Robust Performance Evaluation*

Ashwin Kambhampati†

November 10, 2020
Click for Latest Version

Abstract

I consider a moral hazard in teams model in which a principal knows that the agents she compensates are identical and technologically independent, but does not know all of the actions they can take. I show that any worst-case optimal contract exhibits *joint performance evaluation* and is *nonlinear* in team output. Hence, when robustness is a concern, nonlinear team-based incentive schemes—such as team bonuses and employee stock options—are justified, even if tasks are independent and individual performances are uncorrelated. This result contrasts with the classical theory of incentives, which finds *independent performance evaluation* to be Bayesian optimal, and with the recent literature on robust contracting with unbounded uncertainty, which finds *linear* incentive schemes to be worst-case optimal. Moreover, it reveals a new channel leading to the optimality of joint performance evaluation and formalizes a longstanding idea that interdependent incentive schemes are advantageous due to their flexibility.

*I thank my dissertation committee–Aislinn Bohren, George Mailath, Steven Matthews, and Juuso Toikka–for their time, guidance, and encouragement. I also thank Hal Cole, David Dillenberger, Hanming Fang, Kevin He, Jan Knoepfle, Rohit Lamba, Guillermo Ordoñez, Andrew Postlewaite, Ricardo Serrano-Padial and Rakesh Vohra for helpful discussions and valuable perspective. Last, but not least, I thank my classmates, especially Alejandro Sanchez Becerra, Sherwin Lott, Youngsoo Heo, Mallick Hossain, and Carlos Segura-Rodriguez.

†Department of Economics, University of Pennsylvania; akambh@sas.upenn.edu.

Contents

1 Introduction

Both team-based incentive pay and the use of teams are on the rise (Lazear and Shaw (2007), Deloitte (2016)). This can be rationalized by appealing to *teamwork*: if a worker's success depends on the actions of others, then team-based incentives have the positive effect of encouraging help (Itoh (1991)) and discouraging sabotage (Lazear (1989)). But, such incentives also arise in settings in which work is done independently and individual performances are uncorrelated. For instance, each member of a sales force may make sales calls alone and oversee a distinct market segment; members of a start-up may contribute to the same "vision" of an entrepreneur, but perform independent tasks. In these cases, team-based incentive pay is also common: salespeople receive bonuses for their division's performance and start-up members are compensated using equity or stock options in addition to wages.

This paper shows that team-based incentive pay for independent agents is optimal when *robustness* is a concern. Specifically, if a principal knows that the agents she compensates are identical and technologically independent, but does not know all of the actions they can take, any worst-case optimal contract exhibits *joint performance evaluation*. Furthermore, the optimal form of joint performance evaluation is *nonlinear* in the value the team produces. This result departs from the classical theory of incentives, which finds *independent performance evaluation* to be Bayesian optimal (Holmström (1982)), and from the recent literature on robust contracting in settings with unrestricted productive interdependency, which finds *linear* joint performance evaluation incentive schemes to be worst-case optimal (Dai and Toikka (2018), Walton and Carroll (2019)).

Formally, I study a model in which a principal writes a symmetric contract for two agents who have access to a common set of actions. Actions are costly and unobservable, and affect the probability with which each agent succeeds at her task. There is no productive relationship between the agents and individual success is statistically independent, across agents, conditional on the actions they take. All parties are risk-neutral and the agents are protected by limited liability.

In contrast to standard models, the principal knows only a subset of the com-

mon actions available to the agents, and there may be others she does not know. For instance, a sales manager may know that her sales representatives can follow the company's script. But, there are a myriad of ways in which a sales representative might deviate from this script. Hence, the principal evaluates each contract according to its worst-case performance across all action sets containing the ones that she knows. An optimal contract with respect to this criterion is a *worst-case optimal* contract.

My main result, Theorem 1, is that any worst-case optimal contract exhibits joint performance evaluation—each agent's wage increases in the other's success–and is nonlinear—agents are *not* paid a constant share of the total value of completed tasks. Moreover, a worst-case optimal contract exists.

The logic behind the suboptimality of linear contracts is as follows. Suppose, towards contradiction, that there was a worst-case optimal contract that was linear, i.e. a joint performance evaluation contract in which agents are paid a constant share of team output. Then, each agent would be paid strictly positive wages for the success of the other agent, even when she herself does not succeed. But, given productive independence between the agents, one agent's action cannot affect the probability of the other's success. So, the principal could simply reduce wages by a constant and leave individual incentives unchanged. This adjustment strictly decreases expected wage payments without affecting productivity, thereby strictly increasing the principal's expected payoff.

To understand the intuition behind the optimality of joint performance evaluation, it is instructive to consider the following benchmark contract: Each agent receives $w^* > 0$ if she succeeds and zero otherwise, unconditional on the other's success or failure. Moreover, w^* targets a known action a_0. This contract exhibits independent performance evaluation since it does *not* link one agent's compensation to the performance of the other.

I argue that the principal can improve her worst-case expected payoff by *calibrating* a joint performance evaluation contract to the pair (w^*, a_0). Suppose, relative to w^*, the principal increases the wage of an agent when she succeeds and the other does as well, but reduces her wage when she succeeds and the other fails. If this adjustment keeps her expected wage constant, *conditional on the other agent*

taking the targeted known action a_0, then her incentive to take an unknown, less productive action is exactly the same as if she were offered the unconditional wage w^*. Crucially, however, if both agents take this unknown action, then the principal *reduces* her expected wage payments; joint performance evaluation implies that each is paid less when the other is less productive.

The key finding of my analysis is that the basic *rent-extraction benefit* described in the preceding paragraph can be made to dominate any negative effects such contracts have on *efficiency*, even though such costs can be considerable when there are many unknown actions.[1] Specifically, there always exists a calibrated joint performance evaluation contract that strictly outperforms any independent performance evaluation contract. Moreover, the best joint performance evaluation contract strictly outperforms any other contract, including those exhibiting *relative performance evaluation*.

The contribution of my paper to the literature is threefold. First, it establishes a fundamentally new channel leading to the unique optimality of joint performance evaluation. In the Bayesian contracting paradigm, the Informativeness Principle of Holmström (1979) and Shavell (1979) prescribes independent performance evaluation whenever one agent's performance is statistically uninformative of another's action. Hence, the literature has sought justification for interdependent incentive schemes, such as relative performance evaluation and joint performance evaluation, by introducing productive or informational linkages between agents.[2] My model explicitly rules out these channels in order to isolate the effect of robustness considerations. I thus rationalize empirical evidence documenting firms' preference for joint performance evaluation, such as team bonuses, in cases in which

[1] I exhibit a n-sequence of dominance solvable games with n unknown actions in which agents "undercut" each other as dominated strategies are eliminated, taking progressively less costly and less productive actions. Efficiency losses are maximized as n grows large (Lemma 6).

[2] In the absence of productive interaction, joint performance evaluation may be optimal if agents are affected by a common, negatively correlated productivity shock (Fleckinger (2012)). In the absence of a common shock, joint performance evaluation may be optimal if efforts are complements in production (Alchian and Demsetz (1972)), if it induces help between 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)). See Fleckinger (2012) for a comprehensive analysis of the Bayesian version of the model I study, and Fleckinger and Roux (2012) for an excellent survey of the above literature.

the production and information technologies are independent (Rees, Zax and Herries (2003)). I also offer an explanation for the use of stock options to compensate members of start-ups.[3]

Second, my paper formalizes a longstanding idea that interdependent incentives have an advantage over individual incentives because of their *flexibility*. As Nalebuff and Stiglitz (1983) write,

> "The incentive compensation scheme that is "correct" in one situation will not in general be correct in another. In principle, there could be a different incentive structure for each set of environmental variables. Such a contract would obviously be prohibitively expensive to set up; but more to the point, many of the relevant environmental variables are not costlessly observable to all parties to the contract. Thus, a single incentive structure must do in a variety of circumstances. The lack of flexibility of the piece rate system is widely viewed to be its critical shortcoming: the process of adapting the piece rate is costly and contentious."

In contrast to Nalebuff and Stiglitz (1983), who study a model in which interdependent incentives outperform independent incentives due to their screening ability,[4] I explicitly account for the principal's desire for flexibility by assuming that she uses a max-min criterion to evaluate contracts; a max-min optimal contract performs well across all environments the principal deems feasible. That joint performance evaluation emerges as optimal thus provides a formal justification for the assertion that such schemes are more flexible than individual performance evaluation.

Third, my paper contributes to the growing literature on robust contracting by considering a principal agent model in which the principal has bounded, non-quantifiable uncertainty about the production technology.[5] Carroll (2015) considers a principal-single agent model in which the principal has non-quantifiable

[3]Understood through the Informativeness Principle, such schemes are puzzling given the premise that all members exploit the same underlying technology. If anything, this seems to suggest success should have positive conditional correlation, leading *relative performance evaluation* to be optimal. Fleckinger (2012) develops this point further and offers another explanation based on effort-controlled noise.

[4]For related contributions, see Lazear and Rosen (1981), who consider the optimality of competitive incentives versus piece rates in a setting with a common shock and risk-neutral agents, and Green and Stokey (1983), who consider the case of risk-averse agents.

[5]Related work not discussed in this paragraph include the papers of Hurwicz and Shapiro

uncertainty about the actions available to the agent. His main result is that there exists a worst-case optimal contract that is linear in individual output. My model and analysis enrich that of Carroll (2015) by introducing a seemingly irrelevant agent and showing that multiple agents lead to the optimality of joint incentive schemes.[6] Dai and Toikka (2018) consider a principal-many agent model in which the principal has non-quantifiable, unbounded uncertainty, i.e. she considers all games that the agents might be playing. They find that linear contracts are robustly optimal. This result is driven by the finding that any contract that induces competition between agents is non-robust to a game in which agents sabotage one another, leading the principal to a worst-case payoff of zero. In contrast to Dai and Toikka (2018), I consider a setting in which the principal *knows* that success is independently distributed across agents. This has the immediate effect of ruling out sabotage and ensuring that linear contracts are suboptimal. It also necessitates new techniques to analyze the principal's worst-case payoffs.[7]

The results of this paper complement Dai and Toikka (2018) in terms of their management implications. Agents in Dai and Toikka (2018)'s model are a "real team" in the sense that they work together to produce value for the principal, while agents in my model 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 linear in the value the team generates for the principal, while in the case of co-acting agents it involves nonlinear bonus payments that reward agents

(1978), Garrett (2014), and Frankel (2014), who consider contracting with unknown preferences; Rosenthal (2020) who considers contracting with unknown risk preferences; Marku and Ocampo Diaz (2019), who consider a robust common agency problem; and Chassang (2013) who studies the robust performance guarantees of a different class of calibrated contracts in a dynamic agency problem.

[6]Building upon Carroll (2015)'s single-agent model, Antic (2015) imposes bounds on the principal's uncertainty over unknown actions (see also Section 3.1 of Carroll (2015), which studies lower bounds on costs). In particular, Antic (2015) posits a lower bound on the distribution over output given any unknown technology. In contrast, my model places no restrictions on the technology available to each agent in isolation beyond those of Carroll (2015). Instead, the restrictions I impose concern the relationship between the agents.

[7]For instance, the worst-case payoff of the principal at the optimal contract is achieved by a sequence of games in which the number of actions grows to infinity, rather than one additional action for each agent as in Dai and Toikka (2018).

when all succeed.

The rest of the paper is organized as follows: Section 2 illustrates the mechanism behind the main result using a simple example; Section 3 presents the model; Section 4 states and proves Theorem 1; Section 5 discusses extensions; and Section 6 concludes.

2 Simple Example

In this section, I study a simple example illustrating the rent-extraction benefit of joint performance evaluation relative to independent performance evaluation.

2.1 Set Up

Consider a scenario in which a risk-neutral manager compensates two identical, risk-neutral agents who perform independent tasks and are protected by limited liability. Successful completion of a task yields the manager a utility value of one and failure yields her a utility value of zero. The manager knows that each agent can take an action, call it "work", that results in the successful completion of her task with probability one. However, the manager is concerned about another action available to each agent, call it "shirk", that results in the successful completion of her task with probability $p^* \in [0,1)$, and failure with complementary probability. The manager knows that work incurs a disutility cost of effort of $\frac{1}{4}$ and shirk incurs zero disutility. However, she does not know the value of p^*.

The manager contemplates using two types of contracts, both of which respect the limited liability constraint that ex-post wage payments must be negative.

1. **Independent Performance Evaluation:** Pay each agent $w \in (0,1)$ for success. Pay each agent 0 for failure.

2. **Joint Performance Evaluation:** Pay each agent $w \in (0,1)$ for success when the other agent also succeeds, and $w-\epsilon$, with $\epsilon \in (0,w]$, when the other agent fails. Pay each agent 0 for failure. This contract is calibrated to (w, work) in the following sense: If an agent succeeds and the other agent works, she

obtains the same expected wage as in the case in which she were offered an independent performance evaluation contract with wage w. In particular,

$$1 \times w + 0 \times (w - \epsilon) = w.$$

The manager evaluates any contract according to the same criterion. First, for each value of p^*, she computes her expected payoff in the worst Pareto Efficient Nash equilibrium (from her perspective) in the game induced by the contract she offers. Second, she computes the infimum value of her expected payoff over all values of $p^* \in [0, 1)$. The resulting payoff is called her *worst-case payoff*. Can joint performance evaluation yield the manager a higher worst-case payoff than independent performance evaluation?

2.2 Independent Performance Evaluation

An independent performance evaluation contract with wage w, together with an actual value of p^*, induces the game between the agents depicted in Figure 1.

	work	shirk
work	$w - \frac{1}{4}$, $w - \frac{1}{4}$	$w - \frac{1}{4}$, $p^* w$
shirk	$p^* w$, $w - \frac{1}{4}$	$p^* w$, $p^* w$

Figure 1: *Game induced by IPE w and p^*.*

A naïve intuition is that the worst-case scenario for the principal occurs when $p^* = 0$; if agents take a shirking action with this success probability, then the principal attains an expected payoff of zero. But, this logic ignores incentives. In particular, each agent has a (weak) incentive to shirk if and only if she obtains a higher expected utility from doing so, i.e.

$$p^* w \geq w - \frac{1}{4} \iff p^* \geq 1 - \frac{1}{4w}.$$

So, whenever p^* is strictly smaller than $1 - \frac{1}{4w}$, (work, work) is the unique Nash equilibrium, yielding the principal an expected payoff per agent of $1 - w$.

