Theoretical Economics 18 (2023), 1623–1663 1555-7561/20231623

Robust contracting under double moral hazard

Gabriel Carroll
Department of Economics, University of Toronto

Lukas Bolte
Department of Economics, Stanford University

We study contracting when both principal and agent have to exert noncontractible effort for production to take place. An analyst is uncertain about what actions are available and evaluates a contract by the expected payoffs it guarantees to each party in spite of the surrounding uncertainty. Both parties are risk-neutral; there is no limited liability. Linear contracts, which leave the agent with a constant share of output in exchange for a fixed fee, are optimal. This result holds both in a preliminary version of the model, where the principal only chooses to supply or not supply an input, and in several variants of a more general version, where the principal may have multiple choices of input. The model thus generates nontrivial linear sharing rules without relying on either limited liability or risk aversion.

Keywords. Uncertainty, asymmetric information, principal-agent model, linear contracts, double-sided moral hazard, robustness.

JEL classification. D81, D82, D86.

1. Introduction

Why do profit-sharing rules arise in agency relationships? And what determines the form that such rules take?

The bulk of the literature on principal-agent models, since Holmström (1979) and Grossman and Hart (1983), emphasizes risk aversion, and the importance of the resulting tradeoff between providing incentives and insurance. In these models, typically output results from some costly and unobserved effort provided by the agent. If the agent were risk-neutral, the optimal solution would just be "selling the firm" for a fixed fee, thereby making the agent a full residual claimant to the consequences of his effort. A separate branch of the literature focuses on limited liability constraints (Innes (1990)), which make it impossible for the principal to capture the surplus from selling the firm

Gabriel Carroll: gabriel.carroll@utoronto.ca
Lukas Bolte: lukas.bolte@outlook.com

This work has benefited from comments from audiences at the NSF/NBER/CEME Mathematical Economics Conference, the Econometric Society European Winter Meetings, and internal workshops at Stanford. Authors are listed in random order; both authors contributed equally. The first-listed author thanks the Research School of Economics at ANU for their hospitality during a sabbatical, and the NSF for financial support through a CAREER grant. This paper has also appeared as Chapter 3 of the second-listed author's Ph.D. dissertation at Stanford University.

 https://doi.org/10.3982/TE4916

to the agent; then the principal optimally gives weaker incentives to avoid ceding too much of the surplus.

Yet in many situations, we observe sharing rules between firms, where neither risk aversion nor limited liability seem to be key considerations. We focus here on a different issue: *double-sided moral hazard*, that is, the importance of giving incentives for both the principal and agent to provide noncontractible inputs. A leading application where this arises is in franchising (Bhattacharyya and Lafontaine (1995)): contracts typically specify that a portion of revenues should be returned to the franchisor as royalties; this sharing ensures that the franchisor has incentives to advertise and maintain the reputation of the brand, while the franchisee has incentives to exert effort in local management. Other applications where double moral hazard has been argued to be relevant in determining contract terms include warranties, where both quality provision by the producer and care by the user are subject to moral hazard (Cooper and Ross (1985)); sharecropping (Eswaran and Kotwal (1985)); effort at cost savings in supply chains (Corbett and DeCroix (2001), Corbett, DeCroix, and Ha (2005)); and collaborative business services such as consulting (Roels, Karmarkar, and Carr (2010)). Our study is meant to be general and not geared toward any specific application.

In the context of one-sided moral hazard, there is by now a rich theoretical literature developing various models that generate different functional forms for optimal sharing rules, thereby aiming to understand the advantages of each. This includes linear contracts (Holmström and Milgrom (1987), Diamond (1998)), debt contracts (Innes (1990), Hébert (2018)), and threshold-based bonus contracts (Lopomo, Rigotti, and Shannon (2011), Georgiadis and Szentes (2020)), among others. For double moral hazard, the same questions are much less developed. The seminal model of double moral hazard is that of Bhattacharyya and Lafontaine (1995). That paper was oriented primarily toward franchising applications and noted that in practice franchising contracts are often linear. Their model indeed predicts the existence of an optimal contract that is linear. However, this prediction relies on particular structural assumptions: most importantly, that output, while random, depends on the two parties' effort levels only via a one-dimensional aggregate of the two. The key argument is that, given any candidate contract, a linear contract with appropriately chosen slope can replicate the same first-order conditions for each party. As observed by Kim and Wang (1998), this argument is not specific to linear contracts; there are many optimal contracts, and roughly speaking, any well-behaved one-parameter family of contracts would contain some optimal contract, for the same reason. They further argue that trying to select among the optima by adding a small amount of risk aversion fails to pick out the linear contract. Moreover, even without risk aversion, once we depart from the strong assumption of one-dimensional composite effort, linear contracts can fail to be optimal (see Example 3 in Appendix A).

In this paper, we identify a specific virtue of linear contracts, and do so with minimal structural assumptions. Our argument is based on robustness to uncertainty about details of the environment. The idea is simple: suppose (for example) a contract specifies that $1/4$ of output is left to the agent, with the remaining $3/4$ going to the principal. If an analyst knows that the agent is able to secure an expected payoff of, say, 1000 for himself under such a contract, then she can infer the principal is guaranteed to get at least 3000 in expectation, without needing to know details about exactly what the agent can do or what his optimal action is. This intuition was previously expressed by Carroll (2015), in a one-sided moral hazard model. The principal's guarantee is formalized by a maxmin criterion, and the main result is that the highest possible such guarantee is attained by a linear contract. Essentially, with enough uncertainty about the actions available to the agent, the only thing known about what he will do is a lower bound on his expected payoff; and the only useful tool to turn this into a guarantee on the principal's expected payoff is a linear relationship between the two.

A crucial element of that model, however, is a limited liability constraint. Without such a constraint, the one-sided moral hazard model would again yield the trivial solution of selling the firm to the agent. In the present paper, we show how the same intuition can be expressed in a model with double moral hazard (and no limited liability).

Incorporating maxmin-style uncertainty into a model with noncontractible choices by both parties raises modeling questions. First, how should the possible actions of the agent be modeled, if they might interact with choices by the principal (and vice versa)? Second, what should we assume about how the principal will make her input choice? Note that the latter question must be answered to formulate the agent's participation constraint: We will write the optimal contracting problem as maximizing the principal's guarantee, subject to guaranteeing the agent at least zero. (In our setting, this is equivalent to characterizing the Pareto frontier of contracts, as evaluated by their guarantees to each party.)

Our modeling approach cuts down the difficulties by having the parties move sequentially. In the most basic version of our model, once the contract is signed, the principal moves first and makes a binary choice: either she supplies a costly input or not. If the principal supplies the input, then the agent takes his action; if not, no output can be produced and the relationship ends. This structure allows us to model an action by the agent simply via its effort cost and the resulting probability distribution over output, as in Carroll (2015). For a contract to be able to provide positive guarantees to both parties, it must assure the agent that the principal will have enough incentive to supply the input. With this in mind, we show that linear contracts can provide the optimal guarantee for the principal.

This simple model also delivers some intuitive predictions. The optimal contract is one whose slope (the share paid to the agent) is as high as possible, subject to leaving enough to the principal to incentivize her to provide the input. This maximizes the total surplus, which is then fully extracted by the principal with an appropriately set fixed fee.

The all-or-nothing input assumption is a strong one, so we subsequently introduce the more general version of the model, where the principal has multiple choices of input. The question then arises as to how the principal should choose. One modeling approach, in the spirit of maximizing guarantees, is to assume that the principal chooses her input to maximize her worst-case payoff (over the agent's possible technologies, consistent with the assumed knowledge). With this approach, linear contracts may no longer be optimal: The additional degrees of freedom provided by nonlinear contracts can be a useful tool to commit the principal to choosing a high input rather than a low one, by making the worst-case outcome after a low input especially undesirable for the principal. This commitment can make it easier to ensure the agent's participation.

In fact, it should be no surprise that the linearity result fails in this model: After all, already in a one-sided moral hazard model, with certainty about the agent's technology, the optimal contract would typically be finely tailored to that knowledge; similarly, in a double-sided moral hazard model with certainty about how the *principal* will behave, it is useful to tailor the contract to that knowledge, even if there is uncertainty about the agent's side. This suggests that a model with sufficient uncertainty about the principal's behavior as well might better express the robustness property of linear contracts. More specifically, the argument using uncertainty about the agent is based on the possibility that the agent might produce any output distribution giving him sufficiently high expected payment; this prompts us to consider models in which the principal, too, might have access to any distribution with sufficiently high expected payment for herself. Of course, it remains to show how this idea can be fleshed out, since in our timing, the principal does not directly choose the output distribution.

We present three variations of such a model. In our first formulation, to define a contract's guarantee for the agent, we continue to assume that the principal will choose her input based on a worst-case criterion, but the principal may have additional knowledge of the agent's actions (beyond what is known to the analyst), which can change her optimal choice of input. Our second formulation allows that the principal may have altogether new, unforeseen input choices available. And our third formulation assumes that the principal does fully know the agent's technology when making her input choice, but the analyst still does not have this knowledge. All three models deliver optimality results for linear contracts. Together, they show that there are various ways to fill in the specifics; what is crucial for the argument is to have enough uncertainty about both parties. Linear contracts then turn out to be optimal because they ensure both that every distribution that is high-paying (in expectation) for the agent is also high-paying for the principal *and* vice versa.

The goal of our overall exercise is twofold: to offer a tractable general-purpose model of double moral hazard; and to specifically express the robustness intuition underlying linear contracts, with as little reliance on functional form assumptions as possible. The sequential-move structure is a significant difference from most existing models of double moral hazard, but it has been used before, e.g., Demski and Sappington (1991). Arguably, moving sequentially is no more or less of a gross oversimplification of agency relationships than the one-shot simultaneous structure usually assumed. And this timing has the advantage of allowing for simple (albeit customized) approaches to modeling uncertainty about the principal's behavior that have no obvious counterpart in a simultaneous-move model.

A paper closely related to ours is that of Dai and Toikka (2022). They consider robust incentives for teams of agents who must simultaneously choose costly actions and share the output. (They also consider a model in which there is a residual claimant who is not part of the team; that version of the model is less closely related to ours.) They also derive linear contracts as optimal, but they obtain much stronger conclusions: for *any* nonlinear sharing contract, there is the potential for a "race to the bottom" that leads to no output being produced at all. We discuss further the contrast between their approach and ours in our concluding section.

Aside from this, our work fits into the broader literature on robustness foundations for linear incentive contracts. This includes mostly Bayesian models (Holmström and Milgrom (1987), Diamond (1998), Barron, Georgiadis, and Swinkels (2020)). Chassang (2013) gives a related maxmin-optimality result.

The remainder of the paper proceeds as follows. Section 2 introduces the basic version of our model, with the binary choice by the principal (supply the input or not). We show that linear contracts provide the best guarantee to the principal. Section 3, building on the machinery developed in Section 2, introduces the more general version of the model, where the principal has multiple choices of input, and shows that linear contracts remain optimal under the three variant formulations. Section 4 sums up.

2. Single-input model

2.1 *Setup*

We begin by describing the simple single-input version of the model.

First, some notational conventions: let $\Delta(\mathcal{X})$ denote the space of Borel distributions on $\mathcal{X} \subseteq \mathbb{R}^k$, δ_x the degenerate distribution with weight 1 on x for $x \in \mathcal{X}$, $\text{conv}(\mathcal{D})$ the convex hull of set $\mathcal{D}$, and $\mathbb{R}^+$ the nonnegative reals.

A principal and an agent, who are both risk-neutral, may jointly participate in a production process. The principal may supply some input to the agent at a cost $c_P \in \mathbb{R}^+$. If she does not, then no production takes place, and output is zero (at no cost to either party). If she does supply the input, then the agent can take an action that (stochastically) produces output. Note that, although both parties make costly contributions to production, we use the asymmetric language ("input"/"action" and "principal"/"agent") to reflect their asymmetric roles in the model. There is some set $\mathcal{Y}$ of possible output realizations, which we assume is a compact subset of $\mathbb{R}^+$, with $0 \in \mathcal{Y}$ as the lowest possible output.

An *action* of the agent is modelled as a pair (F, c) where $F \in \Delta(\mathcal{Y})$ is the resulting distribution over output, and $c \in \mathbb{R}^+$ is the cost incurred by the agent. We use the term *technology* to denote a nonempty, compact set of possible actions.

There is an analyst, who wishes to make predictions about the outcome of the interaction; the analyst may or may not be the same person as the principal. We assume that there is some given technology $\widehat{\mathcal{A}}$, representing the actions that the analyst knows the agent can take. The agent's true technology is a superset, $\mathcal{A} \supseteq \widehat{\mathcal{A}}$. The agent knows $\mathcal{A}$; the analyst does not. Both $\widehat{\mathcal{A}}$ and c_P are known by all.

Incentives are provided by a contract that specifies how the output is divided between the principal and agent. Neither the principal's input nor the agent's action are contractible; only the output is. Thus, we define a *contract* w as a continuous function from the output space $\mathcal{Y}$ to the reals.[1] By convention, $w(y)$ is the share received by the

[1] Continuity serves only to ensure existence of best replies and is not a substantive restriction. It can also be relaxed to upper semicontinuity; all results go through, at the cost of requiring some extra verifications. See also Carroll (2015, footnote 1).

agent. A contract w is *linear* if it is of the form $w(y) = \alpha y + \beta$; such a contract will be denoted by $w[\alpha, \beta]$. A couple special cases are worth noting: the *zero contract*, $w[0, 0]$, pays a wage of 0 for any output level; a contract of the form $w[1, -p]$ entails selling the firm to the agent at price p.

We do not explicitly model where the contract comes from. Instead, we study behavior in the game between the principal and agent under a given contract. This allows us to define the contract's guarantee for each party, and from there we can ask what contract maximizes the principal's guarantee, subject to assuring the agent a guarantee at least zero.[2] This is equivalent to studying the Pareto frontier of (principal, agent)-guarantees across all possible contracts, since one can move along this frontier by simply adding or subtracting a constant from w. The approach is thus compatible with the analyst either being the principal herself or being a neutral third party.

The timing of the game is summarized below:

1. the principal chooses whether or not to supply the input. If she does not supply the input, output is 0, so her payoff is $-w(0)$ and the agent's payoff is $w(0)$. If she does supply the input, then
2. the agent chooses an action $(F, c) \in \mathcal{A}$;
3. output $y \sim F$ is realized;
4. payoffs are received: $y - w(y) - c_P$ to the principal and $w(y) - c$ to the agent.[3]

Our analysis of this game will be essentially based on backward induction, but we will have to be precise about what this means, in view of the uncertainty about $\mathcal{A}$.

We will find it useful to define a class of "eligible" contracts, those that guarantee the principal some strictly positive payoff and guarantee at least zero for the agent (the formal definition will appear shortly). Contract w will be eligible if, in the game above:

- at step 2, the agent chooses his action optimally given $\mathcal{A}$,
- anticipating this, at step 1, the principal finds it optimal to supply the input,

and for all $\mathcal{A}$, these strategies give payoffs that are positive for the principal and nonnegative for the agent.

Note that since the total surplus is zero if the principal does not supply the input, the desired payoff guarantees do indeed require us to ensure that the principal supplies the input. We will model this by specifying that the principal's payoff from supplying the input needs to be higher than from not supplying it, regardless of the technology. We emphasize that focusing on contracts with this property is not the same as assuming

[2] We will also shortly require the contract to give the principal a positive guarantee as well. This effectively assumes that both parties' outside options are zero. We could also consider more general outside options $(\underline{u}_P, \underline{u}_A)$; nothing would significantly change, except that if $\underline{u}_P + \underline{u}_A < 0$ then there may be trivial cases where it is optimal to sign a contract but then not supply the input.

[3] An alternative timing would allow the agent to opt for his outside option after the principal chose whether to supply the input. Under this timing, selling the firm at varying prices maps out the Pareto frontier of (principal, agent)-guarantees.

that the principal behaves as a maxmin optimizer at step 1. Indeed, this property also assures the principal's willingness to supply the input under other assumptions of her behavior, for example, if she is actually an expected-utility maximizer with some prior over $\mathcal{A}$ (and perhaps the analyst does not know what this prior is).

