Theoretical Economics 19 (2024), 1151–1184 1555-7561/20241151

Robust performance evaluation of independent agents

Ashwin Kambhampati
Department of Economics, United States Naval Academy

A principal provides incentives for independent agents. The principal cannot observe the agents' actions, nor does she know the entire set of actions available to them. It is shown that an anti-informativeness principle holds: very generally, robustly optimal contracts must link the incentive pay of the agents. In symmetric and binary environments, they must exhibit *joint performance evaluation*—each agent's pay is increasing in the performance of the other.

Keywords. Moral hazard, robustness, teams.

JEL classification. D81, D82, D86.

1. Introduction

Consider a group of economic agents, each of whose individual performance is observable. An agent's contract exhibits *independent performance evaluation* if his pay depends solely on his individual performance. Alternatively, it exhibits *joint performance evaluation* if it is increasing in others' performances and does *not* depend solely on his individual performance. When does joint performance evaluation outperform individual performance evaluation? Conventional economic wisdom builds upon the "Informativeness Principle," which holds that only signals that are statistically informative about a targeted action are valuable to incentivize that action (Holmström (1979), Shavell (1979)). Hence, joint performance evaluation outperforms independent performance evaluation only when others' success is indicative that the agent took the targeted action (Holmström (1982)).

In practice, however, joint performance evaluation appears in settings in which statistical considerations suggest it is suboptimal. For instance, Rees, Zax, and Herries (2003) examined the monthly sales records of 3795 salespeople working in a single company. Despite working alone, 3287 salespeople received contracts exhibiting joint performance evaluation: the performance-based component of monthly income was determined by a weighted average of an increasing function of individual monthly sales

Ashwin Kambhampati: kambhamp@usna.edu
Earlier versions have circulated under the titles "Robust Performance Evaluation" and "Robust Performance Evaluation of Independent and Identical Agents." An earlier draft appears as the first chapter of my dissertation at the University of Pennsylvania and an abstract appears in *EC'21: Proceedings of the 22nd ACM Conference on Economics and Computation*. I thank Nageeb Ali, Aislinn Bohren, Gabriel Carroll, George Mailath, Steven Matthews, Juuso Toikka, and Rakesh Vohra for their time, guidance, and encouragement. I thank two anonymous referees and several seminar and conference participants for helpful comments and suggestions. The views expressed here are solely my own and do not in any way represent the views of the U.S. Naval Academy, U.S. Navy, or the Department of Defense.

© 2024 The Author. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at https://econtheory.org. https://doi.org/10.3982/TE5523

and an increasing function of group sales.[1] In this setting, the canonical, Bayesian principal-agent model predicts that independent performance evaluation is optimal if there is no correlation in productivity across agents conditional on their actions. Otherwise, *relative performance evaluation* is typically optimal. Specifically, if salespeople are subject to common productivity shocks, then it is sensible to reward success in the face of others' failure more than success when others succeed because success in the latter scenario is a stronger indicator of effort.

This paper provides non-Bayesian foundations for joint performance evaluation of independent agents. In the model, a risk-neutral principal compensates a finite number of risk-neutral agents, each of whom is protected by limited liability. Each agent takes a hidden action to stochastically produce observable individual output. There is no correlation in output conditional on agents' actions and no agent can directly affect the output of any of the others.

While the principal knows some actions the agents can take, there may be others she does not know about. For instance, in a sales context, group managers may know some tactics of their sales representatives—representatives can always follow the company's script. But there are a myriad of less costly (but potentially less productive) ways in which a sales representative might deviate from this script. Following Carroll (2015), the principal evaluates each contract according to its expected payoff when the "worst" possible set of action profiles is available to the agents.

The first main result is an anti-informativeness principle: if a contract maximizes the principal's worst-case payoff, then under a wide range of assumptions about agents' behavior, there exists a joint performance evaluation contract that strictly outperforms any independent performance evaluation contract (Theorem 1). The intuition can be understood through a simple example. Suppose there are two agents known to possess a common action set and output is binary (success or failure).[2] Consider first the case in which each receives an independent performance evaluation contract that specifies a positive wage for individual success that does not depend on the other's performance. Then the principal's worst-case payoff is attained when each has available an unknown and costless "shirking" action that is just productive enough to encourage them to deviate from a targeted known action. Now, suppose each agent's wage for individual success is made contingent upon the other's success in a manner consistent with joint performance evaluation. Specifically, reduce each agent's wage when the other fails and increase it when the other succeeds, so that when the other takes their targeted action,

[1]Equation (2) in Rees, Zax, and Herries (2003) shows that the component of monthly income based on performance is determined by the equation

$$0.7 \times f\left(\frac{R_I}{T_I}\right) + 0.3 \times g\left(\frac{R_G}{T_G}\right),$$

where R_I and R_G represent individual and group revenue, and T_I and T_G represent target individual and group revenue. The functions f and g are known to be increasing, nonlinear, and discontinuous with $f(0) = g(0) = 0$. Other properties of f and g are confidential.

[2]Though the symmetry assumption facilitates understanding of the key intuition, actions are not assumed to be common across agents in the baseline model.

expected wages are held constant. Then incentives to shirk are also held constant. Nevertheless, when both agents shirk—the "bad" state of the world that matters under robustness considerations—the principal pays each strictly less than under independent performance evaluation (see Example 1).

The second main result concerns a partial implementation setting in which there are two agents with the same set of actions and two output levels. It is shown that any contract that maximizes the principal's worst-case payoff within the class of symmetric contracts exhibits joint performance evaluation (Theorem 2). As an intermediate step in the proof, it is shown that there does not exist a symmetric relative performance evaluation contract that yields the principal a strictly larger payoff than the optimal independent performance evaluation contract (Lemma 2). In contrast to joint performance evaluation contracts, these contracts increase expected wage payments when agents shirk, eliminating any of their incentive advantages (see Example 2).

It is worth emphasizing that the primary purpose of the independence assumption, while a reasonable approximation of the production environment of many economic agents, is to rule out all known mechanisms leading to the superiority of joint performance evaluation over independent performance evaluation (see Section 1.1 for a detailed discussion). Hence, Theorem 1 and Theorem 2 highlight a rent-extraction benefit of joint performance evaluation that had previously gone unnoticed and which may be of relevance in applications. For instance, in the sales application of Rees, Zax, and Herries (2003), a manager might use joint performance evaluation to hedge against scenarios in which her subordinates discover sufficiently effective shirking tactics. In such applications, however, other advantages of joint performance evaluation might also be present. For instance, Rees, Zax, and Herries (2003) cannot rule out that peer pressure and other-regarding preferences are responsible for the productivity interdependence they observe. Allowing for such channels in the model would presumably reinforce the superiority of joint performance evaluation over independent performance evaluation.

1.1 *Related literature*

This paper makes two main contributions to the theoretical literature. First, it establishes a fundamentally new justification for joint performance evaluation. In the Bayesian contracting paradigm, the Informativeness Principle prescribes independent performance evaluation whenever each agent's performance is statistically uninformative of other agents' actions. To justify joint performance evaluation, the literature has instead introduced informational and productive linkages across agents. In the absence of productive interaction, joint performance evaluation may be optimal if performances are negatively correlated (Fleckinger (2012)). In the absence of correlation, joint performance evaluation may be optimal if efforts are complements in production (Alchian and Demsetz (1972)), if it induces help among agents (Itoh (1991)), or alternatively, if it discourages sabotage (Lazear (1989)). Finally, joint performance evaluation may be optimal if agents are engaged in repeated production and it allows for more effective peer sanctioning (Che and Yoo (2001)). The model studied in this paper explicitly rules out these channels.

Second, this paper contributes to a growing literature on the design of contracts that are robust to unknown preferences or technologies. The use of "maxmin" criteria to identify robustly optimal contracts has a long history dating back to at least Hurwicz and Shapiro (1978), who study the design of single-agent contracts in a setting in which the agent's cost of effort is unknown and the principal minimizes her worst-case "regret" over all possible cost realizations.[3] Within the literature, the modeling framework of this paper is closest to the pioneering work of Carroll (2015), who considers a principal-single agent model in which the principal has nonquantifiable uncertainty about the entire set of actions of the agent. His main result is that there exists a robustly optimal contract that is linear in individual output. The model and analysis in this paper enrich that of Carroll (2015) by introducing seemingly irrelevant agents and showing that multiple agents lead to the optimality of team-based incentive pay.[4]

Dai and Toikka (2022) also extend the analysis of Carroll (2015) to multiagent settings, but consider a model in which the principal deems *any* game the agents might be playing plausible. They find that contracts that are linear in team output are worst-case optimal under partial Nash implementation. This result is driven by the finding that any contract that induces competition among agents is nonrobust to games in which one agent's action can directly influence the productivity of another. In contrast to Dai and Toikka (2022), this paper considers a setting in which the principal *knows* that output is independently distributed across agents. This has the immediate effect of ruling out such games and ensuring that contracts that are linear in team output are suboptimal.[5] Despite these differences, the results and management implications of this paper complement those of Dai and Toikka (2022). Agents in Dai and Toikka (2022)'s model are a "real team" in the sense that they work together to produce value for the principal, while agents in the model of this paper are best thought of as "co-actors" given the assumption of technological independence (Hackman (2002)). Yet, in either case, joint performance evaluation is optimal. What changes is the particular form of the optimal joint performance evaluation contract—in the case of a real team, optimal compensation is always linear in the value the team generates for the principal, while in the case of co-acting agents it is always nonlinear in team output and may involve bonus payments that reward each agent for others' successes.

[3]For more recent work that considers the design of optimal contracts under unknown preferences, see Garrett (2014) and Frankel (2014). Garrett (2014) studies a cost-based procurement model in which the principal has maxmin uncertainty about the agent's cost of procurement. Frankel (2014) considers a delegation problem in which the principal has maxmin uncertainty about the bias of the agent to whom multiple decisions are delegated.

[4]Antic (2021) imposes bounds on the principal's uncertainty over the productivity of unknown actions (see also Section 3.1 of Carroll (2015), which studies lower bounds on costs). In contrast, the model studied here places no restrictions on the technology available to each agent in isolation beyond those of Carroll (2015). Instead, the restrictions concern the relationship among the agents. Rosenthal (2023) extends the original Carroll (2015) model by considering the robust design of single-agent contracts when the principal is both uncertain about the agent's technology and his risk preferences. Marku, Ocampo Diaz, and Tondji (2024) enrich the model by allowing for multiple principals with conflicting preferences over the agent's action. Finally, Chassang (2013) considers a dynamic agency problem in which the principal is uncertain about the stochastic process of returns generated by the agent's actions; he derives the same lower bound on the performance of linear contracts as Carroll (2015).

[5]See Lemma 3 of Kambhampati (2024) for a proof that can be easily adapted to the current setting.

2. Model

2.1 *Environment*

A risk-neutral principal writes a contract for risk-neutral agents, indexed by $i = 1, 2, \dots, n$. Agent i's output, y_i, is observable and belongs to a compact set $Y \subset \mathbb{R}_+$, where $\max(Y) > \min(Y) = 0$. To produce output, agent i chooses an unobservable action, a_i, from a finite set $A_i \subset \mathbb{R}_+ \times \Delta(Y)$, where $\Delta(Y)$ is the set of Borel distributions on Y. Each action a_i is thus identified by an effort cost, $c(a_i) \in \mathbb{R}_+$, and a distribution over output, $F(a_i) \in \Delta(Y)$. Agents are assumed to be independent—there are no informational or productive linkages across agents. Formally, the joint distribution over output vectors induced by any action profile is the product of the marginal distributions over individual outputs,

$$F(a) := F(a_1) \times \cdots \times F(a_n) \in \Delta\big(Y^n\big) \quad \text{for all } a \in A := A_1 \times \cdots \times A_n.$$

2.2 *Contracts*

A *contract* is a function for each agent i,

$$w_i : Y^n \to \mathbb{R}_+,$$

where the nonnegativity restriction in the codomain reflects agent limited liability (no agent can receive negative wages). Direct (revelation) mechanisms[6] and random mechanisms[7] are thus ruled out by assumption. It will be useful to classify contracts according to an extension of the typology of Che and Yoo (2001), who consider binary performance evaluations.

Definition 1 (Performance evaluations). A contract $w = (w_i)_i$ is an

- *independent performance evaluation (IPE)* if, for all i and y_i, $w_i(y_i, y_{-i})$ is constant in y_{-i};
- a *relative performance evaluation (RPE)* if it does not exhibit IPE, and for all i and y_i, $w_i(y_i, y_{-i})$ is decreasing in y_{-i};
- and a *joint performance evaluation (JPE)* if it does not exhibit IPE, and for all i and y_i, $w_i(y_i, y_{-i})$ is increasing in y_{-i}.[8]

[6]It is well known that the principal can partially implement the Bayesian optimal contract technology-by-technology using a revelation mechanism: she can ask agents to report the action set, and if reports disagree, punish them with a contract that always pays zero. The interpretation taken in this paper, however, and in the rest of the literature on robust contracting is that such a mechanism violates the spirit of the robustness exercise. The principal would like to avoid changing the contract she offers as the agents' environment varies. The performance of alternative indirect mechanisms, such as offering a menu of contracts, awaits further study.

[7]Randomizing over contracts is not helpful if the principal believes that Nature selects the agents' action set after her contract is realized. However, if Nature moves simultaneously, then there is scope for randomization to improve the principal's payoff. See Kambhampati (2023) for an analysis of the single-agent case.

[8]Output vectors are equipped with the usual partial order: $y' \geq y$ if y' is weakly larger than y in all components. So, a function of output vectors, f, is increasing if $y' \geq y$ implies $f(y') \geq f(y)$.