Instead, the principal's worst-case payoff from a contract w is attained when $p^* = 1 - \frac{1}{4w}$, just high enough to make shirking attractive to each agent. In this case, (shirk, shirk) is the manager's least-preferred Pareto Efficient Nash equilibrium.[8] In it, the principal obtains a payoff per agent of

$$\underbrace{(1 - \frac{1}{4w})}_{\text{Expected Task Value}} \quad - \quad \underbrace{(1 - \frac{1}{4w})w}_{\text{Expected Wages}} .$$

A simple calculation then shows that $w^* = \frac{1}{2}$ maximizes this expression, yielding the principal a maximum worst-case payoff of $\frac{1}{2}$, as shown by Carroll (2015).

2.3 Joint Performance Evaluation

Can the manager obtain a higher worst-case payoff from a joint performance evaluation contract? Consider the joint performance evaluation contract calibrated to (w^*, work) with wages $(w^*, w^* - \epsilon)$. The game between the agents for a given value of p^* is depicted in Figure 2.

	work	shirk
work	$w^* - \frac{1}{4}$, $w^* - \frac{1}{4}$	$p^* w^* + (1 - p^*)(w^* - \epsilon) - \frac{1}{4}$, $p^* w^*$
shirk	$p^* w^*$, $p^* w^* + (1 - p^*)(w^* - \epsilon) - \frac{1}{4}$	$p^*(p^* w^* + (1 - p^*)(w^* - \epsilon))$, $p^*(p^* w^* + (1 - p^*)(w^* - \epsilon))$

Figure 2: *Game induced by JPE* $(w, w - \epsilon)$ *and* p^*.

The crucial property of calibration to (w^*, work) is that the incentive for each agent to shirk, given that the other agent takes the targeted action work, is identical to the case in which each is offered an independent performance evaluation

[8] There is multiplicity of Nash equilibria, but none are Pareto dominated.

contract with wage w^*. Put differently, as in the case of independent performance evaluation, (work, work) is a Nash equilibrium whenever

$$p^* \leq 1 - \frac{1}{4w^*}.$$

Moreover, whenever (work, work) is a Nash equilibrium it Pareto dominates any other Nash equilibrium; one agent working generates a positive externality on the other since an agent is more likely to receive w^* than $w^* - \epsilon$ conditional on success. So, the principal obtains a payoff per agent of $1 - w^*$, as before.

Given any joint performance evaluation contract $(w^*, w^* - \epsilon)$, the principal's infimum payoff is attained as p^* approaches $1 - \frac{1}{4w^*}$ from above. Along this sequence, (shirk, shirk) is the unique Nash equilibrium. Since $p^* < 1$, a simple calculation shows that the principal's payoff from each agent in the worst-case limit is strictly larger than her worst-case payoff per agent given the worst-case optimal IPE:

$$\underbrace{p^*}_{\text{Expected Task Value}} \quad \underbrace{-p^*(p^*w^* + (1-p^*)(w^* - \epsilon))}_{\text{Expected Wages}} > (1 - \frac{1}{4w^*}) - (1 - \frac{1}{4w^*})w^*.$$

In a nutshell, calibration ensures that the worst-case expected task value is no lower than in the worst-case scenario given the benchmark independent performance evaluation contract. But, the principal is able to pay each agent less in expectation when each shirks, since joint performance evaluation is *responsive* to the other agent's shirking.

3 Model

3.1 Environment

A risk-neutral principal writes a contract for two risk-neutral agents, indexed by $i = 1, 2$. Each agent i chooses an unobservable, costly action, a_i, from a common, finite set $A \subset \mathbb{R}_+ \times [0, 1]$. Each action a_i is identified by its cost, $c(a_i) \in \mathbb{R}_+$, and the probability with which it results in success, $p(a_i) \in [0, 1]$. Let $y_i = 1$ denote success

and $y_i = 0$ denote failure. There are neither informational linkages across agents,

$$Pr(y_i, y_j | a_i, a_j) = Pr(y_i | a_i, a_j) Pr(y_j | a_i, a_j),$$

nor productive linkages across agents,

$$Pr(y_i | a_i, a_j) = Pr(y_i | a_i) = \begin{cases} p(a_i) & \text{if } y_i = 1 \\ 1 - p(a_i) & \text{if } y_i = 0. \end{cases}.$$

A **contract** is a quadruple of non-negative wages,

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

where the first index of each wage indicates an agent's own success or failure and the second indicates the success or failure of the other agent. I impose the assumption that contracts are symmetric throughout, postponing a discussion of asymmetric contracts to Section 5.1.

It will be useful to classify the resulting contracts according to the typology of Che and Yoo (2001).

Definition 1 (Performance Evaluations)
A contract w 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.

While this typology is non-exhaustive (for instance, when $w_{11} > w_{10}$ and $w_{01} < w_{00}$ there is JPE "at the top" and RPE "at the bottom"), I will show later that it is without loss of generality to consider contracts for which $w_{01} = w_{00} = 0$ (Lemma 4). Within this class of contracts, it *is* exhaustive. I now distinguish between linear and nonlinear JPE.

Definition 2 (Linear JPE)
*A JPE is **linear** if*

$$w_{y_i y_j} = \alpha(y_i + y_j) \text{ for some } \alpha \in [0,1]$$

*and **nonlinear** otherwise.*

3.2 Principal's Problem

Agent i's ex post payoff given a contract w, action profile (a_i, a_j), and realization (y_i, y_j) is

$$w_{y_i y_j} - c(a_i),$$

while her expected payoff is

$$U_i(a_i, a_j; w) := \sum_{y_i \in Y} \sum_{y_j \in Y} Pr(y_i, y_j | a_i, a_j) w_{y_i y_j} - c(a_i).$$

Let $\Gamma(w, A)$ denote the normal form game induced by the contract w and $\mathcal{E}(w, A)$ denote its (non-empty) set of mixed strategy Nash equilibria. As non-IPE contracts tie the incentives of agents together, agents may have an incentive to discuss their strategies with one another, even if they cannot make binding commitments. This would deem equilibria that are strictly Pareto dominated implausible, i.e. equilibria $\sigma \in \mathcal{E}(w, A)$ for which there exists another equilibrium $\sigma' \in \mathcal{E}(w, A)$ that makes each agent strictly better off. I thus require that agents play a (weakly) Pareto Efficient Nash equilibrium. Denote the set of such equilibria by $\mathcal{E}_P(w, A)$.

The principal's ex post payoff given a contract w and realization (y_1, y_2) is

$$y_1 + y_2 - w_{y_1 y_2} - w_{y_2 y_1},$$

while her expected payoff is

$$V(w, A) := \min_{\sigma \in \mathcal{E}_P(w, A)} E_\sigma[y_1 + y_2 - w_{y_1 y_2} - w_{y_2 y_1}].$$

In the spirit of a worst-case analysis, I do *not* allow the principal to select her preferred Pareto Efficient Nash Equilibrium when there is multiplicity of equilibria,

i.e. when $\mathcal{E}_P(w, A)$ is not a singleton.[9]

When the principal writes a contract for the agents, she has limited knowledge about the game the agents play. In particular, she knows only a non-empty subset of actions available to them $A^0 \subseteq A$. In the face of her uncertainty, the principal evaluates each contract on the basis of its performance across all finite supersets of her knowledge contained in $\mathbb{R}_+ \times [0,1]$. The **worst-case payoff** she receives from a contract w is thus given by

$$V(w) := \inf_{A \supseteq A^0} V(w, A).$$

The principal's problem is to identify a contract w^* for which

$$V(w^*) = \sup_{w} V(w).$$

Call such a contract a **worst-case optimal contract**.

To rule out trivial cases, I make the following assumptions about A^0.

Assumption 1
The known action set A^0 has the following properties:

1. *(Non-Triviality) There exists an action $a_0 \in A^0$ such that $p(a_0) - c(a_0) > 0$.*
2. *(Known Productive Actions are Costly) If $a_0 \in A^0$ and $p(a_0) > 0$, then $c(a_0) > 0$.*

The first assumption ensures that the principal can possibly obtain a strictly positive worst-case payoff from contracting with the agents. The second ensures that the principal's supremum payoff is never approached by a sequence of contracts converging to one always paying each agent zero.[10]

[9] In the classical Bayesian contracting literature as well as in recent work on robust contracting (e.g., Carroll (2015), Dai and Toikka (2018), and Walton and Carroll (2019)), the principal has the power to select her preferred Nash equilibrium. The primary role of the assumption is technical convenience; it ensures the existence of a worst-case optimal contract. I will not need such an assumption to obtain existence–ruling out Pareto-dominated equilibria is enough.

[10] The assumption that known productive actions are costly is stronger than necessary for my main result (for instance, if there is a zero-cost productive action that the principal does not optimally "target", then the result goes through). Nonetheless, it has the advantage of being easy to interpret.

3.3 Interpretation

The principal's problem can be re-phrased as follows: If the principal must use the same contract, i.e. mapping from successes and failures into wages, given any feasible set of actions the agents might have available, which one does the best in the sense of yielding the highest payoff guarantee? The solution to the problem is a positive description of how a principal might write a contract in the face of structured uncertainty about the agents' environment.

Implicit in this formulation is that contracts can only depend on observable successes and failures and not on the technology of the agents. This raises the question: Can the principal simply ask the agents to report their action set and implement the Bayesian-optimal contract for each action set? One notable feature of my formulation relative to the literature is that I assume that the agents can discuss their strategies with one another before taking an action (captured by the assumption that they play a Pareto Efficient Nash equilibrium). If this interpretation is extended to include the reporting stage, then implementing the Bayesian-optimal contract technology-by-technology would not be incentive compatible; it is easy to construct action sets for which the agents would always prefer to coordinate their reports to exaggerate the cost of the principal's targeted action.

More generally, however, one could ask if it is possible to implement other social choice functions–through either direct or indirect mechanisms– and study whether the principal obtains a better worst-case guarantee. Further study of this important normative problem awaits.

4 Worst-Case Optimal Contracts

4.1 Main Result

My main result shows that the rent-extraction benefit of JPE described in Section 2 is powerful enough to ensure that there exists a contract of this form that is worst-case optimal. Put differently, in spite of the efficiency losses such contracts induce in any game the agents might be playing, no other contract can do better. Moreover, I prove that *any* worst-case optimal contract must be a JPE and that it

must be *nonlinear*.

Theorem 1

Any worst-case optimal contract is a nonlinear JPE. There exists a worst-case optimal contract.

The key intuition behind the result is that by judiciously calibrating a JPE to a benchmark IPE, any (worst-case) efficiency losses such contracts generate can be made approximately the same as those of the benchmark contract. Thus, the reduction in expected wage payments the principal obtains when agents take less productive actions, due to the responsiveness property of JPE outlined in Section 2, causes JPE to outperform the benchmark contract. Of course, to show that only nonlinear JPE can be worst-case optimal, I must also prove strict suboptimality of contracts other than IPE, including those that exhibit RPE and those that reward agents with positive expected wages when they fail.

The remainder of this section is devoted to proving Theorem 1. I first review some preliminaries from the theory of supermodular games, then present the arguments that rule out IPE and RPE contracts as worst-case optimal.

4.2 Preliminaries: Supermodular Games

Equip any action set A with the total order $\succeq$: $a_i \succeq a_j$ if either $p(a_i) > p(a_j)$, or $p(a_i) = p(a_j)$ and $c(a_i) \leq c(a_j)$.[11] In words, a_i is higher than a_j if a_i results in success with a higher probability or if it results in success with the same probability, but at a lower cost. Then, $(A, \succeq)$ is a complete lattice; all subsets of A have both a maximum and a minimum. A supermodular game may thus be defined as follows.[12]

Definition 3 (Supermodular Game)

The game $\Gamma(w, A)$ *is* ***supermodular*** *if* U_i *exhibits increasing differences:* $a_i' \succeq a_i$ *and*

[11] It is easy to verify that this relation is antisymmetric (if $a_i \succeq a_j$ and $a_i \preceq a_j$, then $a_i = a_j$), transitive (if $a_i \succeq a_j$ and $a_j \preceq a_k$, then $a_i \succeq a_k$), and complete ($a_i \succeq a_j$ or $a_j \preceq a_i$).

[12] As all games considered in this paper are finite, I need not introduce any continuity requirements in the definition. The definition of strictly positive spillovers is non-standard, but nevertheless useful. See Vives (1999) for a textbook treatment of supermodular games and Vives (2005) for a survey.

$a_j' \geq a_j$ *implies*

$$U_i(a_i', a_j'; w) - U_i(a_i, a_j'; w) \geq U_i(a_i', a_j; w) - U_i(a_i, a_j; w).$$

If, in addition, $U_i(a_i, a_j; w)$ *is strictly increasing in* $p(a_j)$ *when* $p(a_i) > 0$*, then* $\Gamma(w, A)$ *is said to exhibit* ***strictly positive spillovers****. The game* $\Gamma(w, A)$ *is* ***submodular*** *if* U_i *exhibits decreasing differences:* $a_i' \geq a_i$ *and* $a_j' \geq a_j$ *implies*

$$U_i(a_i', a_j'; w) - U_i(a_i, a_j'; w) \leq U_i(a_i', a_j; w) - U_i(a_i, a_j; w).$$

The important property of supermodular games that I exploit is that best-response dynamics converge to their maximal and minimal equilibria. Moreover, any supermodular game with strictly positive spillovers has a unique Pareto Efficient Nash equilibrium. In particular, let $a_{\max}$ and $a_{\min}$ denote the maximal and minimal elements of A, and $\overline{BR} : A \to A$ and $\underline{BR} : A \to A$ denote the maximal and minimal best-response functions for the agents.[13] Then, the following properties hold.

Lemma 1 (Vives (1990), Milgrom and Roberts (1990))
Suppose $\bar{a}$ *(*$\underline{a}$*) is the limit found by iterating* $\overline{BR}$ *(*$\underline{BR}$*) starting from* $a_{\max}$ *(*$a_{\min}$*). If* $\Gamma(w, A)$ *is supermodular, then it has a maximal Nash equilibrium* $(\bar{a}, \bar{a})$ *and a minimal Nash equilibrium* $(\underline{a}, \underline{a})$*; any other equilibrium* (a_i, a_j) *must satisfy* $\bar{a} \geq a_i \geq \underline{a}$ *and* $\bar{a} \geq a_j \geq \underline{a}$*. If, in addition,* $\Gamma(w, A)$ *exhibits strictly positive spillovers, then* $(\bar{a}, \bar{a})$ *is the unique Pareto Efficient Nash equilibrium.*

A similar property holds for two-player submodular games. Define the mapping

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

Then, the following property holds.