To formalize eligibility, we develop some notation. Consider the agent at step 2, after receiving the input. Denote the actions the agent might optimally choose, and his expected payoff associated with taking them, as

$$\mathcal{A}^*(w|\mathcal{A}) = \underset{(F,c)\in\mathcal{A}}{\arg\max}\left\{E_F\left[w(y)\right] - c\right\} \quad \text{and} \quad V_A(w|\mathcal{A}) = \max_{(F,c)\in\mathcal{A}}\left\{E_F\left[w(y)\right] - c\right\},$$

respectively. If the agent is indifferent between two actions, we will assume that he takes the action that maximizes the principal's payoff.

Since the technology is initially unknown, we evaluate the *principal's guarantee* from a contract w by the worst case over all possible technologies. If the principal supplies the input, then this worst-case expected payoff is

$$V_P(w|\widehat{\mathcal{A}}, c_P) = \inf_{\mathcal{A}\supseteq\widehat{\mathcal{A}}}\left(\max_{(F,c)\in\mathcal{A}^*(w|\mathcal{A})}\left\{E_F\left[y - w(y)\right]\right\} - c_P\right).$$

We may simply write $V_P(w)$ rather than $V_P(w|\widehat{\mathcal{A}}, c_P)$ when there is no ambiguity.

We likewise define the *agent's guarantee*; this is simply $V_A(w|\widehat{\mathcal{A}})$, since it evidently is the agent's payoff under the worst technology for him.

Now, we can give our formal definition of eligibility.

Definition 1. A contract w is *eligible* if

(E1) $V_P(w) > 0$;

(E2) $V_P(w) \geq -w(0)$; and

(E3) $V_A(w|\widehat{\mathcal{A}}) \geq 0$.

(In words: (E1) the principal's guarantee is positive, (E2) she prefers supplying the input to not supplying it, and (E3) the agent's guarantee is nonnegative.)

Let us argue more systematically that this formal criterion corresponds to our backward-induction description.

If the contract is eligible, then backward induction ensures that the principal is willing to supply the input at step 1, and so the parties are indeed guaranteed at least $V_P(w)$ and $V_A(w|\widehat{\mathcal{A}})$. Conversely, an ineligible contract cannot give the required guarantees for both parties: if (E2) fails, the principal may not supply the input; if (E2) holds but (E3) fails, then the agent is not guaranteed at least zero; and if (E2) and (E3) hold but (E1) fails then the principal's guarantee is $V_P(w) \leq 0$.

With this background in mind, we study how to maximize the principal's guarantee over the space of eligible contracts.

2.2 *Analysis*

2.2.1 *Existence of an eligible contract* It is not obvious when an eligible contract exists. In the one-sided moral hazard setting, Carroll (2015) makes the following assumption: there exists $(F, c) \in \widehat{\mathcal{A}}$ such that

$$E_F[y] - c > 0.$$

This assumption is enough to guarantee a positive total surplus and the existence of eligible contracts in that setting. In our setting, accounting for the cost to the principal of supplying the input, a positive total surplus is feasible if there exists $(F, c) \in \widehat{\mathcal{A}}$ such that

$$E_F[y] - c - c_P > 0. \tag{1}$$

Existence of such an action is certainly a necessary condition for existence of an eligible contract (this can be formally seen by adding (E1) and (E3)). Our first question is whether this condition is also sufficient. The answer is *no*. An intuition is that in general, some amount of the output needs to be given to the principal to incentivize her to supply the input. This means that the agent will have to be made a less-than-full residual claimant, and so the surplus available must be large enough that even without receiving all of it, the agent is still motivated to exert effort. Example 1 illustrates this in more detail.

Example 1. Consider a simple environment with $\mathcal{Y} = \{0, \bar{y}\}$, and only one known action, $\widehat{\mathcal{A}} = \{(\delta_{\bar{y}}, c)\}$, where $\bar{y} > c > 0$. Suppose w is an eligible contract in this environment. Denote $w(\bar{y}) = \bar{w}$ and $w(0) = \underline{w}$. It is optimal to set $\bar{w} = c$: if $\bar{w} < c$, then (E3) is violated; if $\bar{w} > c$, then we can reduce both $\bar{w}$ and $\underline{w}$ by some positive amount ϵ, strictly increasing the principal's guarantee while preserving eligibility.

If $\underline{w}$ is nonnegative, the principal does not receive any positive guarantee (this can be seen formally by considering technologies $\mathcal{A} = \{(\delta_{\bar{y}}, c), ((1-\epsilon)\delta_0 + \epsilon\delta_{\bar{y}}, 0)\}$, for small $\epsilon > 0$). Thus, we can focus on $\underline{w} < 0$. For every such $\underline{w}$, we will now determine the principal's guarantee. For any action in any possible technology, the expected output and expected wage lie on the dashed line in Figure 1 connecting $(0, \underline{w})$ and $(\bar{y}, \bar{w})$, which is depicted for

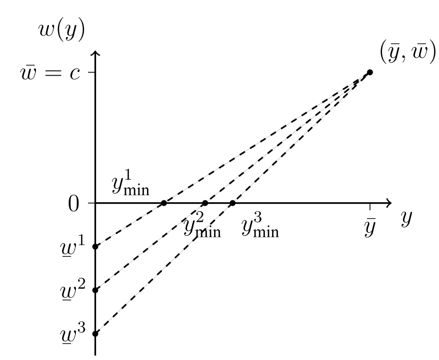


Figure 1. Positive surplus is not enough to guarantee the existence of an eligible contract.

different values of $\underline{w}$: $\underline{w}^1$, $\underline{w}^2$, and $\underline{w}^3$. The agent will never choose an action for which his expected payoff is less than 0, since he can achieve this payoff with the known action. In particular, the expected wage paid to the agent has to be nonnegative. Conversely, any such action is optimal for the agent in some technology. Then the worst-case expected output is given by the intersection of the dashed line and the horizontal axis; we call this worst-case expected output $y_{\min}$, depicted again in Figure 1 for the different values of $\underline{w}$. Algebraically, $y_{\min}$ is given by

$$y_{\min} = -\frac{\underline{w}\bar{y}}{c - \underline{w}}.$$

For w to be eligible, we require that (E1) and (E2) hold. At the worst-case expected output, the expected wage of the agent is 0 so that these conditions become

$$-\frac{\underline{w}\bar{y}}{c - \underline{w}} - c_P > 0 \quad \text{and} \quad -\frac{\underline{w}\bar{y}}{c - \underline{w}} - c_P \geq -\underline{w}.$$

Existence of $\underline{w}$ satisfying these conditions is equivalent to

$$\bar{y} - c - c_P \geq 2\sqrt{cc_P} \quad \text{and} \quad \bar{y} - c - c_P > 0. \tag{2}$$

This is a stronger condition than the existence of positive total surplus as in (1). $\Diamond$

In Example 1, (2) is a necessary and sufficient condition in a stylized environment. However, we can generalize the result to arbitrary environments.

Proposition 1. *An eligible contract exists if and only if there exists $(F, c) \in \widehat{\mathcal{A}}$ such that*

$$E_F[y] - c - c_P \geq 2\sqrt{cc_P} \quad \textit{and} \quad E_F[y] - c - c_P > 0.\text{[4]} \tag{3}$$

The proof of Proposition 1 is postponed to Appendix B, after the proofs of the remaining results in this section (on which it relies). All other proofs provided appear in order by subsection in Appendix B.

2.2.2 *Optimality of linear contracts* The next question we ask is how optimal contracts look like, provided they exist.

Theorem 1. *If an eligible contract exists, then among all eligible contracts there exists a linear contract that maximizes the principal's guarantee.*

There may also exist nonlinear contracts that attain the optimum: in particular, we can start from the linear contract and then change its shape at points outside the support of (known) actions. By adding an assumption to rule out this trivial multiplicity, we can ensure that only linear contracts can be optimal. Specifically, we say that $\widehat{\mathcal{A}}$ *satisfies the full-support condition* if for every action $(F, c) \neq (\delta_0, 0)$ in $\widehat{\mathcal{A}}$, F has full support on $\mathcal{Y}$.

[4]Dai and Toikka (2022) find essentially the same condition for existence of a contract with a positive guarantee in their teams model.

Corollary 1. *If $\widehat{\mathcal{A}}$ satisfies the full-support condition, then every eligible contract that maximizes the principal's guarantee is linear.*

The proof of Theorem 1 builds on the main proof from Carroll (2015), although we organize the ingredients of the proof a bit differently. The arrangement here will allow us to quickly leverage the same tools for the multiple-input versions of the model in Section 3.

For any contract w, we first characterize the *fundamental relationship* it induces between the principal's and the agent's guarantee, as the known technology varies. To do so, we need to introduce some further notation. For a fixed w, write

$$\mathcal{S} = \mathrm{conv}\big(\big\{\big(w(y) - c, y - w(y)\big) : y \in \mathcal{Y}, c \in \mathbb{R}^+\big\}\big). \tag{4}$$

Also, let

$$\mathcal{F} = \big\{(u, v) \in \mathcal{S} : \nexists\big(u', v'\big) \in \mathcal{S} \text{ such that } u' > u, v' < v\big\}. \tag{5}$$

$\mathcal{F}$ is depicted by the solid line in Figure 2 and describes the fundamental relationship between the principal's and the agent's guarantee from contract w as follows.

Let $\mathcal{T}$ be the set of all technologies. Let $\mathcal{R}$ denote the collection of pairs of the agent's guarantee and the principal's guarantee (ignoring the c_P term) as the known technology varies, that is,

$$\mathcal{R} = \big\{\big(V_A\big(w|\widehat{\mathcal{A}}'\big), V_P\big(w|\widehat{\mathcal{A}}', 0\big)\big) : \widehat{\mathcal{A}}' \in \mathcal{T}\big\}.$$

Lemma 1. *For any contract w,*

$$\mathcal{R} = \mathcal{F}.$$

For an intuition behind the lemma, note that the pair of (agent's, principal's) expected payoffs for any possible action must lie in $\mathcal{S}$. Any known technology $\widehat{\mathcal{A}}'$ imposes a lower bound on the payoff that the agent can get. This corresponds to an assurance that the payoff pair lies to the right of some vertical line in the figure. Given this, the worst possible payoff for the principal is determined by the point where this vertical line intersects the lower boundary of $\mathcal{S}$, which is exactly a point on the frontier $\mathcal{F}$.

The proof of Theorem 1 then quickly follows: the lemma shows that the worst case for the principal under w (and known technology $\widehat{\mathcal{A}}$) must involve some action for which

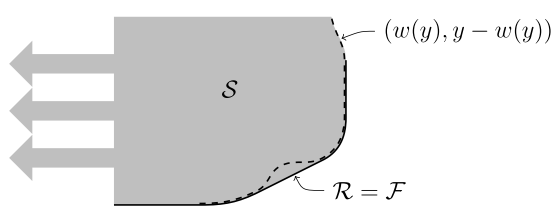


Figure 2. The dashed black line consists of points $(w(y), y - w(y))$ for $y \in \mathcal{Y}$. The solid black line describes the fundamental relationship between the principal's and the agent's guarantee. Set $\mathcal{S}$ is represented by the gray area and extends infinitely far to the left.

the resulting expected (agent, principal)-payoff pair lies on the boundary of the convex hull of w. Hence, either this action is degenerate, or more generally all points in its support lie along some line that is tangent to the convex hull. Replace w with this tangent line, which itself can be viewed as a linear contract w'. We show that w' guarantees at least the same expected payoff for the principal as w. This implies that (E1) for the linear contract w' is satisfied. We also have the comparison $w'(y) \geq w(y)$ for all y which implies conditions (E2) and (E3). Hence, w' is eligible and guarantees at least the same expected payoff as w. The full proofs (of the lemma, theorem, and corollary) are in Appendix B.

One missing detail above is verifying that an optimum among linear contracts actually exists. This is done in the analysis below, which not only shows that the optimum exists but characterizes it.

2.2.3 *Optimal linear contracts* The lemma below allows us to consider linear contracts $w[\alpha, \beta]$ with $\alpha \in [0, 1]$ only and identifies the principal's guarantee for the two boundary cases.

Lemma 2. *Consider any linear contract* $w[\alpha, \beta]$.

(A) $w[\alpha, \beta]$ *can only be eligible if* $0 \leq \alpha \leq 1$.

(B) If $\alpha = 0$ *and* $w[\alpha, \beta]$ *is eligible, then the principal's guarantee is given by* $\max_{(F,0)\in\widehat{\mathcal{A}}} E_F[y] - \beta - c_P$. *(If no action of the form* $(F, 0)$ *exists in* $\widehat{\mathcal{A}}$*, then no contract with* $\alpha = 0$ *is eligible.)*

(C) If $\alpha = 1$ *and* $w[\alpha, \beta]$ *is eligible, then the principal's guarantee is given by* $-\beta$. *(This case corresponds to selling the firm, which is only eligible if* $c_P = 0$.*)*

It remains to evaluate eligible linear contracts $w[\alpha, \beta]$ with $\alpha \in (0, 1)$. Consider any such contract. For any action (F, c) the agent can take, the principal's expected payoff is given by

$$E_F\big[(1-\alpha)y\big] - \beta - c_P,$$

which is increasing in $E_F[y]$. The agent, in turn, takes action (F, c) only if

$$E_F[\alpha y] + \beta - c \geq \max_{(F',c')\in\widehat{\mathcal{A}}} \big\{E_{F'}[\alpha y] + \beta - c'\big\}$$

implying that the expected output produced is bounded below by

$$\frac{1}{\alpha} \max_{(F',c')\in\widehat{\mathcal{A}}} \big\{E_{F'}[\alpha y] - c'\big\}.$$

Note that this bound is attained by some action of the form $(F, 0)$; such an action indeed is possible, that is, the value of the bound is above zero, since otherwise no positive level of total surplus would be guaranteed, and so the contract cannot be eligible.

The principal's guarantee for an eligible linear contract $w[\alpha, \beta]$ with $\alpha \in (0, 1)$ is therefore given by

$$V_P\big(w[\alpha, \beta]\big) = \frac{1-\alpha}{\alpha} \max_{(F,c)\in\widehat{\mathcal{A}}} \big\{E_F[\alpha y] - c\big\} - \beta - c_P. \tag{6}$$

Equation (6) also holds for contracts of the form $w[1, \beta]$ and $w[0, \beta]$ if we define $c/\alpha = 0$ for $\alpha = c = 0$ (and interpret $((1-\alpha)/\alpha)E_F[\alpha y]$ as $E_F[y]$ when $\alpha = 0$).

It follows that a linear contract $w[\alpha, \beta]$ is eligible if and only if

$$\frac{1-\alpha}{\alpha} \max_{(F,c)\in\widehat{\mathcal{A}}} \{E_F[\alpha y] - c\} - \beta - c_P > 0; \tag{7}$$

$$\frac{1-\alpha}{\alpha} \max_{(F,c)\in\widehat{\mathcal{A}}} \{E_F[\alpha y] - c\} - \beta - c_P \geq -\beta; \tag{8}$$

$$\max_{(F,c)\in\widehat{\mathcal{A}}} \{E_F[\alpha y] + \beta - c\} \geq 0. \tag{9}$$

Notice that for any given α, we can decrease β until (9) binds; doing so will increase V_P and will not break (7)–(8). Hence, we define

$$\beta(\alpha) = -\max_{(F,c)\in\widehat{\mathcal{A}}} \{E_F[\alpha y] - c\} \tag{10}$$

and focus on eligible contracts of the form $w[\alpha, \beta(\alpha)]$. Note also that for such contracts, the principal's guarantee is given by

$$V_P\big(w\big[\alpha, \beta(\alpha)\big]\big) = \frac{1}{\alpha} \max_{(F,c)\in\widehat{\mathcal{A}}} \{E_F[\alpha y] - c\} - c_P = \max_{(F,c)\in\widehat{\mathcal{A}}} \left\{E_F[y] - \frac{c}{\alpha}\right\} - c_P. \tag{11}$$

Evidently, this expression is weakly increasing in α, and strictly increasing wherever the relevant maximizer satisfies $c > 0$.