2.3 *Payoffs*

Agent i's ex post payoff given a contract w, action profile a, and output vector y is

$$w_i(y) - c(a_i),$$

while his expected payoff is

$$U_i(a; w) := \mathbb{E}_{F(a)}\big[w_i(y)\big] - c(a_i).$$

The principal's ex post payoff given a contract w and output vector y is

$$\sum_{i=1}^{n}\big(y_i - w_i(y)\big).$$

Given a contract w and action set A, the first part of the analysis assumes only that the principal believes that the agents' behavior is consistent with common knowledge of rationality. That is, she assumes play of a (correlated) rationalizable action profile.[9] Let $\mathcal{R}(w, A) \subseteq \Delta(A)$ be the set of Borel distributions over rationalizable action profiles in the game $\Gamma(w, A)$. Then the nonempty set of expected payoffs obtainable under some distribution over rationalizable action profiles is

$$V(w, A) := \left\{\mathbb{E}_\sigma\left[\sum_{i=1}^{n}\big(y_i - w_i(y)\big)\right] : \sigma \in \mathcal{R}(w, A)\right\}.$$

2.4 *Uncertainty*

When the principal writes a contract for the agents, she has limited knowledge about the action set available to each agent. In particular, she knows only a nonempty subset of actions available to each agent $A_i^0 \subseteq A_i$. To rule out uninteresting cases, it is assumed that, for each agent i, there exists an action $a_i^0 \in A_i^0$ such that $\mathbb{E}_{F(a_i^0)}[y_i] - c(a_i^0) > 0$. This ensures that the principal obtains a strictly positive payoff from contracting with agent i. In addition, it is assumed that if $a_i^0 \in A_i^0$, then $c(a_i^0) > 0$. This ensures that the principal's optimal independent performance evaluation contract for agent i is different from one that always pays zero. In the face of her uncertainty, the principal evaluates each contract on the basis of its performance across all finite supersets of her knowledge, collected in the feasible set of uncertainty $\mathcal{A} := \{A \subset (\mathbb{R}_+ \times \Delta(Y))^n : A_i \supseteq A_i^0 \text{ and } |A| < \infty\}$.

3. Anti-informativeness principle

Let

$$v_{\text{IPE}} := \max_{w:\, w \text{ is an IPE}} \inf_{A \in \mathcal{A}} \max\big(V(w, A)\big) > 0$$

[9] Correlated rationalizable action profiles are those obtained by iterated elimination of strictly dominated actions; see Brandenburger and Dekel (1987).

be the principal's highest payoff obtainable under rationalizable behavior when restricted to use IPE contracts.[10] The first main result is that there exists a JPE contract whose rationalizable payoffs robustly dominate those obtained under any IPE contract.

THEOREM 1 (Anti-informativeness principle). *For each agent i, there exists a base share, $\phi_i > 0$, and bonus factor, $\beta_i > 0$, such that the JPE contract, w^{JPE}, with*

$$w_i^{\mathrm{JPE}}(y_i, y_{-i}) = \left(\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n} y_j\right) y_i \quad \textit{for } i = 1, \dots, n$$

has rationalizable payoffs that robustly dominate those obtained under any IPE:

$$\inf_{A\in\mathcal{A}} \max\big(V\big(w^{\mathrm{JPE}}, A\big)\big) > \inf_{A\in\mathcal{A}} \min\big(V\big(w^{\mathrm{JPE}}, A\big)\big) = v_{\mathrm{IPE}}.$$

PROOF. The proof is in Appendix A.1. □

The JPE contract considered in the statement of Theorem 1, w^{JPE}, induces a supermodular game among the agents under an appropriately defined order on action profiles (see Observation 2 in Appendix A.1). Specifically, when other agents choose actions with higher expected output, the marginal benefit of choosing an action with higher expected output increases. Hence, by standard results in the literature on supermodular games (see, e.g., Vives (1990) and Milgrom and Roberts (1990)), there is a maximal and minimal rationalizable action profile and each profile is a Nash equilibrium. When the parameters $(\phi_i, \beta_i)_i$ are chosen so that the principal's payoff is increasing in the expected output of each agent, it follows that the principal's rationalizable payoffs are an interval,[11]

$$I\big(w^{\mathrm{JPE}}, A\big) = \big[\min\big(V\big(w^{\mathrm{JPE}}, A\big)\big), \max\big(V\big(w^{\mathrm{JPE}}, A\big)\big)\big].$$

Theorem 1 establishes that, for any action set $A \supseteq A^0$, all payoffs in $I(w^{\mathrm{JPE}}, A)$ are weakly larger than v_{IPE} and there are a continuum of payoffs strictly larger than v_{IPE}. Formally, the set of worst-case rationalizable payoffs dominate $\{v_{\mathrm{IPE}}\}$ in the interval order $\succeq_I$: for closed intervals $X \subseteq \mathbb{R}$ and $Y \subseteq \mathbb{R}$, $X \succeq_I Y$ if $x \in X$ and $y \in Y$ implies $x \geq y$. This dominance is also retained when taking limits; under any sequence of action sets in which the principal's smallest rationalizable (Nash) payoff is either equal to or approaches v_{IPE}, there is a sequence of rationalizable (Nash) payoffs bounded away from v_{IPE}.

[10] Carroll (2015) establishes the existence of an optimal IPE under principal-preferred action selection. Under the assumption that known actions are costly, there continues to exist an optimal contract even under principal least-preferred action selection. Moreover,

$$v_{\mathrm{IPE}} = \max_{w:\, w \text{ is an IPE}} \inf_{A\in\mathcal{A}} \min\big(V(w, A)\big).$$

[11] To see that all payoffs in the interval are attainable, it suffices to randomize over the maximal and minimal action profiles.

Theorem 1 has a number of important implications for the selection of robustly optimal contracts under single-valued solution concepts. For instance, under principal-preferred Nash equilibrium selection (the assumption in the literature discussed in Section 1.1), JPE strictly outperforms IPE.

Remark 1 (λ-maxmin Nash equilibrium selection). Suppose that, given a contract w and action set A, the principal has ambiguity aversion over the Nash equilibrium the agents will play as captured by the Hurwicz (1951) criterion

$$V_\lambda(w, A) := \lambda \min_{\sigma \in \mathcal{E}(w,A)} \mathbb{E}_\sigma\left[\sum_{i=1}^n \big(y_i - w_i(y)\big)\right] + (1-\lambda) \max_{\sigma \in \mathcal{E}(w,A)} \mathbb{E}_\sigma\left[\sum_{i=1}^n \big(y_i - w_i(y)\big)\right],$$

where $\lambda \in [0, 1]$ and $\mathcal{E}(w, A)$ is the set of Nash equilibria in $\Gamma(w, A)$ (see Ghirardato, Maccheroni, and Marinacci (2004) for an axiomatization). Then, under the JPE contract identified in Theorem 1, w^{JPE},

$$\inf_{A \in \mathcal{A}} V_\lambda\big(w^{\mathrm{JPE}}, A\big) \geq \sup_{w:\, w \text{ is an IPE}} \inf_{A \in \mathcal{A}} V_\lambda(w, A),$$

where the inequality is strict for any $\lambda < 1$. Notice that $\lambda = 0$ corresponds to partial Nash implementation.

In addition, the maximal rationalizable (Nash) action profile Pareto dominates all other rationalizable (Nash) action profiles. So, under any selection of rationalizable action profiles satisfying weak Pareto efficiency for the agents, JPE strictly outperforms IPE.

Remark 2 (Cooperative solutions). Let $f : \mathcal{W} \times \mathcal{A} \to \Delta((\mathbb{R}_+ \times \Delta(Y))^n)$ be any selection of a distribution over rationalizable action profiles that is weakly Pareto efficient for the agents. Specifically, for any contract w and action set A, $f(w, A) = \sigma \in \mathcal{R}(w, A)$ and there does not exist a distribution $\sigma' \in \mathcal{R}(w, A)$ such that, for all agents i, $\mathbb{E}_\sigma[U_i(\cdot; w)] < \mathbb{E}_{\sigma'}[U_i(\cdot; w)]$. Given a contract w and action set A, let

$$V_f(w, A) := \mathbb{E}_{f(w,A)}\left[\sum_{i=1}^n \big(y_i - w_i(y)\big)\right].$$

Then, under the JPE contract identified in Theorem 1, w^{JPE},

$$\inf_{A \in \mathcal{A}} V_f\big(w^{\mathrm{JPE}}, A\big) > \sup_{w:\, w \text{ is an IPE}} \inf_{A \in \mathcal{A}} V_f(w, A).$$

A simple example illustrates the key intuition behind Theorem 1.

Example 1 (JPE versus IPE). There are two agents ($n = 2$) and output is binary ($Y = \{0, 1\}$). There is a single, common known action ($A_1^0 = A_2^0 = \{a^0\}$). The known action results in success, $y = 1$, with probability $p(a^0) > 0$ and failure, $y = 0$, with probability $1 - p(a^0)$. Its effort cost is $c(a^0) \in (0, p(a^0))$. The principal is concerned about unknown

action sets of the form $A = \{a^0, a^*\} \times \{a^0, a^*\}$, where a^* is a "shirking" action available to both agents. She knows that shirking entails zero effort cost, $c(a^*) = 0$, and is less productive than the known action, that is, the probability of success is $p(a^*) < p(a^0)$. Moreover, she assumes that she can select her most-preferred Nash equilibrium in case of multiplicity.

Consider first the principal's payoff guarantee from an optimal IPE contract, which pays each agent a share of output $\alpha \in (c(a^0), 1)$. A naive intuition is that the principal's worst-case payoff is obtained when $p(a^*) = 0$; if agents take a shirking action with this success probability, then the principal obtains an expected payoff of zero. But, this logic ignores incentives, as pointed out by Carroll (2015). In particular, each agent has a strict incentive to shirk only if she obtains a higher expected utility from doing so. Hence, (a^0, a^0) is a Nash equilibrium whenever

$$p(a^*)\alpha \leq p(a^0)\alpha - c(a^0) \quad \Longleftrightarrow \quad p(a^*) \leq p(a^0) - \frac{c(a^0)}{\alpha},$$

yielding the principal a payoff per agent of

$$p(a^0)(1-\alpha).$$

The principal's worst-case payoff is instead obtained as p^* approaches $p(a^0) - c(a^0)/\alpha$ from above. Along this sequence, (a^*, a^*) is the unique Nash equilibrium and the principal's payoff per agent becomes arbitrarily close to

$$\left(p(a^0) - \frac{c(a^0)}{\alpha}\right)(1-\alpha).$$

Now, consider a contract of the form described in the statement of Theorem 1, parameterized by $\phi := \phi_1 = \phi_2 > 0$ and $\beta := \beta_1 = \beta_2 > 0$. Specifically, choose ϕ so that it is strictly smaller than the benchmark IPE share, α, and choose

$$\beta = \frac{\alpha - \phi}{p(a^0)}.$$

This contract is calibrated to the optimal IPE contract in the following sense: if an agent succeeds at his task, then his expected wage payment remains the same conditional on the other agent working. That is,

$$\phi + \beta p(a^0) = \alpha.$$

Hence, (a^0, a^0) is again a Nash equilibrium whenever

$$p(a^*) \leq p(a^0) - \frac{c(a^0)}{\alpha}.$$

Moreover, the principal's worst-case payoff is again obtained as $p(a^*)$ approaches $p(a^0) - c(a^0)/\alpha$ from above. (Along this sequence, (a^*, a^*) is the unique Nash equilibrium.) However, a simple calculation shows that the principal obtains a strictly higher

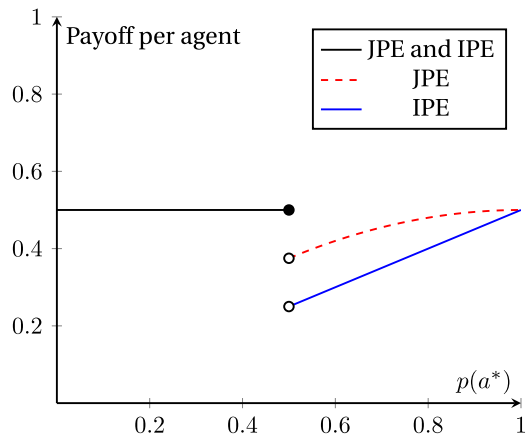

FIGURE 1. Principal's expected payoff per agent as a function of $p(a^*)$. Parameters: $p(a^0) = 1$, $c(a^0) = 1/4$, $\alpha = 1/2$, $\phi = 0$, and $\beta = 1/2$.

payoff per agent in worst-case scenarios:

$$\left(p(a^0) - \frac{c(a^0)}{\alpha}\right)\left(1 - \left(\phi + \beta\left(p(a^0) - \frac{c(a^0)}{\alpha}\right)\right)\right) > \left(p(a^0) - \frac{c(a^0)}{\alpha}\right)(1-\alpha),$$

where the inequality follows from $\phi + \beta p(a^*) < \phi + \beta p(a^0) = \alpha$ for $p(a^*) < p(a^0)$. See Figure 1 for an illustration. ◊

The intuition behind Example 1 is simple. By constructing a mean-preserving spread of an agent's wage with respect to the targeted action of the other, worst-case productivity is held constant. But, under joint performance evaluation, the principal pays agents less in expectation in worst-case scenarios. Each is punished for the shirking of the other.

While Example 1 identifies the key rent-extraction advantage of JPE over IPE, the class of games considered is *with* loss of generality. Appendix A.2 establishes that, under partial Nash implementation and the assumption that agents possess a common action set, the worst-case payoff for the principal is attained in the limit of a sequence of dominance solvable games in which the cardinality of the common action set grows to infinity. Such games can harm the principal because joint performance evaluation generates a free-riding problem. When others are less productive, each agent is willing to take an even less productive action—on average, they receive a smaller share of the output they produce. Hence, with a larger uncertainty set, there are two effects of joint performance evaluation: a rent-extraction benefit and a free-riding cost.

Theorem 1 shows that there nevertheless exists a JPE contract whose rent-extraction benefit outweighs its free-riding cost under partial Nash implementation and other, more permissive, solution concepts. To mitigate the free-riding problem and increase the entire set of rationalizable payoffs, the proof utilizes a more conservative calibration argument than illustrated in Example 1. Specifically, it identifies an improved JPE contract under which each agent's contract is calibrated to an IPE contract yielding him

a larger share of individual output than optimal. Despite encouraging greater productivity, these larger IPE contracts are suboptimal on their own because they leave too much rent to each agent. But, under the calibrated JPE contract, the principal reduces expected wage payments in worst-case scenarios. This reduction is shown to be large enough that the calibrated JPE contract strictly outperforms both the larger IPE contract and the smaller, optimal IPE contract. Hence, the superiority of JPE over IPE is retained even when agents may be playing more complicated games than those considered in Example 1 and under a broader class of solution concepts than principal-preferred Nash equilibrium.