Lemma 2 (Vives (1990), Milgrom and Roberts (1990))
Suppose $(\bar{a}, \underline{a})$ *is the limit found by iterating* $\widetilde{BR}$ *starting from the action profile* $(a_{\max}, a_{\min})$*.*

[13] Formally, if $a_i = \overline{BR}(a_j)$, then a_i is a best-response to a_j and $a_i \geq a_i'$ for any other best-response a_i'. Similarly, if $a_i = \underline{BR}(a_j)$, then a_i is a best-response to a_j and $a_i \leq a_i'$ for any other best-response a_i'. Both $\overline{BR}$ and $\underline{BR}$ are well-defined by Corollary 4.1 of Topkis (1978).

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 $\underline{a}$.

4.3 Proof of Main Result

Say that a contract w is **eligible** if $V(w) > 0$.[14] It is without loss of generality to restrict attention to eligible contracts; Carroll (2015) already identifies that

$$V^*_{IPE} := \sup_{w:\ w \text{ is an IPE}} V(w) = 2 \max_{w \in [0,1], a_0 \in A^0} \left[(p(a_0) - \frac{c(a_0)}{w})(1-w) \right] > 0$$

by an argument that generalizes the one sketched in Section 2.[15] Hence, any contract w for which $V(w) \leq 0$ cannot be worst-case optimal.

The proof has five steps. First, I show that linear contracts are strictly suboptimal (Lemma 3) and that any contract can be (weakly) improved by a contract w for which $w_{01} = w_{00} = 0$ or yields a worst-case payoff smaller than V^*_{IPE} (Lemma 4). Second, I show that there does not exist an RPE that yields the principal a strictly larger payoff than V^*_{IPE} (Lemma 5). Third, I compute the principal's worst-case payoff given any JPE (Lemma 6). Fourth, I show that there exists a (calibrated) JPE that yields a strictly higher payoff than V^*_{IPE} (Lemma 7). Fifth, I establish existence of a worst-case optimal JPE with $w_{01} = w_{10} = 0$ and re-examine the proof of Lemma 4 to show that no other class of contracts can be optimal.

4.3.1 Suboptimality of Linear and Related Contracts

I provide a simple proof that any eligible linear contract is strictly suboptimal.

Lemma 3 (Linear Contracts are Suboptimal)
For any eligible linear contract w, there exists a nonlinear contract w' that yields the principal a strictly higher worst-case payoff.

[14] This definition implies eligibility in the sense of Carroll (2015), who requires that, in addition, $V(w)$ yields a higher worst-case payoff than the contract paying zero wages for all pairs (y_i, y_j). By the assumption of costly known productive actions, such a contract yields the principal a worst-case payoff of zero.

[15] Due to adversarial equilibrium selection, V^*_{IPE} may only be approached by a sequence of contracts, in contrast to the setting of Carroll (2015).

Proof. Let $\alpha \in [0,1]$ parameterize the eligible linear contract w. If $\alpha = 0$, then the assumption that known productive actions are costly ensures that w cannot guarantee the principal more than zero in the game $\Gamma(w, A^0 \cup \{a_\emptyset\})$, where $p(a_\emptyset) = c(a_\emptyset) = 0$, because in this game each agent has a strict incentive to choose $a_\emptyset$. So, since w is eligible, it must be that $\alpha > 0$. Under w, $w_{00} = 0$, $w_{10} = w_{01} = \alpha > 0$, and $w_{11} = 2\alpha > 0$. Define a contract w' with $w'_{01} = 0$ and $w'_{11} = \alpha$ that is otherwise equal to w. Then, the incentives of the agents are unchanged; a constant shift in an agent's payoff given any action of the other does not affect her optimal choice of action. Hence, for any $A \supseteq A^0$, an equilibrium under w is also an equilibrium under w'. By eligibility of w, however, some agent must succeed at her task with strictly positive probability in any equilibrium $\sigma \in \mathcal{E}_P(w, A)$. But, conditional on this event, the principal's wage payments must decrease. Hence, her expected wage payments strictly decrease in any equilibrium. It follows that $V(w') > V(w)$. □

More generally, any eligible contract w with $w_{00} > 0$ or $w_{01} > 0$ can be improved upon by another contract w' with $w'_{00} = w_{01} = 0$ or, alternatively, cannot yield a payoff higher than V^*_{IPE}. The following Lemma is proved in Appendix A.1.

Lemma 4 (Positive Wages for Failure is Suboptimal)
For any eligible contract w with $w_{00} > 0$ or $w_{01} > 0$, there either exists a contract w' with $w'_{01} = w'_{00} = 0$ and $V(w') \geq V(w)$, or $V^*_{IPE} \geq V(w)$.

While the "shifting" argument used in the proof of Lemma 3 rules out many contracts, there are two cases that require different arguments. When $w_{11} > 0$ and $w_{00} > 0$ (with $w_{01} = w_{00} = 0$), I exploit supermodularity of the payoff function and a comparative statics result of Milgrom and Roberts (1990) to argue that the probability of success given any equilibrium action decreases in w_{00}. When $w_{10} > 0$ and $w_{01} > 0$ (with $w_{01} = w_{00} = 0$), I must elaborate upon the proof idea in Lemma 5 to rule out asymmetric and mixed equilibria that might be beneficial for the principal. I therefore encourage the interested reader to review it only upon reading the rest of Section 4.

An immediate corollary of Lemma 4 is that to find a worst-case optimal contract it suffices to consider nonlinear JPE satisfying $w_{11} > w_{10}$, IPE satisfying $w_{11} = w_{10}$, and RPE satisfying $w_{11} < w_{10}$. When $w_{11} > w_{10}$ ($w_{11} < w_{10}$), so that w is a

JPE (RPE), it is easy to show that $U_i(a_i, a_j; w)$ exhibits increasing (decreasing) differences: If $a_i' \geq a_i$, and $a_j' \geq a_j$, so that $p(a_i') \geq p(a_i)$ and $p(a_j') \geq p(a_j)$, then

$$\begin{aligned} U_i(a_i', a_j'; w) - U_i(a_i, a_j'; w) &= (p(a_i') - p(a_i))\Big[p(a_j')w_{11} + (1 - p(a_j'))w_{10}\Big] - (c(a_i') - c(a_i)) \\ &\geq (p(a_i') - p(a_i))\Big[p(a_j)w_{11} + (1 - p(a_j))w_{10}\Big] - (c(a_i') - c(a_i)) \\ &= U_i(a_i', a_j; w) - U_i(a_i, a_j; w). \end{aligned}$$

A similar calculation establishes that when $w_{11} < w_{10}$ payoff functions exhibit decreasing differences. Intuitively, the marginal benefit of taking a higher action for agent i is increasing (decreasing) in the action of agent j in the case of JPE (RPE).

Moreover, if w is a JPE, any game $\Gamma(w, A)$ with $A \supseteq A^0$ exhibits strictly positive spillovers:

$$U_i(a_i, a_j; w) = p(a_i)\Big[p(a_j)w_{11} + (1 - p(a_j))w_{10}\Big] - c(a_i)$$

is strictly increasing in $p(a_j)$ when $p(a_i) > 0$. I thus make the following observation.

Observation 1
If w is an RPE for which $w_{00} = w_{01} = 0$ and $A \supseteq A^0$, then $\Gamma(w, A)$ is a submodular game. If w is a JPE for which $w_{00} = w_{01} = 0$ and $A \supseteq A^0$, then $\Gamma(w, A)$ is a supermodular game exhibiting strictly positive spillovers.

4.3.2 RPE Cannot Outperform IPE

I now establish that no RPE can yield a higher payoff than V^*_{IPE}.

Lemma 5 (IPE Outperforms RPE)
No RPE with $w_{01} = w_{00} = 0$ can yield the principal a higher worst-case payoff than V^*_{IPE}.

The proof of the Lemma is in Appendix A.2. I sketch the proof for the case in which there is a single known action, i.e. $A^0 := \{a_0\}$. Suppose each agent has available a single additional zero-cost action a^* that results in success with probability $p(a^*) < p(a_0)$. Then, a^* is a strict best response to a^* if and only if

$$\underbrace{p(a^*)\left(p(a^*)w_{11} + (1 - p(a^*))w_{10}\right)}_{\text{Payoff } a^* \text{ against } a^*} > \underbrace{p(a_0)\left(p(a^*)w_{11} + (1 - p(a^*))w_{10}\right) - c_0}_{\text{Payoff } a_0 \text{ against } a^*} \iff$$

$$p(a^*) > p(a_0) - \frac{c(a_0)}{p(a^*)w_{11} + (1 - p(a^*))w_{10}}.$$

This condition also ensures that a^* is a strictly dominant strategy. Intuitively, if a^* is a strict best response to a^*, which is less productive than a_0, then it must also be a strict best response to a_0 since the marginal benefit of shirking against a more productive action is higher (because $w_{10} > w_{11}$); this property is a direct consequence of the submodularity of the game induced by RPE. Hence, (a^*, a^*) is the unique Nash equilibrium. The principal's payoff as p^* approaches the value at which the incentive constraint binds is therefore

$$\underbrace{2\left(p(a_0) - \frac{c(a_0)}{p(a^*)w_{11} + (1 - p(a^*))w_{10}}\right)}_{\text{Probability Success}} \times \underbrace{\left[1 - (p(a^*)w_{11} + (1 - p(a^*))(1 - w_{10}))\right]}_{\text{Conditional Expected Surplus}}.$$

Letting $\hat{w} := p(a^*)w_{11} + (1 - p(a^*))(1 - w_{10})$, it is immediate that she can do no better than V^*_{IPE}:

$$2(p(a_0) - \frac{c(a_0)}{\hat{w}})(1 - \hat{w}) \leq 2 \max_{w \in [0,1]} \left[(p(a_0) - \frac{c(a_0)}{w})(1 - w)\right] = V^*_{IPE}.$$

The proof for general known action sets builds upon this idea. In particular, I consider a worst-case action set with a zero-cost action a^* that results in success with a high enough probability that (a^*, a^*) is a strict Nash equilibrium. I then argue that this equilibrium is unique and that, in it, the principal obtains a payoff no higher than V^*_{IPE}.

4.3.3 JPE Worst-Case Payoffs

Within the class of contracts setting $w_{00} = w_{01} = 0$, the only contracts left to consider are nonlinear JPE for which $w_{11} > w_{10}$. Lemma 6 states the principal's worst-case payoff guarantee from any contract of this form. Its proof is in Appendix A.3.[16]

[16] The characterization holds for *any* JPE if I replace w_{11} with $w_{11} - w_{01}$ and w_{10} with $w_{10} - w_{00}$ in Equation 1 and change $\bar{p}\left[\bar{p}(1 - w_{11}) + (1 - \bar{p})(1 - w_{10})\right]$ to $\bar{p}\left[\bar{p}(1 - w_{11}) + (1 - \bar{p})(1 - w_{10})\right] + (1 - \bar{p})\left[\bar{p}(-w_{01}) + (1 - \bar{p})(-w_{00})\right]$.

Lemma 6 (JPE Worst-Case Payoffs)
Suppose w is a JPE with $w_{00} = w_{01} = 0$ and, for each $a_0 \in A^0$, $\hat{p}(\cdot|a_0) : [0, \hat{t}(a_0)] \to [0, p(a_0)]$ is the unique solution to the initial value problem

$$\begin{aligned} \hat{p}'(t) =& f(\hat{p}(t)) := \frac{-1}{\hat{p}(t)w_{11} + (1-\hat{p}(t))w_{10}} \quad \textit{with} \\ \hat{p}(0) =& p(a_0), \end{aligned} \tag{1}$$

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,

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

where

$$\bar{p} := \max_{a_0 \in A^0} \ \hat{p}(\hat{t}(a_0)|a_0).$$

The principal's worst-case payoff, $V(w)$, is two times the minimum of two terms. The first term

$$1 - w_{11}$$

is the principal's payoff from each agent when the worst-case action set induces a game between the agents in which there is an equilibrium in which both succeed with probability one. The second term

$$\bar{p}\,[\bar{p}(1 - w_{11}) + (1 - \bar{p})(1 - w_{10})]$$

is the principal's payoff when the worst-case action set induces a game between the agents with a "shirking equilibrium" in which each succeeds with a probability $\bar{p}$ as low as possible. (Both are required because, for high enough w_{11}, the principal may prefer the shirking equilibrium.) The solution to each differential equation, $\hat{p}(\cdot|a_0)$, characterizes best-response dynamics in the limit of a sequence of discrete games in which a_0 is the only known action; $\hat{p}(\hat{t}(a_0)|a_0)$ is the limit of the equilibrium probability of success in this sequence of games; and $\bar{p}$ is the maximum of these limits.

I discuss the proof of Lemma 6 in two parts. First, I describe the sequence of games that leads to the worst-case probability $\bar{p}$. Second, I describe why $\bar{p}$ is, in

fact, a lower bound.

The Worst-Case Sequence of Games For simplicity, suppose there is a single known action a_0 with success probability $p(a_0) = 1$ and cost $c(a_0) = \frac{1}{4}$, as in Section 2. As shown in that section, the optimal IPE puts $w^* = w_{11} = w_{10} = \frac{1}{2}$. Given w^*, the worst-case success probability approaches

$$p(a_0) - \frac{c(a_0)}{w^*} = \frac{1}{2}.$$

Now, suppose I reduce w_{10} to zero (corresponding to setting $\epsilon = \frac{1}{2}$ in the example), but keep all other wages the same. This contract is calibrated to w^* and the known action a_0:

$$p(a_0)w_{11} + (1 - p(a_0))w_{10} = w_{11} = w^*.$$

So, according to the analysis of Section 2, there is ostensibly *no* efficiency loss generated by this modification.

In particular, if I consider only the class of games with action sets of the form $A^1 := A^0 \cup \{a_1^1\}$, for some action a_1^1 with success probability $p(a_1^1) < p(a_0)$, then the worst-case for the principal occurs as $p(a_1^1)$ approaches the value at which the best-response condition binds:

$$p(a_1^1)[p(a_0)w_{11} + (1 - p(a_0))w_{10}] - c(a_1^1) = p(a_0)[p(a_0)w_{11} + (1 - p(a_0))w_{10}] - c(a_0)$$

$$\iff p(a_1^1) = p(a_0) - \frac{c(a_0) - c(a_1^1)}{p(a_0)w_{11} + (1 - p(a_0))w_{10}} \geq \frac{1}{2}.$$

Figure 3 depicts the best-response response path starting from the known (maximal) action a_0. The dashed line may be interpreted as an *indifference curve* with slope $m = -1/(p(a_0)w_{11} + (1 - p(a_0))w_{10})$ and intercept $b = p(a_0)$: each action on the line, a, is identified by its cost relative to $c(a_0)$, $x = c(a_0) - c(a)$, and its success probability, $y = p(a)$. Since the slope of the indifference curve is negative, the maximal reduction in success probability occurs when the cost reduction is as large as possible, i.e. when $c(a_1^1) = 0$ so that $x = \frac{1}{4}$.