It remains to choose α to maximize this expression, subject to eligibility of the contract $w[\alpha, \beta(\alpha)]$. It then suffices to check (8), since (7) automatically holds at the maximum as long as some eligible contract exists. Because (8) carves out a closed set of possible values of α, and V_P is weakly increasing in α, it is now immediate that the maximum does indeed exist, and we have the explicit characterization, stated in the following result.[5]

Proposition 2. *If an eligible linear contract exists, then either the zero contract is an optimal eligible linear contract or the unique optimum in the class of eligible linear contracts is given by* $w[\alpha^*, \beta(\alpha^*)]$, *where*

$$\alpha^* = \max\left\{\alpha \in [0,1] : \frac{1-\alpha}{\alpha} \max_{(F,c)\in\widehat{\mathcal{A}}} \{E_F[\alpha y] - c\} - c_P \geq 0\right\}. \tag{12}$$

Furthermore,

$$\frac{1-\alpha^*}{\alpha^*} \max_{(F,c)\in\widehat{\mathcal{A}}} \{E_F\big[\alpha^* y\big] - c\} - c_P = 0 \quad \textit{and} \quad V_P\big(w\big[\alpha^*, \beta\big(\alpha^*\big)\big]\big) = -\beta\big(\alpha^*\big). \tag{13}$$

[5]Although the result is stated as optimizing over linear contracts, recall from Theorem 1 that the resulting contract is then optimal among *all* eligible contracts.

We note in passing a couple of implications of the results so far. First, with a small change in the specification of the environment, the outcome can change discontinuously between having a contract that provides a large positive guarantee to the principal and not having an eligible contract exist at all. Indeed, (13) and (10) imply that the principal's guarantee equals the amount that the agent gains by taking his best known action instead of producing zero output. When the condition for existence of an eligible contract (see Proposition 1) is just barely met, this gain must be bounded away from zero: otherwise, if the agent had an action available where he could produce very low output at zero cost, he would deviate to do so; but then total surplus would be negative (due to the input cost c_P), which is impossible.

Second, we consider some comparative statics. If the principal's input cost c_P increases, then the inequality in (12) becomes tighter, and so the optimal slope α^* decreases. If the costs c of all known actions increase, then again α^* decreases for the same reason. This contrasts with previous wisdom on double moral hazard, as summarized in Lafontaine (1992), which holds that the agent's share should be decreasing in the size of the principal's moral hazard problem and increasing in the size of the agent's moral hazard problem. If we think of c_P and c as representing these two "sizes," respectively, then our model recovers the first comparative static but not the second. To reconcile these ideas, note that in this model, it is optimal to maximize surplus (which the principal then extracts via appropriate choice of β) by making α as large as possible, subject only to the principal's own incentive constraint to provide the input; thus, the latter constraint is binding at the optimal contract. An increase in c makes it easier for the agent to be tempted by less-productive actions if they turn out to be available, thus reducing the principal's gain from supplying the input. Thus, an increase in c effectively *also increases the principal's moral hazard problem*, explaining why α^* decreases.

In Appendix C.1, we transform formula (12) into a more explicit characterization of the optimal linear contract $w[\alpha^*, \beta(\alpha^*)]$, which is used in the proof of Proposition 1 and may also be useful in computing examples.

3. Formulations for multiple inputs

The single-input assumption is a strong one, and it delivers a correspondingly extreme conclusion: the optimal contract is such that the principal's incentive to supply the input is binding. We now extend the model to allow the principal more choices, which may be interpreted as different types of input, or different quantities or qualities of input. For each choice that the principal makes, there is some cost to herself, and some resulting set of (known) actions $\widehat{\mathcal{A}}$ the agent can take in response. To the extent that $\widehat{\mathcal{A}}$ varies across inputs, this can be interpreted as variation in the set of physical actions that the agent can take, or as variation in the consequences (and perhaps the costs) of a given action by the agent; the difference in interpretation is immaterial. Thus, we will model an *input* directly as an ordered pair $(\widehat{\mathcal{A}}, c_P)$, describing the resulting known actions available to the agent, and the principal's cost of supplying the input. We will use the phrase *input space* to denote a nonempty, finite set of such pairs, interpreted as the set of inputs from which the principal can choose.

Let $\mathcal{W}$ be an input space. For each $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$, the agent's true technology is given by some $\mathcal{A}$ that is a superset of $\widehat{\mathcal{A}}$.

The timing of the game, under contract w, is summarized below:

1. the principal chooses whether to supply an input $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$, and if so, which one. If she does not supply any input, her payoff is $-w(0)$ and the agent's payoff is $w(0)$. If she does supply input $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$, then
2. the agent chooses an action $(F, c) \in \mathcal{A}$, where $\mathcal{A} \supseteq \widehat{\mathcal{A}}$ is the agent's corresponding technology;
3. output $y \sim F$ is realized;
4. payoffs are received: $y - w(y) - c_P$ to the principal and $w(y) - c$ to the agent.

We will use some of our analysis from the single-input environment (Section 2).

Definition 2. A contract w is *locally eligible via* $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$, if w is eligible in the single-input environment where the known technology is $\widehat{\mathcal{A}}$ and the cost of supplying the input is c_P.

3.1 *Weakly eligible contracts*

We would like to extend the backward-induction approach from the single-input model to this multiple-input model. Thus, we are interested in studying contracts w for which the following behavior is consistent with backward induction and guarantees a positive payoff for the principal and nonnegative payoff for the agent:

- at step 1, the principal chooses some particular input $(\widehat{\mathcal{A}}, c_P)$;
- at step 2, the agent chooses his action optimally given the corresponding $\mathcal{A}$.

However, without specifying how the principal's choice of input at step 1 is made, this description is evidently incomplete. Given our focus on maximizing the principal's guarantee V_P, a natural option is for the analyst to assume that the principal chooses whichever input $(\widehat{\mathcal{A}}, c_P)$ gives her the highest guarantee. This leads to the definition of *weak eligibility* that we formalize below. However, we shall subsequently propose several other, more preferred notions of eligibility (beginning in Section 3.2).

Definition 3. Let w be a contract. Define

$$V_P(w|\mathcal{W}) = \max_{(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}} V_P(w|\widehat{\mathcal{A}}, c_P).$$

We say that an input $(\widehat{\mathcal{A}}^*, c_P^*)$ is *an optimal input (given w)* if $V_P(w|\widehat{\mathcal{A}}^*, c_P^*) = V_P(w|\mathcal{W})$.

Definition 4. A contract w is *weakly eligible* if it is locally eligible via some input $(\widehat{\mathcal{A}}^*, c_P^*)$ that is optimal given w.

We define the principal's *guarantee* from such a contract w as the corresponding value of $V_P(w|\mathcal{W})$.

The fact that $(\widehat{\mathcal{A}}^*, c_P^*)$ is optimal given w implies that the guarantee-maximizing principal is willing to supply $(\widehat{\mathcal{A}}^*, c_P^*)$ at step 1. With this behavior by the principal, local eligibility is indeed the criterion to assure a positive guarantee for the principal and a nonnegative guarantee for the agent.

How does an optimal contract look like in this environment? As Example 2 below shows, even if weakly eligible linear contracts exist, nonlinear contracts may be preferable. The intuition is that nonlinear contracts may provide the principal with a form of commitment power over the choice of input that linear contracts cannot. In particular, a contract can specify a low flat wage over part of the output space, giving the agent insufficient incentives to exert effort following some inputs. This in turn will make those input choices unappealing to the principal, ensuring that she chooses a higher input instead.

Example 2. Suppose that there is a costly "high" input and a cheap "low" input: $\mathcal{W} = \{(\widehat{\mathcal{A}}^h, c_P^h), (\widehat{\mathcal{A}}^l, c_P^l)\}$. Let $\mathcal{Y} = [0, 30]$. To be concrete, let

$$\widehat{\mathcal{A}}^h = \{(\delta_{24}, 8)\} \text{ and } c_P^h = 4 \quad \text{and} \quad \widehat{\mathcal{A}}^l = \{(\delta_{12}, 3)\} \text{ and } c_P^l = 2.$$

We want to show that, among weakly eligible contracts, linear ones do not attain the optimum. To this end, we find an upper bound on the principal's guarantee for such contracts, show that no linear contract attains this upper bound, and finally construct a nonlinear contract that does and is thus optimal.

The principal's guarantee from any weakly eligible contract can be bounded above by considering all contracts that are locally eligible for some input (not necessarily an optimal input). Thus, let us consider contracts that are locally eligible for input $\widehat{\mathcal{A}}^h$ at cost c_P^h. By Proposition 2 and (10), an optimal contract, and the unique optimum that is linear, is given by $w[\alpha, \beta]$ with $(\alpha, \beta) = (2/3, -8)$. The principal's guarantee from supplying $(\widehat{\mathcal{A}}^h, c_P^h)$ given contract $w[\alpha, \beta]$ is

$$V_P\big(w[\alpha, \beta] | \widehat{\mathcal{A}}^h, c_P^h\big) = -\beta = 8.$$

This is an upper bound for the principal's guarantee provided by any weakly eligible contract. (A contract that is locally eligible for $(\widehat{\mathcal{A}}^l, c_P^l)$ cannot do better, since the total known surplus there is only $7 < 8$.) Can this upper bound be achieved by a linear contract? The only candidate contract is $w[\alpha, \beta]$. By construction, $w[\alpha, \beta]$ is locally eligible via $(\widehat{\mathcal{A}}^h, c_P^h)$. However, the principal's guarantee from supplying $(\widehat{\mathcal{A}}^l, c_P^l)$ given contract $w[\alpha, \beta]$ is

$$\begin{aligned} V_P\big(w[\alpha, \beta] | \widehat{\mathcal{A}}^l, c_P^l\big) &= \frac{1-\alpha}{\alpha} \max_{(F,c) \in \widehat{\mathcal{A}}^l} \big\{E_F[\alpha y] - c\big\} - \beta - c_P^l \\ &= \frac{1-2/3}{2/3}\{2/3 \cdot 12 - 3\} + 8 - 2 \\ &= \frac{1}{2}(8-3) + 8 - 2 > V_P\big(w[\alpha, \beta] | \widehat{\mathcal{A}}^h, c_P^h\big). \end{aligned}$$

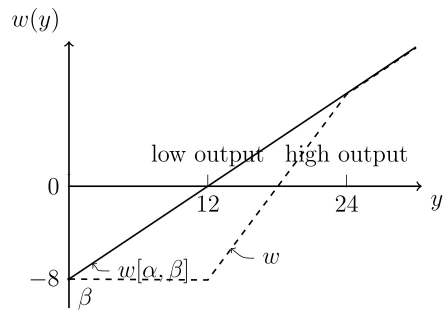


FIGURE 3. Contract $w[\alpha, \beta]$ is not weakly eligible because the principal has an incentive to supply the low input. By decreasing the wage for low levels of output, the resulting contract, w, does not guarantee any positive output after the low input. As a result, the principal optimally supplies the high input and w is weakly eligible.

Hence, $w[\alpha, \beta]$ is not weakly eligible, as $(\widehat{\mathcal{A}}^h, c_P^h)$ is not an optimal input given w, and under the principal's optimal choice of $(\widehat{\mathcal{A}}^l, c_P^l)$, the agent's guarantee is negative.

Finally, we design a nonlinear contract that is weakly eligible and achieves the upper bound, and that therefore is optimal. We will take the contract $w[\alpha, \beta]$ above and modify it. Specifically, let contract w be given by

$$w(y) = \begin{cases} \beta & \text{for } y \le 12 \\ \beta + \dfrac{8-\beta}{24-12}(y-12) & \text{for } 12 \le y \le 24 \\ \alpha y + \beta & \text{for } y \ge 24. \end{cases}$$

Contract w decreases the wage of the agent for low levels of output. Contracts w and $w[\alpha, \beta]$ are depicted in Figure 3.

We now check that the input $\widehat{\mathcal{A}}^h$ is now optimal and still gives the principal a guarantee of 8, so that the contract attains the upper bound.

Let $\mathcal{A}^l = \widehat{\mathcal{A}}^l \cup \{(\delta_0, 0)\}$ be the technology of the agent when the principal supplies input $\widehat{\mathcal{A}}^l$. Given contract w, the agent chooses action $(\delta_0, 0)$ and the principal's payoff is given by $-\beta - c_P^l = 8 - 2 = 6$ providing an upper bound for the principal's guarantee when supplying the low input. However, by supplying the high input, the principal guarantees the agent a payoff at least 0 (since he can get this from action $(\delta_{24}, 8)$) and thus guarantees herself at least 8, by a calculation similar to that used to derive (6): The agent's optimal action must pay him at least 0 in expectation; since $w(y) \le 2y/3 - 8$ for all y, this can happen only if expected output is at least 12; since the principal is paid $y - w(y) \ge y/3 + 8$, her expected payment is at least $12/3 + 8 = 12$, so her overall expected payoff is at least $12 - c_P^h = 8$. Thus, $(\widehat{\mathcal{A}}^h, c_P^h)$ is indeed optimal for the principal to choose, and w is weakly eligible. ◊

3.2 *Three proposed notions of eligibility*

We have seen that the weak eligibility notion we have put forward does not identify linear contracts as optimal. This could be interpreted as finding that Theorem 1 does not generalize to the multiple-input model. However, arguably weak eligibility is not really a generalization of eligibility from the single-input model, because it relies on the principal choosing the input to maximize her guarantee $V_P(w|\cdot,\cdot)$ at step 1. A contract may be weakly eligible yet fail to guarantee the agent a nonnegative payoff if the principal actually uses some other criterion to make her input choice. The solution concept thus demands a much more specific interpretation than we had in the single-input model: recall from the discussion in Section 2.1 that in that model, any eligible contract would still motivate the principal to provide the input (and thus still provide a nonnegative guarantee for the agent) under alternative assumptions about her decision-making criterion.

In the following, we relax the assumption that the principal chooses the input that maximizes $V_P(w|\cdot,\cdot)$. Instead, we postulate three alternative modeling approaches, each of which adds some uncertainty about the principal's behavior, and each of which will deliver results on optimality of linear contracts. In particular, the three different approaches allow that: (1) the principal maximizes her guarantee but might have additional knowledge of the agent's available actions beyond the $\widehat{\mathcal{A}}$'s known to the analyst; (2) the principal still maximizes her guarantee but might have access to input choices unforeseen by the analyst; or (3) the principal has full knowledge of the agent's technology associated with each input and maximizes her expected utility accordingly. Each of these approaches leads to a different test of whether any given contract guarantees a nonnegative payoff for the agent, and thus, they lead to three different notions of eligibility. We refer to contracts satisfying each as *eligible with further actions*, *eligible with further inputs*, *eligible with full knowledge*, respectively.

To see more concretely how each approach matters, return to Example 2, supposing that the parties share output according to the nonlinear contract w that we proposed there. We discuss the three proposed approaches in turn; in each case, we shall see that w no longer gives the agent a nonnegative guarantee.

(1) *Eligibility with further actions:* Suppose that the agent's technology $\mathcal{A}^l$ following the low input also includes an action $(\delta_{24}, 9)$, in addition to the $\widehat{\mathcal{A}}^l$ that was originally assumed known. Suppose, moreover, that the principal knows of the existence of this additional action. In this case, a lower bound on the principal's guarantee from choosing the low input can be calculated via (again) an analogue of the argument for (6): the agent is able to obtain a payoff of $8-9=-1$; since he always receives at most $2y/3-8$, any action he could take that gives him at least this high a payoff would have to generate expected output at least $(-1+8)/(2/3)=10.5$, and so would give the principal at least $(10.5/3+8)-c_P^l=9.5$. This is higher than the guarantee of 8 from choosing the high input. Thus, the principal would rather choose the low input at step 1. But this would leave the agent with a payoff of -1. So, the contract fails to give a nonnegative guarantee to the agent.

(2) *Eligibility with further inputs:* Now suppose that, when the principal has to choose her input, she has access to a new input that includes an action $(\delta_{24}, 9)$ and is costless for her to supply. By repeating the calculations above, replacing c_P^l with 0, the principal's guarantee associated with providing this input is now given by 11.5

whereas the agent may still only receive a payoff of -1. Thus, if the analyst allows that the principal's input space might extend beyond the original $\mathcal{W}$, then again contract w does not guarantee a nonnegative payoff to the agent.

(3) *Eligibility with full knowledge:* Lastly, suppose that at the input choice stage, the principal has full knowledge about the agent's technology associated with each input. For example, the principal knows that agent's technology is given by $\mathcal{A}^l = \widehat{\mathcal{A}}^l \cup \{(\delta_{24}, 9)\}$ and $\mathcal{A}^h = \widehat{\mathcal{A}}^h$ associated with the low input and high input, respectively.

With the low input provided, the agent would prefer his higher action over his lower one. Thus, as the output is 24 for both inputs, the principal would again choose the low (and cheaper) input, leaving the agent with a negative payoff.

