$ (if it is available) is at least as good for the agent as (F, c), and so under technology $\mathcal{A} = \mathcal{A}^P \cup \{(F', 0)\}$, this action becomes optimal for the agent, that is, $F' \in \Gamma_A^S(w, \beta w, \mathcal{A})$. *Q.E.D.*

Proofs of Results in Supplementary Material

PROOF OF PROPOSITION S-1: *(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 $\mathcal{A}_1 \supseteq \mathcal{A}_1^P$ and $\mathcal{A}_2 \supseteq \mathcal{A}_2^P$, 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 $w_{A1} = w_{A2} = w_0$ (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

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

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

$$\begin{aligned}
&\mathbb{E}_{F'}\big[w(y)\big] \\
&\quad < \mathbb{E}_{H(\sigma_1',\sigma_2')}\big[w(y)-w_{A1}'(y)-w_{A2}'(y)\big] \\
&\quad = \pi\mathbb{E}_{F'}\big[w(y)-w_{A1}'(y)-w_{A2}'(y)\big]+(1-\pi)\mathbb{E}_{H(\sigma_1',\sigma_2'')}\big[w(y)-w_{A1}'(y)-w_{A2}'(y)\big].
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}_{F'}\big[w(y)\big] &< \mathbb{E}_{H(\sigma_1',\sigma_2'')}\big[w(y)-w_{A1}'(y)-w_{A2}'(y)\big] \\
&\le \mathbb{E}_{H(\sigma_1',\sigma_2'')}\big[w(y)-(1-\pi)w_{A1}'(y)-w_{A2}'(y)\big],
\end{aligned}$$

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 $\mathbb{E}_{F'}[w(y)]\ge\mathbb{E}_F[w(y)]$, 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 $\mathcal{A}_1\supseteq\mathcal{A}_1^P$ and $\mathcal{A}_2\supseteq\mathcal{A}_2^P$ 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 $F'\in\Gamma_A^S(w,w_{A1},w_{A2},\mathcal{A}_1,\mathcal{A}_2)$. Since $F\notin\Phi^{ST}(w)$, under $\mathcal{A}_1$ and $\mathcal{A}_2$, either (a) there do 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, $F\notin\Gamma_S^i(w',\mathcal{A}_1,\mathcal{A}_2)$. Suppose (b) holds under w. Swapping w for w', observe that

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

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 $F\notin\Gamma_S^i(w',\mathcal{A}_1,\mathcal{A}_2)$. Hence, $F\notin\bigcup_{\mathcal{A}_1,\mathcal{A}_2}\Gamma_S^i(w',\mathcal{A}_1,\mathcal{A}_2)=\Phi^{ST}(w')$, and Responsiveness holds. *Q.E.D.*

PROOF OF LEMMA S-2: 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,

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

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 $(\mathcal{A}, H, c) \supseteq (\mathcal{A}^P, H^P, c^P)$ 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. (We abuse notation slightly by writing I for the set of agents as well as the number of agents, and likewise for subsets $J \subseteq I$.)

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 $a_{-J} \in \mathcal{A}_{-J}^P$, let $p^J(a_{-J}) = \max_{a_J\in\mathcal{A}_J^P}\{\mathbb{E}_{H(a_J,a_{-J})}[w(y)] - I\sum_{j\in J} c_j(a_j)\}$. Note that $p^J(a_{-J}) < \bar{w}$, since the second term inside the max operator is strictly positive for all $a_J \in \mathcal{A}_J^P$. We also observe the recursive relationship $p^{J+i}(a_{-(J+i)}) = \max_{a_i\in\mathcal{A}_i^P}[p^J(a_i, a_{-(J+i)}) - Ic_i(a_i)]$ 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 $a_i \in \mathcal{A}_i^P$, and also $\varepsilon_I < \underline{w} - w^0$.