But what if there are two unknown actions? Consider the action set $A^2 := A^0 \cup \{a_1^2, a_2^2\}$, where a_1^2 has a positive cost of $c(a_1^2) = \frac{c(a_0)}{2} = \frac{1}{8}$ and $c(a_2^2) = 0$. A simple

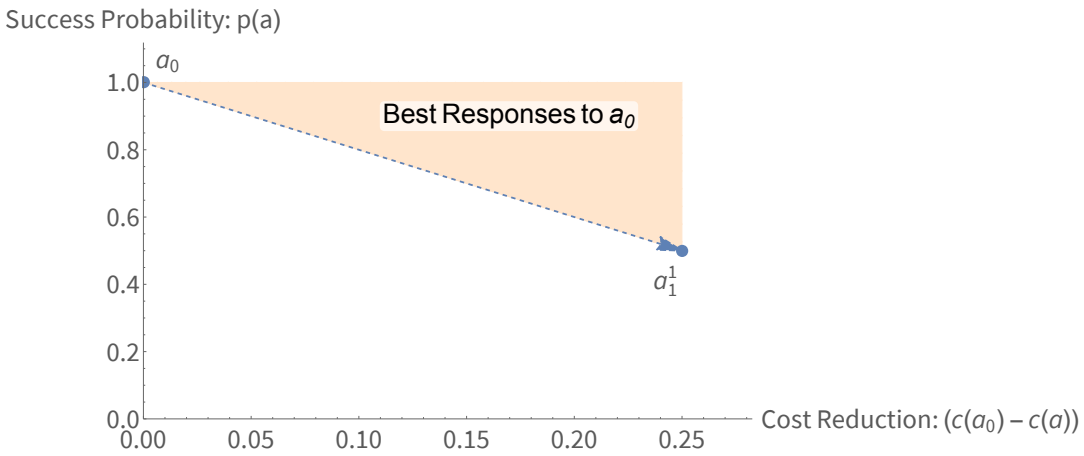

Figure 3: A^1 *best-response path.*

calculation shows that for a_1^2 to be a strict best-response to a_0, it must be the case that

$$p(a_1^2)\left[p(a_0)w_{11} + (1-p(a_0))w_{10}\right] - c(a_1^2) > p(a_0)\left[p(a_0)w_{11} + (1-p(a_0))w_{10}\right] - c(a_0)$$

$$\iff p(a_1^2) > p(a_0) - \frac{c(a_0) - c(a_1^2)}{p(a_0)w_{11} + (1-p(a_0))w_{10}} = \frac{3}{4}.$$

Furthermore, for a_2^2 to be a best-response to a_1^2, it must be the case that

$$p(a_2^2) > p(a_1^2) - \frac{c(a_1^2) - c(a_2^2)}{p(a_1^2)w_{11} + (1-p(a_1^2))w_{10}} = p(a_1^2) - \frac{1}{4p(a_1^2)}.$$

If $p(a_1^2)$ is close to $\frac{3}{4}$ and $p(a_2^2)$ is close to $p(a_1^2) - 1/(4p(a_1^2))$, then, in addition, a_1^2 is the unique best-response to a_0 and a_2^2 is the unique best-response to a_1^2. Hence, best-response dynamics converge to (a_1^2, a_1^2). Since $\Gamma(w, A^1)$ is supermodular (Observation 1), Lemma 1 thus implies that (a_1^2, a_1^2) is the unique Nash (and therefore, Pareto Efficient Nash) equilibrium. In it, each agent's success probability can be made arbitrarily close to

$$\frac{3}{4} - \frac{1}{4\frac{3}{4}} = \frac{5}{12} < \frac{1}{2}.$$

See Figure 4, which now depicts a second indifference curve, with a steeper slope,

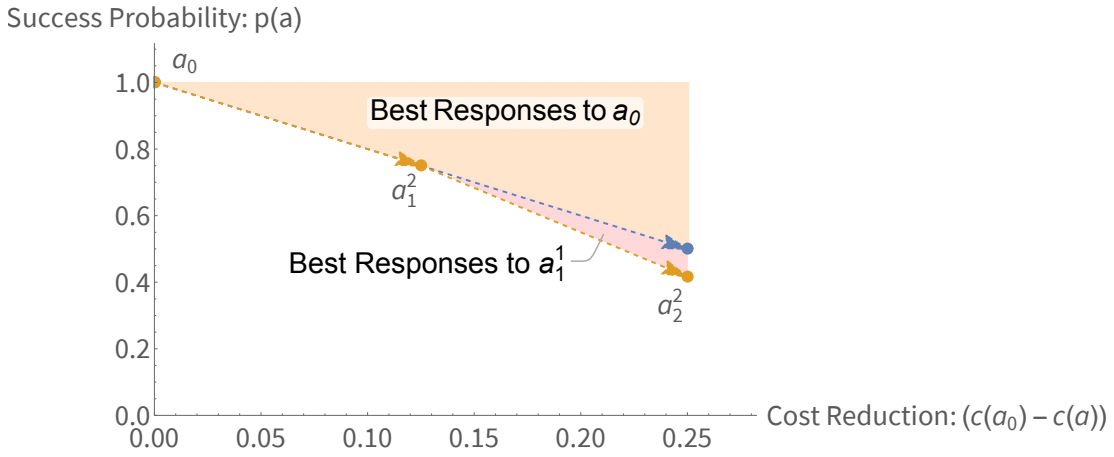

Figure 4: A^2 *best-response path.*

corresponding to weak best-responses to a_1^2.

I now generalize this construction to drive the equilibrium probabilities of success even lower. Let $A^n := A^0 \cup \{a_1^n, ..., a_n^n\}$ be an action set with $c(a_k^n) = (n-k)\frac{c(a_0)}{n}$, so that costs are evenly distributed on a grid between zero and $c(a_0)$. For each $k = 1, ..., n$, choose $p(a_k)$ so that a_k is a best-response to a_{k-1}, i.e. set

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

where $\epsilon(n) := \frac{c(a_0)}{n}$ and $\rho(n) > 0$.[17] For $\rho(n)$ small, a_k is a maximal best-response to a_{k-1} for all k. It follows that the unique Nash equilibrium of $\Gamma(w, A^n)$ is (a_n^n, a_n^n), found again by iterating best-responses. Hence, the equilibrium probability of success for each agent is $p(a_n^n)$.

What is the limit of $p(a_n^n)$ as $n \to \infty$? The key observation is that Equation E is an Euler approximation of Equation 1, where $\frac{c(a_0)}{n}$ is the step size of the approximation and $\rho(n)$ is a "rounding error". Hence, as n grows large, if the rounding error $\rho(n)$ approaches zero at an appropriately fast rate relative to $\epsilon(n)$, agents' best-response dynamics are well-described by the solution to Equation 1, $\hat{p}(\cdot|a_0)$, under the interpretation that time t is "cost-reduction relative to a_0".[18] In the

[17] To see why this is an equivalent condition, multiply both sides of the equation by $p(a_{k-1})w_{11} + (1 - p(a_{k-1}))w_{10}$.

[18] See, for instance, Theorem 6.3 of Atkinson (1989) and the proceeding discussion.

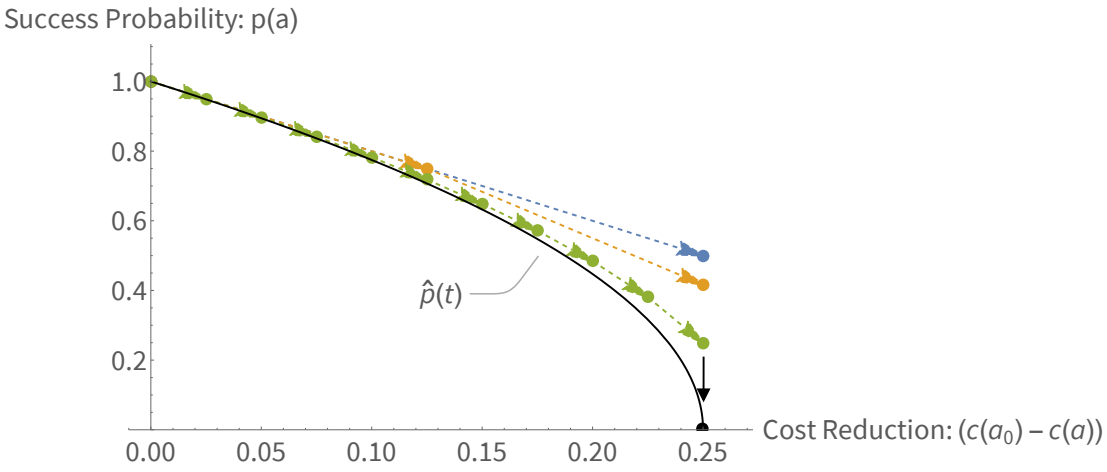

Figure 5: $p(a_n^n)$ *as* $n \to \infty$.

example considered here, the limit is

$$\bar{p} = \hat{p}(\hat{t}(a_0)|a_0) = \hat{p}(c(a_0)|a_0) = \hat{p}(0.25|a_0) = 0,$$

as depicted in Figure 5.

Why is $\bar{p}$ a Lower Bound? Since the law of motion in Equation 1 is controlled by the wages the principal offers, the principal can increase $\hat{p}(\hat{t}(a_0)|a_0)$ above zero. For instance, Figure 6 shows that if the principal increases w_{11} to $\frac{2}{3}$, while keeping w_{10} at 0, then she increases $\bar{p} = \hat{p}(\hat{t}(a_0)|a_0)$ back to $\frac{1}{2}$, the worst-case probability of success given the optimal IPE. In this case, Lemma 6 then dictates that there does not exist a game that drives the equilibrium probability of success below $\frac{1}{2}$.

I outline the proof that $\frac{1}{2}$ is a lower bound. By Observation 1, for any action set $A \supseteq A^0 = \{a_0\}$, the game $\Gamma(w, A)$ is supermodular and exhibits strictly positive spillovers. Hence, by Lemma 1, its unique Pareto Efficient Nash equilibrium can be found by iterating the maximal best-response function $\overline{BR}$ starting from the maximal element of A, $a_{\max}$. There are two possible cases: (i) $a_0 = a_{\max}$ and (ii) $a_{\max} \geq a_0$ and $a_0 \not\geq a_{\max}$. I argue that the equilibrium probability of success cannot be below $\bar{p}$ in either case.

Suppose first that $a_0 = a_{\max}$. It suffices to show that any best-response path $(a_0, ..., a_n)$, beginning at a_0 and ending at a_n, satisfies $p(a_n) \geq \hat{p}(c(a_0)|a_0) = \bar{p}$. If

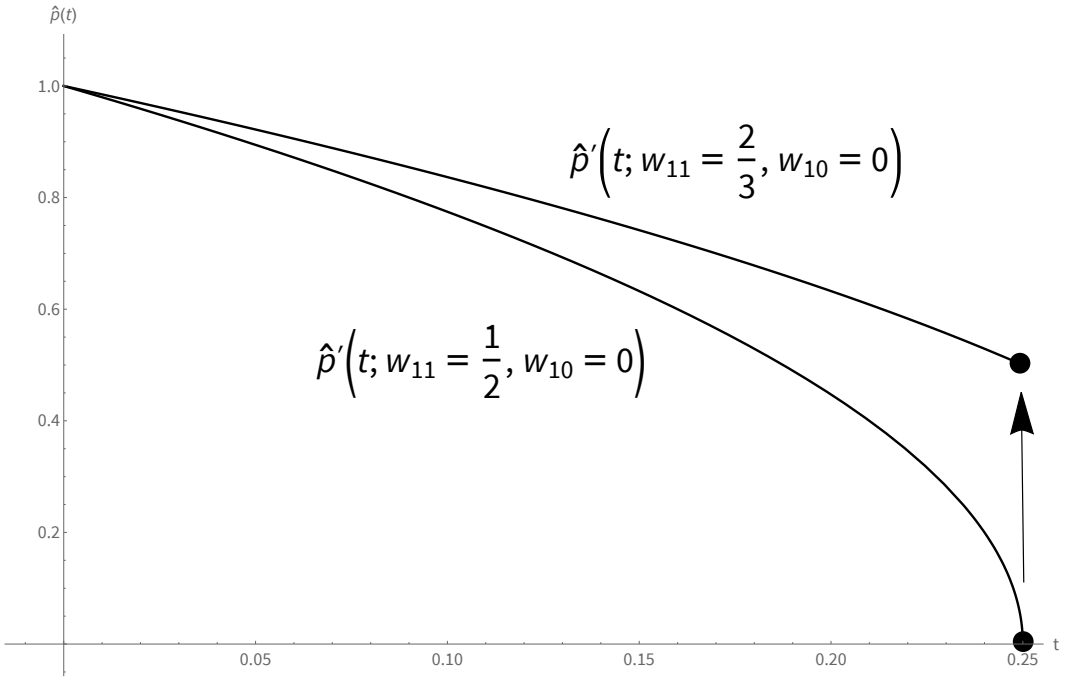

Figure 6: *Increasing* w_{11}.

$a_1 = \overline{BR}(a_0)$, then it must be the case that

$$\begin{aligned} p(a_1) &> p(a_0) - \frac{c(a_0) - c(a_1)}{p(a_0)w_{11} + (1 - p(a_0))w_{10}} \\ &= \hat{p}(0|a_0) - \epsilon \hat{p}'(\epsilon|a_0) \\ &\geq \hat{p}(\epsilon|a_0), \end{aligned}$$

where $\epsilon := c(a_0) - c(a_1) > 0$ and the inequality follows from concavity of $\hat{p}(\cdot|a_0)$. By induction, it can then be shown that

$$p(a_k) \geq \hat{p}(\sum_{\ell=1}^{k} \epsilon_\ell | a_0) \quad \text{for all } k = 1, ..., n,$$

where $\epsilon_k := c(a_k) - c(a_{k-1}) > 0$. As $\sum_{\ell=1}^{n} \epsilon_\ell = c(a_0)$, this means that $p(a_n) \geq \hat{p}(c(a_0)|a_0)$ as desired.