4. Optimality of joint performance evaluation

The second main result is that, in a canonical setting, any optimal contract exhibits joint performance evaluation.[12] Specifically, it is assumed that there are two agents, two individual output levels (success, $y_i = 1$, and failure, $y_i = 0$), and both agents share a common action set. In the notation of the model, $n = 2$, $Y = \{0, 1\}$, $A_1^0 = A_2^0$, and the feasible set of uncertainty is $\mathcal{A}_s := \{A \subset (\mathbb{R}_+ \times \Delta(Y))^2 : \ A_i \supseteq A_i^0, \ A_1 = A_2, \text{ and } |A| < \infty\}$. Moreover, the principal assumes the agents play her preferred Nash equilibrium (as in the literature discussed in Section 1.1). Then, from contract w, the principal obtains a payoff of

$$V(w) := \inf_{A \in \mathcal{A}_s} \max_{\sigma \in \mathcal{E}(w, A)} \mathbb{E}_\sigma \left[\sum_{i=1}^{n} \big(y_i - w_i(y)\big) \right],$$

where $\mathcal{E}(w, A)$ denotes the set of Nash equilibria in game $\Gamma(w, A)$.

Before the result is stated, the contract space and notion of optimality are formally defined. If there are two agents and two output levels, a contract for agent i can be represented as a quadruple of nonnegative wages,

$$w_i := \big(w_i^{11}, w_i^{10}, w_i^{01}, w_i^{00}\big) \in \mathbb{R}_+^4,$$

where the first number of a wage's superscript indicates agent i's own success ($y_i = 1$) or failure ($y_i = 0$) and the second indicates the success or failure of the other agent. If, in addition, contracts are assumed to be symmetric, then the subscript can be dropped and the set of all contracts is simply the set of all nonnegative quadruples. The typology of performance evaluations thus simplifies considerably.

Definition 2 (Binary performance evaluations). A symmetric contract, $w \in \mathbb{R}_+^4$, is

- an *independent performance evaluation (IPE)* if $(w^{11}, w^{01}) = (w^{10}, w^{00})$;
- a *relative performance evaluation (RPE)* if $(w^{11}, w^{01}) < (w^{10}, w^{00})$;
- and a *joint performance evaluation (JPE)* if $(w^{11}, w^{01}) > (w^{10}, w^{00})$,

where $>$ and $<$ indicate strict inequality in at least one component and weak in both.

[12] The Bayesian analog of this setting is thoroughly analyzed by Fleckinger (2012) and Fleckinger, Martimort, and Roux (2024), who show that symmetric IPE contracts are optimal for independent and identical agents.

A contract is said to be *s-optimal* if it maximizes $V(\cdot)$ within the class of symmetric contracts, $w \in \mathbb{R}^4_+$.

The main result follows below.

THEOREM 2 (Optimality of JPE). *Suppose there are two agents, $n = 2$, with a common set of actions and the set of output levels is binary, $Y = \{0, 1\}$. Then any s-optimal contract is a JPE contract with $w^{11} > w^{10}$ and $w^{01} = w^{00} = 0$. Moreover, there exists an s-optimal contract.*

PROOF. The proof is in Appendix A.3. □

The result is a consequence of two important lemmas. The first lemma establishes that it is without loss of generality to consider contracts that do not reward failure.

LEMMA 1 (Suboptimality of positive wages for failure). *For any contract w with $w^{00} > 0$ or $w^{01} > 0$, there exists an IPE, RPE, or JPE contract $\hat{w}$ with $\hat{w}^{01} = \hat{w}^{00} = 0$ that yields the principal a higher payoff.*

PROOF. The proof is in Appendix A.4. □

Though the result is familiar, the proof is surprisingly nontrivial. Specifically, while reducing agent payoffs by a constant yields an improvement for the principal in many cases, other cases require different arguments. The most difficult case involves proving that no contract with $w^{10} > w^{11} = 0$ and $w^{01} > w^{00} = 0$ can yield a higher payoff than v_{IPE}. While it is intuitive that reducing w^{01} improves individual incentives to take more productive actions, characterizing the effect of a change in the contract parameter on the principal's payoff under asymmetric and mixed equilibria is nontrivial.

The second lemma establishes that no RPE contract can outperform the best IPE contract.

LEMMA 2 (IPE outperforms RPE). *No RPE contract with $w^{01} = w^{00} = 0$ can yield the principal a higher payoff than the optimal IPE contract.*

PROOF. The proof is in Appendix A.5. □

Symmetric RPE contracts such as salesperson-of-the-year awards are commonly utilized in practice. So, it is worthwhile to describe the economic intuition behind their suboptimality under robustness considerations. Under RPE, agents are discouraged to take more productive actions when others are more productive. When one agent is productive, the other agent has less of an incentive to take a productive action because his chance of outperforming the other decreases. Given that one agent is willing to shirk, it is then possible to provide incentives for the other agent to shirk. In the resulting equilibrium, expected wage payments actually *increase*; weight is shifted from w^{11} to w^{10} and $w^{10} > w^{11}$, in contrast to the case of JPE. The corresponding increase in expected wage payments offsets the advantage of encouraging productivity by one of the two agents. The mechanics of the argument are illustrated in an elaboration of Example 1.

Example 2 (RPE versus IPE). Suppose the environment and space of uncertainty are the same as in Example 1. Consider the performance guarantee of an RPE contract with $w^{11} > w^{10} > 0$. Observe that a^* is a strict best response to a^* if and only if

$$\underbrace{p(a^*)\big(p(a^*)w^{11} + (1 - p(a^*))w^{10}\big)}_{\text{Payoff } a^* \text{ against } a^*} > \underbrace{p(a^0)\big(p(a^*)w^{11} + (1 - p(a^*))w^{10}\big) - c(a^0)}_{\text{Payoff } a^0 \text{ against } a^*}$$

$$\iff \quad p(a^*) > p(a^0) - \frac{c(a^0)}{p(a^*)w^{11} + (1 - p(a^*))w^{10}}.$$

That is, if one agent shirks, the other has a strict incentive to shirk. If this inequality is satisfied, then a^* is also a strict best response to a^0; the incentive to shirk is larger when the other works because $w^{11} < w^{10}$. So, a^* is a strictly dominant strategy and (a^*, a^*) is the unique Nash equilibrium.

Now, suppose the productivity of the shirking action approaches from above the value, $p(a^*)$, at which

$$p(a^*) = p(a^0) - \frac{c(a^0)}{p(a^*)w^{11} + (1 - p(a^*))w^{10}}.$$

Then (a^*, a^*) is the unique Nash equilibrium along the sequence and the principal's payoff per agent approaches

$$\left(p(a^0) - \frac{c(a^0)}{p(a^*)w^{11} + (1 - p(a^*))w^{10}}\right)\big(1 - \big(p(a^*)w^{11} + (1 - p(a^*))w^{10}\big)\big).$$

This payoff can be no higher than what is obtained from an IPE contract with share $\alpha := p(a^*)w^{11} + (1 - p(a^*))(1 - w^{10})$, whose payoff is derived in Example 1. ◊

To conclude the sketch of the proof of Theorem 2, observe that from Remark 1 (setting $\lambda = 0$) and the proof of Theorem 1, there exists a symmetric JPE contract with $w^{01} = w^{00} = 0$ that strictly outperforms any IPE contract. Hence, if there exists an s-optimal contract, then there exists an s-optimal JPE contract with $w^{01} = w^{00} = 0$. Existence follows because the closed-form expression for the principal's worst-case payoff under a JPE contract with $w^{01} = w^{00} = 0$ is continuous in the contract parameters w_{11} and w_{10} (see Lemma 6 in Appendix A.2). Moreover, the JPE contract parameters $w_{11} > w_{10}$ can be taken to lie in a compact set by relaxing the strict inequality constraint and bounding w_{11} from above. Uniqueness follows from inspecting the improvements in Lemma 1.

The s-optimal contracts identified in Theorem 2 resemble those observed in the sales force studied by Rees, Zax, and Herries (2003). Adapted to the binary setting, each salesperson i's performance-based compensation is

$$w_i(y_i, y_j) = P \times \left(0.7 \times f\left(\frac{y_i}{T_I}\right) + 0.3 \times g\left(\frac{y_i + y_j}{T_G}\right)\right),$$

where P is the portion of monthly income contingent on performance, T_I and T_G are target individual and group output levels, f is increasing in individual output with $f(0) = 0$, and g is an increasing and nonlinear function of total output with $g(0) = 0$. To rationalize $w_i(1, 1) = w^{11} > w^{10} = w_i(1, 0)$, $w_i(0, 0) = w^{00} = 0$, and $w_i(0, 1) = w^{01} = 0$, set $T_I = 1$ and $T_G = 2$ (these values are not reported in the data). Then f must be increasing in individual output: it must put $f(1/T_I) \geq f(0/T_I)$. Moreover, $f(0/T_I) = 0$. In addition, g must be increasing and nonlinear in the ratio of total output to the group target: it must put $g((y_i + y_j)/T_G) > 0$ if and only if $y_i + y_j = 2$. Moreover, $g(0/T_G) = 0$.

A qualifying remark is in order regarding the application. If output is binary, any function of individual output, f, with $f(0) = 0$ must trivially be linear. In contrast, Rees, Zax, and Herries (2003) document that f is nonlinear. Because the form of the optimal contract in the model with a general set of output levels has not been determined, it remains an open question whether the precise compensation formula used in practice is optimal in the model.

5. Final remarks

This paper identifies nonstatistical foundations for team-based incentive pay. Very generally, it is shown that linking the pay of independent agents is robustly optimal. Moreover, in a canonical environment, joint performance evaluation contracts are optimal. Such contracts approximate the incentive properties of benchmark independent performance evaluation contracts, while flexibly reducing expected wage payments when agents are less productive than the principal anticipates. The worst-case analysis draws attention to these scenarios, uncovering an economic intuition that had previously gone unnoticed.

Appendix: Proofs

A.1 *Proof of Theorem 1*

From Carroll (2015), there exists an optimal IPE contract such that, for each i,

$$w_i(y_i, y_{-i}) = \alpha_i y_i,$$

where $\alpha_i \in (0, 1)$ and $\alpha_i = \sqrt{c(a_i^0)}/\sqrt{\mathbb{E}_{F(a_i^0)}[y_i]}$ for some $a_i^0 \in A_i^0$. The infimum payoff from agent i is attained in the limit of a sequence of action sets $(A_i(k))_k$ in which the agent's unique rationalizable action has expected output converging to

$$\bar{p}_i := \mathbb{E}_{F(a_i^0)}[y_i] - \frac{c(a_i^0)}{\alpha_i}, \tag{1}$$

yielding the principal a (worst-case) payoff of

$$\bar{p}_i(1 - \alpha_i).$$

When compensating n agents using only optimal IPE contracts, the principal's payoff is thus

$$v_{\text{IPE}} = \sum_{i=1}^{n} \bar{p}_i(1-\alpha_i).$$

Fix a collection of optimal IPE shares $(\alpha_i)_i$, where $\alpha_i \in (0,1)$ and $\alpha_i = \sqrt{c(a_i^0)}/\sqrt{\mathbb{E}_{F(a_i^0)}[y_i]}$ for some $a_i^0 \in A_i^0$. Consider a JPE contract such that, for each agent i,

$$w_i^{\text{JPE}}(y_i, y_{-i}) = \left(\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n} y_j\right) y_i, \tag{2}$$

where $1 > \alpha_i > \phi_i > c(a_i^0)/\mathbb{E}_{F(a_i^0)}[y_i]$ and $c(a_i^0)/(\mathbb{E}_{F(a_i^0)}[y_i])^2 > \beta_i > 0$ are chosen so that

$$\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n} \bar{p}_j = \alpha_i, \tag{3}$$

where $\bar{p}_i$ is defined in (1). In addition, require that $(\phi_i, \beta_i)_{i=1}^n$ are chosen so that, for all i,

$$\phi_i + \beta_i \max(Y) + \sum_{j\neq i}^{n} \frac{\beta_j}{n-1} < 1. \tag{4}$$

Two observations about the constructed JPE contract follow below.

Observation 1. The principal's expected payoff under the JPE contract and an arbitrary action profile a is

$$\sum_{i=1}^{n} \mathbb{E}_{F(a_i)}[y_i]\left(1 - \left(\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j)}[y_j]\right)\right).$$

Hence, by (3), if $\mathbb{E}_{F(a_i)}[y_i] = \bar{p}_i$ for all i, then the principal's expected payoff under w^{JPE} is exactly v_{IPE}. Moreover, the principal's expected payoff is strictly increasing in agent i's expected output whenever

$$\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j)}[y_j] + \sum_{j\neq i}^{n} \frac{\beta_j}{n-1} < 1.$$

Any feasible output distribution $F(a_j) \in \Delta(Y)$ satisfies $\mathbb{E}_{F(a_j)}[y_j] \leq \max(Y)$. So, (4) ensures that the principal's expected payoff is strictly increasing in the expected output of agent i given any action profile of the other agents, a_{-i}.