The three extensions thus show various ways in which a weakly eligible contract can fail to ensure a nonnegative guarantee for the agent once we add some uncertainty about the principal's behavior. These extensions give rise to three new notions of eligibility, each capturing a different approach to modeling the uncertainty about the principal.

We do not view any one of the above variations as being a more natural eligibility notion than the others: the first perhaps hews closest to weak eligibility as we have presented it; but the second arguably treats the uncertainty about the principal more symmetrically to the uncertainty about the agent (by allowing each of them to have initially-unforeseen choices); and the third treats their *behavior* more symmetrically, by making both of them Bayesian maximizers, and having only the analyst use the worst-case criterion. For this reason, rather than single out one of these eligibility notions, we will consider each in turn, giving results on how the optimality of linear contracts is restored in each case. Intuitively, in a one-sided moral hazard environment, linear contracts are optimally robust when the set of output distributions the agent might generate is sufficiently large; the corresponding argument carries through under double moral hazard if the principal *also* potentially has access to a large enough space of output distributions. The three notions of robustness then provide different ways of formalizing what "access" means.

The different proofs demonstrating optimality of linear contracts in each context will build heavily on the machinery developed to prove Theorem 1.

3.2.1 *Eligibility with further actions* Our first alternative notion of eligibility allows that the principal has additional knowledge about the inputs while maintaining the assumption that she maximizes her guarantee. In particular, the principal may know of additional actions the agent has. Note that this remains consistent with the interpretation that the principal and the analyst are the same person, as long as we envision that the analyst's predictions take place at an ex ante stage before this additional knowledge is acquired. (An analogous remark applies for the other two notions of eligibility below.)

To develop a notion of eligibility in this context, we define the notion of a *worrisome input*. This is an input choice such that, for some realization of additional knowledge the principal might have about the agent's possible actions, the principal would choose this input, *and* this input choice then fails to guarantee a nonnegative payoff for the agent.

Definition 5. Given contract w, input $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ is *FA-worrisome* if there exists $\widehat{\mathcal{A}}' \supseteq \widehat{\mathcal{A}}$ such that

$$V_P(w|\widehat{\mathcal{A}}', c_P) > V_P(w|\mathcal{W}) \quad \text{and} \quad V_A(w|\widehat{\mathcal{A}}') < 0.$$

This allows us to define our first strengthened version of eligibility.

Definition 6. A contract w is *eligible with further actions (EFA)* if it is locally eligible via some optimal input and no $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ is FA-worrisome.

We define the principal's *guarantee* from such a contract w as the corresponding value of $V_P(w|\mathcal{W})$.

Notice that this definition of the principal's guarantee is the same as was used under weak eligibility. Indeed, we wish to say that an EFA contract guarantees the principal a certain level of payoff if the principal can choose her subsequent input in a way that ensures her that payoff no matter what the agent's technology is; and this definition of the principal's guarantee captures that condition. Again, depending on what she knows, the principal may or may not actually choose this particular input.

We now have the following proposition.

Proposition 3. *If an EFA contract exists, then among all EFA contracts there exists a linear contract that maximizes the principal's guarantee.*

The proof of Proposition 3 builds on the fundamental relationship and other machinery developed earlier to prove Theorem 1. A key step is to characterize EFA contracts in terms of this machinery, as we now describe. Given a contract w, we can define a frontier for each choice of input $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$:

$$\mathcal{F}_{(\widehat{\mathcal{A}}, c_P)} = \{(u, v) : u \geq V_A(w|\widehat{\mathcal{A}}), v \geq V_P(w|\widehat{\mathcal{A}}, c_P), (u, v + c_P) \in \mathcal{F}\}$$

where $\mathcal{F}$ is as was defined in (5).

Note that for any $(u, v) \in \mathcal{F}_{(\widehat{\mathcal{A}}, c_P)}$, there exists $\widehat{\mathcal{A}}' \supseteq \widehat{\mathcal{A}}$ such that

$$\big(V_A(w|\widehat{\mathcal{A}}'), V_P(w|\widehat{\mathcal{A}}', c_P)\big) = (u, v).$$

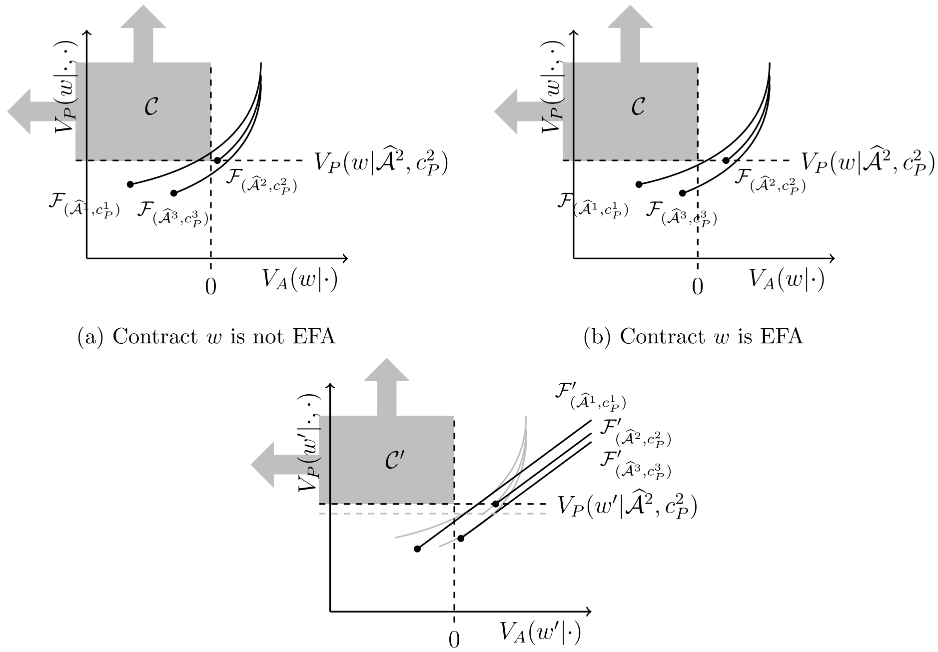
Indeed, the existence of some technology $\widehat{\mathcal{A}}'$ with these guarantees is given by Lemma 1, and then we can ensure $\widehat{\mathcal{A}}' \supseteq \widehat{\mathcal{A}}$ by simply replacing $\widehat{\mathcal{A}}'$ by $\widehat{\mathcal{A}}' \cup \widehat{\mathcal{A}}$ if necessary; the fact that this does not change V_A or V_P follows from the bounds on u and v.

Define the set of *feasible outcomes* $\mathcal{U}_{\mathcal{W}}$ as

$$\mathcal{U}_{\mathcal{W}} = \bigcup_{(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}} \mathcal{F}_{(\widehat{\mathcal{A}}, c_P)}.$$

Define the *critical region* $\mathcal{C}$ as

$$\mathcal{C} = \{(u, v) : u < 0, v > V_P(w|\mathcal{W})\}. \tag{14}$$



(a) Contract w is not EFA (b) Contract w is EFA

(c) Contract w' is EFA if contract w is

Figure 4. In all three figures, jointly the black lines represent the sets of feasible outcomes; and the grey regions represent the critical regions. The contract in Figure 4a is not EFA as input $(\widehat{\mathcal{A}}^1, c_P^1)$ is worrisome because $\mathcal{F}_{(\mathcal{A}^1, c_P^1)}$ intersects the critical region $\mathcal{C}$. Contract w in Figure 4b is EFA as long as w is locally eligible via $(\widehat{\mathcal{A}}^2, c_P^2)$. Figure 4c depicts contract w from Figure 4b replaced by a linear one as in the proof of Theorem 1. Observe that the new critical region, $\mathcal{C}'$, is smaller than the old one; and that all input-specific frontiers, $\mathcal{F}'_{(\mathcal{A}^1, c_P^1)}$, $\mathcal{F}'_{(\mathcal{A}^2, c_P^2)}$, and $\mathcal{F}'_{(\mathcal{A}^3, c_P^3)}$, moved downward and rightward.

This region consists of all the payoff pairs such that the principal would be willing to choose them if she knew she could, but such that the agent would then not be assured a nonnegative payoff. More simply put, payoff pairs in this region are the ones that could make an input worrisome.

We display two examples in Figures 4a and 4b. Here, $\mathcal{W} = \{(\widehat{\mathcal{A}}^1, c_P^1), (\widehat{\mathcal{A}}^2, c_P^2), (\widehat{\mathcal{A}}^3, c_P^3)\}$; $c_P^1 < c_P^2 < c_P^3$. The principal's optimal input is $(\widehat{\mathcal{A}}^2, c_P^2)$, which delineates the lower boundary to the critical region.

Our characterization is then the following.

Lemma 3. *The contract w is EFA if and only if (i) it is locally eligible via some optimal input and (ii) $\mathcal{U}_{\mathcal{W}} \cap \mathcal{C} = \emptyset$.*

Moreover, if w is linear, then (i) can be weakened to require only that w be locally eligible via some *input.*

For the two cases depicted in Figure 4a and Figure 4b, Lemma 3 then implies that the contract shown in Figure 4a is not EFA, as input $(\widehat{\mathcal{A}}^1, c_P^1)$ is worrisome, whereas the contract shown in Figure 4b is EFA provided it is locally eligible via input $(\widehat{\mathcal{A}}^2, c_P^2)$.

Once this is established, the proof that linear contracts are optimal follows the same lines as the proof of Theorem 1. We begin with an arbitrary EFA contract w, and we replace it by a new, linear contract w' as in that earlier proof. The new contract dominates the old one pointwise and also gives the principal a better guarantee if the choice of input is held fixed. From these properties, it quickly follows that the critical region $\mathcal{C}'$ for the new contract is smaller than the old one, while the input-specific frontiers $\mathcal{F}'_{(\widehat{\mathcal{A}}, c_P)}$ can only have moved downward and rightward; consequently, if the feasible outcomes and the critical region did not intersect previously, they still do not intersect, and the contract remains EFA by Lemma 3. (Note that the original input may no longer be optimal, but it does not need to be, by the second clause of Lemma 3.) This is shown in Figure 4c.

The full proof of Proposition 3 (including Lemma 3) is in Appendix B.

Also appearing in the Appendix is a characterization of when an EFA contract exists, and how the optimal one can be found; the discussion is mainly in Appendix C.2, though it draws on some calculations contained within the proof of Proposition 3. Very briefly, to identify the optimal such contract, for any candidate slope α, we use the definition of EFA to determine the lowest possible β; we then choose α to maximize the resulting guarantee for the principal.

One finding that emerges from this characterization is that the EFA formulation can escape the counterintuitive comparative statics we saw in the single-input model. In particular, at least for some parameterizations (Case 2 in Appendix C.2), the optimal slope is given by

$$\alpha = \frac{\sqrt{c}}{\sqrt{c} + \sqrt{c_P - \underline{c}_P}},$$

where c_P is the cost of an optimal input, c is the cost of the agent's optimal action within that input, and $\underline{c}_P$ is the cost of the cheapest input (see (37)). This expression is increasing in c and decreasing in c_P (the "sizes" of the two parties' moral hazard problems), as one might expect. This formula for α comes from maximizing an objective (given in (35)) that consists of total surplus, minus two terms representing agency frictions: one representing the amount the principal could lose if the agent deviates to a less-productive, lower-cost action, and the other representing the agent's losses from the principal's possible deviation to a cheaper input. When the optimal α is interior, it is determined by a tradeoff between these two loss terms, and changes in c or c_P can make one term or the other predominate.

A caveat is that changes in the environment may make a different input, or a different action within that input, become optimal; if so, this would lead to new values of c_P and c, which could either enhance or reverse the above comparative statics on α.

3.2.2 *Eligibility with further inputs* In our second variation, we return to the backward induction approach with the assumptions of no new actions and maxmin decision-making at step 1, but we now introduce uncertainty about the *principal's* available choices. In evaluating a contract w, the analyst knows that the principal can choose inputs from $\mathcal{W}$, but now there may be other inputs available as well: the true input space is $\widetilde{\mathcal{W}} \supseteq \mathcal{W}$. The principal can supply any input $(\widehat{\mathcal{A}}, c_P) \in \widetilde{\mathcal{W}}$. The principal only supplies inputs that maximize her guarantee, as in Section 3.1, and we again focus on contracts that guarantee a nonnegative payoff for the agent.

This idea leads to the following definition.

Definition 7. A contract w is *eligible with further inputs (EFI)* if it is locally eligible via some optimal input and, for all $\widetilde{\mathcal{W}} \supseteq \mathcal{W}$ and $(\widehat{\mathcal{A}}, c_P) \in \widetilde{\mathcal{W}}$,

$$\text{if} \quad V_P(w|\widehat{\mathcal{A}}, c_P) > V_P(w|\mathcal{W}), \quad \text{then} \quad V_A(w|\widehat{\mathcal{A}}) \geq 0.$$

We define the principal's *guarantee* from such a contract w as the corresponding value of $V_P(w|\mathcal{W})$.

It turns out that eligibility with further inputs is formally equivalent to a special case of eligibility with further actions. Write $\mathcal{A}_{\text{triv}}$ to denote the "trivial" input $\{(\delta_0, 0)\}$.

Proposition 4. *A contract w that is locally eligible via some optimal input is EFI if and only if it is EFA under input space $\mathcal{W}' = \mathcal{W} \cup \{(\mathcal{A}_{\text{triv}}, 0)\}$.*

Furthermore, the principal's guarantee is equal in these two environments.

Thus, by combining Proposition 4 and Proposition 3, it is without loss to optimize over the space of EFI contracts that are linear, and the analysis in Appendix C.2 can be used to identify an optimal contract. This leads to the following corollary, whose proof is omitted.

Corollary 2. *If an EFI contract exists, then among all EFI contracts there exists a linear contract that maximizes the principal's guarantee.*

Furthermore, the optimum among EFI linear contracts is given by the optimum among EFA linear contracts under input space $\mathcal{W}' = \mathcal{W} \cup \{(\mathcal{A}_{\text{triv}}, 0)\}$ instead of $\mathcal{W}$.

3.2.3 *Eligibility with full knowledge* In our last variation, we consider the case when the principal fully knows the agent's technology associated to each input, and she chooses the input based on this knowledge.

Similar to EFA, to develop a notion of eligibility, we define a notion of inputs that are chosen under some full knowledge resulting in a negative payoff for the agent.

Definition 8. Given contract w, input $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ is *FK-worrisome* if there exists a technology $\mathcal{A} \supseteq \widehat{\mathcal{A}}$ such that

$$\max_{(F,c) \in \mathcal{A}^*(w|\mathcal{A})} \left\{E_F\left[y - w(y)\right]\right\} - c_P > V_P(w|\mathcal{W}) \quad \text{and} \quad V_A(w|\mathcal{A}) < 0.$$

We then define our final alternative version of eligibility as follows.

Definition 9. A contract w is *eligible with full knowledge (EFK)* if it is locally eligible via some optimal input and no $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ is FK-worrisome.

We define the principal's *guarantee* from such a contract w as the corresponding value of $V_P(w|\mathcal{W})$.

How does an optimal contract look like in this environment? When contracts are restricted to be *monotone* (i.e., w weakly increasing in y), then linear contracts are optimal.

Proposition 5. *If a monotone EFK contract exists, then among all monotone EFK contracts there exists a linear contract that maximizes the principal's guarantee.*

When a monotone EFK contract exists and how the optimal monotone EFK contract can be found is described in Appendix C.3.

Why do we need the restriction to monotone contracts? Nonlinear contracts, when they are nonmonotone, can help provide commitment power to prevent the principal from choosing inputs that are bad for the agent, by making sure that any additional actions that could potentially make those inputs tempting for the principal will be so low-paying that the agent would never choose them. Example 4 in Appendix A illustrates this more concretely.

The proof of Proposition 5 proceeds similarly to that of Proposition 3, building on tools developed to prove Theorem 1.

In Section 3.2.1, the set of feasible outcomes consisted of the union of frontiers for each choice of input. With full knowledge of the technology, the outcome need not lie on the frontier; thus, the analogous set of feasible outcomes builds on $\mathcal{S}$, as defined in (4), as opposed to its frontier only.