Suppose, instead, that $a_{\max} \geq a_0$ and $a_0 \not\geq a_{\max}$. Then, $p(a_{\max}) = 1$ and $c(a_{\max}) < \frac{1}{4} = c(a_0)$. Any best-response path starting at a_{max} and ending at a_n must have $p(a_n) \geq \hat{p}(c(a_{max})|a_{max})$ by the argument just outlined. Plotting $\hat{p}(\cdot|a_{max})$ and $\hat{p}(\cdot|a_0)$

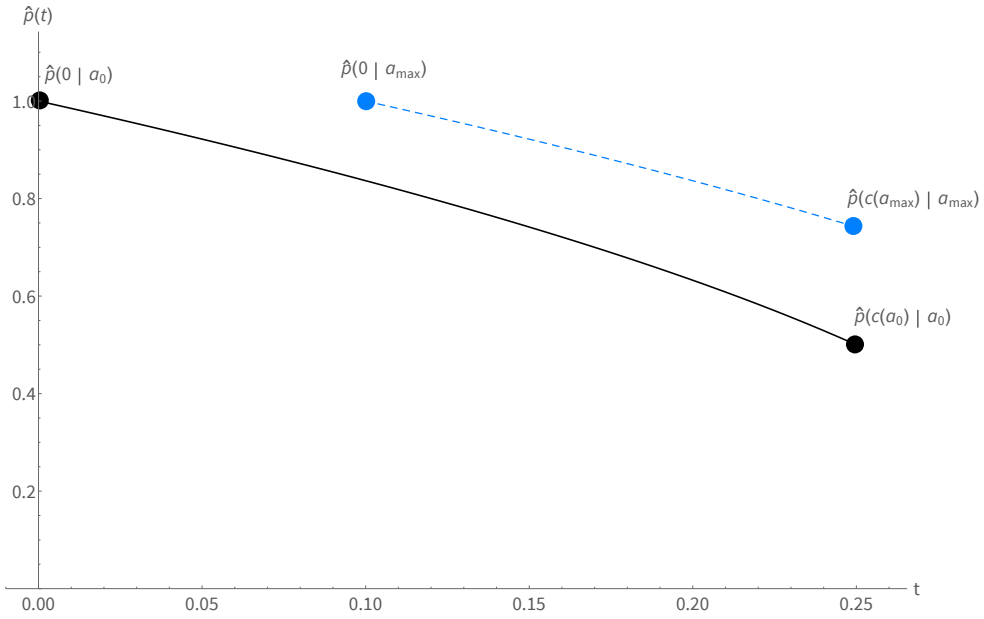

Figure 7: $\hat{p}(t|a_{\max})$ *lies above* $\hat{p}(t|a_0)$.

on a cost-adjusted axis, however, it is clear that $\hat{p}(\cdot|a_{max})$ lies above $\hat{p}(\cdot|a_0)$ (see Figure 7).[19] Hence, $p(a_n) \geq \hat{p}(c(a_{max})|a_{max}) \geq \hat{p}(c(a_0)|a_0)$, establishing the result.

The full proof of Lemma 6 extends the previous arguments to the case of an arbitrary known action set A^0. This entails showing that the lowest probability of success, $\bar{p}$, is the maximum of $\hat{p}(\hat{t}(a_0)|a_0)$ for all $a_0 \in A^0$ rather than, say, the minimum. It also involves ruling out best-response paths originating from unknown actions that succeed with strictly higher probability than any known action. To prove these claims, I show that any path of actions containing an action "beneath" a differential equation associated with a known action *cannot* be a best-response path.

4.3.4 Existence of a Calibrated JPE Outperforming IPE

While I demonstrated in the previous section that not *every* calibrated JPE outperforms a benchmark (optimal) IPE, I prove that there must exist one that does. Thus, I obtain the following Lemma, proved in Appendix A.4.

[19] More formally, since both are solutions to the same initial value problem with distinct initial conditions, their paths can never cross. Since there is a time period t at which $\hat{p}(c(a_0) - c(a_{\max}) + t|a_{max})$ is above $\hat{p}(t|a_0)$, the result follows.

Lemma 7 (JPE Outperforms IPE)
There exists a JPE with $w_{00} = w_{10} = 0$ yielding the principal a strictly higher worst-case payoff than V^*_{IPE}.

I illustrate the argument using the running example with a single known action a_0 that results in success with probability $p(a_0) = 1$ and has an effort cost of $c(a_0) = \frac{1}{4}$. As previously pointed out, the optimal IPE given this action puts $w^* = w_{11} = w_{10} = \frac{1}{2}$. Consider the calibrated JPE setting $w_{10} = w^* - \epsilon = \frac{1}{2} - \epsilon$ for small $\epsilon > 0$ and setting

$$p(a_0)w_{11} + (1 - p(a_0))w_{10} = w^* \iff w_{11} = \frac{1}{2}.$$

I show that this contract strictly increases the principal's worst-case payoff.

Elementary methods show that the solution to the differential equation defining $\bar{p}$ in Lemma 6 is

$$\bar{p}(\epsilon) := \frac{\sqrt{\frac{1}{2}(\frac{1}{2} - \epsilon)} - (\frac{1}{2} - \epsilon)}{\epsilon}.$$

A simple application of L'Hôpital's rule confirms that as $\epsilon \to 0^+$, so that the wage scheme I constructed approaches the optimal IPE, $\bar{p}(\epsilon)$ approaches $\frac{1}{2}$, the worst-case equilibrium probability of success given the optimal IPE. Differentiating $\bar{p}(\epsilon)$ and taking its limit as $\epsilon \to 0^+$, I identify a local calibration effect on the worst-case probability of success:

$$\lim_{\epsilon \to 0^+} \bar{p}'(\epsilon) = -\frac{1}{4}.$$

I now compute the local effect of calibration on the principal's profit from each agent in the shirking equilibrium.[20] For any $\epsilon > 0$, the principal's payoff per agent in the shirking equilibrium is

$$\pi(\epsilon) := \underbrace{\bar{p}(\epsilon)}_{\text{Expected Task Value}} \times \underbrace{[1 - (\bar{p}(\epsilon)w_{11} + (1 - \bar{p}(\epsilon))w_{10})]}_{\text{Conditional Expected Surplus}}.$$

[20] As the principal's profit in the shirking equilibrium at the optimal IPE is strictly lower than in the equilibrium in which both agents succeed probability one, it suffices to show that the principal benefits from such a decrease to exhibit a strict increase in the principal's payoff.

Using the chain rule and taking limits,

$$\lim_{\epsilon\to 0^+}\pi'(\epsilon) = \lim_{\epsilon\to 0^+}\underbrace{\bar{p}'(\epsilon)[1-(\bar{p}(\epsilon)w_{11}+(1-\bar{p}(\epsilon))w_{10})]}_{\text{Efficiency Loss}}+\underbrace{\bar{p}(\epsilon)\frac{d}{d\epsilon}[1-(\bar{p}(\epsilon)w_{11}+(1-\bar{p}(\epsilon))w_{10})]}_{\text{Gain in Rents}}$$

$$= (\lim_{\epsilon\to 0^+}\bar{p}'(\epsilon))(1-w^*)+(\lim_{\epsilon\to 0^+}\bar{p}(\epsilon))w^*$$

$$= -\frac{1}{4}\times\frac{1}{2}+\frac{1}{4} > 0.$$

This establishes the desired result.

4.3.5 Existence, Uniqueness, and Optimal Wages

I summarize the preceding arguments. Lemma 4 establishes that, for the purposes of finding a weakly optimal contract, it suffices to consider those setting $w_{00} = w_{01} = 0$. Any such contract is either an RPE, JPE, or IPE. Lemma 5 establishes that no RPE with $w_{00} = w_{01} = 0$ can outperform V^*_{IPE}, the supremum payoff attainable within the class of IPE. On the other hand, Lemma 7 establishes that there *does* exist a JPE with $w_{00} = w_{01} = 0$ that yields the principal a strictly higher payoff than V^*_{IPE}. Hence, if there exists a JPE with $w_{00} = w_{01} = 0$ that maximizes Equation 2, the principal's worst-case payoff given an arbitrary JPE, then it is a worst-case optimal contract.

To establish existence of a worst-case optimal contract, I simply observe that the search for an optimal JPE with $w_{00} = w_{01} = 0$ can be recast as a maximization problem of a continuous function over a compact set. To establish that any worst-case optimal contract must be a nonlinear JPE with $w_{00} = w_{01} = 0$, I need only strengthen the proof of Lemma 4 to show that any contract w with either $w_{00} > 0$ or $w_{01} > 0$ is either weakly outperformed by an IPE or RPE, or strictly outperformed by a JPE. I leave these last details to Appendix A.5, thereby completing the proof of Theorem 1.

To conclude the analysis, notice that the optimal values of w_{11} and w_{10} can be found by solving the following maximization problem:

$$\max_{w_{11}>w_{10}\geq 0}\ \min\{1-w_{11}, \bar{p}\,[\bar{p}(1-w_{11})+(1-\bar{p})(1-w_{10})]\},$$

where $\bar{p}$ is defined in the statement of Lemma 6. In the running example I have considered, the optimal wages are $w_{11} = \frac{2}{3}$ and $w_{10} = w_{01} = w_{00} = 0$; the principal increases w_{11} above the optimal IPE wage, $\frac{1}{2}$, to mitigate the efficiency loss I illustrated when $w_{11} = \frac{1}{2}$ and $w_{10} = 0$. As shown in Figure 6, by doing so, she increases $\bar{p}$ to $\frac{1}{2}$, the worst-case equilibrium probability of success given the optimal IPE.[21]

5 Discussion

I briefly sketch how the model might be enriched, describe how the analysis changes, and draw attention to some open questions.

5.1 Asymmetric Contracts

Symmetric contracts are attractive from a normative perspective: Any asymmetric contract is *discriminatory* in the sense of treating equals unequally. Hence, they may be ruled out by either legal considerations or–if the principal randomizes–ex post fairness considerations. However, it is natural to wonder whether the "anti"-Informativeness Principle finding of my paper holds when asymmetric contracts are permitted. In particular, is it in general optimal to link the incentives of identical, technologically independent agents? I provide an affirmative answer to this question.[22]

Formally, an **asymmetric contract** is a quadruple $w^i = (w^i_{11}, w^i_{10}, w^i_{01}, w^i_{00}) \in \mathbb{R}^4_+$ for each agent $i = 1, 2$, where the first index of each wage indicates agent i's success or failure and the second indicates agent j's success or failure. An asymmetric contract is **linear** if there exist parameters $\alpha_i \in [0, 1]$ for each agent $i = 1, 2$ such that $w^i_{y_i y_j} = \alpha_i(y_i + y_j)$ (and **nonlinear**, otherwise). It is an **independent performance**

[21] Incidentally, $1 - w_{11} = \bar{p}\,[\bar{p}(1 - w_{11}) + (1 - \bar{p})(1 - w_{10})]$ at the optimal wage scheme, as well. I remark that this is not a general property, nor it is a general property that the principal exactly offsets the efficiency loss generated by JPE by increasing w_{11}. It is, however, a general property that at any worst-case optimal contract the principal's payoff in the equilibrium in which agents succeed with probability one is greater than in the shirking equilibrium.

[22] In a Bayesian environment in which a principal demands effort as a unique Nash equilibrium, Winter (2004) shows that asymmetric contracts can be optimal even when agents are symmetric. As Winter (2004) points out, however, if agents were restricted to play Pareto Efficient Nash equilibria, then any optimal contract is symmetric. His argument therefore appears to have no relevance to the model I study.

evaluation (IPE) if $w^i_{y1} = w^i_{y0}$ for each agent $i = 1, 2$ and success or failure $y \in \{0, 1\}$. It is a **dependent performance evaluation (DPE)**, otherwise. A nearly immediate corollary of the proof of Theorem 1 is the following.

Corollary 1

If there exists an asymmetric contract that outperforms the optimal symmetric nonlinear JPE, then it must be nonlinear and it must be a DPE.

That any worst-case optimal asymmetric contract must be nonlinear is immediate from the argument in Lemma 3. If, towards contradiction, some agent were compensated linearly, then, outside of trivial cases, the principal can simply shift their wages down by a constant and strictly increase her payoff given any action set available to the agents.

That any worst-case optimal asymmetric contract must be a DPE is immediate upon observing that, within the class of all IPE, if there exists a worst-case optimal IPE, then there exists a worst-case optimal symmetric IPE. Indeed, given the absence of productive or informational linkages between agents, any optimal contract for agent i, w^i, is also an optimal contract for agent j; if not, then the contract offered to agent i could not have been optimal in the first place. Since, by Lemma 7, there exists a JPE that strictly outperforms the optimal symmetric IPE, this implies that there exists a JPE that strictly outperforms any IPE–symmetric or asymmetric.

Though I have *not* found an asymmetric contract that outperforms the optimal symmetric nonlinear JPE, proving that no such contract exists is non-trivial.[23] I therefore leave as an unproven conjecture that the optimal symmetric contract I

[23]To understand the difficulties involved in constructing a proof, it is instructive to consider how such a result is proved in standard Bayesian contracting models. In these models, if the principal considers implementing each possible action profile and then chooses the implementation that maximizes her profits. For symmetric action profiles, if there exists an incentive compatible asymmetric contract that minimizes expected wage payments and if agents are symmetric, then a "flipped" contract in which the agents labels are exchanged is also incentive compatible and minimizes expected wage payments. As incentive constraints are linear in probabilities, it then follows that randomizing over asymmetric contracts produces a symmetric contract that satisfies the incentive constraints and also minimizes the principal's expected payments. Hence, asymmetry does not pose a problem if the principal wants to implement symmetric profiles. In the robust contracting setting, this argument does not work because the principal does not solve her problem by fixing an action profile that she wants to implement and then maximizing over all implementations.

have identified is also optimal when the principal is permitted to use asymmetric contracts.

5.2 Multiple Levels of Success and Multiple Agents

Now, suppose there are finitely many agents $i = 1, 2, ..., n$ and individual output can take on any value in a compact set $Y \subset \mathbb{R}_+$ with $\min(Y) = 0$. Each action is now described by an effort cost and probability distribution over Y. A **linear** contract in this model is a function

$$w : Y^N \to \mathbb{R}_+$$
$$(y_1, ..., y_n) \mapsto \alpha \sum_{i=1}^{n} y_i,$$

for some value $\alpha \in [0, 1]$. Otherwise, it is **nonlinear**. The result that worst-case optimal compensation is nonlinear readily generalizes to this setting by again modifying the argument establishing Lemma 3.