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

$$\mathbb{E}_{H(a_J',a_{-J})}\big[w(y)\big] = \max\big\{\underline{w}, p^J(a_{-J})\big\} + \varepsilon_J. \tag{S-2}$$

The assumptions on the ε's ensure that the right side of (S-2) always lies in the interval $[\underline{w}, \bar{w}]$, so that the desired distribution indeed exists. Notice also that (S-2) 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 $a_i \in \mathcal{A}_i^P$. If $a \in \mathcal{A}^P$, we have

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

$$> \mathbb{E}_{H(a_i,a_{-i})}\big[w(y)/I\big] - c_i(a_i),$$

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

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

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 (S-2) 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

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

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 $(\mathcal{A}, H, c) \supseteq (\mathcal{A}^P, H^P, c^P)$ 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 $F \in \Phi^{UT}(w)$. *Q.E.D.*

PROOF OF PROPOSITION S-3: *(Richness)* This property follows directly from Lemma S-2.

(Responsiveness) Assume the conditions of Responsiveness on $w, w' \in C^+(Y)$ and F, with $F \notin \Phi^{UT}(w)$. By Lemma S-2, $\mathbb{E}_F[w(y)] \leq w^0$. Let a^0 be the maximizer of P (the potential when contract w is given) over $\mathcal{A}^P$, and let $F' = H^P(a^0)$, and let $C = I\sum_i c_i^P(a_i^0)$, 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 S-2. Let P' be the potential under contract w', and w'^0 its corresponding maximum on $\mathcal{A}^P$. Then

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

Again applying Lemma S-2, $F \notin \Phi^{UT}(w')$.

With Richness and Responsiveness proven, applying Theorem 1 yields the result. *Q.E.D.*

PROOF OF PROPOSITION S-4: ($\Phi^{PA}(w) \subseteq \Phi^{PSA(i)}(w)$) Consider $F \in \Phi^{PA}(w)$. There exists some technology $\mathcal{A} \supseteq \mathcal{A}^P$ 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 $F \in \Gamma^i_S(w, \mathcal{A})$. When the supervisor offers $w_A = w_0$, $F \in \Gamma_A(w_A, \mathcal{A})$, 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

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

and, therefore, $w_A = w_0$ is a maximizer of the supervisor's payoff, and $F \in \Gamma^S_A(w, w_0, \mathcal{A})$. 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 $\Gamma^S_A(w, w_0, \mathcal{A})$. Then $F \in \Gamma^{PS}_A(w, w_0, \mathcal{A})$ follows from F being principal-preferred in the P-A model, so $F \in \Phi^{PSA(i)}(w)$.

($\Phi^{PSA(i)}(w) \subseteq \Phi^{PSA(ii)}(w)$) Consider $F \in \Phi^{PSA(i)}(w)$. From Lemma 7, there exists a PSA(i)-certificate $(\mathcal{A}, 0, w_0)$ for F under w. Let $\mathcal{A}^S = \mathcal{A}$. The certificate implies that $\mathbb{E}_F[w(y)] \geq \mathbb{E}_{\Gamma^S_A(w,w_A,\mathcal{A})}[w(y) - w_A(y)]$ for every $w_A \in \mathcal{S}$. We want to show that $F \in \Gamma^u_S(w, \mathcal{A}^S, \mathcal{A})$ for hierarchical model (ii). We know

$$\inf_{\tilde{\mathcal{A}} \supseteq \mathcal{A}^S} \mathbb{E}_{\Gamma^S_A(w,w_0,\tilde{\mathcal{A}})}\big[w(y)\big] = \mathbb{E}_F\big[w(y)\big].$$

It is without loss of generality to consider only contracts w_A of the form βw for $\beta \in [0, 1]$, since Proposition 5 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 $w_A = \beta w \in \mathcal{S}$,

$$\begin{aligned} V^u_S(w_0 | w, \mathcal{A}^S) &= \mathbb{E}_F\big[w(y)\big] \geq \mathbb{E}_{\Gamma^S_A(w,w_A,\mathcal{A})}\big[w(y) - w_A(y)\big] \\ &\geq \inf_{\tilde{\mathcal{A}} \supseteq \mathcal{A}^S} \mathbb{E}_{\Gamma^S_A(w,w_A,\tilde{\mathcal{A}})}\big[w(y) - w_A(y)\big] = V^u_S(w_A | w, \mathcal{A}^S), \end{aligned}$$