Given a contract w, define for each choice of input $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$:

$$\mathcal{S}_{(\widehat{\mathcal{A}}, c_P)} = \left\{(u, v) : u \geq V_A(w|\widehat{\mathcal{A}}), v \geq V_P(w|\widehat{\mathcal{A}}, c_P), (u, v + c_P) \in \mathcal{S}\right\}.$$

Similar to before, we can consider the set of all payoff pairs realized as $(u, v) = (V_A(w|\mathcal{A}), \max_{(F,c)\in\mathcal{A}^*(w|\mathcal{A})}\{E_F[y - w(y)]\} - c_P)$, as the technology $\mathcal{A} \supseteq \widehat{\mathcal{A}}$ varies. The set $\mathcal{S}_{(\widehat{\mathcal{A}}, c_P)}$ is the closure of the set of all such payoff pairs (it may not exactly coincide with the set of such payoff pairs due to boundary issues from the tie-breaking assumption). The set of feasible outcomes is now the union of such sets: $\tilde{\mathcal{U}}_{\mathcal{W}} = \bigcup_{(\widehat{\mathcal{A}}, c_P)\in\mathcal{W}} \mathcal{S}_{(\widehat{\mathcal{A}}, c_P)}$; the critical region $\mathcal{C}$ is defined as in (14) and an analogue of Lemma 3 is given by the following.

Lemma 4. *The contract w is EFK if and only if it is locally eligible via some input and*

$$\tilde{\mathcal{U}}_{\mathcal{W}} \cap \mathcal{C} = \emptyset.$$

We omit the detailed proofs of Lemma 4 and Proposition 5; the arguments proceed almost identically to the proof of Proposition 3, which is given in detail in Appendix B. The restriction to monotonicity is needed to ensure that the sets $\mathcal{S}_{(\widehat{\mathcal{A}}, c_P)}$ move rightward and downward when the initial contract w is replaced with a linear contract w': without this restriction, the sets may become taller, so that the change of contracts creates an intersection with the critical region where none existed previously.

4. Conclusion

We have studied a contracting problem with moral hazard on both sides: the principal and the agent both need to exert effort for production to take place. Our interest is in developing insights into what forms of contract can perform well, in parallel with the literature on this question in one-sided moral hazard models; and more specifically, in seeing whether the idea that linear contracts are robust to uncertainty about the agent's possible actions can be expressed in such a setting. We have captured this focus on robustness by seeking contracts that maximize the worst-case payoff guarantee for the principal, subject to a guarantee of at least zero for the agent; and we have presented several versions of a model in which the maximum such guarantee is indeed attained by a linear contract.[6]

Defining the guarantees of a contract poses modeling challenges: how should the unknown actions of the agent be modeled, and what should we assume about the principal's behavior in the face of this uncertainty? Our approach has been to model the game as sequential, with the principal moving first. This allows us to model actions of the agent as in the one-sided moral hazard model of Carroll (2015), to give a simple intuition parallel to that model about how linear contracts can provide guarantees for the principal, and to give a simple definition for the principal's guarantee from any contract. However, this leaves us with challenges in modeling the principal's behavior, and thus delineating the set of contracts that provide a nonnegative guarantee for the agent. We offered a simple way to delineate this set (eligibility), based on backward induction, in our preliminary model with only a binary choice of input. For our more general model, we saw that a direct generalization based on the maxmin objective for the principal's behavior led to a notion (weak eligibility) under which linear contracts were not optimal. However, we proposed three alternative approaches that incorporate more uncertainty about the principal's behavior, leading to three different notions of eligibility—varying the objective and knowledge the principal is assumed to have at the input choice stage—and restoring optimality of linear contracts in each.

It may be useful to briefly compare our overall approach with that of Dai and Toikka (2022). They also consider a robust moral hazard problem with multiple parties taking costly actions. The parties play symmetric roles in their model; what we have described as double moral hazard would correspond to a team of two agents in their setting. Their agents play a simultaneous-move game, where each agent has some known actions but may also have unknown ones, and the distribution of output produced by each unknown action profile may be arbitrary. They seek to identify an output-sharing rule to maximize surplus in the worst case over the unknown actions the agents may have. They find, as we do, that a linear sharing rule is optimal. However, their model delivers a much starker version of this conclusion: a nonlinear sharing rule cannot deliver *any* positive surplus guarantee. A brief explanation is that each agent may have incentives to take "sabotage" actions that decrease expected surplus while shifting the output

[6]The arguments can also be generalized beyond the variations we have given here. For example, one can formulate the model without risk neutrality, and extend the arguments to show that optimal contracts are then linear in utility space; details are available from the authors.

distribution toward realizations where her own share is larger. As long as agents' incentives are not perfectly aligned, one can construct a game with a chain of such actions, in which iterated dominance reasoning leads agents to sabotage more and more until all surplus is destroyed. In contrast, in our framework, even when there is uncertainty about both parties' actions (as in our EFI formulation), the sequential-move structure prevents long chains of iterated dominance. Consequently, nonlinear contracts can still deliver some guarantees, although they turn out not to be optimal.

Although the various approaches we have developed to modeling uncertainty, and the corresponding eligibility concepts, are tailored specifically to our setting of double moral hazard, the modeling approach we have taken may provide future inspiration for other models of robust contracting that require interaction among multiple agents.

Appendix A: Additional examples

Here, we give the example referred to in the Introduction, showing that in a model without uncertainty (so that the production technology is known), linear contracts are typically not optimal without imposing specific functional form assumptions. We retain here the timing of the single-input model in Section 2, but one could easily give a similar example in a simultaneous-move setup as in Bhattacharyya and Lafontaine (1995).

Example 3. Suppose that output can range between 0 and 2, and the agent's choice of effort e (if the principal has supplied the input) is a number between 1 and 2. For each such e, let $F(y|e)$ be a distribution on $[1, 2]$ with mean equal to e. Assume the agent's cost of effort is given by an increasing, differentiable, convex function $c(e)$ with $c(1) = 0$ and $c'(1) < 1/2 < c'(2)$. Assume the cost of providing the input, c_P, is small but positive.

If the principal supplies the input, then the agent chooses e, and output y is determined as follows: with probability $1/2$, output is drawn from a uniform distribution on $[0, 1]$; with complementary probability $1/2$, output is drawn from $F(y|e)$ on $[1, 2]$. (If the principal does not supply the input, then output y is zero.)

Evidently, the first-best outcome is generated when the agent chooses e^{FB} given by $c'(e^{\mathrm{FB}}) = 1/2$. If output is shared according to a piecewise-linear contract of the form $w(y) = \beta$ for $y \leq 1$ and $w(y) = (y - 1) + \beta$ for $y > 1$, then the agent is a full residual claimant for his effort, so this contract induces the first-best effort e^{FB}. An appropriate choice of β then gives the full surplus to the principal, leaving the agent with a payoff of zero. Moreover, as long as c_P is small, the principal will indeed be willing to supply the input, since she is the residual claimant for output until it surpasses 1.

In contrast, a linear contract $w(y) = \alpha y + \beta$ cannot induce the first-best: by the first-order condition, the agent cannot be made to choose e^{FB} unless $\alpha = 1$, but a contract with a slope of 1 cannot motivate the principal to supply the input. $\diamondsuit$

Next, here is the example showing that nonmonotone contracts can outperform linear contracts in the multiple-input model under eligibility with full knowledge.

Example 4. Let $\mathcal{W} = \{(\widehat{\mathcal{A}}^1, c_P^1), (\widehat{\mathcal{A}}^2, c_P^2)\}$, with $\widehat{\mathcal{A}}^1 = \{(\delta_{24}, 8)\}$, $c_P^1 = 4$, $\widehat{\mathcal{A}}^2 = \{(\delta_0, 2)\}$ and $c_P^2 = 48$, and $\mathcal{Y} = [0, 150]$. As in Example 2, the optimal contract subject only to local eligibility via some input is given by $w[\alpha, \beta]$ with $(\alpha, \beta) = (2/3, -8)$, and this input $(\widehat{\mathcal{A}}^1, c_P^1)$ guarantees the principal a payoff of 8. However, $w[\alpha, \beta]$ is not EFK: if the agent's technology for input $(\widehat{\mathcal{A}}^2, c_P^2)$ is given by $\mathcal{A}^2 = \widehat{\mathcal{A}}^2 \cup \{(\delta_{150}, 100)\}$, the agent chooses the high action, giving a greater payoff to the principal than her guarantee from the first input while leaving himself still with a negative payoff. Consider now a nonlinear, and in particular nonmonotone, variation of contract $w[\alpha, \beta]$. Let contract w be given by

$$w(y) = \begin{cases} \alpha y + \beta & \text{for } y \leq 24 \\ \alpha \cdot 24 + \beta - (y - 24) & \text{for } y > 24. \end{cases}$$

With the second input provided, the agent only chooses actions for which his expected payoff is at least as high as the expected payoff he is guaranteed through action $(\delta_0, 2)$, $2/3 \cdot 0 - 8 - 2 = -10$. The (nonmonotone) construction of w implies that $w(y) \leq 32 - y$, so expected payoff at least -10 requires expected output of at most 42. But since the cost of providing this input is $c_P^2 = 48$, it cannot be that the agent receives an expected payoff of at least -10 and the principal of at least 8, whereas the first input does still guarantee at least 8 for the principal (by logic similar to that in Example 2). Hence, contract w is EFK.

Note that the linear contract $w[\alpha, \beta]$ is EFA: for example, if the principal is aware of action $(\delta_{150}, 100)$ available to the agent, she would still not provide the second input as the agent may also have action $(\delta_1, 0)$ available which he prefers. $\diamondsuit$

Appendix B: Proofs

We begin with notation that will be useful in several of the proofs.

Given a contract w, define

$$\underline{\mathcal{Y}} = \arg\min_{y \in \mathcal{Y}} \{y - w(y)\}; \qquad \bar{\mathcal{Y}} = \arg\max_{y \in \mathcal{Y}} w(y),$$

and let

$$y_0 = \max(\underline{\mathcal{Y}}), \qquad y_1 = \min(\bar{\mathcal{Y}}), \qquad y_2 = \max(\bar{\mathcal{Y}}).$$

In Figure 2, the rightmost point on the horizontal segment is $(w(y_0), y_0 - w(y_0))$, and the lowest and highest points on the vertical segment likewise correspond to y_1 and y_2. (In general, y_0, y_1, y_2 may not be all distinct.)

Proof of Lemma 1. Take any $\widehat{\mathcal{A}} \in \mathcal{T}$. (For brevity, we will write $\widehat{\mathcal{A}}$ throughout this proof rather than $\widehat{\mathcal{A}}'$ as in the definition of $\mathcal{R}$; note the $\widehat{\mathcal{A}}$ from the main model is never needed for this lemma.)

We make the following three claims:

(i) $V_P(w|\widehat{\mathcal{A}}, 0)$ is bounded below as

$$\begin{aligned} V_P(w|\widehat{\mathcal{A}}, 0) \geq \min E_F\big[y - w(y)\big] & \\ \text{over } F \in \Delta(\mathcal{Y}) \text{ such that } E_F\big[w(y)\big] \geq V_A(w|\widehat{\mathcal{A}}); & \end{aligned} \tag{15}$$

(ii) if $V_A(w|\widehat{\mathcal{A}}) < w(y_1)$, then (15) holds with equality; and

(iii) if $V_P(w|\widehat{\mathcal{A}}, 0) > y_0 - w(y_0)$, then whenever F attains the minimum in (15),

$$E_F\big[w(y)\big] = V_A(w|\widehat{\mathcal{A}}).$$

To show (i), note that for any $F \in \Delta(\mathcal{Y})$, action (F, c) is only chosen by the agent if his expected payoff is at least the expected payoff from choosing an optimal action in $\widehat{\mathcal{A}}$, that is, only if

$$E_F\big[w(y)\big] - c \geq V_A(w|\widehat{\mathcal{A}}).$$

As $c \geq 0$, it follows that $V_P(w|\widehat{\mathcal{A}}, 0)$ cannot be smaller than the minimum of $E_F[y - w(y)]$ over $F \in \Delta(\mathcal{Y})$ such that $E_F[w(y)] \geq V_A(w|\widehat{\mathcal{A}})$.

Now suppose that $V_A(w|\widehat{\mathcal{A}}) < w(y_1)$. Suppose that F achieves the minimum in (15). We need to show that $V_P(w|\widehat{\mathcal{A}}, 0)$ cannot be strictly greater than $E_F[y - w(y)]$. If $\text{supp}(F) \nsubseteq \bar{\mathcal{Y}}$, then let $\mathcal{A}$ be given by $\mathcal{A} = \widehat{\mathcal{A}} \cup \{(F', 0)\}$ where $F' = \epsilon\delta_{y_1} + (1 - \epsilon)F$. For any $\epsilon > 0$, the agent chooses action $(F', 0)$ and as $\epsilon \to 0$, $E_{F'}[y - w(y)] \to E_F[y - w(y)]$ implying that $V_P(w|\widehat{\mathcal{A}}, 0) \leq E_F[y - w(y)]$. Suppose now that $\text{supp}(F) \subseteq \bar{\mathcal{Y}}$. Then by assumption, $V_A(w|\widehat{\mathcal{A}}) < w(y_1) = E_F[w(y)]$. Thus, for $\mathcal{A}$ given by $\mathcal{A} = \widehat{\mathcal{A}} \cup \{(F, 0)\}$, the agent chooses $(F, 0)$ and the principal's expected payoff is $E_F[y - w(y)]$, again, bounding $V_P(w|\widehat{\mathcal{A}}, 0)$ by $E_F[y - w(y)]$ from above.

Now suppose that $V_P(w|\widehat{\mathcal{A}}, 0) > y_0 - w(y_0)$. Let F attain the minimum in (15). Suppose that $E_F[w(y)] > V_A(w|\widehat{\mathcal{A}})$. If $\text{supp}(F) \nsubseteq \underline{\mathcal{Y}}$, then let $\mathcal{A}$ be given by $\mathcal{A} = \widehat{\mathcal{A}} \cup \{(F', 0)\}$ where $F' = \epsilon\delta_{y_0} + (1 - \epsilon)F$. For ϵ small enough, the agent chooses $(F', 0)$ contradicting minimality. Suppose now that $\text{supp}(F) \subseteq \underline{\mathcal{Y}}$. Then the agent chooses action $(F, 0)$ if $\mathcal{A}$ is given by $\mathcal{A} = \widehat{\mathcal{A}} \cup \{(F, 0)\}$, bounding $V_P(w|\widehat{\mathcal{A}}, 0)$ above by $y_0 - w(y_0)$, a contradiction. Thus (i)–(iii) are shown.

Now we can show that every point in $\mathcal{R}$ is in $\mathcal{F}$. Each point in $\mathcal{R}$ is of the form $(u, v) = (V_A(w|\widehat{\mathcal{A}}), V_P(w|\widehat{\mathcal{A}}, 0))$ for some technology $\widehat{\mathcal{A}}$. If the conclusions in (ii) and (iii) hold, then for any F attaining the minimum in (15), $V_A(w|\widehat{\mathcal{A}}) = E_F[w(y)]$ and $V_P(w|\widehat{\mathcal{A}}, 0) = E_F[y - w(y)]$. Hence, $(u, v) \in \mathcal{F}$. If the conclusion in (ii) holds but $V_P(w|\widehat{\mathcal{A}}, 0) = y_0 - w(y_0)$, then $V_A(w|\widehat{\mathcal{A}}) \leq w(y_0)$ and again $(u, v) \in \mathcal{F}$. If the conclusion in (iii) holds but $V_A(w|\widehat{\mathcal{A}}) = w(y_1)$, then $V_P(w|\widehat{\mathcal{A}}, 0)$ is bounded below by $y_1 - w(y_1)$ and above by $y_2 - w(y_2)$, as the agent does not take any action (F, c) where $\text{supp}(F) \nsubseteq \bar{\mathcal{Y}}$, and again $(u, v) \in \mathcal{F}$. Lastly, if $V_A(w|\widehat{\mathcal{A}}) = w(y_1)$ and $V_P(w|\widehat{\mathcal{A}}, 0) = y_0 - w(y_0)$, then there exists action $(F, 0) \in \widehat{\mathcal{A}}$ such that $\text{supp}(F) \subseteq \bar{\mathcal{Y}}$, but for all such actions $\text{supp}(F) \subseteq \underline{\mathcal{Y}}$ as otherwise the agent would choose the action preferred by the principal. For such F, $V_A(w|\widehat{\mathcal{A}}) = E_F[w(y)]$ and $V_P(w|\widehat{\mathcal{A}}, 0) = E_F[y - w(y)]$. Hence, $(u, v) \in \mathcal{F}$. Thus, $\mathcal{R} \subseteq \mathcal{F}$.