Corollary 2
If $Y \subset \mathbb{R}_+$ is a compact set with $\min(Y) = 0$ and there are n agents, then any worst-case optimal contract must be nonlinear.

Showing that dependent performance evaluation is optimal in the case of multiple agents when effort is binary is immediate from the main analysis, which shows that the principal would stand to benefit from using JPE with any two agents rather than offering each the optimal IPE. However, showing that JPE is optimal is more challenging because RPE no longer induces a supermodular game between the agents (it no longer suffices to "reverse" the order given to one agent's action set when there are more than two of them). Hence, Lemma 5 must be extended.

Proving that optimal compensation involves dependent performance evaluation when there are multiple output levels is non-trivial. The key challenge is that the order $\succeq$ defined on action sets is no longer total. Hence, upon perturbing an optimal IPE in the direction of JPE, the characterization result of Lemma 6 must be extended.

A complete analysis of this more general model awaits further research.

6 Final Remarks

I study a moral hazard in teams model in which a principal compensates identical, independent agents. In contrast to the classical model, however, I assume that the principal has non-quantifiable uncertainty about the common actions available to the agents. The worst-case optimal contracts that arise–nonlinear, joint performance evaluation contracts–contrast strikingly with what arises if the principal has either unbounded, non-quantifiable uncertainty–in which case linear contracts are worst-case optimal–or if she is fully Bayesian–in which case independent performance evaluation is optimal. I thereby provide a novel robustness foundation for nonlinear joint performance evaluation contracts observed in practice, such as team bonuses and employee stock options in start-ups.

I conclude by commenting on a broader theme in the literature. Over the last decades, a growing number of papers have investigated the "robustness" of classical game-theoretic predictions and mechanisms to various relaxations of the agents' environment. For instance, Bergemann and Morris (2005) consider robust implementation across all type spaces; Chen et al. (2017) propose a metric on the Universal Type Space to capture the strategic impact of relaxing higher-order beliefs in *all* possible games the agents might play; and, as discussed, Dai and Toikka (2018) study moral hazard in teams in a robust contracting setting in which the principal deems all possible unknown action profiles to be plausible.

While these papers make important methodological contributions, the uncertainty faced by the designer (or modeler) in these settings appears to be *too extreme* for many applications. My paper contributes to a small, but growing, research agenda exploring the robustness of predictions and mechanisms in the "intermediate" cases between fully Bayesian and fully Knightian uncertainty. For recent work in this spirit, see Antic (2015), who imposes bounds on the principal's uncertainty over unknown actions in a single-agent robust principal-agent model; Ollar and Penta (2019), who consider robust implementation in the case in which it is common knowledge that agents' types are identically distributed; Gensbittel, Peski and Renault (2020), who consider robustness to higher-order beliefs within the class of zero-sum games; and Malenko and Tsoy (2020), who study optimal

project financing when the financier has bounded, non-quantifiable uncertainty about a project's cash flows.

References

Alchian, Armen A, and Harold Demsetz. 1972. "Production, information costs, and economic organization." *American Economic Review*, 62(5): 777–795.

Antic, Nemanja. 2015. "Contracting with unknown technologies." *Unpublished manuscript, Princeton University*.

Atkinson, Kendall E. 1989. *An introduction to numerical analysis.* . 2nd ed., New York: Wiley.

Bergemann, Dirk, and Stephen Morris. 2005. "Robust mechanism design." *Econometrica*, 1771–1813.

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

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

Chen, Yi-Chun, Alfredo Di Tillio, Eduardo Faingold, and Siyang Xiong. 2017. "Characterizing the strategic impact of misspecified beliefs." *The Review of Economic Studies*, 84(4): 1424–1471.

Che, Yeon-Koo, and Seung-Weon Yoo. 2001. "Optimal Incentives for Teams." *American Economic Review*, 91(3): 525–541.

Coddington, Earl A., and Norman Levinson. 1955. *Theory of Ordinary Differential Equations.* McGraw-Hill Book Company, Inc.

Dai, Tianjiao, and Juuso Toikka. 2018. "Robust incentives for teams." *Unpublished manuscript, Massachusetts Institute of Technology*.

Deloitte. 2016. "Global Human Capital Trends." https://www2.deloitte.com/content/dam/Deloitte/global/Documents/HumanCapital/gx-dup-global-human-capital-trends-2016.pdf.

Fleckinger, Pierre. 2012. "Correlation and relative performance evaluation." *Journal of Economic Theory*, 147(1): 93 – 117.

Fleckinger, Pierre, and Nicolas Roux. 2012. "Collective versus relative incentives: the agency perspective." *Working paper*.

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

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

Gensbittel, Fabien, Marcin Peski, and Jerome Renault. 2020. "Value-Based Distance Between Information Structures." *Unpublished manuscript*.

Green, Jerry R., and Nancy L. Stokey. 1983. "A Comparison of Tournaments and Contracts." *Journal of Political Economy*, 91(3): 349–364.

Hackman, J Richard. 2002. *Leading Teams: Setting the Stage for Great Performances*. Harvard Business Press.

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

Holmström, Bengt. 1982. "Moral Hazard in Teams." *Bell Journal of Economics*, 13(2): 324–340.

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

Itoh, Hideshi. 1991. "Incentives to Help in Multi-Agent Situations." *Econometrica*, 59(3): 611–636.

Lazear, Edward P. 1989. "Pay equality and industrial politics." *Journal of Political Economy*, 97(3): 561–580.

Lazear, Edward P, and Kathryn L Shaw. 2007. "Personnel economics: The economist's view of human resources." *Journal of economic perspectives*, 21(4): 91–114.

Lazear, Edward P, and Sherwin Rosen. 1981. "Rank-order tournaments as optimum labor contracts." *Journal of Political Economy*, 89(5): 841–864.

Malenko, Andrey, and Anton Tsoy. 2020. "Asymmetric Information and Security Design under Knightian Uncertainty." *Unpublished manuscript*.

Marku, Keler, and Sergio Ocampo Diaz. 2019. "Robust Contracts in Common Agency." *Unpublished manuscript, University of Minnesota*.

Milgrom, Paul, and John Roberts. 1990. "Rationalizability, learning, and equilib-

rium in games with strategic complementarities." *Econometrica*, 1255–1277.

Nalebuff, Barry J., and Joseph E. Stiglitz. 1983. "Prizes and Incentives: Towards a General Theory of Compensation and Competition." *Bell Journal of Economics*, 14(1): 21–43.

Ollar, Mariann, and Antonio Penta. 2019. "Implementation via Transfers with Identical but Unknown Distributions." *Barcelona GSE Working Paper Series, Working Paper nº 1126*.

Rees, Daniel I, Jeffrey S Zax, and Joshua Herries. 2003. "Interdependence in worker productivity." *Journal of Applied Econometrics*, 18(5): 585–604.

Rosenthal, Maxwell. 2020. "Robust Incentives for Risk." *Unpublished manuscript, Georgia Institute of Technology*.

Shavell, Steven. 1979. "On Moral Hazard and Insurance." *Quarterly Journal of Economics*, 93(4): 541–562.

Sundaram, Rangarajan K. 1996. *A First Course in Optimization Theory.* Cambridge University Press.

Topkis, Donald M. 1978. "Minimizing a Submodular Function on a Lattice." *Operations Research*, 26(2): 305–321.

Vives, Xavier. 1990. "Nash equilibrium with strategic complementarities." *Journal of Mathematical Economics*, 19(3): 305 – 321.

Vives, Xavier. 1999. *Oligopoly pricing: old ideas and new tools.* MIT press.

Vives, Xavier. 2005. "Complementarities and Games: New Developments." *Journal of Economic Literature*, 43(2): 437–479.

Walton, Daniel, and Gabriel Carroll. 2019. "When are Robust Contracts Linear?" *Unpublished manuscript, Stanford University*.

Winter, Eyal. 2004. "Incentives and Discrimination." *American Economic Review*, 94(3): 764–773.

A Proofs

A.1 Proof of Lemma 4

Given an eligible contract w, agent i's expected payoff is

$$\begin{aligned} U_i(a_i, a_j; w) &= p(a_i)\Big[p(a_j)w_{11} + (1-p(a_j))w_{10}\Big] \\ &\quad + (1-p(a_i))\Big[p(a_j)w_{01} + (1-p(a_j))w_{00}\Big] - c(a_i) \\ &= p(a_i)\Big[p(a_j)(w_{11}-w_{01}) + (1-p(a_j))(w_{10}-w_{00})\Big] \\ &\quad + \Big[p(a_j)w_{01} + (1-p(a_j))w_{00}\Big] - c(a_i). \end{aligned}$$

Hence, if $w_{11} > w_{01}$ ($w_{10} > w_{00}$), setting $w'_{11} = w_{11} - w_{01}$ and $w'_{01} = 0$ ($w'_{10} = w_{10} - w_{00}$ and $w'_{00} = 0$) shifts each agent's payoff by a constant. Similarly, if $w_{11} < w_{01}$ ($w_{10} < w_{00}$), setting $w'_{01} = w_{01} - w_{11}$ and $w'_{11} = 0$ ($w'_{00} = w_{00} - w_{10}$ and $w'_{10} = 0$) shifts each agent's payoff by a constant. It follows that any equilibrium under w is also an equilibrium under w'. Since the principal's ex post payment decreases, these adjustments must (weakly) increase her worst-case payoff.

The argument in the previous paragraph immediately establishes that if $w_{11} > 0$ and $w_{10} > 0$, then there exists an improved contract w' for which $w'_{00} = w'_{01} = 0$. There are three other cases to consider: (i) $w_{01} > 0$ and $w_{00} > 0$ (with $w_{11} = w_{10} = 0$); (ii) $w_{11} > 0$ and $w_{00} > 0$ (with $w_{01} = w_{10} = 0$); and (iii) $w_{01} > 0$ and $w_{10} > 0$ (with $w_{11} = w_{00} = 0$). I discuss each case in turn.

$w_{01} > 0$ and $w_{00} > 0$

If $w_{01} > 0$ and $w_{00} > 0$, then w cannot be eligible. To wit, consider the action set $A := A^0 \cup \{a_\emptyset\}$ 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 $-2w_{00} < 0$.

$w_{11} > 0$ and $w_{00} > 0$

I first argue that if w is eligible, then it must have $w_{11} \geq w_{00}$. Suppose, towards contradiction, that $w_{00} > w_{11}$. Consider the action set $A := A^0 \cup \{a_\emptyset\}$ where $p(a_\emptyset) =$

$0 = c(a_\emptyset)$. Then, $(a_\emptyset, a_\emptyset)$ is the only Pareto Efficient Nash Equilibrium because each agent obtains the maximum wage w_{00} at no effort cost and any other Nash Equilibrium results in this wage with a probability strictly less than one. In the equilibrium $(a_\emptyset, a_\emptyset)$, however, the principal obtains a payoff $-2w_{00} < 0$.

If $w_{11} \geq w_{00} > 0$, then agent i's payoff is

$$U_i(a_i, a_j; w) = p(a_i)p(a_j)w_{11} + (1 - p(a_i))(1 - p(a_j))w_{00} - c(a_i),$$

and satisfies increasing differences in (a_i, a_j). Hence, any game this contract induces is supermodular. Moreover, fixing a_j, (a_i, w_{00}) satisfies decreasing differences. Theorem 6 of Milgrom and Roberts (1990) then implies that the maximal equilibrium of any game $\Gamma(w, A)$, $A \supseteq A^0$, is decreasing in w_{00}.

Now, suppose agent i produces succeeds with probability p_i. The principal's payoff given (p_i, p_j) is

$$\pi(p_i, p_j) := p_i p_j(2 - 2w_{11}) + p_i(1 - p_j) + p_j(1 - p_i) + (1 - p_i)(1 - p_j)(0 - 2w_{00}).$$

Profits are therefore increasing in p_i if and only if,

$$\frac{\partial \pi}{\partial p_i} = p_j(2 - 2w_{11}) + (1 - 2p_j) + (1 - p_j)2w_{00} \geq 0 \iff$$

$$p_j \leq \frac{1 + 2w_{00}}{2w_{11} + 2w_{00}}.$$

If $w_{11} \leq \frac{1}{2}$, then the right-hand side expression is greater than one and profits are strictly increasing in p_i and p_j on their whole domain (for any w_{00}). The principal's worst-case payoff thus strictly increases when w_{00} decreases to zero. If $1 > w_{11} > \frac{1}{2}$, then the principal's payoff is increasing in p_i and p_j when both are less than $\frac{1+2w_{00}}{2w_{11}+2w_{00}}$ and decreasing above it. If $A := A^0 \cup \{a_1\}$, where $p(a_1) = 1 > 0 = c(a_1)$, however, then (a_1, a_1) is the maximal equilibrium. Since the principal may only obtain a strictly lower payoff than $2 - 2w_{11}$ if the maximal equilibrium of some game is in the region in which profits are strictly increasing in both p_i and p_j, it is once again in the principal's interest to increase the maximal equilibrium by setting $w_{00} = 0$. Last, I need not consider the case in which $w_{11} \geq 1$ since no such

contract is eligible.

$w_{01} > 0$ **and** $w_{10} > 0$

Notice, if $w_{01} > 0$ and $w_{10} > 0$ and all other wages are zero, agent i's payoff from an action profile (a_i, a_j) is

$$\begin{aligned} U_i(a_i, a_j; w) &= p(a_i)(1 - p(a_j))w_{10} + (1 - p(a_i))p(a_j)w_{01} - c(a_i) \\ &= p(a_i)\Big[w_{10} - p(a_j)(w_{10} + w_{01})\Big] + p(a_j)w_{01} - c(a_i), \end{aligned}$$

which satisfies decreasing differences. I show that the principal's payoff under such a contract cannot exceed 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

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

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 Brouwer's Fixed Point Theorem, it thus has at least one fixed point.