where the first inequality is by the PSA(i)-certificate for F and the second by definition of infimum. Hence, $w_A = w_0$ is a maximizer of $V^u_S(\cdot | w, \mathcal{A}^S)$, and $F \in \Gamma^S_A(w, w_0, \mathcal{A})$. Moreover, since $\Gamma^S_A(w, w_0, \mathcal{A})$ is the same in both model (i) and (ii), if F survives principal-preferred tie-breaking within this set in model (i) then it also survives the tie-breaking in model (ii). Thus $F \in \Phi^{PSA(ii)}(w)$.

For the last statement of the proposition: Whenever w is linear, our requirement on $\mathcal{S}$ is satisfied, and so we have shown that $\Phi^{PA}(w) \subseteq \Phi^{PSA(i)}(w) \subseteq \Phi^{PSA(ii)}(w)$. So, taking the infima over the respective sets, $V^{PA}_P(w) \geq V^{PSA(i)}_P(w) \geq V^{PSA(ii)}_P(w)$. Taking maxima over w, and noting that each maximum is attained for a linear w by our earlier results, completes the proof. *Q.E.D.*

PROOF OF PROPOSITION S-5: Let $w' : \text{co}(Y) \to \mathbb{R}$ be the concavification of w. We can view w' as a contract as well (by restricting to Y). For any $y \in Y$, we can define the points $l(y), u(y)$ (for "lower" and "upper") as the endpoints of the relevant interval of concavification (which may be degenerate); thus $l(y), u(y) \in Y$ with $l(y) \leq y \leq u(y)$; w' is an affine function on $[l(y), u(y)]$; and w' coincides with w at points $l(y)$ and $u(y)$.

Take any $F \in \Phi(w')$. Randomly generate an output level $y' \in Y$ as follows: draw $y \sim F$ and set $y' = l(y), u(y)$ with respective probabilities $(u(y) - y)/(u(y) - l(y))$ and

$(y - l(y))/(u(y) - l(y))$. (If $u(y) = l(y) = y$, then just put $y' = y$.) Let $F' \in \Delta(Y)$ be the resulting distribution of y'. Since $\mathbb{E}[y'|y] = y$, F' is a mean-preserving spread of F. Since w' is affine on each interval $[l(y), u(y)]$, we also have $\mathbb{E}[w'(y')|y] = w'(y)$; hence, the expected value of w' under F' is the same as under F. Spread-Richness then implies that $F' \in \Phi(w')$ also.

Every distribution over output generates at least as high an expected payment under w' as under w (since $w' \geq w$ pointwise); but equality holds for F', since by construction, F' is supported on points where w' and w coincide. Then, if $F' \notin \Phi(w)$ then Responsiveness would imply $F' \notin \Phi(w')$, contradicting the previous paragraph. So, $F' \in \Phi(w)$. This implies

$$V_P(w) \leq \mathbb{E}_{F'}\big[y - w(y)\big] = \mathbb{E}_{F'}\big[y - w'(y)\big] = \mathbb{E}_F\big[y - w'(y)\big].$$

Since $F \in \Phi(w')$ was arbitrary, we conclude $V_P(w) \leq V_P(w')$. *Q.E.D.*

REFERENCES

ALIPRANTIS, CHARALAMBOS D., AND KIM C. BORDER (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide* (Third Ed.). Springer. [1]

BARRON, DANIEL, GEORGE GEORGIADIS, AND JEROEN SWINKELS (2020): "Optimal Contracts With a Risk-Taking Agent," *Theoretical Economics*, 15, 715–761. [6]

DAI, TIANJIAO, AND JUUSO TOIKKA (2022): "Robust Incentives for Teams," *Econometrica* 90 (4), 1583–1613. [1,3,4]

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. [2]

Co-editor Joel Sobel handled this manuscript.

Manuscript received 10 June, 2019; final version accepted 22 October, 2021; available online 24 February, 2022.