Observation 2. Under the JPE contract, the only payoff relevant attribute of an individual distribution over output is its mean. So, let each agent's individual action belong to the set $\mathbb{R}_+ \times [0, \max(Y)]$, where the second component of an action is its mean,

and equip any $A_i \supseteq A_i^0$ with the total order $\succeq_i$: $a_i \succeq_i a_i'$ if either $\mathbb{E}_{F(a_i)}[y_i] > \mathbb{E}_{F(a_i')}[y_i]$, or $\mathbb{E}_{F(a_i)}[y_i] = \mathbb{E}_{F(a_i')}[y_i]$ and $c(a_i) \leq c(a_i')$. Then $\Gamma(w^{\mathrm{JPE}}, A)$ is a supermodular game under the corresponding product order on action profiles: $a' \succeq a$ implies $\mathbb{E}_{F(a_i')}[y_i] \geq \mathbb{E}_{F(a_i)}[y_i]$ for all i, and hence

$$\begin{aligned}
U_i(a_i', a_{-i}'; w^*) - U_i(a_i, a_{-i}'; w^*) &= \left(\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j')}[y_j]\right)\left(\mathbb{E}_{F(a_i')}[y_i] - \mathbb{E}_{F(a_i)}[y_i]\right) \\
&\quad - \left(c(a_i') - c(a_i)\right) \\
&\geq \left(\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j)}[y_j]\right)\left(\mathbb{E}_{F(a_i')}[y_i] - \mathbb{E}_{F(a_i)}[y_i]\right) \\
&\quad - \left(c(a_i') - c(a_i)\right) \\
&= U_i(a_i', a_{-i}; w^*) - U_i(a_i, a_{-i}; w^*).
\end{aligned}$$

The following lemma establishes that, in any game, every agent i plays an action with expected output weakly larger than $\bar{p}_i$ in any rationalizable action profile. Moreover, there exists a game with a rationalizable action profile in which each agent i produces expected output equal to $\bar{p}_i$.

Lemma 3. *Given any action set A satisfying $A_i \supseteq A_i^0$, any rationalizable action for agent i in $\Gamma(w^{\mathrm{JPE}}, A)$ has expected output weakly larger than $\bar{p}_i$. However, there exists an action set A satisfying $A_i \supseteq A_i^0$ such that $\Gamma(w^{\mathrm{JPE}}, A)$ has a rationalizable action profile in which each agent i produces expected output exactly equal to $\bar{p}_i$.*

Proof. Given the presence of a_i^0, under any conjecture about other agents' actions, agent i is unwilling to play an action with expected output smaller than

$$p_i^1 := \mathbb{E}_{F(a_i^0)}[y_i] - \frac{c(a_i^0)}{\phi_i} > 0$$

because $\min(Y) = 0$. If agent i knows each agent $j \neq i$ is unwilling to play an action with expected output smaller than p_j^1, then he is unwilling to play an action with expected output smaller than

$$p_i^2 := \mathbb{E}_{F(a_i^0)}[y_i] - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} p_j^1} > p_i^1.$$

Iterating yields a strictly increasing and bounded sequence, $(p_1^k, \ldots, p_n^k)_k$. Hence, its limit, $(p_1^\infty, \ldots, p_2^\infty) \in [0, \max(Y)]^n$, exists by the monotone convergence theorem and

must satisfy

$$p_i^\infty = \mathbb{E}_{F(a_i^0)}[y_i] - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} p_j^\infty} \quad \text{for all } i = 1, \dots, n.$$

By (3), $p_i^\infty = \bar{p}_i$ for all i is a solution to the system of equations. It is the unique solution because the map $T : \mathbb{R}_+^n \to \mathbb{R}_+^n$ with ith component

$$T_i(p) = \mathbb{E}_{F(a_i^0)}[y_i] - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} p_j}$$

is a contraction on $(\mathbb{R}_+^n, d)$, where $d(x, y) := \max_i |x_i - y_i|$ is the supremum (Chebyshev) distance. To prove this, observe that for any vectors $p, p' \in \mathbb{R}_+^n$,

$$\begin{aligned}
\left|T_i(p) - T_i(p')\right| &= \left| \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} p_j} - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} p'_j} \right| \\
&= \left| \frac{c(a_i^0)\beta_i\left(\dfrac{1}{n-1}\left(\displaystyle\sum_{j\neq i} p_j - \sum_{j\neq i} p'_j\right)\right)}{\left(\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} p_j\right)\left(\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} p'_j\right)} \right| \\
&\leq \left| \frac{c(a_i^0)\beta_i}{\phi_i^2} \right| d(p, p'),
\end{aligned}$$

with $|c(a_i^0)\beta_i/\phi_i^2| \leq |\beta_i((\mathbb{E}_{F(a_i^0)}[y_i])^2/c(a_i^0))| < 1$ by $\phi_i > c(a_i^0)/\mathbb{E}_{F(a_i^0)}[y_i]$ and $\beta_i < c(a_i^0)/(\mathbb{E}_{F(a_i^0)}[y_i])^2$. So,

$$d(T(p), T(p')) = \max_i \left|T_i(p) - T_i(p')\right| \leq \kappa d(p, p'),$$

where $\kappa := \min_i |c(a_i^0)\beta_i/\phi_i^2| < 1$.

Now, consider the action set $A := \times_{i=1}^n A_i^0 \cup \{a_i^*\}$, where $c(a_i^*) = 0$ and $\mathbb{E}_{F(a_i^*)}[y_i] = \bar{p}_i$. In $\Gamma(w, A)$, $(a_1^*, \dots, a_n^*)$ is rationalizable because a_i^* is a best response to a_{-i}^* (it yields the same payoff as i's targeted known action, which is a best response to a_{-i}^* in A_i^0). So, there exists an action set with a rationalizable action profile in which each agent i produces expected output $\bar{p}_i$. □

In addition, there exists some $\epsilon > 0$ such that, in any game played by the agents, there exists a rationalizable action profile in which there is some agent i with expected output weakly larger than $\bar{p}_i + \epsilon$.

Lemma 4. *There exists an $\epsilon > 0$ such that, in any game $\Gamma(w^{\mathrm{JPE}}, A)$ satisfying $A_i \supseteq A_i^0$, there is at least one rationalizable action profile in which some agent i plays an action with expected output weakly larger than $\bar{p}_i + \epsilon$.*

Proof. For each i, define

$$\hat{p}_i := \mathbb{E}_{F(a_i^0)}[y_i] - \frac{\big(c(a_i^0)/2\big)}{\phi_i + \dfrac{\beta_i}{(n-1)} \displaystyle\sum_{j \neq i}^{n} \bar{p}_j} > \bar{p}_i,$$

a lower bound on the expected output of any rationalizable action with cost weakly greater than $c(a_i^0)/2$ by Lemma 3. Let $\delta \in (0, \frac{1}{3} \min_i (\mathbb{E}_{F(a_i^0)}[y_i] - \bar{p}_i))$ satisfy both

$$\frac{1}{3} \max_i \left(\frac{\phi_i + \dfrac{\beta_i}{n-1} \displaystyle\sum_{j \neq i}^{n} (\bar{p}_j + 3\delta)}{\phi_i + \dfrac{\beta_i}{n-1} \displaystyle\sum_{j \neq i}^{n} \bar{p}_j} \right) < \frac{1}{2} \tag{5}$$

and

$$\delta < \frac{1}{3} \min_i \left(\frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1} \displaystyle\sum_{j \neq i}^{n} \bar{p}_j} - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1} \displaystyle\sum_{j \neq i}^{n} \hat{p}_j} \right). \tag{6}$$

Define

$$\epsilon := \min_i \left(\frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1} \displaystyle\sum_{j \neq i}^{n} \bar{p}_j} - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1} \displaystyle\sum_{j \neq i}^{n} (\bar{p}_j + \delta)} \right) > 0. \tag{7}$$

Toward contradiction, suppose that there exists an action set with $A_i \supseteq A_i^0$ in which all rationalizable action profiles involve each agent producing expected output strictly smaller than $\bar{p}_i + \epsilon$. It suffices to consider action sets with maximal action profile equal to the targeted known action profile, $(a_1^0, \ldots, a_n^0)$. Let $(a_i^k)_{k=0}^{\infty}$ be the best-response path for agent i obtained from infinite iteration of maximal best-response functions. Let $a_i^{\infty} := \lim_{k \to \infty} a_i^k$. Then $(a_1^{\infty}, \ldots, a_n^{\infty})$ is a rationalizable action profile (see, e.g., Vives (1990) and Milgrom and Roberts (1990)). Hence, it must satisfy $\mathbb{E}_{F(a_i^{\infty})}[y_i] < \bar{p}_i + \epsilon$ for all i, or

$$\mathbb{E}_{F(a_i^{\infty})}[y_i] < \mathbb{E}_{F(a_i^0)}[y_i] - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1} \displaystyle\sum_{j \neq i}^{n} \bar{p}_j} + \epsilon \tag{8}$$

for all i using the definition of $\bar{p}_i$. For any $k \geq 1$, a_i^k is in agent i's maximal best-response path only if

$$\mathbb{E}_{F(a_i^k)}[y_i] \geq \mathbb{E}_{F(a_i^{k-1})}[y_i] - \frac{c(a_i^{k-1}) - c(a_i^k)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j^{k-1})}[y_j]}.$$

So,

$$\mathbb{E}_{F(a_i^\infty)}[y_i] \geq \mathbb{E}_{F(a_i^0)}[y_i] - \sum_{k=1}^{\infty} \frac{c(a_i^{k-1}) - c(a_i^k)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j^{k-1})}[y_j]}.$$

Hence, from (8),

$$\mathbb{E}_{F(a_i^0)}[y_i] - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} \bar{p}_j} + \epsilon > \mathbb{E}_{F(a_i^0)}[y_i] - \sum_{k=1}^{\infty} \frac{c(a_i^{k-1}) - c(a_i^k)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j^{k-1})}[y_j]},$$

which holds if and only if

$$\epsilon > \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} \bar{p}_j} - \sum_{k=1}^{\infty} \frac{c(a_i^{k-1}) - c(a_i^k)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j^{k-1})}[y_j]}.$$

Rearranging and using the definition of $\epsilon > 0$ yields

$$\sum_{k=1}^{\infty} \frac{c(a_i^{k-1}) - c(a_i^k)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j^{k-1})}[y_j]} > \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} (\bar{p}_j + \delta)} \tag{9}$$

for all i. Let k_i be the largest iteration at which $\sum_{j\neq i}^{n} \mathbb{E}_{F(a_j^{k_i-1})}[y_j] \geq \sum_{j\neq i}^{n} (\bar{p}_j + 3\delta)$. Observe that $k_i \geq 1$ because $\delta < \frac{1}{3}\min_i(\mathbb{E}_{F(a_i^0)}[y_i] - \bar{p}_i)$ and $\mathbb{E}_{F(a_i^{k-1})}[y_i] \geq \mathbb{E}_{F(a_i^k)}[y_i] \geq \bar{p}_i$ for all $k \geq 1$. Moreover, $k_i < \infty$. If not, then (9) would be violated because $c(a_i^{k-1}) \geq c(a_i^k) \geq 0$ for all $k \geq 1$. In addition, $c(a_i^{k_i}) \geq c(a_i^0)/2$. If not, then (9) would be violated because $\mathbb{E}_{F(a_i^{k-1})}[y_i] \geq \mathbb{E}_{F(a_i^k)}[y_i] \geq \bar{p}_i$ for all $k \geq 1$, and the value of x that solves

$$\frac{x}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} (\bar{p}_j + 3\delta)} + \frac{c(a_i^0) - x}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} \bar{p}_j} = \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n} (\bar{p}_j + \delta)}$$

$$\iff \quad x = \frac{c(a_i^0)}{3}\left(\frac{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n}(\bar{p}_j + 3\delta)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n}\bar{p}_j}\right)$$

is smaller than $\frac{1}{2}c(a_i^0)$ by (5). Choose $K := \min_i k_i$. Then, in iteration K, there must exist some agent i and action a_i^K in i's best-response path satisfying $\mathbb{E}_{F(a_i^K)}[y_i] < \bar{p}_i + 3\delta$ and $c(a_i^K) \geq 0$ when $\mathbb{E}_{F(a_i^{K-1})}[y_j] \geq \hat{p}_j$ for all j. That is, it must be that

$$\left(\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n}\hat{p}_j\right)(\bar{p}_i + 3\delta) > \left(\phi_i + \frac{\beta_i}{n-1}\sum_{j\neq i}^{n}\hat{p}_j\right)\mathbb{E}_{F(a_i^0)}[y_i] - c(a_i^0).$$

But rearranging and using the definition of $\bar{p}_i$ yields

$$\delta > \frac{1}{3}\left(\frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n}\bar{p}_j} - \frac{c(a_i^0)}{\phi_i + \dfrac{\beta_i}{n-1}\displaystyle\sum_{j\neq i}^{n}\hat{p}_j}\right),$$

which contradicts (6). □

Observation 1, Lemma 3, and Lemma 4 together establish the desired result:

$$\inf_{A\in\mathcal{A}} \max\big(V(w^{\mathrm{JPE}}, A)\big) > \inf_{A\in\mathcal{A}} \min\big(V(w^{\mathrm{JPE}}, A)\big) = v_{\mathrm{IPE}}.$$

A.2 *Worst-case symmetric game*

In this subsection, let $n = 2$, $Y = \{0, 1\}$, $A_1^0 = A_2^0$, and suppose that the feasible set of uncertainty is $\mathcal{A}_s := \{A \subset (\mathbb{R}_+ \times \Delta(Y))^2 : \ A_i \supseteq A_i^0, \ A_1 = A_2, \text{ and } |A| < \infty\}$. Moreover, suppose that the principal can select her preferred Nash equilibrium in case of multiplicity. Then, from contract w, the principal obtains a payoff of

$$V(w) := \inf_{A\in\mathcal{A}_s} \max_{\sigma\in\mathcal{E}(w,A)} \mathbb{E}_\sigma\left[\sum_{i=1}^{n}\big(y_i - w_i(y)\big)\right],$$

where $\mathcal{E}(w, A)$ is the set of Nash equilibria in game $\Gamma(w, A)$.

Consider a JPE contract with $w^{00} := w_1(0,0) = w_2(0,0) = 0$, $w^{01} := w_1(0,1) = w_2(1,0) = 0$, and $w^{11} := w_1(1,1) = w_2(1,1) > w^{10} := w_1(1,0) = w_2(1,0)$. Let $A_i \subseteq \mathbb{R}_+ \times [0,1]$, where $c(a_i) \in \mathbb{R}_+$ is the cost of action $a_i \in A_i$ and $p(a_i) \in [0,1]$ is the probability with which it results in success, $y_i = 1$. Equip any $A_i \supseteq A_i^0$ with the total order $\succeq_i$: $a_i \succeq_i a_i'$ if either $p(a_i) > p(a_i')$, or $p(a_i) = p(a_i')$ and $c(a_i) \leq c(a_i')$. Then $\Gamma(w^{\mathrm{JPE}}, A)$ is a supermodular game under the corresponding product order on action profiles (see Section A.1 for a more general proof).