By construction, $(a^*_\epsilon, a^*_\epsilon)$ is a Nash equilibrium of $\Gamma(w, A_\epsilon)$, where $A_\epsilon := A^0 \cup \{a^*_\epsilon, a_\emptyset\}$. Now, consider a sequence of strictly positive values ϵ_1, ϵ_2,... that converges to zero and for which there is a convergent sequence of fixed points $p(a^*_{\epsilon_1})$, $p(a^*_{\epsilon_2})$,... of the mappings T_{ϵ_1}, T_{ϵ_2},... . Since $[0,1]$ is a compact set, such a convergent sequence must exist. Moreover, its limit is the distribution

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

I show that the principal's worst-case payoff in the limit can be no larger than

what she obtains from the optimal IPE. If p^* equals zero, then the principal attains less than zero profits and so lower profits than under the optimal IPE. Otherwise, let $\hat{a}_0$ denote a maximizer of $p(a) - \frac{c(a)}{w_{10} - p^*(w_{10}+w_{01})}$ over $A^0 \cup \{a_\emptyset\}$, let $\hat{\alpha} := (1-p^*)w_{10}$, and notice that the principal attains a payoff of

$$\begin{aligned}
&2\Big[(p^*)^2 + p^*(1-p^*)(1-w_{01}-w_{10})\Big] \\
&= 2\left[p(\hat{a}_0) - \frac{c(\hat{a}_0)}{(1-p^*)(w_{10}+w_{01})}\right][1-(1-p^*)(w_{10}+w_{01})] \\
&\leq 2\left[p(\hat{a}_0) - \frac{c(\hat{a}_0)}{(1-p^*)w_{10}}\right][1-(1-p^*)w_{10}] \\
&= 2\left[p(\hat{a}_0) - \frac{c(\hat{a}_0)}{\hat{\alpha}}\right][1-\hat{\alpha}].
\end{aligned}$$

But,

$$\begin{aligned}
2\left[p(\hat{a}_0) - \frac{c(\hat{a}_0)}{\hat{\alpha}}\right](1-\hat{\alpha}) &\leq 2 \max_{\alpha\in[0,1], a_0 \in A^0 \cup \{a_\emptyset\}} \left[(1-\alpha)(p(a_0) - \frac{c(a_0)}{\alpha})\right] \\
&= 2 \max_{\alpha\in[0,1], a_0 \in A^0} \left[(1-\alpha)(p(a_0) - \frac{c(a_0)}{\alpha})\right] \\
&= V^*_{IPE},
\end{aligned}$$

where the inequality follows because $p(\hat{a}_0) - \frac{c(\hat{a}_0)}{\hat{\alpha}} \geq 0$ for all $\hat{\alpha} \geq 0$ 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 $k \geq K$, $(a^*_{\epsilon_k}, a^*_{\epsilon_k})$ 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, even as k grows large (so that ϵ grows small). I now address this issue.

First, consider the case in which the limit of $(a^*_{\epsilon_k})$ is $a_\emptyset$ and 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, by Lemma 1, there would exist a K such that for all $k \geq K$, $(a^*_{\epsilon_k}, a^*_{\epsilon_k})$ is the maximal Nash equilibrium of $\Gamma(w, A_{\epsilon_k})$ and hence the unique Nash equilibrium. If this action is less than $\frac{w_{10}}{w_{10}+w_{01}}$, then the principal's payoff in an

equilibrium in which it is played is less than zero:

$$p(a_0)(1-w_{10}-w_{01}) \leq \frac{w_{10}}{w_{10}+w_{01}} - w_{10} < 0.$$

If this action is strictly larger than $\frac{w_{10}}{w_{10}+w_{01}}$, then I can add to each A_{ϵ_k} the action a_0' for which $c(a_0') = 0$ and $p(a_0') = p(a_0) - \frac{c(a_0)}{w_{10}}$ if $p(a_0) - \frac{c(a_0)}{w_{10}} > \frac{w_{10}}{w_{10}+w_{01}}$ and $p(a_0') = \frac{w_{10}}{w_{10}+w_{01}} + \epsilon_k$ otherwise. In the first case, the principal attains a payoff of

$$\left[p(a_0) - \frac{c(a_0)}{w_{10}}\right](1-w_{10}-w_{01}) \leq 2 \max_{\alpha \in [0,1], a_0 \in A^0} \left[(1-\alpha)(p(a_0) - \frac{c(a_0)}{\alpha})\right] = V^*_{IPE}.$$

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 since $a_\emptyset$ is a strict best response to any action larger than $\frac{w_{10}}{w_{10}+w_{01}}$ (the marginal benefit of producing succeeding with higher probability is less than zero).

Second, consider the case in which the limit of $(a^*_{\epsilon_k})$ is $p^* > 0$. Any other pure or mixed Nash equilibrium of $\Gamma(w, A_{\epsilon_k})$ must involve one agent succeeding with probability $\hat{p} \geq \frac{w_{10}}{w_{10}+w_{01}} > p^*$. 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 $p(a^*_\epsilon)$ 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 $p(a_0) > \frac{w_{10}}{w_{10}+w_{01}}$), $a^*_{\epsilon_k}$ (only if $p(a_0) = \frac{w_{10}}{w_{10}+w_{01}}$), or a mixture between the two (again, only if $p(a_0) = \frac{w_{10}}{w_{10}+w_{01}}$); known productive actions are costly and the marginal benefit of succeeding with higher probability is less than zero (strictly so if $p(a_0) > \frac{w_{10}}{w_{10}+w_{01}}$).

It suffices to show that the principal's payoff in the equilibrium in which one agent chooses $a_\emptyset$ is less than V^*_{IPE}; none of the other equilibria can Pareto dominate it as the mixing player is indifferent between $a_\emptyset$ and $a^*_{\epsilon_k}$ and I have already argued that the symmetric equilibrium I constructed yields the principal a worse payoff

than V^*_{IPE}. To show this, it suffices to consider any action, $a_0 \in A^0$, satisfying $p(a_0) \geq \frac{w_{10}}{w_{10}+w_{01}}$ in the support of the strategy succeeding with probability $\hat{p}$. Mirroring the argument in the previous case, I can then add to each A_{ϵ_k} the action a'_0 for which $c(a'_0) = 0$ and $p(a'_0) = p(a_0) - \frac{c(a_0)}{w_{10}} + \epsilon_k$ if $p(a_0) - \frac{c(a_0)}{w_{10}} > \frac{w_{10}}{w_{10}+w_{01}}$ and $p(a'_0) = \frac{w_{10}}{w_{10}+w_{01}} + \epsilon_k$ otherwise. These adjustments ensure that a'_0 is the unique best response to $a_\emptyset$ for every k and so, mirroring the steps in the proof of the previous case, the principal attains a payoff no larger than V^*_{IPE}.

A.2 Proof of Lemma 5

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

$$T_\epsilon(p) := \max_{a_0 \in A^0 \cup \{a_\emptyset\}} \left[p(a_0) - \frac{c(a_0)}{pw_{11} + (1-p)w_{10}} \right] + \epsilon,$$

where $\epsilon > 0$ is small.[24] 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 $[0,1]$. By Brouwer's Fixed Point Theorem, it thus has at least one fixed point.

Now, define an action space $A_\epsilon := A^0 \cup \{a^*_\epsilon, a_\emptyset\}$. 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 $p(\bar{a}_n) = 1 - \frac{\gamma(\epsilon)}{2}$ for $\gamma(\epsilon) := \frac{\epsilon(p(a^*_\epsilon)w_{11} + (1-p(a^*_\epsilon))w_{10}}{2}$. 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 $\bar{a}_\epsilon$.

I show 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 $\geq$. So, remove any actions in $\Gamma(w, A_\epsilon)$ below a^*_ϵ and denote the resulting action space

[24] Interpret $-\frac{c(a_0)}{pw_{11}+(1-p)w_{10}}$ as zero if the denominator is zero and $c(a_0) = 0$ and $-\infty$ if the denominator is zero and $c(a_0) > 0$.

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 $\hat{a}_0 \succ a^*_\epsilon$ and

$$U_i(\hat{a}_0, a_0; w) - U_i(a^*_\epsilon, a_0; w) \geq U_i(\hat{a}_0, \bar{a}; w) - U_i(a^*_\epsilon, \bar{a}; w) > 0 \text{ for any } a_0 \in \hat{A},$$

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 $(a^*_\epsilon, \bar{a})$ in $\Gamma(w, \hat{A})$. By Lemma 2, 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 ϵ_1, ϵ_2,... that converges to zero and for which there is a convergent sequence of fixed points $p(a^*_{\epsilon_1})$, $p(a^*_{\epsilon_2})$,... of the mappings T_{ϵ_1}, T_{ϵ_2},... . 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

$$2\,[p(a^*)]\,[p(a^*)(1 - w_{11}) + (1 - p(a^*))(1 - w_{10})] =$$

$$2\left[p(\hat{a}_0) - \frac{c(\hat{a}_0)}{\hat{\alpha}} \right](1 - \hat{\alpha}) \leq 2 \max_{\alpha \in [0,1], a_0 \in A^0 \cup \{a_\emptyset\}} \left[(1 - \alpha)(p(a_0) - \frac{c(a_0)}{\alpha}) \right],$$

where the inequality follows because $p(\hat{a}_0) - \frac{c(\hat{a}_0)}{\hat{\alpha}} \geq 0$ for all $\hat{\alpha} \geq 0$ and so I need only consider values of α between zero and one to maximize $(1 - \alpha)(p(a_0) - \frac{c(a_0)}{\alpha})$ for any $a_0 \in A^0 \cup \{a_\emptyset\}$. But,

$$2 \max_{\alpha \in [0,1], a_0 \in A^0 \cup \{a_\emptyset\}} \left[(1 - \alpha)(p(a_0) - \frac{c(a_0)}{\alpha}) \right] = 2 \max_{\alpha \in [0,1], a_0 \in A^0} \left[(1 - \alpha)(p(a_0) - \frac{c(a_0)}{\alpha}) \right] = V^*_{IPE}$$

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 $\alpha \in [0,1]$.

A.3 Proof of Lemma 6

Comparative Statics in Principal's Payoff

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

$$\pi(p_i, p_j) := p_i p_j (2 - 2w_{11}) + \Big[p_i(1-p_j) + (1-p_i)p_j\Big](1 - w_{10}).$$

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

$$\frac{\partial \pi(p)}{\partial p_i} = p_j(2 - 2w_{11}) + (1 - 2p_j)(1 - w_{10}) \geq 0 \iff$$

$$p_j \leq \frac{1}{2}\left[\frac{1 - w_{10}}{w_{11} - w_{10}}\right].$$

The shape of $\pi(p_i, p_j)$ on $[0,1]$ thus depends on w: (i) if $w_{10} \geq 1$, then π is decreasing on $[0,1]$ in p_i and p_j; (ii) if $w_{10} < 1$ and $w_{11} \leq \frac{1+w_{10}}{2}$, then $\pi(p)$ is increasing on $[0,1]$ in p_i and p_j; and, (iii) if $w_{10} < 1$ and $w_{11} > \frac{1+w_{10}}{2}$, then $\pi(p)$ is increasing in p_i if $p_j \in [0, \frac{1}{2}\left[\frac{1-w_{10}}{w_{11}-w_{10}}\right]]$ and decreasing in p_i if $p_j \in [\frac{1}{2}\left[\frac{1-w_{10}}{w_{11}-w_{10}}\right], 1]$.

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 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 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 $\Gamma(w, A)$ succeeds with strictly positive probability, for any $A \supseteq A^0$, is as small as possible (by Observation 1 and Lemma 1 there always exists such an action). Letting $\bar{p}$ denote the greatest lower bound on such probabilities, the principal's payoff is,

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

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

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

I identify $\bar{p}$ to complete the proof of the Lemma.

Defining $\bar{p}$

Consider an arbitrary 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

$$\hat{p}'(t|a) = f(\hat{p}(t|a)) := \frac{-1}{\hat{p}(t|a)w_{11} + (1 - \hat{p}(t|a))w_{10}} \text{ with}$$
$$\hat{p}(0|a) = p(a)$$

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

$$\bar{p} := \max_{a_0 \in A^0} \hat{p}(\hat{t}(a_0)|a_0).$$

$\bar{p}$ is a lower bound

I show that $\bar{p}$ is a lower bound on the probability of the maximal equilibrium action of any game $\Gamma(w, A)$, where $A \supseteq A^0$. I begin with the following claim.

Claim 1 (Lower Bound of a $\overline{BR}$ Path)
Fix some game $\Gamma(w, A)$, where $A \supseteq A^0$. Let $(a_1, a_2, ..., a_n)$ 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 $\ell = 1, ..., n$, then

$$p(a_n) \geq \hat{p}(\hat{t}(a)|a).$$

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 $k = \ell + 1, ..., n$ only if $p(a_{k-1}) > p(a_k)$ and,

$$p(a_k)[p(a_{k-1})w_{11} + (1 - p(a_{k-1}))w_{10}] - c(a_k) > p(a_{k-1})[p(a_{k-1})w_{11} + (1 - p(a_{k-1}))w_{10}] - c(a_{k-1})$$

$$\iff p(a_k) > p(a_{k-1}) - \frac{c(a_{k-1}) - c(a_k)}{p(a_{k-1})w_{11} + (1 - p(a_{k-1}))w_{10}}.$$

Hence, $\epsilon_k := c(a_{k-1}) - c(a_k) > 0$ for any $k = \ell + 1, ..., n$. 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, I need only 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 $\hat{p}(c(a)|a) \leq \hat{p}(\sum_{k=\ell+1}^{n} \epsilon_k|a)$.

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

$$\begin{aligned} p(a_{\ell+1}) &\geq p(a_\ell) - \frac{\epsilon_1}{p(a_\ell)w_{11} + (1 - p(a_\ell))w_{10}} \\ &= \hat{p}(0|a) + \hat{p}'(0|a)\epsilon_1 \\ &\geq \hat{p}(\epsilon_{\ell+1}|a), \end{aligned}$$

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 $m = \ell + 1, ..., K$. I show 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,

$$\begin{aligned} p(a_{K+1}) &\geq p(a_K) - \frac{\epsilon_{K+1}}{p(a_K)w_{11} + (1 - p(a_K))w_{10}} \\ &\geq \hat{p}\Big(\sum_{k=\ell+1}^{K} \epsilon_k|a\Big) + \hat{p}'\Big(\sum_{k=\ell+1}^{K} \epsilon_k|a\Big)\epsilon_{K+1} \\ &\geq \hat{p}\Big(\sum_{k=\ell+1}^{K} \epsilon_k + \epsilon_{K+1}|a\Big), \end{aligned}$$

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

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

initial value problem,

$$\tilde{p}'(t|a) = f(\tilde{p}(t|a)) = \frac{-1}{\tilde{p}(t|a)w_{11} + (1-\tilde{p}(t|a))w_{10}}$$
$$\tilde{p}(\bar{c} - c(a)|a) = p(a),$$

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 $\tilde{p}(\bar{c} - c(a) + t|a) = \hat{p}(t|a)$ 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.[25]

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, towards 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, $(a_1, ..., a_n)$, for which (i) a_1 is the maximal element of A and $p(a_n) = p$, (ii) $p(a_1) > ... > p(a_n)$, and (iii) $a_k \in \overline{BR}(a_{k-1})$ (so that $c(a_1) > ... > c(a_n)$) for $k = 2, ..., n$. It suffices to consider the case in which $\bar{p} > 0$, so that for any $\bar{a}_0 \in \underset{a_0}{argmax}\ \hat{p}(\hat{t}(a_0)|a_0)$, $\tilde{p}'(\cdot|\bar{a}_0)$ 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 $(a_1, ..., a_n)$ 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,

$$\tilde{p}(\bar{c} - c(a_1)|a_1) = p(a_1) \geq p(\bar{a}_0) = \tilde{p}(\bar{c} - c(\bar{a}_0)|\bar{a}_0) > \tilde{p}(\bar{c} - c(a_1)|\bar{a}_0),$$

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.