Now we show $\mathcal{F} \subseteq \mathcal{R}$. Take any $(u, v) \in \mathcal{F}$. If $v = y_0 - w(y_0)$, then $u \leq w(y_0)$ so that $c := w(y_0) - u \geq 0$. Let $\widehat{\mathcal{A}} = \{(\delta_{y_0}, c)\}$. Clearly, $V_A(w|\widehat{\mathcal{A}}) = w(y_0) - c = u$. Furthermore, $V_P(w|\widehat{\mathcal{A}}, 0)$ is bounded above by $y_0 - w(y_0)$, e.g., for $\mathcal{A} = \widehat{\mathcal{A}}$, and bounded below by $y_0 - w(y_0)$ by definition of y_0. Hence, $V_P(w|\widehat{\mathcal{A}}, 0) = y_0 - w(y_0) = v$ and $(u, v) \in \mathcal{R}$.

If $u = w(y_1)$, then $y_2 - w(y_2) \geq v \geq y_1 - w(y_1)$ so that there exists $x \in [0, 1]$ for which $F := x\delta_{y_2} + (1 - x)\delta_{y_1}$ satisfies $E_F[y - w(y)] = v$. Let $\widehat{\mathcal{A}} = \{(F, 0)\}$. Clearly, $V_A(w|\widehat{\mathcal{A}}) = w(y_1) = u$. Furthermore, $V_P(w|\widehat{\mathcal{A}}, 0)$ is bounded above by $E_F[y - w(y)]$, e.g., for $\mathcal{A} = \widehat{\mathcal{A}}$, and bounded below by $E_F[y - w(y)]$ as the agent will always choose the action preferred by the principal if he is indifferent. Hence, $V_P(w|\widehat{\mathcal{A}}, 0) = E_F[y - w(y)] = v$ and $(u, v) \in \mathcal{R}$.

This leaves the case $v > y_0 - w(y_0)$ and $u < w(y_1)$. Pick

$$F^* \in \arg\min E_F\big[y - w(y)\big] \quad \text{over } F \in \Delta(\mathcal{Y}) \text{ such that } E\big[w(y)\big] = u. \tag{16}$$

(Note that this can only be done if we know that $u \geq \min_y w(y)$, but this is indeed the case: in fact $u \geq w(y_0)$ since otherwise the existence of $(u', v') = (w(y_0), y_0 - w(y_0))$ would contradict $(u, v) \in \mathcal{F}$.)

Let $\widehat{\mathcal{A}} = \{(F^*, 0)\}$. Clearly, $V_A(w|\widehat{\mathcal{A}}) = E_{F^*}[w(y)] = u$. Let us characterize $V_P(w|\widehat{\mathcal{A}}, 0)$.

Note that F^* still attains the minimum in (16) when the constraint $E[w(y)] = u$ is replaced by $E[w(y)] \geq u$: if it did not, then (u, v) would not belong to $\mathcal{F}$. Thus, F^* attains the minimum in (15).

Lastly, if (15) does not hold with equality, then by (ii), it must be that $u = V_A(w|\widehat{\mathcal{A}}) \geq w(y_1)$ contrary to our initial assumption. Thus, (15) holds with equality implying that $V_P(w|\widehat{\mathcal{A}}, 0) = E_{F^*}[y - w(y)] = v$ and $(u, v) \in \mathcal{R}$. □

Proof of Theorem 1. Consider any eligible contract w. Note that $V_P(w|\widehat{\mathcal{A}}, c_P) = V_P(w|\widehat{\mathcal{A}}, 0) - c_P$. Let $\widetilde{\mathcal{S}}$ consist of points (u, v) such that $u > V_A(w|\widehat{\mathcal{A}})$ and $v < V_P(w|\widehat{\mathcal{A}}, 0)$. Lemma 1 tells us that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are disjoint.

By the separating hyperplane theorem, there exist constants κ, λ, and μ with $(\lambda, \mu) \neq (0, 0)$ such that

$$\kappa + \lambda u - \mu v \leq 0 \quad \text{for all } (u, v) \in \mathcal{S}, \tag{17}$$

$$\kappa + \lambda u - \mu v \geq 0 \quad \text{for all } (u, v) \in \widetilde{\mathcal{S}}. \tag{18}$$

Inequality (18) implies that λ, μ are nonnegative as otherwise the inequality is not satisfied for large u or small v. Rearranging (17) implies that

$$\mu\big(y - w(y)\big) \geq \kappa + \lambda w(y) \quad \text{for all } y \in \mathcal{Y}. \tag{19}$$

The point $(V_A(w|\widehat{\mathcal{A}}), V_P(w|\widehat{\mathcal{A}}, 0))$ is in the closures of both $\mathcal{S}$ and $\widetilde{\mathcal{S}}$, implying that

$$\mu V_P(w|\widehat{\mathcal{A}}, 0) = \kappa + \lambda V_A(w|\widehat{\mathcal{A}}). \tag{20}$$

Define a linear contract w' by

$$w'(y) = \frac{\mu}{\mu + \lambda} y - \frac{\kappa}{\mu + \lambda}. \tag{21}$$

Contract w' satisfies (19) as an equality. For any technology $\mathcal{A} \supseteq \widehat{\mathcal{A}}$, let (F, c) be an action that the agent takes under contract w'. Taking expectation over y distributed according to F, (19) for w' implies

$$\mu E_F\big[y - w'(y)\big] \geq \kappa + \lambda E_F\big[w'(y)\big].$$

Inequality (19) implies $w' \geq w$ pointwise. Hence, the agent's expected payoff if his technology is just $\widehat{\mathcal{A}}$ is at least as large under w' as under w. As $c \geq 0$, we have

$$\begin{aligned}
\mu E_F\big[y - w'(y)\big] &\geq \kappa + \lambda E_F\big[w'(y)\big] \\
&\geq \kappa + \lambda\big(E_F\big[w'(y)\big] - c\big) \\
&\geq \kappa + \lambda V_A\big(w'|\widehat{\mathcal{A}}\big) \\
&\geq \kappa + \lambda V_A(w|\widehat{\mathcal{A}}).
\end{aligned}$$

Combining the above with (20) gives

$$\mu E_F\big[y - w'(y)\big] \geq \mu V_P(w|\widehat{\mathcal{A}}, 0)$$

for any (F, c) the agent might choose given w'.

If $\mu > 0$, this implies

$$V_P\big(w'|\widehat{\mathcal{A}}, 0\big) \geq V_P(w|\widehat{\mathcal{A}}, 0).$$

If $\mu = 0$, then it must have been the case that $V_A(w|\widehat{\mathcal{A}}) = w(y_1)$. In this case, w' is constant with value $-\kappa/\lambda = w(y_1)$. Given $\widehat{\mathcal{A}}$, the agent chooses an action $(F, 0)$ with F having support on $\bar{\mathcal{Y}}$. Among all such actions, he chooses the one that gives the principal the highest expected payoff. Thus, increasing all wages to $w(y_1)$, as w' does, only increases the principal's guarantee. Thus, in either case

$$V_P\big(w'|\widehat{\mathcal{A}}, 0\big) \geq V_P(w|\widehat{\mathcal{A}}, 0). \tag{22}$$

Furthermore, w' is eligible because

$$\begin{aligned}
V_P(w|\widehat{\mathcal{A}}, c_P) > 0 \quad &\Longrightarrow \quad V_P\big(w'|\widehat{\mathcal{A}}, c_P\big) > 0 \quad \text{by (22);} \\
V_P(w|\widehat{\mathcal{A}}, c_P) \geq -w(0) \quad &\Longrightarrow \quad V_P\big(w'|\widehat{\mathcal{A}}, c_P\big) \geq -w'(0) \quad \text{by (22) and as } w'(0) \geq w(0); \\
V_A(w|\widehat{\mathcal{A}}) \geq 0 \quad &\Longrightarrow \quad V_A\big(w'|\widehat{\mathcal{A}}\big) \geq 0 \quad \text{because } w' \geq w \text{ pointwise.}
\end{aligned}$$

Thus, we have an eligible linear contract w' delivering at least as high a guarantee for the principal as w.

Finally, it remains to show that an optimal eligible contract exists. The arguments in Section 2.2.3 show that the optimum among eligible linear contracts exists. This contract must then be optimal among *all* eligible contracts, since the preceding arguments imply that no nonlinear eligible contract can be better than all linear eligible contracts. □

Proof of Corollary 1. Assume that w is an optimal eligible contract but is not linear. We repeat the initial steps of the proof of Theorem 1, defining w' as in (21). Suppose first that $\mu > 0$ and $\lambda > 0$. Again, w' satisfies (19) as an equality. Thus,

$$\mu V_P\big(w'|\widehat{\mathcal{A}}, 0\big) \geq \kappa + \lambda V_A\big(w'|\widehat{\mathcal{A}}\big) = \mu V_P(w|\widehat{\mathcal{A}}, 0) + \lambda\big(V_A\big(w'|\widehat{\mathcal{A}}\big) - V_A(w|\widehat{\mathcal{A}})\big). \tag{23}$$

We have $w' \geq w$ pointwise, with strict inequality for some output levels. Thus, $V_A(w'|\widehat{\mathcal{A}}) > V_A(w|\widehat{\mathcal{A}})$, because $\widehat{\mathcal{A}}$ satisfies the full-support condition. As μ, λ in (23) are strictly positive, it follows that

$$V_P\big(w'|\widehat{\mathcal{A}}, 0\big) > V_P(w|\widehat{\mathcal{A}}, 0) \quad \text{and so} \quad V_P\big(w'|\widehat{\mathcal{A}}, c_P\big) > V_P(w|\widehat{\mathcal{A}}, c_P),$$

contradicting the assumption that w is optimal.

Next, if $\lambda = 0$, then $V_P(w|\widehat{\mathcal{A}}, 0) = y_0 - w(y_0)$. This case can only arise if $c_P = 0$, since otherwise the principal would not have the incentive to supply the input. In this case, note that the principal can extract the full (known) surplus via a linear contract of slope 1, so $V_P(w|\widehat{\mathcal{A}}, 0)$ must equal this full surplus in order for w to be optimal. But then, since $V_A(w'|\widehat{\mathcal{A}}) > V_A(w|\widehat{\mathcal{A}}) \geq 0$ and $V_P(w'|\widehat{\mathcal{A}}, 0) \geq V_P(w|\widehat{\mathcal{A}}, 0)$, it follows that the sum of the two parties' guarantees under w', $V_A(w'|\widehat{\mathcal{A}}) + V_P(w'|\widehat{\mathcal{A}}, 0)$, exceeds the full surplus, which is impossible.

Finally, if $\mu = 0$ and $\lambda > 0$, then $V_A(w|\widehat{\mathcal{A}}) = w(y_1)$. By the full support assumption, it must be that $\bar{\mathcal{Y}} = \mathcal{Y}$ implying that contract $w(y)$ is constant, again a contradiction. □

Proof of Lemma 2. (A) We know α cannot be strictly greater than 1 because (E2) would be violated; take $\mathcal{A} = \widehat{\mathcal{A}} \cup \{(\delta_{\bar{y}}, 0)\}$ for $\bar{y} = \max(\mathcal{Y})$. Also, α cannot be strictly less than 0 because the agent would take action $(\delta_0, 0)$ if available so that (E2) would imply that $c_P = 0$, $V_P(w[\alpha, \beta]) = -\beta$, and $V_A(w[\alpha, \beta]|\widehat{\mathcal{A}}) \leq \beta$, which implies that not both (E1) and (E3) can be satisfied.

(B) If $\alpha = 0$, there are two cases to consider. If there are no actions of the form $(F, 0)$ in $\widehat{\mathcal{A}}$, then for $\mathcal{A} = \widehat{\mathcal{A}} \cup \{(\delta_0, 0)\}$ the agent chooses action $(\delta_0, 0)$, thus failing to generate a positive total surplus; this contradicts eligibility. If there are actions of the form $(F, 0)$ in $\widehat{\mathcal{A}}$, the principal's guarantee is given by $V_P(w[0, \beta]) = \max_{(F,0)\in\widehat{\mathcal{A}}} E_F[y] - \beta - c_P$.

(C) If $\alpha = 1$, (E2) implies that $c_P = 0$ for such contract to be eligible; if it is costly to supply the input and the principal does not receive any share of the output, she will abstain from supplying the input. The principal's guarantee is thus given by $V_P(w[1, \beta]) = -\beta$. □

Proof of Proposition 2. The arguments preceding the proposition statement show that, if any eligible linear contract exists, then $w[\alpha^*, \beta(\alpha^*)]$ is an optimal eligible linear contract. If for some $(F^*, c^*) \in \arg\max_{(F,c)\in\widehat{\mathcal{A}}}\{E_F[y] - c/\alpha^*\}$, we have $c^* = 0$, then the zero contract is an optimal contract and, in fact, $V_P(w[\alpha, \beta(\alpha)]) = V_P(w[\alpha^*, \beta(\alpha^*)])$ for all $\alpha \in [0, \alpha^*]$ by (11); and further $w[\alpha, \beta(\alpha)]$ is eligible. If for all such (F^*, c^*), we have $c^* > 0$, then for all $\alpha < \alpha^*$, $V_P(w[\alpha, \beta(\alpha)]) < V_P(w[\alpha^*, \beta(\alpha^*)])$ and uniqueness follows.

Next, we show (13). For $\alpha = 1$,

$$\frac{1-\alpha}{\alpha} \max_{(F,c)\in\widehat{\mathcal{A}}} \big\{E_F[\alpha y] - c\big\} - c_P = -c_P \leq 0$$

implying the first part of (13); this in turn implies

$$V_P\big(w\big[\alpha^*,\beta\big(\alpha^*\big)\big]\big)=\frac{1-\alpha^*}{\alpha^*}\max_{(F,c)\in\widehat{\mathcal{A}}}\big\{E_F\big[\alpha^* y\big]-c\big\}-\beta\big(\alpha^*\big)-c_P=-\beta\big(\alpha^*\big).$$

□

Proof of Proposition 1. Suppose there exists $(F, c)\in\widehat{\mathcal{A}}$ satisfying condition (3).

We draw on the alternative characterization of the optimal linear contract in Appendix C.1. Let r be given by (30) and let

$$\alpha=r(F,c)\quad\text{and}\quad\beta=-\big\{E_F[\alpha y]-c\big\}.$$

Note that $\alpha\in(0,1]$. Contract $w[\alpha,\beta]$ is eligible if and only if (7)–(9) hold. Conditions (8) and (9) are satisfied because

$$\frac{1-\alpha}{\alpha}\max_{(F',c')\in\widehat{\mathcal{A}}}\big\{E_{F'}[\alpha y]-c'\big\}-c_P\geq\frac{1-\alpha}{\alpha}\big\{E_F[\alpha y]-c\big\}-c_P=0 \tag{24}$$

and

$$\max_{(F',c')\in\widehat{\mathcal{A}}}\big\{E_{F'}[\alpha y]+\beta-c'\big\}\geq E_F[\alpha y]+\beta-c=0.$$

Condition (7) is satisfied if $\beta<0$. If, instead, $\beta\geq 0$, then $c_P=0$ by (24). But $c_P=0$ implies that selling the firm guarantees an expected payoff of $\max_{(F',c')\in\widehat{\mathcal{A}}}\{E_{F'}[y]-c'\}-c_P\geq E_F[y]-c-c_P>0$ to the principal and is thus eligible. Thus, in either case, an eligible contract exists.

Now for the converse, suppose an eligible contract exists. In particular, one such contract is the optimal eligible linear contract given by Proposition 2, namely $w[\alpha^*,\beta(\alpha^*)]$ defined by (12).

Let

$$(F,c)\in\arg\max_{(F',c')\in\widehat{\mathcal{A}}}\big\{E_{F'}\big[\alpha^* y\big]-c'\big\}.$$

It is immediate that $E_F[y]-c-c_P>0$, as $E_F[y]-c-c_P\geq(1/\alpha^*)(E_F[\alpha^* y]-c)-c_P$ is an upper bound on the guaranteed output minus the input cost and, therefore, on the total guaranteed surplus, which is strictly positive for an eligible contract.