In what follows, it will be useful to abuse notation and let A^0 denote the set of common known actions (instead of the set of common known action profiles), A denote an arbitrary set of common actions (instead of an arbitrary set of common action profiles), and $\Gamma(w, A)$ denote a game with common action set A. Let $\overline{BR} : A \to A$ denote the maximal best-response function for each agent. Then the following lemma holds.

Lemma 5 (Vives (1990), Milgrom and Roberts (1990)). *Suppose $(\bar{a}_i, \bar{a}_j)$ is the limit found by iterating $\overline{BR}$ starting from a maximal action profile in A in the order $\succeq$. If $\Gamma(w, A)$ is supermodular, then it has a maximal Nash equilibrium $(\bar{a}_i, \bar{a}_j)$; any other equilibrium (a_i, a_j) must satisfy $\bar{a} \succeq_i a_i$ and $\bar{a} \succeq_i a_j$.*

The principal's worst-case payoff from a JPE contract with $w^{00} = w^{01} = 0$ is computed in closed form by constructing an n-sequence of symmetric dominance solvable games (which can be solved by iterating $\overline{BR}$) in which there are n unknown actions.

Lemma 6. *Suppose w is a JPE contract with $w^{00} = w^{01} = 0$ and $w^{11} > w^{10}$. For each $a^0 \in A^0$, let $\hat{p}(\cdot|a^0) : [0, \hat{t}(a^0)] \to [0, p(a^0)]$ be the unique solution to the initial value problem*

$$\begin{aligned} &\hat{p}'(t) = f\big(\hat{p}(t)\big) := -\big(\hat{p}(t)w^{11} + \big(1 - \hat{p}(t)\big)w^{10}\big)^{-1} \quad \textit{with} \\ &\hat{p}(0) = p\big(a^0\big), \end{aligned} \tag{10}$$

where $[0, \hat{t}(a^0)] \subseteq [0, c(a^0)]$ is the largest interval on which $\hat{p}(t) > 0$ for all $t \in [0, \hat{t}(a^0))$. Then

$$V(w) = 2\ \min\big\{1 - w^{11},\ \bar{p}\big(\bar{p}\big(1 - w^{11}\big) + (1 - \bar{p})\big(1 - w^{10}\big)\big)\big\}, \tag{11}$$

where

$$\bar{p} := \max_{a^0 \in A^0} \hat{p}\big(\hat{t}\big(a^0\big)|a^0\big).$$

The proof follows below.

Comparative statics in principal's payoff Suppose agent i succeeds with probability p_i. The principal's payoff given (p_i, p_j) is

$$\pi(p_i, p_j) := p_i p_j\big(2 - 2w^{11}\big) + \big(p_i(1 - p_j) + (1 - p_i)p_j\big)\big(1 - w^{10}\big).$$

The principal's payoff is therefore increasing in p_i if and only if

$$p_j \le \frac{1}{2}\left(\frac{1 - w^{10}}{w^{11} - w^{10}}\right).$$

Monotonicity of $\pi(p_i, p_j)$ on $[0, 1]$ thus depends on w: (i) if $w^{10} \ge 1$, then π is decreasing on $[0, 1]$ in p_i and p_j; (ii) if $w^{10} < 1$ and $w^{11} \le \frac{1}{2}(1 + w^{10})$, then $\pi(p)$ is increasing on $[0, 1]$ in p_i and p_j; and, (iii) if $w^{10} < 1$ and $w^{11} > \frac{1}{2}(1 + w^{10})$, then $\pi(p)$ is increasing in p_i if $p_j \in [0, \frac{1}{2}(1 - w^{10})/(w^{11} - w^{10})]$ and decreasing in p_i if $p_j \in [\frac{1}{2}(1 - w^{10})/(w^{11} - w^{10}), 1]$.

In case (i), π is minimized when $p_i = p_j = 1$, yielding the principal a payoff of

$$2 - 2w^{11}.$$

This payoff can be achieved exactly: Consider the common action set $A := A^0 \cup \{\hat{a}\} \supseteq A^0$, where $p(\hat{a}) = 1$ and $c(\hat{a}) = 0$. Then, because $w^{11} > w^{10} \geq 1$, $\hat{a}$ is a strictly dominant strategy for each agent and so the unique Nash equilibrium of $\Gamma(w, A)$ is $(\hat{a}, \hat{a})$. In case (ii), π is minimized when the probability with which the maximal equilibrium action of each agent is as small as possible. Letting $\bar{p}$ denote the greatest lower bound on such probabilities, the principal's payoff is

$$\bar{p}^2\big(2 - 2w^{11}\big) + \bar{p}(1 - \bar{p})\big(2 - 2w^{10}\big).$$

In case (iii), the principal's payoff is the minimum of the payoff in case (i) and case (ii),

$$V(w) = \min\big\{2 - 2w^{11},\, \bar{p}^2\big(2 - 2w^{11}\big) + \bar{p}(1 - \bar{p})\big(2 - 2w^{10}\big)\big\}.$$

To complete the proof of the lemma, the value of $\bar{p}$ is identified.

Defining $\bar{p}$ Consider an arbitrary common action $a \in A$ with cost $c(a)$ and probability $p(a)$. Let $\hat{p}(\cdot|a)$ be a solution to the initial value problem

$$\begin{aligned} &\hat{p}'(t|a) = f\big(\hat{p}(t|a)\big) := -\big(\hat{p}(t|a)w^{11} + \big(1 - \hat{p}(t|a)\big)w^{10}\big)^{-1} \quad \text{with} \\ &\hat{p}(0|a) = p(a) \end{aligned}$$

on $D = [0, \hat{t}(a)] \times [0, p(a)]$, where $[0, \hat{t}(a)] \subseteq [0, c(a)]$ is the largest interval on which $\hat{p}(t|a) > 0$ for all $t \in [0, \hat{t}(a))$. Notice $\hat{p}'(t|a)$ exists on $(0, \hat{t}(a))$, $\hat{p}'(t|a) < 0$, and $\hat{p}''(t|a) < 0$. So, $\hat{p}(\cdot|a)$ is strictly decreasing and strictly concave. Now, define

$$\bar{p} := \max_{a^0 \in A^0} \hat{p}\big(\hat{t}\big(a^0\big)|a^0\big).$$

$\bar{p}$ is a lower bound It is next shown that $\bar{p}$ is a lower bound on the probability of the maximal equilibrium action of any game $\Gamma(w, A)$, where $A \supseteq A^0$ is a common set of actions.

Claim 1 (Lower bound of a $\overline{BR}$ path). Fix some game $\Gamma(w, A)$, where $A \supseteq A^0$ is a common set of actions. Let $(a^1, a^2, \ldots, a^n)$ be the path starting from the maximal element of A, a^1, to the maximal equilibrium action, a^n, obtained by iterating $\overline{BR}$. If $a = a^\ell$ for some $\ell = 1, \ldots, n$, then

$$p\big(a^n\big) \geq \hat{p}\big(\hat{t}(a)|a\big).$$

Proof. Consider the truncated path starting at $a = a^\ell$ and ending at a^n. Notice that $a^k \in \overline{BR}(a^{k-1})$ for $k = \ell + 1, \ldots, n$ only if $p(a^{k-1}) > p(a^k)$ and

$$p\big(a^k\big) > p\big(a^{k-1}\big) - \frac{c\big(a^{k-1}\big) - c\big(a^k\big)}{p\big(a^{k-1}\big)w^{11} + \big(1 - p\big(a^{k-1}\big)\big)w^{10}}.$$

Hence, $\epsilon_k := c(a^{k-1}) - c(a^k) > 0$ for any $k = \ell+1, \dots, n$. This implies that $\sum_{k=\ell+1}^n \epsilon_k \leq c(a)$, since $c(a^n) \geq 0$.

To show that $p(a^n) \geq \hat{p}(\hat{t}(a)|a)$, it suffices to consider the case in which $f(t, \hat{p}(t)|a)$ exists for all $t \in [0, c(a)]$ (it must always be the case that $p(a^n) \geq 0$). To show this, it suffices to show that $p(a^n) \geq \hat{p}(\sum_{k=\ell+1}^n \epsilon_k|a)$ because $\hat{p}(\cdot|a)$ is decreasing and so $\hat{p}(c(a)|a) \leq \hat{p}(\sum_{k=\ell+1}^n \epsilon_k|a)$.

The inequality is proven by induction. For the base case, recall that $p(a_{\ell+1})$ must satisfy the best-response condition

$$\begin{aligned} p\big(a^{\ell+1}\big) &\geq p\big(a^{\ell}\big) - \frac{\epsilon_{\ell+1}}{p\big(a^{\ell}\big)w^{11} + \big(1 - p(a\ell)\big)w^{10}} \\ &= \hat{p}(0|a) + \hat{p}'(0|a)\epsilon_{\ell+1} \\ &\geq \hat{p}(\epsilon_{\ell+1}|a), \end{aligned}$$

where the last inequality follows because $\hat{p}(\cdot|a)$ is concave.

For the inductive step, suppose $\hat{p}(\sum_{k=\ell+1}^m \epsilon_k|a) \leq p(a^m)$ for $m = \ell+1, \dots, K$. It is shown that $\hat{p}(\sum_{k=\ell+1}^K \epsilon_k + \epsilon_{K+1}|a) \leq p(a^{K+1})$. Once again, a^{K+1} is a best response to a^K only if

$$\begin{aligned} p\big(a^{K+1}\big) &\geq p\big(a^{K}\big) - \frac{\epsilon_{K+1}}{p\big(a^{K}\big)w^{11} + \big(1 - p\big(a^{K}\big)\big)w^{10}} \\ &\geq \hat{p}\left(\sum_{k=\ell+1}^{K} \epsilon_k|a\right) + \hat{p}'\left(\sum_{k=\ell+1}^{K} \epsilon_k|a\right)\epsilon_{K+1} \\ &\geq \hat{p}\left(\sum_{k=\ell+1}^{K} \epsilon_k + \epsilon_{K+1}|a\right), \end{aligned}$$

where the second inequality follows from the induction hypothesis and the last follows because $\hat{p}(\cdot|a)$ is concave. □

Consider any finite set of common actions $A \supseteq A^0$. Let $\tilde{c}$ be the maximal cost of any action in A and $\tilde{p}$ be the maximal probability of any action in A. For any action $a \in A$, let $\tilde{p}(\cdot|a)$ be the solution to the initial value problem,

$$\begin{aligned} \tilde{p}'(t|a) = f\big(\tilde{p}(t|a)\big) &= -\big(\tilde{p}(t|a)w^{11} + \big(1 - \tilde{p}(t|a)\big)w^{10}\big)^{-1}, \\ \tilde{p}\big(\tilde{c} - c(a)|a\big) &= p(a), \end{aligned}$$

on $D = [0, \tilde{t}(a)] \times [0, \tilde{p}]$, where $[0, \tilde{t}(a)] \subseteq [0, \tilde{c}]$ is the largest interval on which $\hat{p}(t|a) > 0$ for all $t \in [0, \hat{t}(a))$. Notice that $\tilde{p}(\tilde{c} - c(a) + t|a) = \hat{p}(t|a)$ for any $t \in [0, \hat{t}(a)]$, $\tilde{p}'(\cdot|a) < 0$ for all $t \in [0, \tilde{t}(a))$, and $\tilde{p}''(\cdot|a) < 0$ for all $t \in [0, \tilde{t}(a))$. Moreover, the following "no crossing" property holds; its proof is immediate upon observing that the solution to the initial value problem is unique on any interval $[0, \bar{t}]$ for $\bar{t} < \tilde{c}$, since $f'(\hat{p}(t|a))$ is bounded and exists (see, for instance, Theorem 2.2 of Coddington and Levinson (1955)).

Claim 2 (No crossing). If $\tilde{p}(t|a) > \tilde{p}(t|a')$ for some $t \in [0, \tilde{t}(a)] \cap [0, \tilde{t}(a')]$, then $\tilde{p}(t'|a) \geq \tilde{p}(t'|a')$ for any other $t' \in [0, \tilde{t}(a)] \cap [0, \tilde{t}(a')]$ and so $\hat{p}(\hat{t}(a)|a) \geq \hat{p}(\hat{t}(a')|a')$.

Suppose, toward contradiction, that there was a game with a maximal equilibrium action distribution p satisfying $p < \bar{p}$. Then there must exist a finite path of actions in A, $(a^1, \dots, a^n)$, for which (i) a^1 is the maximal element of A and $p(a^n) = p$, (ii) $p(a^1) > \cdots > p(a^n)$, and (iii) $a^k \in \overline{BR}(a^{k-1})$ (so that $c(a^1) > \cdots > c(a^n)$) for $k = 2, \dots, n$. It suffices to consider the case in which $\bar{p} > 0$, so that for any $\bar{a}^0 \in \arg\max_{a^0} \hat{p}(\hat{t}(a^0)|a^0)$, $\tilde{p}'(\cdot|\bar{a}^0)$ is defined on $[0, \tilde{c}]$. Otherwise, it could never be that $p < \bar{p}$.

Now, let a_k be the first action in the path $(a^1, \dots, a^n)$ at which $c(a^k) < c(\bar{a}^0)$. Such an action must exist. If not, then $c(a^n) \geq c(\bar{a}^0)$. So, if $p = p(a^n) < \bar{p} < p(\bar{a}^0)$, then (a^n, a^n) could not be a Nash equilibrium; $\bar{a}^0$ would be a strict best response to a^n.