[25] See, for instance, Theorem 2.2 of Coddington and Levinson (1955).

Hence, by Claim 1,

$$p = p(a_n) \geq \hat{p}(\hat{t}(a_1)|a_1) \geq \hat{p}(\hat{t}(\bar{a}_0)|\bar{a}_0) = \bar{p}.$$

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

$$p(a_k) < \tilde{p}(\bar{c} - c(a_k)|\bar{a}_0) \leq \tilde{p}(\bar{c} - c(\bar{a}_0)|\bar{a}_0) = p(\bar{a}_0).$$

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.

I show 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,

$$p(\bar{a}_0)[p(a_{k-1})w_{11} + (1 - p(a_{k-1}))w_{10}] - c(\bar{a}_0) \geq p(a_k)[p(a_{k-1})w_{11} + (1 - p(a_{k-1}))w_{10}] - c(a_k),$$

$$\iff -\left[\frac{p(\bar{a}_0) - p(a_k)}{c(\bar{a}_0) - c(a_k)}\right] \leq -\left[\frac{1}{p(a_{k-1})w_{11} + (1 - p(a_{k-1}))w_{10}}\right].$$

Notice that,

$$-\left[\frac{p(\bar{a}_0) - p(a_k)}{c(\bar{a}_0) - c(a_k)}\right] \leq \frac{\tilde{p}(\bar{c} - c(\bar{a}_0)|\bar{a}_0) - \tilde{p}(\bar{c} - c(a_k)|\bar{a}_0)}{(\bar{c} - c(\bar{a}_0)) - (\bar{c} - c(a_k))} \leq \tilde{p}'(\bar{c} - c(a_k)|\bar{a}_0),$$

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

I need only exhibit a sequence of action spaces (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. 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 $a_0 \in A^0$.[26]

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:

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

and,

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

Now, consider a sequence of action spaces (A_n), with $A_n := \{a_1^n, a_2^n, ..., a_n^n\} \cup A^0$. 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 $c(a_{k-1}^n) - c(a_k^n) = \frac{\tilde{c}}{n} := \epsilon(n)$ for $k = 2, .., n$, $\rho(n) := \frac{1}{n^2}\frac{\tilde{c}}{w_{11}+1}$, 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 $k = 2, ..., n$. Notice,

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

for $k = 2, ..., n$ so that $a_1^n > a_2^n > ... > a_n^n$. 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

[26] 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 time and therefore, invoking Claim 2, is always above the differential equations associated with other known actions.

bounded by $\kappa_1 > 0$, and $\tilde{p}''(\cdot|a)$ is bounded by $\kappa_2 > 0$, it must be the case that

$$|p(\bar{a}_n) - \tilde{p}(\bar{c}|\bar{a}_0)| \leq \left[\frac{e^{c(a)\kappa_1} - 1}{\kappa_1}\right]\left[\frac{\epsilon(n)}{2}\kappa_2 + \frac{\rho(n)}{\epsilon(n)}\right].$$

Since $\epsilon(n) \to 0$ as $n \to \infty$ and $\frac{\rho(n)}{\epsilon(n)} = \frac{1}{n}\frac{1}{w_{11}+1} \to 0$ as $n \to \infty$, the right-hand side approaches zero. Hence, $p(\bar{a}_n)$ becomes arbitrarily close to $\tilde{p}(\bar{c}|\bar{a}_0) = \bar{p}$ as $n \to \infty$.

I need only 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 $t \in [\underline{t}, \bar{c}]$ 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 $k = 2, ...n$. 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 $\hat{t} < \bar{t}$, I can 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 $c(a_{k-1}^n) - c(a_k^n) = \frac{\hat{t}}{n} := \epsilon(n)$ for all $k = 1, .., n$ and $\rho(n) := \frac{1}{n^2}\frac{\hat{t}}{w_{11}+1}$ 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 distribution $p(a_n^n)$ converging to a point in $[0, \epsilon)$ as n approaches infinity. This establishes that $\bar{p} = 0$ is the greatest lower bound.

A.4 Proof of Lemma 7

Let

$$(w^*, a_0^*) \in \underset{w \in [0,1], a_0 \in A^0}{\arg\max} (1 - w)(p(a_0) - \frac{c(a_0)}{w}),$$

$p^* := p(a_0^*)$, and $c^* := c(a_0^*)$. By the assumption of non-triviality, $p^* > c^*$ since choosing any action in A^0 that does not satisfy this property results in at most zero profit. By the assumption that productive known actions are costly, $c^* > 0$ and so

$w^* = \sqrt{\frac{c^*}{p^*}} \in (0,1)$. Moreover,

$$V^*_{IPE} = (1-w^*)(p^* - \frac{c^*}{w^*}) < 1 - w^*.$$

Now, consider the JPE setting $w_{10} = w^* - \epsilon$, for $\epsilon > 0$ small, and

$$p^* w_{11} + (1-p^*) w_{10} = w^*.$$

I show that the principal obtains a strictly higher profit than V^*_{IPE}. Since $V^*_{IPE} = (1-w^*)(p^* - \frac{c^*}{w^*}) < 1 - w^*$, I need only show that the principal obtains a higher payoff in the worst-case shirking equilibrium.

Elementary methods show that the solution to the differential equation in Lemma 6 associated with a^*_0 evaluated at c^* is:

$$\begin{aligned}\bar{p}(\epsilon) :&= \frac{\sqrt{(p^* w_{11} + (1-p^*) w_{10})^2 - 2c^*(w_{11} - w_{10})} - w_{10}}{w_{11} - w_{10}} \\ &= \frac{\sqrt{(w^*)^2 - 2\frac{c^*}{p^*}\epsilon} - (w^* - \epsilon)}{\epsilon / p^*}.\end{aligned}$$

Moreover, it is easy to show that

$$\lim_{\epsilon \to 0^+} \bar{p}(\epsilon) = p^* - \frac{c^*}{w^*},$$

and

$$\lim_{\epsilon \to 0^+} \bar{p}'(\epsilon) = -\frac{1}{2} p^* w^*.$$

Notice, if both agents choose an action that results in success with probability $p(\epsilon)$, the principal's payoff from each agent in the shirking equilibrium is

$$\pi(\epsilon) := \bar{p}(\epsilon)\left[1 - (\bar{p}(\epsilon) w_{11} + (1 - \bar{p}(\epsilon)) w_{10})\right]$$

and

$$\lim_{\epsilon \to 0^+} \pi(\epsilon) = (p^* - \frac{c^*}{w^*})(1 - w^*),$$

the least upper bound payoff the principal obtains from each agent within the class

of IPE. Since $\bar{p}$ (as defined in Lemma 6) is weakly larger than $\bar{p}(\epsilon)$ for every $\epsilon > 0$ and profits are strictly increasing in the probability with each worker succeeds when $\epsilon > 0$ is small, I need only show that $\pi(\epsilon)$ increases in ϵ at zero to demonstrate the existence of an improvement in the principal's payoff.[27]

It suffices to show that

$$\partial_+ \pi(0) > 0,$$

where ∂_+ is the right derivative of $\pi(\epsilon)$ at 0. For $\epsilon > 0$, the derivative of π is well-defined and equals

$$\pi'(\epsilon) = \bar{p}'(\epsilon)(1 - (\bar{p}(\epsilon)w_{11} + (1 - \bar{p}(\epsilon))w_{10})) - \bar{p}(\epsilon) \underbrace{\frac{d}{d\epsilon}[\bar{p}(\epsilon)w_{11} + (1 - \bar{p}(\epsilon))w_{10}]}_{= \frac{-c^*}{\sqrt{p^*(c^* - 2c^*\epsilon)}}}.$$

Hence,

$$\begin{aligned}
\partial_+ \pi(0) = \lim_{\epsilon \to 0^+} \pi'(\epsilon) &= (\lim_{\epsilon \to 0^+} \bar{p}'(\epsilon))(1 - w^*) + (\lim_{\epsilon \to 0^+} \bar{p}(\epsilon))w^* \\
&= (-\frac{1}{2} p^* w^*)(1 - w^*) + (p^* - \frac{c^*}{w^*})w^* \\
&= \frac{1}{2}(p^* w^* - c^*).
\end{aligned}$$

So,

$$\partial_+ \pi(0) > 0 \iff p^* w^* > c^*,$$

which holds because $w^* > 0$ and $p^* > c^*$, establishing the desired result.

[27] Simply observe that, for $\epsilon > 0$ small,

$$\frac{\partial}{\partial p}[p(1 - w^*) + p(1 - p)\epsilon] = (1 - w^*) + (1 - 2p)\epsilon > 0,$$

since $w^* < 1$.

A.5 Proofs for Section 4.3.5

Existence

A worst-case optimal JPE with $w_{10} = w_{00} = 0$ is one that solves following maximization problem:

$$\max_{w_{11},w_{10}} \min\{1 - w_{11}, \bar{p}\,[\bar{p}(1 - w_{11}) + (1 - \bar{p})(1 - w_{10})]\}$$

$$\text{subject to}$$

$$\bar{p} = \max_{a_0 \in A^0} \hat{p}(\hat{t}(a_0; w_{11}, w_{10})|a_0; w_{11}, w_{10}),$$

$$w_{11} > w_{10} \geq 0.$$

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 (I now make explicit the terms that depend on the wage scheme).

I argue that the solution set of the latter problem coincides with that of the following:

$$\max_{w_{11},w_{10}} \min\{1 - w_{11}, \bar{p}\,[\bar{p}(1 - w_{11}) + (1 - \bar{p})(1 - w_{10})]\}$$

$$\text{subject to}$$

$$\bar{p} = \max_{a_0 \in A^0} \hat{p}(\hat{t}(a_0; w_{11}, w_{10})|a_0; w_{11}, w_{10})$$

$$1 \geq w_{11} \geq w_{10} \geq 0.$$

I may bound w_{11} above by 1 without altering the solution set because any larger wage cannot be eligible (it yields the principal a profit of at most zero by the first argument of the objective function). I may relax the strict inequality between w_{11} and w_{10} to be a weak relationship without altering the solution set since I have already shown that there exist wages $w_{11} > w_{10}$ that yield the principal strictly higher profits than any wage scheme setting $w_{11} = w_{10}$.

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 function

$$f : \mathcal{D} \to \mathbb{R}$$

$$(w_{11}, w_{10}) \mapsto \min\{1 - w_{11}, \bar{p}\,[\bar{p}(1 - w_{11}) + (1 - \bar{p})(1 - w_{10})]\},$$

with,

$$\bar{p} = \max_{a_0 \in A^0} \hat{p}(\hat{t}(a_0; w_{11}, w_{10})|a_0; w_{11}, w_{10})$$

is continuous.[28] Hence, the Weierstrass Theorem (Theorem 3.1 of Sundaram (1996)) ensures the existence of a solution.

Uniqueness

The proof of Lemma 4 shows that any contract that is not a JPE and does not set $w_{11} > 0$, $w_{00} > 0$, and $w_{10} = w_{01} = 0$ is weakly improved upon by an IPE or RPE. Lemma 5 and Lemma 7 then establish that such contracts are strictly suboptimal. So, all that is left to show is that (i) any JPE with either $w_{00} > 0$ or $w_{01} > 0$ is strictly suboptimal and (ii) any contract setting $w_{11} > 0$ and $w_{00} > 0$ (with $w_{10} = w_{01} = 0$) is strictly suboptimal.

For case (i), notice that the characterization of the principal's worst-case payoff given a JPE identified in Lemma 6 holds when replacing w_{11} with $w_{11} - w_{01}$ and w_{10} with $w_{10} - w_{00}$ in Equation 1 and setting

$$V(w) = 2\min\{1-w_{11}, \bar{p}\,[\bar{p}(1-w_{11}) + (1-\bar{p})(1-w_{10})] + (1-\bar{p})\,[\bar{p}(-w_{01}) + (1-\bar{p})(-w_{00})]\}.$$

If $1 - w_{11}$ is strictly smaller than the principal's payoff in the shirking equilibrium, then the contract could not have been optimal; the principal could reduce w_{11} by a small amount and strictly increase her payoffs (because $\bar{p}$ is continuous in w_{11}). If the principal's payoff in the shirking equilibrium is larger than $1-w_{11}$, then setting $w'_{01} = 0$, $w'_{11} = w_{11} - w_{01}$, $w'_{00} = 0$, and $w'_{10} = w_{10} - w_{00}$ leaves $\bar{p}$ unchanged, thereby strictly increasing the principal's profits in the shirking equilibrium in all cases in which $\bar{p} > 0$. If w_{11} is affected by this adjustment, then this ensures that the principal's payoff strictly increases. If not, then decreasing w'_{11} by a small amount strictly increases the principal's payoff in the case that $1 - w_{11}$ is strictly smaller than that in the shirking equilibrium.

[28] 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.

For case (ii), the characterization of the principal's worst-case payoff given a JPE identified in Lemma 6 holds when replacing the law of motion in Equation 1 with

$$\hat{p}'(t) = f(\hat{p}(t)) := \frac{-1}{\hat{p}(t)w_{11} - (1-\hat{p}(t))w_{00}}$$

and setting

$$V(w) = 2\min\{1 - w_{11}, \bar{p}^2(1 - w_{11}) + (1-\bar{p})^2(-w_{00})\}.$$

The proof of Lemma 4 establishes that setting $w_{00} = 0$ yields a weak improvement for the principal. It also establishes that this improvement is strict if, given this adjustment, the principal's payoff (from each agent) in the shirking equilibrium is smaller than $1 - w_{11}$. So, I need only consider the case in which $1 - w_{11}$ is strictly smaller than the principal's payoff in the shirking equilibrium. In this case, the resulting contract is strictly suboptimal; the principal could reduce w_{11} by a small amount and strictly increase her payoff (because $\bar{p}$ is continuous in w_{11}). Hence, the original contract with $w_{00} > 0$ is strictly suboptimal as well.