By definition of α^* and (F, c), and using (13),

$$\frac{1-\alpha^*}{\alpha^*}\big\{E_F\big[\alpha^* y\big]-c\big\}-c_P=0$$

implying further that $E_F[y]-c-c_P\geq 2\sqrt{cc_P}$ for α^* to be real. □

Proof of Lemma 3. For the first statement, suppose that w satisfies the stated conditions but is not EFA. Then there exists some FA-worrisome input $(\widehat{\mathcal{A}}, c_P)\in\mathcal{W}$ and $\mathcal{A}\supseteq\widehat{\mathcal{A}}$ such that

$$V_P(w|\mathcal{A},c_P)>V_P(w|\mathcal{W})\quad\text{and}\quad V_A(w|\mathcal{A})<0. \tag{25}$$

As $\mathcal{A}\supseteq\widehat{\mathcal{A}}$, $(V_A(w|\mathcal{A}),V_P(w|\mathcal{A},c_P))\in\mathcal{F}_{(\widehat{\mathcal{A}},c_P)}$. Furthermore, (25) implies that $(V_A(w|\mathcal{A}),V_P(w|\mathcal{A},c_P))\in\mathcal{C}$, contradicting $\mathcal{U}_{\mathcal{W}}\cap\mathcal{C}=\emptyset$.

For the converse, suppose that w is EFA. Condition (i) is immediate, so we show (ii). Suppose that $\mathcal{U}_{\mathcal{W}} \cap \mathcal{C} \neq \emptyset$ and consider $(u, v) \in \mathcal{U}_{\mathcal{W}} \cap \mathcal{C}$. Then $(u, v) \in \mathcal{F}_{(\widehat{\mathcal{A}}, c_P)}$ for some $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$. Using Lemma 1, there exists $\mathcal{A}'$ such that $(V_A(w|\mathcal{A}'), V_P(w|\mathcal{A}', c_P)) = (u, v)$. Let $\mathcal{A} = \widehat{\mathcal{A}} \cup \mathcal{A}'$ and note that $V_A(w|\mathcal{A}) = \max\{V_A(w|\widehat{\mathcal{A}}), V_A(w|\mathcal{A}')\} = u < 0$, while $V_P(w|\mathcal{A}, c_P) \geq V_P(w|\mathcal{A}', c_P) = v > V_P(w|\mathcal{W})$. Thus, $(\widehat{\mathcal{A}}, c_P)$ is FA-worrisome given contract w.

Finally, for the last statement, suppose that $w = w[\alpha, \beta]$ is linear, is locally eligible via some nonoptimal input $(\widehat{\mathcal{A}}, c_P)$, and $\mathcal{U}_{\mathcal{W}} \cap \mathcal{C} = \emptyset$. We need to show that w is locally eligible via an optimal input as well. Clearly, (E1) and (E2) still hold, so we only need to show that (E3) holds for an optimal input. Given Lemma 2 part (A), we know $\alpha \in [0, 1]$, so consider two cases:

- If $\alpha \in [0, 1)$, let $(\widehat{\mathcal{A}}^*, c_P^*)$ be an optimal input, and assume for contradiction that $V_A(w|\widehat{\mathcal{A}}^*) < 0$. First, suppose $V_A(w|\widehat{\mathcal{A}}^*) \geq w(0)$. Let F be a worst-case distribution for the principal under this input, so that $V_A(w|\widehat{\mathcal{A}}^*) = E_F[w(y)]$ and $V_P(w|\widehat{\mathcal{A}}^*, c_P^*) = E_F[y - w(y)] - c_P^*$. Take ϵ small, let $F' = (1 - \epsilon)F + \epsilon\delta_{\bar{y}}$ where $\bar{y} = \max(\mathcal{Y})$, and let $\mathcal{A} = \widehat{\mathcal{A}}^* \cup \{(F', 0)\}$. Note that $F \neq \delta_{\bar{y}}$ since otherwise the agent could not get a positive payoff under w at all. Then, under $(\mathcal{A}, c_P^*)$, the principal's guarantee is strictly higher than $V_P(w|\widehat{\mathcal{A}}^*, c_P^*)$ (since the principal receives a share $1 - \alpha > 0$ of improved output relative to F) and the agent's payoff is still below 0; this gives a point lying in $\mathcal{F}_{(\widehat{\mathcal{A}}^*, c_P^*)} \cap \mathcal{C}$, contradicting the assumption that this intersection was empty.

  This leaves the possibility $V_A(w|\widehat{\mathcal{A}}^*) < w(0)$. In this case, the agent would produce δ_0 if it comes at cost 0, so the principal's guarantee is $V_P(w|\widehat{\mathcal{A}}^*, c_P^*) = -\beta - c_P^* \leq -\beta$. But we know that the contract is locally eligible via $(\widehat{\mathcal{A}}, c_P)$, so the guarantee from this input is $\geq -\beta$, so this input is also optimal and we are done.

- If $\alpha = 1$, then as we saw in Section 2.2.3, w can only be locally eligible via $(\widehat{\mathcal{A}}, c_P)$ if $c_P = 0$, and then any input with $c_P = 0$ is optimal, since the principal's payoff is always $-\beta$. □

The proof of Proposition 3 also uses the following lemma for the multiple-input setting.

Lemma 5. *Let $(\underline{\widehat{\mathcal{A}}}, \underline{c}_P) \in \arg\min_{(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}} c_P$, and let w be any contract. Then w is EFA if and only if it is locally eligible via some optimal input and $(\underline{\widehat{\mathcal{A}}}, \underline{c}_P)$ is not a FA-worrisome input.*

(Thus, we can test whether a contract is EFA without concerning ourselves about whether any input other than $(\underline{\widehat{\mathcal{A}}}, \underline{c}_P)$ is FA-worrisome.)

Proof. Suppose that $(\underline{\widehat{\mathcal{A}}}, \underline{c}_P)$ is not a FA-worrisome input but $(\widehat{\mathcal{A}}, c_P)$ is. Then there exists $\widehat{\mathcal{A}}' \supseteq \widehat{\mathcal{A}}$ such that $V_A(w|\widehat{\mathcal{A}}') < 0$ and $V_P(w|\widehat{\mathcal{A}}', c_P) > V_P(w|\mathcal{W})$. But then let $\mathcal{A} = \widehat{\mathcal{A}}' \cup \underline{\widehat{\mathcal{A}}}$ and note that $V_P(w|\mathcal{A}, \underline{c}_P) = V_P(w|\widehat{\mathcal{A}}', \underline{c}_P) \geq V_P(w|\widehat{\mathcal{A}}', c_P) > V_P(w|\mathcal{W})$, where the first equality follows as otherwise $V_P(w|\underline{\widehat{\mathcal{A}}}, \underline{c}_P) > V_P(w|\mathcal{W})$, a contradiction. Thus, the two points on the frontier $\mathcal{F}_{(\underline{\widehat{\mathcal{A}}}, \underline{c}_P)}$ corresponding to technologies $\mathcal{A}$ and $\widehat{\mathcal{A}}'$ have the same value of V_P. If they have the same value of V_A as well, we have $V_A(w|\mathcal{A}) = V_A(w|\widehat{\mathcal{A}}') < 0$ and $(\widehat{\underline{\mathcal{A}}}, \underline{c}_P)$ is a FA-worrisome input, a contradiction. Otherwise, both points must lie on a flat segment of $\mathcal{F}_{(\widehat{\underline{\mathcal{A}}}, \underline{c}_P)}$, and then the minimum of V_P along the frontier is also attained on this segment, implying $V_P(w|\widehat{\underline{\mathcal{A}}}, \underline{c}_P) = V_P(w|\mathcal{A}, \underline{c}_P) > V_P(w|\mathcal{W})$, a contradiction. $\square$

Proof of Proposition 3. We begin by showing the following claim: For any contract w that is EFA, there is a linear contract w' that is EFA and guarantees the principal a weakly greater payoff than w does.

To show this, let $(\widehat{\mathcal{A}}^*, c_P^*)$ be an optimal input given w. Define κ, λ, μ, y_0, y_1, y_2, and w' as in the proof of Theorem 1. As in that earlier proof, w' is locally eligible (via the same input), and it gives the principal at least as high a guarantee as w. It remains to show that w' is EFA. Note that the critical region for w' is contained in that of w.

Similar to before, define

$$\mathcal{S}' = \operatorname{conv}\bigl(\bigl\{\bigl(w'(y) - c, y - w'(y)\bigr) : y \in \mathcal{Y}, c \in \mathbb{R}^+\bigr\}\bigr).$$

The fundamental relationship between the principal's and the agent's guarantee given w' is now given by

$$\mathcal{F}' = \bigl\{(u, v) \in \mathcal{S}' : \nexists\bigl(u', v'\bigr) \in \mathcal{S}', u' > u, v' < v\bigr\}.$$

For any input $(\widehat{\mathcal{A}}, c_P)$, let $\mathcal{F}'_{(\widehat{\mathcal{A}}, c_P)}$ be defined as

$$\mathcal{F}'_{(\widehat{\mathcal{A}}, c_P)} = \bigl\{(u, v) : u \geq V_A\bigl(w'|\widehat{\mathcal{A}}\bigr), v \geq V_P\bigl(w'|\widehat{\mathcal{A}}, c_P\bigr), (u, v + c_P) \in \mathcal{F}'\bigr\}.$$

Take any $(u', v') \in \mathcal{F}'$ with $u' < 0$. The frontier $\mathcal{F}$ contains some point (u'', v'') with $u'' \leq u'$ (e.g., it contains all such points with $v'' = y_0 - w(y_0)$ and u'' sufficiently low). But as w is locally eligible, $\max_y w(y) \geq 0$, so $\mathcal{F}$ also contains some point whose first coordinate is positive. Hence, there exists some intermediate $(u, v) \in \mathcal{F}$ with $u = u'$.

Let $F, F' \in \Delta(\mathcal{Y})$ and $c, c' \geq 0$ satisfy

$$\begin{aligned} (u, v) &= \bigl(E_F\bigl[w(y)\bigr] - c, E_F\bigl[y - w(y)\bigr]\bigr) \quad \text{and} \\ \bigl(u', v'\bigr) &= \bigl(E_{F'}\bigl[w'(y)\bigr] - c', E_{F'}\bigl[y - w'(y)\bigr]\bigr). \end{aligned}$$

If $c' > 0$, then $\operatorname{supp}(F') \subseteq \arg\min\{y - w'(y)\}$ and since $w' \geq w$ it follows that $v \geq v'$. Suppose now that $c' = 0$. We must have $\mu > 0$, since otherwise w' is constant at $-\kappa/\lambda = w(y_1) > 0$, contradicting our earlier statements $E_{F'}[w'(y)] = u' < 0$. Since point (u, v) is in $\mathcal{S}$, it satisfies the inequality (17) from the proof of Theorem 1. Combining this with (21) gives

$$v \geq \frac{\kappa + \lambda u}{\mu} = \frac{\kappa + \lambda u'}{\mu} = v'$$

in this case, too.

Now consider any $(u', v') \in \mathcal{F}'_{(\widehat{\mathcal{A}}, c_P)}$ with $u' < 0$. By translating the argument above, there exists $(u, v) \in \mathcal{F}_{(\widehat{\mathcal{A}}, c_P)}$ with $u = u'$ and $v \geq v'$. (Note that the requirement $u \geq V_A(w|\widehat{\mathcal{A}})$ holds by $u = u' \geq V_A(w'|\widehat{\mathcal{A}}) \geq V_A(w|\widehat{\mathcal{A}})$, since $w' \geq w$ pointwise.)

It follows that the new critical region and the feasible set given w' still do not intersect: if they did intersect at a point (u', v'), then the above argument would give us a point (u, v) where the critical region and the feasible set for the original contract w intersected, contradicting EFA for w. Now Lemma 3 assures EFA for w'. (Note that because w' is linear, we need not know whether $(\widehat{\mathcal{A}}^*, c_P^*)$ remains optimal under w'.)

This completes the proof of the claim.

Thus, the principal's maximum guarantee among EFA contracts (assuming it exists) is attained by a linear contract. To complete the proof of the proposition, it suffices to show that the maximum over linear contracts $w[\alpha, \beta]$ is attained. To this end, we consider each possible slope α that can arise in some EFA contract and identify the optimal β for which $w[\alpha, \beta]$ is EFA, then write the resulting optimization over α explicitly to show it has a solution.

Consider a linear contract $w[\alpha, \beta]$. By Lemma 5, a contract is EFA if it is locally eligible via an optimal input and the cheapest input $(\underline{\widehat{\mathcal{A}}}, \underline{c}_P)$ is not FA-worrisome, that is, for any $\mathcal{A} \supseteq \underline{\widehat{\mathcal{A}}}$,

$$V_P\big(w[\alpha, \beta]|\mathcal{A}, \underline{c}_P\big) > V_P\big(w[\alpha, \beta]|\mathcal{W}\big) \tag{26}$$

implies

$$V_A\big(w[\alpha, \beta]|\mathcal{A}\big) \geq 0.$$

A decrease in β has no effect on the former of these conditions and tightens the latter, so given α, it is optimal to decrease β until the implication is just about to fail, that is, until there exists $\mathcal{A} \supseteq \underline{\mathcal{A}}$ with (26) an equality and $V_A(w[\alpha, \beta]|\mathcal{A}) = 0$. (This assumes that (E3) from local eligibility is not the binding constraint on the choice of β; momentarily we shall verify that this is the case.)

To calculate the principal's guarantee, first note that (26) implies a nonnegative worst-case expected output, that is, $(1/\alpha)\max_{(F,c)\in\mathcal{A}}\{E_F[\alpha y] - c\} \geq 0$; otherwise, $V_P(w[\alpha, \beta]|\mathcal{A}, \underline{c}_P) = V_P(w[\alpha, \beta]|\underline{\widehat{\mathcal{A}}}, \underline{c}_P)$ contradicting the definition of $V_P(w[\alpha, \beta]|\mathcal{W})$. Hence, by calculations similar to those used to derive (6), the worst-case expected output can equivalently be written as $(1/\alpha)(V_A(w[\alpha, \beta]|\mathcal{A}) - \beta)$ and the principal's and the agent's guarantee are related as

$$V_P\big(w[\alpha, \beta]|\mathcal{A}, c_P\big) = \frac{1-\alpha}{\alpha}\big(V_A\big(w[\alpha, \beta]|\mathcal{A}\big) - \beta\big) - \beta - \underline{c}_P.$$

Thus, the optimal β, call it $\beta_{\text{EFA}}(\alpha)$, satisfies

$$\frac{1-\alpha}{\alpha}\cdot 0 - \frac{\beta_{\text{EFA}}(\alpha)}{\alpha} - \underline{c}_P = V_P\big(w\big[\alpha, \beta_{\text{EFA}}(\alpha)\big]|\mathcal{W}\big).$$

Let $(\widehat{\mathcal{A}}^*, c_P^*)$ be an optimal input. The worst-case expected output must again be nonnegative; otherwise, $w[\alpha, \beta]$ cannot be locally eligible via this optimal input (for any β). Thus, the principal's and agent's guarantee are (again) related as

$$V_P\big(w[\alpha, \beta]|\mathcal{W}\big) = V_P\big(w[\alpha, \beta]|\widehat{\mathcal{A}}^*, c_P^*\big) = \frac{1-\alpha}{\alpha}\max_{(F,c)\in\widehat{\mathcal{A}}^*}\big\{E_F[\alpha y] - c\big\} - c_P^* - \beta,$$

and the optimal β is given by

$$\beta_{\text{EFA}}(\alpha) = \left(c_P^* - \underline{c}_P\right)\frac{\alpha}{1-\alpha} - \max_{(F,c)\in\widehat{\mathcal{A}}^*}\left\{E_F[\alpha y] - c\right\}.$$

Note that, as promised above, this β does assure the agent a nonnegative guarantee under $(\mathcal{A}^*, c_P^*)$, so we were safe in ignoring (E3). With this choice of β for each α, then, the principal's guarantee is

$$\frac{1}{\alpha}\max_{(F,c)\in\widehat{\mathcal{A}}^*}\left\{E_F[\alpha y] - c\right\} - \frac{\alpha}{1-\alpha}\left(c_P^* - \underline{c}_P\right) - c_P^*. \tag{27}$$

We maximize this over all choices of α that arise in some EFA contract, taking $(\widehat{\mathcal{A}}^*, c_P^*)$ to be the optimal input for given α. Notice, moreover, that for any fixed α, the choice of input $(\widehat{\mathcal{A}}^*, c_P^*)$ that is optimal is in fact the same one that maximizes (27). So, the problem is equivalent to maximizing

$$\frac{1}{\alpha}\max_{(F,c)\in\widehat{\mathcal{A}}}\left\{E_F[\alpha y] - c\right\} - \frac{\alpha}{1-\alpha}(c_P - \underline{c}_P) - c_P$$

over $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ and $\alpha \in [0, 1]$ such that

$$\frac{1-\alpha}{\alpha}\max_{(F,c)\in\widehat{\mathcal{A}}}\left\{E_F[\alpha y] - c\right\} - c_P \geq 0$$

to satisfy local eligibility (recall (8)).