Consider the case in which $k = 1$, so that $c(a^1) < c(\bar{a}^0)$. Then

$$\tilde{p}\big(\bar{c} - c(a^1)|a^1\big) = p(a^1) \geq p(\bar{a}^0) = \tilde{p}\big(\bar{c} - c(\bar{a}^0)|\bar{a}^0\big) > \tilde{p}\big(\bar{c} - c(a^1), |\bar{a}^0\big),$$

where the first inequality follows because a^1 is maximal in A and the second because $\tilde{p}(\cdot|\bar{a}^0)$ is strictly decreasing. But then $\hat{p}(\hat{t}(a^1)|a^1) \geq \hat{p}(\hat{t}(\bar{a}^0)|\bar{a}^0)$ by Claim 2. Hence, by Claim 1,

$$p = p(a^n) \geq \hat{p}\big(\hat{t}(a^1)|a^1\big) \geq \hat{p}\big(\hat{t}(\bar{a}^0)|\bar{a}^0\big) = \bar{p}.$$

Consider the case in which $k > 1$. Then there exist two actions a^{k-1} and a^k for which $c(a^{k-1}) \geq c(\bar{a}^0) > c(a^k)$. Notice, $p(a^{k-1}) \geq p(\bar{a}^0)$; if not and $k = 2$, then a^{k-1} could not have been a maximal element and, if $k > 2$, then a^{k-1} could not have been a best response to a^{k-2} because $\bar{a}^0$ would have yielded a strictly higher payoff. Notice also that it must be the case that

$$p(a^k) < \tilde{p}\big(\bar{c} - c(a^k)|\bar{a}^0\big) \leq \tilde{p}\big(\bar{c} - c(\bar{a}^0)|\bar{a}^0\big) = p(\bar{a}^0).$$

If the first inequality did not hold, then $\tilde{p}(\bar{c} - c(a^k)|\bar{a}^0) \leq p(a^k) = \tilde{p}(\bar{c} - c(a^k)|a^k)$, in which case Claim 2 implies that $\hat{p}(\hat{t}(a^k)|a^k) \geq \hat{p}(\hat{t}(\bar{a}^0)|\bar{a}^0)$. Hence, by Claim 1, it must be that $p = p(a^n) \geq \hat{p}(\hat{t}(a^k)|a^k) \geq \hat{p}(\hat{t}(\bar{a}^0)|\bar{a}^0) = \bar{p}$. The second inequality follows because $\tilde{p}(\cdot|\bar{a}^0)$ is decreasing.

It is now shown that $\bar{a}^0$ is a weakly better response to a^{k-1} than a^k, contradicting the claim that $a^k \in \overline{BR}(a^{k-1})$ (since $\bar{a}^0 > a^k$). This is equivalent to showing that

$$\begin{aligned}
&p(\bar{a}^0)\big(p(a^{k-1})w^{11} + (1 - p(a^{k-1}))w^{10}\big) - c(\bar{a}^0) \\
&\quad \geq p(a^k)\big(p(a^{k-1})w^{11} + (1 - p(a^{k-1}))w^{10}\big) - c(a^k) \\
&\iff \quad -\left(\frac{p(\bar{a}^0) - p(a^k)}{c(\bar{a}^0) - c(a^k)}\right) \leq -\left(\frac{1}{p(a^{k-1})w^{11} + (1 - p(a^{k-1}))w^{10}}\right).
\end{aligned}$$

Notice that

$$-\left(\frac{p(\bar{a}^0) - p(a^k)}{c(\bar{a}^0) - c(a^k)}\right) \leq \frac{\tilde{p}\big(\bar{c} - c(\bar{a}^0)|\bar{a}^0\big) - \tilde{p}\big(\bar{c} - c(a^k)|\bar{a}^0\big)}{\big(\bar{c} - c(\bar{a}^0)\big) - \big(\bar{c} - c(a^k)\big)} \leq \tilde{p}'\big(\bar{c} - c(a^k)|\bar{a}^0\big),$$

where the first inequality follows because $p(a^k) < \tilde{p}(\bar{c} - c(a^k)|\bar{a}^0)$ and the second inequality follows because $\tilde{p}(\cdot|\bar{a}^0)$ is concave. Further,

$$-\left(\frac{1}{p(a^{k-1})w^{11} + (1 - p(a^{k-1}))w^{10}}\right) \geq -\left(\frac{1}{p(\bar{a}^0)w^{11} + (1 - p(\bar{a}^0))w^{10}}\right)$$
$$= \tilde{p}'(\bar{c} - c(\bar{a}^0)|\bar{a}^0),$$

where the first inequality follows from $p(a^{k-1}) \geq p(\bar{a}^0)$. But, since $c(\bar{a}^0) \geq c(a^k)$,

$$\tilde{p}'(\bar{c} - c(a^k)|\bar{a}^0) \leq \tilde{p}'(\bar{c} - c(\bar{a}^0)|\bar{a}^0),$$

again by concavity of $\tilde{p}(\cdot|\bar{a}^0)$.

$\bar{p}$ is the greatest lower bound It suffices to exhibit sequence of common action sets (A^n) for which $A^n \supseteq A^0$, $\bar{a}^n$ is the maximal Nash equilibrium action of $\Gamma(w, A^n)$, and

$$p(\bar{a}^n) \to \bar{p} \quad \text{as } n \to \infty.$$

Let $\tilde{c}$ be the maximal cost of any action in A^0 and $\tilde{p}$ be the maximal probability of any action in A^0. Then define $\tilde{p}(\cdot|a)$ as before. Finally, let $\bar{a}^0 \in \arg\max_{a^0} \hat{p}(\hat{t}(a^0)|a^0)$ be chosen so that $\tilde{t}(\bar{a}^0) \geq \tilde{t}(a^0)$ for all $a^0 \in A^0$.[13]

Suppose first that $f(t, \tilde{p}(t|\bar{a}^0))$ exists for all $t \in [0, \tilde{c}]$ so that $\tilde{p}'(\cdot|a)$ and $\tilde{p}''(\cdot|a)$ are bounded:

$$|\tilde{p}'(t|a)| \leq \left|\frac{p'(t|a)(w^{11} - w^{10})}{(\hat{p}(\hat{t}|a)w^{11} + (1 - \hat{p}(\hat{t}|a))w^{10})^2}\right| := \kappa_1 > 0$$

and

$$|\hat{p}''(t|a)| \leq \left|\kappa_1 \frac{(w^{11} - w^{10})}{(\hat{p}(\hat{t}|a)w^{11} + (1 - \hat{p}(\hat{t}|a))w^{10})^2}\right| := \kappa_2 > 0.$$

Now, consider a sequence of common action spaces (A^n), with $A^n := \{a_1^n, a_2^n, \dots, a_n^n\} \cup A^0$. Set $a_1^n = \tilde{p}(\underline{t}|\bar{a}^0)$, where $\underline{t} \in [0, \tilde{c}]$ is such that $\tilde{p}(\underline{t}|\bar{a}^0) = 1$, and $\bar{a}_n := a_n^n$ for each n. Set $c(a_{k-1}^n) - c(a_k^n) = \tilde{c}/n := \epsilon(n)$ for $k = 2, .., n$, $\rho(n) := (1/n^2)(\tilde{c}/(w^{11} + 1))$, and

$$p(a_k^n) = p(a_{k-1}^n) - \frac{\epsilon(n)}{p(a_{k-1}^n)w^{11} + (1 - p(a_{k-1}^n))w^{10}} + \rho(n) \tag{E}$$

for $k = 2, \dots, n$. Notice

$$-\frac{1}{n}\frac{c(a)}{p(a_{k-1}^n)w^{11} + (1 - p(a_{k-1}^n))w^{10}} + \frac{1}{n^2}\frac{c(a)}{w^{11} + 1} < 0$$

[13]Intuitively, $\tilde{p}(\hat{t}(a^0)|a^0)$ may equal zero for many $a^0 \in A^0$. The selection of $\bar{a}^0$ ensures that $\tilde{p}(\cdot|\bar{a}^0)$ hits zero at the largest value of t and, therefore, invoking Claim 2, is always above the differential equations associated with other known actions.

for $k = 2, \dots, n$ so that $a_1^n > a_2^n > \cdots > a_n^n$. Equation (E) approximates $\tilde{p}(t|\bar{a}^0)$ on $[\underline{t}, \bar{c}] \times [0, \bar{p}]$ using Euler's method with rounding error term $\rho(n)$. By the rounding error analysis of Atkinson (1989) (see Theorem 6.3 and Equation (6.2.3)), since $\tilde{p}'(\cdot|a)$ is bounded by $\kappa_1 > 0$, and $\tilde{p}''(\cdot|a)$ is bounded by $\kappa_2 > 0$, it must be the case that

$$\left|p(\bar{a}_n) - \tilde{p}\big(\bar{c}|\bar{a}^0\big)\right| \leq \left(\frac{e^{c(a)\kappa_1} - 1}{\kappa_1}\right)\left(\frac{\epsilon(n)}{2}\kappa_2 + \frac{\rho(n)}{\epsilon(n)}\right).$$

Since $\epsilon(n) \to 0$ as $n \to \infty$ and $\rho(n)/\epsilon(n) = (1/n)(1/(w^{11}+1)) \to 0$ as $n \to \infty$, the right-hand side approaches zero. Hence, $p(\bar{a}^n)$ becomes arbitrarily close to $\tilde{p}(\tilde{c}|\bar{a}^0) = \bar{p}$ as $n \to \infty$.

It remains to argue that (a_n^n, a_n^n) is the maximal Nash equilibrium of $\Gamma(w, A^n)$. For any $a^0 \in A^0$, $\hat{p}(\hat{t}(\bar{a}^0)|\bar{a}^0) \geq \hat{p}(\hat{t}(a^0)|a^0)$. Claim 2 thus ensures that $\tilde{p}(t|\bar{a}^0) \geq \tilde{p}(t|a^0)$ for any $t \in [\underline{t}, \tilde{c}]$ for which both $\tilde{p}(t|\bar{a}^0)$ and $\tilde{p}(t|a^0)$ are defined. Hence, $a_1^n = \bar{a}^0$ is the maximal element of A_n; if there is another action in A^0 that succeeds with probability one, it must have a higher cost. Finally, as Euler's method approximates $\tilde{p}(\cdot|\bar{a}^0)$ from above and there does not exist an element $a^0 \in A^0$ for which $\tilde{p}(t|a^0) > \tilde{p}(t|\bar{a}^0)$ for any $t \in [\underline{t}, \bar{c}]$, $a_k^n \in \overline{BR}(a_{k-1}^n)$ for each n and $k = 2, \dots, n$. This implies that a_n^n is the maximal Nash equilibrium action of $\Gamma(w, A_n)$.

In the case in which $f(t, \tilde{p}(t)|\bar{a}^0)$ does *not* exist for all $t \in [0, \bar{c}]$, there exists some $\bar{t} \in [0, \bar{c}]$ at which $\hat{p}(\bar{t}|\bar{a}^0) = 0$, where $\tilde{p}(\bar{t}|\bar{a}^0)$ is the solution to the differential equation on $[0, \bar{t}] \times [0, p(a)]$. For any interval $[0, \hat{t}]$ such that $\hat{t} < \bar{t}$, mirror the argument in the case in which $f(t, \tilde{p}(t)|\bar{a}^0)$ is well-defined for all $t \in [0, \bar{c}]$ by setting $c(a_{k-1}^n) - c(a_k^n) = \hat{t}/n := \epsilon(n)$ for all $k = 1, .., n$ and $\rho(n) := (1/n^2)(\hat{t}/(w^{11}+1))$ to show that $p(a_n^n)$ approaches $\tilde{p}(\hat{t}|\bar{a}^0)$ as n goes to infinity. But $\hat{t}$ can be chosen arbitrarily close to $\bar{t}$, in which case $\tilde{p}(\hat{t}|\bar{a}^0)$ becomes arbitrarily close to $\tilde{p}(\bar{t}|\bar{a}^0) = 0$. Hence, for any $\epsilon > 0$, there exists a sequence of games with a maximal equilibrium action probability $p(a_n^n)$ converging to a point in $[0, \epsilon)$ as n approaches infinity. This establishes that $\bar{p} = 0$ is the greatest lower bound.

A.3 *Proof of Theorem 2*

Carroll (2015)'s analysis shows that there is a symmetric IPE contract that is optimal within the class of all IPE contracts when $A_1^0 = A_2^0$: $w^{10} = w^{11} = \alpha$, where $\alpha = \sqrt{c(a^0)}/\sqrt{\mathbb{E}_{F(a^0)}[y]} > 0$ for some $a^0 \in A_1^0 = A_2^0$, and $w^{00} = w^{01} = 0$. From (2)–(3), and the rest of the proof of Theorem 1, there thus exists a symmetric JPE contract parameterized by $\phi \in (0, \alpha)$ and $\beta > 0$ with $w^{11} = \phi + \beta$, $w^{10} = \phi$, and $w^{01} = w^{00} = 0$ that strictly outperforms the optimal IPE contract.[14]

Lemma 1, proved in Appendix A.4, establishes that it suffices to compare RPE contracts satisfying $w^{00} = w^{01} = 0$ to JPE contracts. Lemma 2, proved in Appendix A.5, establishes that there is no such RPE contract that outperforms the best symmetric IPE contract. Hence, from the preceding paragraph, if an s-optimal contract exists, then there exists one that exhibits JPE.

[14]A previous working paper, Kambhampati (2024), directly establishes a strict improvement using (11).

Now, observe that an optimal JPE contract with $w^{00} = w^{01} = 0$ solves

$$\max_{w^{11}, w^{10}} \min\{1 - w^{11}, \bar{p}(w^{11}, w^{10})(\bar{p}(w^{11}, w^{10})(1 - w^{11}) + (1 - \bar{p}(w^{11}, w^{10}))(1 - w^{10}))\}$$

subject to

$$\bar{p}(w^{11}, w^{10}) = \max_{a^0 \in A_1^0 = A_2^0} \hat{p}(\hat{t}(a^0; w^{11}, w^{10}) | a^0; w^{11}, w^{10})$$

$$1 \geq w^{11} \geq w^{10} \geq 0,$$

where $\hat{p}(\hat{t}(a^0; w^{11}, w^{10})|a^0; w^{11}, w^{10})$ is defined in the statement of Lemma 6 (the terms that depend on the wage scheme are now made explicit). It is without loss of generality to bound w^{11} above by 1 without altering the solution set because any larger wage yields the principal a profit of at most zero by the first argument of the objective function. Similarly, the strict inequality between w^{11} and w^{10} can be made weak without altering the solution set, because for any wage scheme setting $w^{11} = w^{10}$, there exist wages $w^{11} > w^{10}$ that yield the principal strictly higher profits. As $\mathcal{D} := \{(w^{11}, w^{10}) : 0 \leq w^{10} \leq w^{11} \leq 1\}$ is a closed and bounded subset of $\mathbb{R}^2$, it is compact. Moreover, the objective function is continuous.[15] Hence, the Weierstrass theorem ensures the existence of a solution.

The proof of Lemma 1 shows that any contract setting $w^{11} > 0$ and $w^{00} > 0$ (with $w^{10} = w^{01} = 0$) is strictly suboptimal. Moreover, any other contract is weakly improved upon by an IPE or RPE contract. The uniqueness result follows.

A.4 *Proof of Lemma 1*

In the proof, it will be useful to abuse notation and let A^0 denote the set of common known actions (instead of the set of common known action profiles), A denote an arbitrary set of common actions (instead of an arbitrary set of common action profiles), and $\Gamma(w, A)$ denote a game with common action set A.

If $w^{11} \geq w^{01}$ ($w^{10} \geq w^{00}$), setting $\hat{w}_{11} = w^{11} - w^{01}$ and $\hat{w}_{01} = 0$ ($\hat{w}_{10} = w^{10} - w^{00}$ and $\hat{w}_{00} = 0$) shifts each agent's payoff by a constant. Similarly, if $w^{11} \leq w^{01}$ ($w^{10} \leq w^{00}$), setting $\hat{w}_{01} = w^{01} - w^{11}$ and $\hat{w}_{11} = 0$ ($\hat{w}_{00} = w^{00} - w^{10}$ and $\hat{w}_{10} = 0$) shifts each agent's payoff by a constant. It follows that any Nash equilibrium under w is also a Nash equilibrium under $\hat{w}$. Since the principal's ex post payment decreases, these adjustments must (weakly) increase her payoff.

The argument in the previous paragraph immediately establishes that if $w^{11} \geq w^{01}$ and $w^{10} \geq w^{00}$, then there exists an improved contract $\hat{w}$ in which $\hat{w}_{00} = \hat{w}_{01} = 0$. There are three other cases to consider: (i) $w^{01} \geq w^{11}$ and $w^{00} \geq w^{10}$ (in which case it suffices to set $w^{11} = w^{10} = 0$); (ii) $w^{11} \geq w^{01}$ and $w^{00} > w^{10}$ (in which case it suffices to set $w^{01} = w^{10} = 0$); and (iii) $w^{01} > w^{11}$ and $w^{10} \geq w^{00}$ (in which case it suffices to set $w^{11} = w^{00} = 0$).

[15]This follows from continuity of $\hat{p}(\hat{t}(a^0; w^{11}, w^{10})|a^0; w^{11}, w^{10})$ (see Theorem 4.1 of Coddington and Levinson (1955)), which in turn implies that $\bar{p}$ is continuous (since the maximum of continuous functions is continuous), which in turn implies that $\bar{p}(\bar{p}(1 - w^{11}) + (1 - \bar{p})(1 - w^{10}))$ is continuous. As $1 - w^{11}$ is continuous and the minimum of two continuous functions is continuous, the result follows.

If $w^{01} \geq 0$ and $w^{00} \geq 0$, then w cannot yield the principal a positive payoff (and hence does not outperform the best IPE). To wit, let $A := A^0 \cup \{a^\emptyset\}$ where $p(a^\emptyset) = 0 = c(a^\emptyset)$. Then $a^\emptyset$ is a strictly dominant strategy and so $(a^\emptyset, a^\emptyset)$ is the unique Nash equilibrium. In this equilibrium, the principal obtains a payoff $-2w^{00} \leq 0$.

If $w^{11} \geq w^{01} = 0$ and $w^{00} > w^{10} = 0$, then it must be that $w^{11} > 0$ or the principal could not attain a positive payoff by the argument in the preceding paragraph. Under such a contract, agent i's payoffs satisfy increasing differences in (a_i, a_j) when agent i's action set A_i is equipped with partial order $\succeq_i$: $a_i \succeq_i a_i'$ if either $\mathbb{E}_{F(a_i)}[y_i] > \mathbb{E}_{F(a_i')}[y_i]$, or $\mathbb{E}_{F(a_i)}[y_i] = \mathbb{E}_{F(a_i')}[y_i]$ and $c(a_i) \leq c(a_i')$. Hence, any game this contract induces is supermodular. Moreover, fixing a_j, (a_i, w^{00}) satisfies decreasing differences and (a_i, w^{11}) satisfies increasing differences. Theorem 6 of Milgrom and Roberts (1990) then implies that the maximal and minimal equilibria of any game $\Gamma(w, A)$, where $A \supseteq A^0$ is a common action set, are decreasing in w^{00} and increasing in w^{11}. Since the principal's worst-case payoff either occurs when both agents succeed with probability one or in a region in which increasing the maximal equilibrium action strictly increases the principal's payoff, reducing either w^{11} or w^{00} by a small amount constitutes a strict payoff increase.

If $w^{01} > w^{11} = 0$ and $w^{10} \geq w^{00} = 0$, agent i's payoffs satisfy decreasing differences. It is shown that the principal's payoff under such a contract cannot exceed the principal's payoff under the best IPE contract, v_{IPE}. Let $a^\emptyset$ be the action satisfying $c(a^\emptyset) = p(a^\emptyset) = 0$. Let a_ϵ^* be an action for which $c(a_\epsilon^*) = 0$ and for which $p(a_\epsilon^*)$ is a fixed point of

$$T_\epsilon(p) := \begin{cases} \max\limits_{a \in A^0 \cup \{a^\emptyset\}} \left[p(a) - \dfrac{c(a)}{w^{10} - p\left(w^{10} + w^{01}\right)} \right] + \epsilon & \text{if } w^{10} - p\left(w^{10} + w^{01}\right) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\epsilon > 0$ is small. To see that T_ϵ has a fixed point, notice that, for any $p \in [0, 1]$, $T_\epsilon(p)$ is larger than zero (because $a^\emptyset \in A^0 \cup \{a^\emptyset\}$) and less than one if ϵ is small enough (because A^0 does not contain a zero-cost action that results in success with probability one by the assumption of costly known productive actions). Hence, T_ϵ is a continuous function mapping $[0, 1]$ into $[0, 1]$.

By construction, $(a_\epsilon^*, a_\epsilon^*)$ is a Nash equilibrium of $\Gamma(w, A_\epsilon)$, where $A_\epsilon = A^0 \cup \{a_\epsilon^*, a^\emptyset\}$ is a common action set. Now, consider a sequence of strictly positive values $\epsilon_1, \epsilon_2, \ldots$ that converges to zero and for which there is a convergent sequence of fixed points $p(a_{\epsilon_1}^*), p(a_{\epsilon_2}^*), \ldots$ of the mappings $T_{\epsilon_1}, T_{\epsilon_2}, \ldots$. (Because $[0, 1]$ is a compact set, such a convergent sequence must exist.) Moreover, if the limit p^* satisfies $w^{10} - p^*(w^{10} + w^{01}) > 0$, then it must equal

$$p^* := \max_{a \in A^0 \cup \{a^\emptyset\}} \left[p(a) - \frac{c(a)}{w^{10} - p^*\left(w^{10} + w^{01}\right)} \right].$$

It is shown that the principal's worst-case payoff in the limit can be no larger than what she obtains from the optimal IPE contract. If p^* equals zero, then the principal attains less than zero profits and so lower profits than under the optimal IPE contract.

Otherwise, let $\hat{a}^0$ denote a maximizer of $p(a) - c(a)/(w^{10} - p^*(w^{10} + w^{01}))$ over $A^0 \cup \{a^\emptyset\}$, let $\hat{\alpha} := (1 - p^*)w^{10}$, and notice that the principal attains a payoff of

$$\begin{aligned}
&2\big((p^*)^2 + p^*(1-p^*)(1 - w^{01} - w^{10})\big) \\
&\quad = 2\left(p(\hat{a}^0) - \frac{c(\hat{a}^0)}{(1-p^*)(w^{10}+w^{01})}\right)\big(1 - (1-p^*)(w^{10}+w^{01})\big) \\
&\quad \le 2\left(p(\hat{a}^0) - \frac{c(\hat{a}^0)}{(1-p^*)w^{10}}\right)\big(1 - (1-p^*)w^{10}\big) \\
&\quad = 2\left(p(\hat{a}^0) - \frac{c(\hat{a}^0)}{\hat{\alpha}}\right)(1-\hat{\alpha}).
\end{aligned}$$

But

$$\begin{aligned}
2\left(p(\hat{a}^0) - \frac{c(\hat{a}^0)}{\hat{\alpha}}\right)(1-\hat{\alpha}) &\le 2 \max_{\alpha\in[0,1],\, a^0 \in A^0 \cup \{a^\emptyset\}} \left[(1-\alpha)\left(p(a^0) - \frac{c(a^0)}{\alpha}\right)\right] \\
&= 2 \max_{\alpha\in[0,1],\, a^0 \in A^0} \left[(1-\alpha)\left(p(a^0) - \frac{c(a^0)}{\alpha}\right)\right] \\
&= v_{\mathrm{IPE}},
\end{aligned}$$

where the inequality follows because $p(\hat{a}^0) - c(\hat{a}^0)/\hat{\alpha} \ge 0$ for all $\hat{\alpha} \ge 0$, and the equality follows because setting $\alpha = 1$ yields the principal a payoff of zero given any action in A^0, the payoff attained from choosing $a^\emptyset$ and any $\alpha \in [0, 1]$.

The previous argument establishes that if there exists a K such that, for all $k \ge K$, $(a^*_{\epsilon_k}, a^*_{\epsilon_k})$ is the unique Nash equilibrium of $\Gamma(w, A_{\epsilon_k})$, then the principal's worst-case payoff is no higher than v_{IPE}. But other pure and mixed strategy equilibria may exist that benefit the principal, even as k grows large. First, consider the case in which the limit of $(a^*_{\epsilon_k})$ is $a^\emptyset$. If multiplicity arises, then there exists an action $a^0 \in A^0$ that results in success with strictly positive probability and is a weak best response to any action that succeeds with zero probability; if not, then there would exist a K such that for all $k \ge K$, $(a^*_{\epsilon_k}, a^*_{\epsilon_k})$ is the maximal Nash equilibrium of $\Gamma(w, A_{\epsilon_k})$, and hence the unique Nash equilibrium. If $p(a^0) \le w^{10}/(w^{10} + w^{01})$, then the principal's payoff in any equilibrium in which such an action is played with positive probability is less than zero. This follows from

$$p(a^0)(1 - w^{10} - w^{01}) \le \frac{w^{10}}{w^{10} + w^{01}} - w^{10} < 0.$$

If, on the other hand, $p(a^0) > w^{10}/(w^{10} + w^{01})$, then add to each A_{ϵ_k} the action a'_0 for which $c(a'_0) = 0$ and $p(a'_0) = p(a^0) - c(a^0)/w^{10}$ if $p(a^0) - c(a^0)/w^{10} > w^{10}/(w^{10} + w^{01})$ and $p(a'_0) = w^{10}/(w^{10} + w^{01}) + \epsilon_k$ otherwise. In the first case, the principal attains a payoff of

$$\left(p(a^0) - \frac{c(a^0)}{w^{10}}\right)(1 - w^{10} - w^{01})$$

$$\leq 2 \max_{\alpha \in [0,1],\, a^0 \in A^0} \left[(1-\alpha) \left(p\big(a^0\big) - \frac{c\big(a^0\big)}{\alpha} \right) \right] = v_{\mathrm{IPE}}.$$

In the second case, there exists a K such that for all $k \geq K$, the principal's payoff in the equilibrium $(a_0', a^*_{\epsilon_k})$ is less than zero because the inequality in the previous displayed equation is strict. Finally, no mixed equilibria can exist in any of the cases considered because $a^{\emptyset}$ is a strict best response to any action larger than $w^{10}/(w^{10}+w^{01})$ (the marginal benefit of succeeding with higher probability is less than zero).

Second, consider the case in which the limit productivity is $p^* > 0$. Any other pure or mixed Nash equilibrium of $\Gamma(w, A_{\epsilon_k})$ must involve one agent succeeding with probability $\hat{p} \geq w^{10}/(w^{10}+w^{01}) > p^*$. If not, then $p(a^*_{\epsilon_k})$ would be a best response to the distribution $\hat{p}$ and, if $p(a^*_{\epsilon_k})$ is played, then any distribution $\hat{p}$ could not be a best response. The first statement follows because $p(a^*_{\epsilon_k})$ has zero cost, profits would still be increasing in the probability with which the agent succeeds, and there are strictly decreasing differences. The second follows because $p(a^*_{\epsilon_k})$ is a strict best response to $p(a^*_{\epsilon_k})$ by construction. However, any equilibrium in which one agent generates a distribution $\hat{p}$ must have the other play either $a^{\emptyset}$ (if $\hat{p} > w^{10}/(w^{10}+w^{01})$), $a^*_{\epsilon_k}$ (only if $\hat{p} = w^{10}/(w^{10}+w^{01})$), or a mixture between the two (again, only if $\hat{p} = w^{10}/(w^{10}+w^{01})$); known productive actions are costly and the marginal benefit of succeeding with higher probability is less than zero (strictly so if $\hat{p} > w^{10}/(w^{10}+w^{01})$). It suffices to consider the case in which $\hat{p} > w^{10}/(w^{10}+w^{01})$ and one agent plays $a^{\emptyset}$. In the other two cases, if $\hat{p} < 1$, introducing an action that has a marginally larger productivity than the most productive action in the support of the player's strategy that succeeds with probability $\hat{p}$ reduces the problem to this case. Otherwise, introducing an action that has a marginally smaller cost than the lowest-cost action with distribution $\hat{p} = 1 > w^{10}/(w^{10}+w^{01})$ reduces the problem to this case. So, consider a known common action, $a^0 \in A^0$, satisfying $p(a^0) > w^{10}/(w^{10}+w^{01})$ in the support of the strategy succeeding with probability $\hat{p} > w^{10}/(w^{10}+w^{01})$. Mirroring the argument in the preceding paragraph, add to each A_{ϵ_k} the action a_k' for which $c(a_k') = 0$ and $p(a_k') = p(a^0) - c(a^0)/w^{10} + \epsilon_k$ if $p(a^0) - c(a^0)/w^{10} > w^{10}/(w^{10}+w^{01})$ and $p(a_k') = w^{10}/(w^{10}+w^{01}) + \epsilon_k$ otherwise. These adjustments ensure that a_k' is the unique best response to $a^{\emptyset}$ for every k and so, mirroring the steps in the previous paragraph, the principal attains a payoff no larger than v_{IPE}.

A.5 *Proof of Lemma 2*

In the proof, it will be useful to abuse notation and let A^0 denote the set of common known actions (instead of the set of common known action profiles), A denote an arbitrary set of common actions (instead of an arbitrary set of common action profiles), and $\Gamma(w, A)$ denote a game with common action set A.

The proof will utilize the following result from the theory of supermodular games. As in the preceding proofs, equip A_i with the order $\succeq_i$: $a_i \succeq_i a_i'$ if either $\mathbb{E}_{F(a_i)}[y_i] > \mathbb{E}_{F(a_i')}[y_i]$, or $\mathbb{E}_{F(a_i)}[y_i] = \mathbb{E}_{F(a_i')}[y_i]$ and $c(a_i) \leq c(a_i')$. Let $a_{\max}$ and $a_{\min}$ denote the maximal and minimal elements of A in the corresponding product order, and $\overline{BR} : A \to A$

and $\underline{BR} : A \to A$ denote the maximal and minimal best-response functions for an agent with the common action set of A. Define the mapping

$$\widetilde{BR} : A \times A \to A \times A$$
$$(a_i, a_j) \mapsto \big(\overline{BR}(a_j), \underline{BR}(a_i)\big).$$

Then the following lemma holds.

Lemma 7 (Vives (1990), Milgrom and Roberts (1990)). *Suppose $(\bar{a}, \underline{a})$ is the limit found by iterating $\widetilde{BR}$ starting from the action profile $(a_{\max}, a_{\min})$. If $\Gamma(w, A)$ is submodular, then both $(\bar{a}, \underline{a})$ and $(\underline{a}, \bar{a})$ are Nash equilibria and any other Nash equilibrium action must be smaller than $\bar{a}$ and larger than $\underline{a}$.*

Now, let $a^{\emptyset}$ be the action satisfying $c(a^{\emptyset}) = p(a^{\emptyset}) = 0$. Let a^*_ϵ be an action for which $c(a^*_\epsilon) = 0$ and for which $p(a^*_\epsilon)$ is a fixed point of

$$T_\epsilon(p) := \max_{a^0 \in A^0 \cup \{a^{\emptyset}\}} \left[p\big(a^0\big) - \frac{c\big(a^0\big)}{pw^{11} + (1-p)w^{10}} \right] + \epsilon,$$

where $\epsilon > 0$ is small.[16] To see that T_ϵ has a fixed point, notice that, for any $p \in [0, 1]$, $T_\epsilon(p)$ is larger than zero (because $a^{\emptyset} \in A^0 \cup \{a^{\emptyset}\}$) and less than one if ϵ is small enough (because A^0 does not contain a zero-cost action that results in success with probability one). Hence, T_ϵ is a continuous function mapping $[0, 1]$ into $[0, 1]$.

Now, define a common action set $A_\epsilon = A^0 \cup \{a^*_\epsilon, a^{\emptyset}\}$. If A^0 contains an action producing $y_i = 1$ with probability one, consider the least costly among all of them, $\bar{a}^0$, and add to A_ϵ the action $\bar{a}_\epsilon$, where $c(\bar{a}_\epsilon) = c(\bar{a}^0) - \gamma(\epsilon)$ and $p(\bar{a}_n) = 1 - \gamma(\epsilon)/2$ for $\gamma(\epsilon) := (\epsilon(p(a^*_\epsilon)w^{11} + (1 - p(a^*_\epsilon))w^{10})/2$. Then $\bar{a}_\epsilon$ strictly dominates $\bar{a}^0$ (and so any other action producing $y_i = 1$ with probability one is as well) and a^*_ϵ is a strictly better reply to a^*_ϵ than $\bar{a}_\epsilon$.

It is shown that $(a^*_\epsilon, a^*_\epsilon)$ is the unique Nash equilibrium of $\Gamma(w, A_\epsilon)$. Notice, by construction, $(a^*_\epsilon, a^*_\epsilon)$ is a strict Nash equilibrium. Now, remove all actions producing $y_i = 1$ with probability one since they are strictly dominated by $\bar{a}_\epsilon$. Upon removing these actions, a^*_ϵ strictly dominates any action smaller than it in the order $\succeq_i$. So, remove any actions in $\Gamma(w, A_\epsilon)$ below a^*_ϵ and denote the resulting action space by $\hat{A}$. Now, consider the profile $(\bar{a}, a^*_\epsilon)$, where $\bar{a}$ is the largest element of $\hat{A}$. Since a^*_ϵ is the unique best response to a^*_ϵ (because $(a^*_\epsilon, a^*_\epsilon)$ is a strict Nash equilibrium), the maximal best response to a^*_ϵ is a^*_ϵ. This also implies that a^*_ϵ is the minimal best response to $\bar{a}$; if not, there exists some $\hat{a}^0 \in \hat{A}$ such that $\hat{a}^0 \succ_i a^*_\epsilon$ and $U_i(\hat{a}^0, a^0; w) - U_i(a^*_\epsilon, a^0; w) \geq U_i(\hat{a}^0, \bar{a}; w) - U_i(a^*_\epsilon, \bar{a}; w) > 0$ for any $a^0 \in \hat{A}$, where the first inequality follows from the property of decreasing differences and the second from a^0 being the smallest best response to $\bar{a}$. Hence, $\hat{a}^0$ strictly dominates a^*_ϵ, contradicting the previous observation that a^*_ϵ is a best response to a^*_ϵ. As $(a^*_\epsilon, a^*_\epsilon)$ is a fixed point of $\widetilde{BR}$, $(a^*_\epsilon, a^*_\epsilon)$ is the limit found by iterating $\widetilde{BR}$ from $(\bar{a}, a^*_\epsilon)$ or

[16]Interpret $-c(a^0)/(pw^{11} + (1-p)w^{10})$ as zero if the denominator is zero and $c(a^0) = 0$ and $-\infty$ if the denominator is zero and $c(a^0) > 0$.

$(a^*_\epsilon, \bar{a})$ in $\Gamma(w, \hat{A})$. By Lemma 7, it follows that $(a^*_\epsilon, a^*_\epsilon)$ is the unique Nash equilibrium of $\Gamma(w, \hat{A})$, and hence of $\Gamma(w, A_\epsilon)$.

Now, consider a sequence of strictly positive values $\epsilon_1, \epsilon_2, \ldots$ that converges to zero and for which there is a convergent sequence of fixed points $p(a^*_{\epsilon_1}), p(a^*_{\epsilon_2}), \ldots$ of the mappings $T_{\epsilon_1}, T_{\epsilon_2}, \ldots$. Since $[0, 1]$ is a compact set, such a convergent sequence must exist. Moreover, its limit is the distribution

$$p(a^*) = \max_{a^0 \in A^0 \cup \{a^\emptyset\}} \left[p(a^0) - \frac{c(a^0)}{p(a^*)w^{11} + (1 - p(a^*))w^{10}} \right].$$

Let $\hat{a}^0 \in A^0 \cup \{a^\emptyset\}$ denote the maximizer on the right-hand side and define $\hat{\alpha} := p(a^*)w^{11} + (1 - p(a^*))w^{10}$. The principal's payoff in the unique equilibrium $(a^*_{\epsilon_k}, a^*_{\epsilon_k})$ of $\Gamma(w, A_{\epsilon_k})$ as k grows large becomes arbitrarily close to

$$\begin{aligned} &2\big(p(a^*)\big)\big(p(a^*)(1 - w^{11}) + (1 - p(a^*))(1 - w^{10})\big) \\ &\quad = 2\left(p(\hat{a}^0) - \frac{c(\hat{a}^0)}{\hat{\alpha}}\right)(1 - \hat{\alpha}) \le 2 \max_{\alpha \in [0,1], a^0 \in A^0 \cup \{a^\emptyset\}} \left[(1 - \alpha)\left(p(a^0) - \frac{c(a^0)}{\alpha}\right)\right], \end{aligned}$$

where the inequality follows because $p(\hat{a}^0) - c(\hat{a}^0)/\hat{\alpha} \ge 0$ for all $\hat{\alpha} \ge 0$ and so it suffices to consider values of α between zero and one to maximize $(1 - \alpha)(p(a^0) - c(a^0)/\alpha)$ for any $a^0 \in A^0 \cup \{a^\emptyset\}$. But

$$\begin{aligned} &2 \max_{\alpha \in [0,1], a^0 \in A^0 \cup \{a^\emptyset\}} \left[(1 - \alpha)\left(p(a^0) - \frac{c(a^0)}{\alpha}\right)\right] \\ &\quad = 2 \max_{\alpha \in [0,1], a^0 \in A^0} \left[(1 - \alpha)\left(p(a^0) - \frac{c(a^0)}{\alpha}\right)\right] \\ &\quad = v_{\mathrm{IPE}}, \end{aligned}$$

where v_{IPE} is the principal's payoff under the best IPE, because setting $\alpha = 1$ yields the principal a payoff of zero given any action in A^0, the same payoff attained from choosing $a^\emptyset$ and any $\alpha \in [0, 1]$.

References

Alchian, Armen A. and Harold Demsetz (1972), "Production, information costs, and economic organization." *American Economic Review*, 62, 777–795. [1153]

Antic, Nemanja (2021), "Contracting with unknown technologies." Unpublished manuscript, Northwestern University. [1154]

Atkinson, Kendall E. (1989), *An Introduction to Numerical Analysis*, second edition. Wiley, New York. [1176]

Brandenburger, Adam and Eddie Dekel (1987), "Rationalizability and correlated equilibria." *Econometrica*, 1391–1402. [1156]

Carroll, Gabriel (2015), "Robustness and linear contracts." *American Economic Review*, 105, 536–563. [1152, 1154, 1157, 1159, 1164, 1176]

Chassang, Sylvain (2013), "Calibrated incentive contracts." *Econometrica*, 81, 1935–1971. [1154]

Che, Yeon-Koo and Seung-Weon Yoo (2001), "Optimal incentives for teams." *American Economic Review*, 91, 525–541. [1153, 1155]

Coddington, Earl A. and Norman Levinson (1955), *Theory of Ordinary Differential Equations*. McGraw-Hill Book Company, Inc. [1173, 1177]

Dai, Tianjiao and Juuso Toikka (2022), "Robust incentives for teams." *Econometrica*, 90, 1583–1613. [1154]

Fleckinger, Pierre (2012), "Correlation and relative performance evaluation." *Journal of Economic Theory*, 147, 93–117. [1153, 1161]

Fleckinger, Pierre, David Martimort, and Nicolas Roux (2024), "Should they compete or should they cooperate? The view of agency theory." *Journal of Economic Literature*. [1161]

Frankel, Alexander (2014), "Aligned delegation." *American Economic Review*, 104, 66–83. [1154]

Garrett, Daniel F. (2014), "Robustness of simple menus of contracts in cost-based procurement." *Games and Economic Behavior*, 87, 631–641. [1154]

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci (2004), "Differentiating ambiguity and ambiguity attitude." *Journal of Economic Theory*, 118, 133–173. [1158]

Hackman, J. Richard (2002), *Leading Teams: Setting the Stage for Great Performances*. Harvard Business Press. [1154]

Holmström, Bengt (1979), "Moral hazard and observability." *Bell Journal of Economics*, 10, 74–91. [1151]

Holmström, Bengt (1982), "Moral hazard in teams." *Bell Journal of Economics*, 13, 324–340. [1151]

Hurwicz, Leonid (1951), "A class of criteria for decision-making under uncertainty." Cowles Commission Discussion Paper, Statistics No. 356. [1158]

Hurwicz, Leonid and Leonard Shapiro (1978), "Incentive structures maximizing residual gain under incomplete information." *Bell Journal of Economics*, 180–191. [1154]

Itoh, Hideshi (1991), "Incentives to help in multi-agent situations." *Econometrica*, 59, 611–636. [1153]

Kambhampati, Ashwin (2023), "Randomization is optimal in the robust principal-agent problem." *Journal of Economic Theory*, 207, 105585. [1155]

Kambhampati, Ashwin (2024), "Robust performance evaluation of independent and identical agents." arXiv:2401.16542. [1154, 1176]

Lazear, Edward P. (1989), "Pay equality and industrial politics." *Journal of Political Economy*, 97, 561–580. [1153]

Marku, Keler, Sergio Ocampo Diaz, and Jean-Baptiste Tondji (2024), "Robust contracts in common agency." *The RAND Journal of Economics*, 55, 199–229. [1154]

Milgrom, Paul and John Roberts (1990), "Rationalizability, learning, and equilibrium in games with strategic complementarities." *Econometrica*, 1255–1277. [1157, 1168, 1171, 1178, 1181]

Rees, Daniel I., Jeffrey S. Zax, and Joshua Herries (2003), "Interdependence in worker productivity." *Journal of Applied Econometrics*, 18, 585–604. [1151, 1152, 1153, 1163, 1164]

Rosenthal, Maxwell (2023), "Robust incentives for risk." *Journal of Mathematical Economics*, 109, 102893. [1154]

Shavell, Steven (1979), "On moral hazard and insurance." *Quarterly Journal of Economics*, 93, 541–562. [1151]

Vives, Xavier (1990), "Nash equilibrium with strategic complementarities." *Journal of Mathematical Economics*, 19, 305–321. [1157, 1168, 1171, 1181]

Co-editor Simon Board handled this manuscript.

Manuscript received 9 December, 2022; final version accepted 14 April, 2024; available online 16 April, 2024.