(In the special case $\alpha = 1$, local eligibility is possible only if $c_P = \underline{c}_P = 0$, and then the contract is automatically EFA; the formula remains valid in this case with the $(\alpha/(1 - \alpha))(c_P - \underline{c}_P)$ term interpreted as zero. In the case $\alpha = 0$, the formula is also correct as long as the contract is locally eligible, as discussed in Section 2.2.3.)

Because this objective function is continuous in α, the maximum is attained. $\square$

Proof of Proposition 4. Take any contract w that is locally eligible via some optimal input. Suppose w is not EFI. Then there exists $\widetilde{\mathcal{W}} \supseteq \mathcal{W}$ and $(\widehat{\mathcal{A}}, c_P) \in \widetilde{\mathcal{W}}$ such that

$$V_P(w|\widehat{\mathcal{A}}, c_P) > V_P(w|\mathcal{W}) \quad \text{and} \quad V_A(w|\widehat{\mathcal{A}}) < 0. \tag{28}$$

Let $\mathcal{A} = \widehat{\mathcal{A}} \cup \mathcal{A}_{\text{triv}}$. Consider $(\mathcal{A}_{\text{triv}}, 0) \in \mathcal{W}'$. Clearly, $\mathcal{A} \supseteq \mathcal{A}_{\text{triv}}$. If $V_A(w|\mathcal{A}) \geq 0$, then the agent must be choosing action $(\delta_0, 0)$ implying that the principal's guarantee from $(\widehat{\mathcal{A}}, c_P)$ is at most 0, contradicting (28) as $V_P(w|\mathcal{W}) > 0$.

Thus,

$$V_A(w|\mathcal{A}) < 0 \quad \text{and} \quad V_P(w|\mathcal{A}, 0) \geq V_P(w|\widehat{\mathcal{A}}, c_P) > V_P(w|\mathcal{W}),$$

that is, $(\mathcal{A}_{\text{triv}}, 0) \in \mathcal{W}'$ is a FA-worrisome input given contract w.

Conversely, suppose w is not EFA under input space $\mathcal{W}'$. Then there exists some $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}'$ that is FA-worrisome given w, that is, there exists $\mathcal{A} \supseteq \widehat{\mathcal{A}}$ such that

$$V_P(w|\mathcal{A}, c_P) > V_P\left(w|\mathcal{W}'\right) \geq V_P(w|\mathcal{W}) \quad \text{and} \quad V_A(w|\mathcal{A}) < 0. \tag{29}$$

Let $\widetilde{\mathcal{W}} = \mathcal{W} \cup \{(\mathcal{A}, c_P)\}$. Clearly, $\widetilde{\mathcal{W}} \supseteq \mathcal{W}$ and $(\mathcal{A}, c_P) \in \widetilde{\mathcal{W}}$. By (29), this implies w is not EFI. $\square$

Appendix C: Characterizations of optimal contracts

C.1 *Eligibility in single-input model*

We give an alternative characterization of the optimal linear contract in the single-input model. Let the function $r : \widehat{\mathcal{A}} \to \{-1\} \cup [0, 1]$ be defined as

$$r(F, c) = \begin{cases} \dfrac{E_F[y] + c - c_P}{2E_F[y]} + \sqrt{\left(\dfrac{E_F[y] + c - c_P}{2E_F[y]}\right)^2 - \dfrac{c}{E_F[y]}} \\ \qquad \text{if } E_F[y] - c - c_P \geq 2\sqrt{cc_P} \text{ and } E_F[y] - c - c_P > 0 \\ -1 \quad \text{otherwise.} \end{cases} \tag{30}$$

Fixing an action (F, c) in $\widehat{\mathcal{A}}$, function r returns the larger root of the equation

$$\frac{1-\alpha}{\alpha}\big\{E_F[\alpha y] - c\big\} - c_P = 0.$$

(The second branch of (30) corresponds to the case where there are no real roots.) Combining this with (12) leads us to the following result.

Lemma 6. *If an eligible linear contract exists, then*

$$\max_{(F,c)\in\widehat{\mathcal{A}}} r(F, c) = \alpha^*.$$

Proof. First, suppose $c_P = 0$. Then an eligible contract exists as long as there is some action with strictly positive surplus, and for any such action, the formula (30) for r simplifies to 1, which indeed is the value of α^*.

Now suppose $c_P > 0$. Let

$$\alpha = \max_{(F',c')\in\widehat{\mathcal{A}}} r\big(F', c'\big) \quad \text{and} \quad (F, c) \in \operatorname*{arg\,max}_{(F',c')\in\widehat{\mathcal{A}}} \big\{E_{F'}[\alpha y] - c'\big\}.$$

(Note that the max in defining α exists: it could only fail to exist if the sup were approached by a sequence of actions whose limit fails to satisfy the strict inequality constraint in (30), that is, $E_F[y] - c - c_P = 0$, but that still satisfy the weak inequality constraint. This requires the limit to satisfy $c = 0$. In this case, the limiting value of the formula (30) is zero, which cannot be the supremum.)

To show that $\alpha \geq \alpha^*$, note that

$$\alpha = \max_{(F',c')\in\widehat{\mathcal{A}}} r\big(F', c'\big) \geq r\big(F^*, c^*\big) = \alpha^*$$

where $(F^*, c^*) \in \arg\max_{(F',c')\in\widehat{\mathcal{A}}}\{E_{F'}[\alpha^* y] - c'\} - c_P$.

To show that $\alpha^* \geq \alpha$, note that

$$0 = \frac{1-\alpha}{\alpha}\big\{E_F[\alpha y] - c\big\} - c_P \leq \frac{1-\alpha}{\alpha} \max_{(F',c')\in\widehat{\mathcal{A}}} \big\{E_{F'}[\alpha y] - c'\big\} - c_P. \qquad \square$$

C.2 *Eligibility with further actions*

Here and for the rest of the Appendix, we turn to the multiple-input model.

The optimal EFA linear contract can be identified as follows. First, for a given α, we derive the lowest β that satisfies the constraint that the agent's guarantee should be nonnegative. Second, we maximize the principal's guarantee over $\alpha \in [0, 1]$ and $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ subject to the requirement of local eligibility.

The first step was done in the proof of Proposition 3, so rather than repeat the calculations here, we just restate their conclusions: for each α, if there exists a β that makes the contract EFA, the optimal such β is given by

$$\beta_{\text{EFA}}(\alpha) = (c_P - \underline{c}_P)\frac{\alpha}{1-\alpha} - \max_{(F,c)\in\widehat{\mathcal{A}}}\bigl\{E_F[\alpha y] - c\bigr\}, \tag{31}$$

and the principal's guarantee from the resulting linear contract is

$$\frac{1}{\alpha}\max_{(F,c)\in\widehat{\mathcal{A}}}\bigl\{E_F[\alpha y] - c\bigr\} - \frac{\alpha}{1-\alpha}(c_P - \underline{c}_P) - c_P \tag{32}$$

where $(\widehat{\mathcal{A}}, c_P)$ is the input that maximizes the value of (32), and this input is also optimal in the sense of Definition 3. Here, $\underline{c}_P$ is the lowest cost among all inputs in $\mathcal{W}$, and the first term in (31) is interpreted as zero if $\alpha = 1$ and $c_P = \underline{c}_P$. The local eligibility constraint (8) rewrites as

$$\frac{1-\alpha}{\alpha}\max_{(F,c)\in\widehat{\mathcal{A}}}\bigl\{E_F[\alpha y] - c\bigr\} - c_P \geq 0.$$

Thus, our problem is equivalent to the following: simultaneously choose $\alpha \in [0, 1]$, $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$, and $(F, c) \in \widehat{\mathcal{A}}$ to maximize

$$\frac{1}{\alpha}\bigl(E_F[\alpha y] - c\bigr) - \frac{\alpha}{1-\alpha}(c_P - \underline{c}_P) - c_P \tag{33}$$

subject to the constraint

$$\frac{1-\alpha}{\alpha}\bigl(E_F[\alpha y] - c\bigr) - c_P \geq 0. \tag{34}$$

Refer to the optimal α in the above problem as α^*. (As noted in the proof of Proposition 3, continuity ensures the optimum exists.)

Then an EFA contract exists if and only if this maximum value of (33) is positive. If so, an optimal EFA contract is given by $w[\alpha^*, \beta_{\text{EFA}}(\alpha^*)]$, and (33) gives the corresponding optimal value of V_P.

We now proceed to a more explicit analysis of the optimization problem in (33)–(34).

We proceed as follows. For any fixed choice of input and action, we optimize (33) over α. These artificial optimization problems are easy to solve as the objectives will be concave and subject to an interval constraint (namely (34)). Thus, their solution falls into one of two cases: (1) the constraint holds with equality or (2) the solution is characterized by a first-order condition.

If the solution to the overall optimization falls under case (1), then we can equivalently find it by pretending the constraint holds with equality for *all* input-action pairs (because this will only decrease the objective for other such pairs while leaving it unchanged for the optimal pair). Otherwise, if it falls under case (2), then we can use the first-order condition to identify a candidate α for each input-action pair, and characterize the solution by maximizing (33) over the set of pairs for which the corresponding α satisfies the constraint (34).

To begin carrying out the above analysis, fix $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ and $(F, c) \in \widehat{\mathcal{A}}$. We can rewrite the objective function (33) as

$$\big(E_F[y] - c - c_P\big) - \frac{1-\alpha}{\alpha}c - \frac{\alpha}{1-\alpha}(c_P - \underline{c}_P). \tag{35}$$

The constraint is still (34). The objective is concave and α is constrained to lie in some interval. Thus, we can divide into two cases as above.

Case 1: First, suppose that the constraint binds in which case

$$\frac{1-\alpha}{\alpha}\big\{E_F[\alpha y] - c\big\} - c_P = 0.$$

The objective function evaluated at this contract is then given by

$$\frac{\alpha}{1-\alpha}\underline{c}_P,$$

which is increasing in α. Thus, we can identify the optimal contract and input, if the constraint binds, by a variation of function r defined in (30) and Lemma 6 in Appendix C.1.

Define $r^m((F, c), c_P)$, for the multiple-input environment, by the same formula as r, in the single-input environment (see (30)); the only difference is that we now write c_P as an argument rather than a constant.

Let

$$\big(\widehat{\mathcal{A}}^b, c_P^b\big) \in \operatorname*{arg\,max}_{(\widehat{\mathcal{A}}, c_P)\in\mathcal{W}} \max_{(F,c)\in\widehat{\mathcal{A}}} r^m\big((F, c), c_P\big) \quad \text{and} \quad \alpha^b = \max_{(F,c)\in\widehat{\mathcal{A}}^b} r^m\big((F, c), c_P^b\big)$$

where superscript b stands for "binding."

The optimal contract is given by $w[\alpha^b, \beta_{\text{EFA}}(\alpha^b)]$, the corresponding optimal input is $(\widehat{\mathcal{A}}^b, c_P^b)$ and the principal's guarantee is given by

$$\frac{\alpha^b}{1-\alpha^b}\underline{c}_P. \tag{36}$$

Case 2: Now, suppose that at the optimal contract, the constraint does not bind and can be ignored in the maximization over α. The first-order condition yields

$$\alpha = \frac{\sqrt{c}}{\sqrt{c} + \sqrt{c_P - \underline{c}_P}}. \tag{37}$$

Evaluating (35) at this α gives

$$E_F[y] - c - c_P - 2\sqrt{c}\sqrt{c_P - \underline{c}_P} \tag{38}$$

and the formula (31) for β (using this (F, c) in place of the agent's optimal action) gives

$$\sqrt{c}\sqrt{c_P - \underline{c}_P} - \left\{E_F\left[\frac{\sqrt{c}}{\sqrt{c} + \sqrt{c_P - \underline{c}_P}}y\right] - c\right\},$$

respectively.

To check whether the constraint (34) holds, rewrite it as

$$\frac{1}{\alpha}\{E_F[\alpha y] - c\} - \frac{1}{1-\alpha}c_P \geq 0$$

and note that

$$\begin{aligned}\frac{1}{\alpha}\{E_F[\alpha y] - c\} - \frac{1}{1-\alpha}c_P &= \frac{1}{\alpha}\{E_F[\alpha y] - c\} - \frac{\alpha}{1-\alpha}(c_P - \underline{c}_P) - c_P - \frac{\alpha}{1-\alpha}\underline{c}_P \\ &= E_F[y] - c - c_P - 2\sqrt{c}\sqrt{c_P - \underline{c}_P} - \frac{\sqrt{c}}{\sqrt{c_P - \underline{c}_P}}\underline{c}_P.\end{aligned}$$

Thus, if

$$E_F[y] - c - c_P - 2\sqrt{c}\sqrt{c_P - \underline{c}_P} - \frac{\sqrt{c}}{\sqrt{c_P - \underline{c}_P}}\underline{c}_P \geq 0, \tag{39}$$

then local eligibility is satisfied.

Thus, the optimal contract in this case can be found by maximizing (38) over $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ and $(F, c) \in \widehat{\mathcal{A}}$ such that (39) is satisfied.

Finally, comparing this value of (38) to (36) determines which of the two cases identifies the global optimum.

C.3 *Eligibility with full knowledge*

The optimal EFK linear contract can be identified as in the EFA case. First, for a given α, we derive the corresponding optimal β. Second, we maximize the objective function over $\alpha \in [0, 1]$ and $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ while ensuring local eligibility. Thus, write $\bar{y} = \max(\mathcal{Y})$, and then for any $\alpha \in [0, 1]$ and $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$, define $\mathcal{W}^w(\alpha, (\widehat{\mathcal{A}}, c_P)) \subseteq \mathcal{W}$ as $\mathcal{W}^w(\alpha, (\widehat{\mathcal{A}}, c_P)) := \{(\widehat{\mathcal{A}}', c'_P) \in \mathcal{W} : (1-\alpha)\bar{y} - c'_P > V_P(w[\alpha, 0]|\widehat{\mathcal{A}}, c_P)\} \cup \{(\widehat{\mathcal{A}}, c_P)\}$; this is the set of inputs the principal may supply if the comparison point to define worrisome inputs is $(\widehat{\mathcal{A}}, c_P)$ and the slope of the contract is α.

Under contract $w[\alpha, \beta]$, if the principal uses any such input $(\widehat{\mathcal{A}}', c'_P)$, the agent's subsequent guarantee is just $\max_{(F,c)\in\widehat{\mathcal{A}}'}\{E_F[\alpha y] - c\} + \beta$. (Unlike the EFA case, the fact that the principal's knowledge leads her to choose this input does not imply any better payoff guarantee for the agent: it could be that he has a very productive action, but all the extra payment it earns him is dissipated by a high effort cost.) Thus, the lowest β to guarantee the agent at least zero is

$$\beta_{\text{EFK}}(\alpha, (\widehat{\mathcal{A}}, c_P)) = -\min_{(\widehat{\mathcal{A}}', c'_P)\in\mathcal{W}^w(\alpha, (\widehat{\mathcal{A}}, c_P))} \max_{(F,c)\in\widehat{\mathcal{A}}'}\{E_F[\alpha y] - c\}.$$

Let α^* and $(\widehat{\mathcal{A}}^*, c_P^*)$ jointly maximize

$$V_P\big(w\big[\alpha, \beta_{\text{EFK}}\big(\alpha, (\widehat{\mathcal{A}}, c_P)\big)\big]|\widehat{\mathcal{A}}, c_P\big) \tag{40}$$

over $\alpha \in [0, 1]$ and $(\widehat{\mathcal{A}}, c_P) \in \mathcal{W}$ such that

$$\frac{1-\alpha}{\alpha} \max_{(F,c)\in\widehat{\mathcal{A}}} \big\{E_F[\alpha y] - c\big\} - c_P \geq 0.$$

Then an optimal monotone EFK contract (if any such contract exists) is given by $w[\alpha^*, \beta_{\text{EFK}}(\alpha^*, (\widehat{\mathcal{A}}^*, c_P^*))]$, and it is locally eligible for $(\widehat{\mathcal{A}}^*, c_P^*)$. (Note that, for a given α, β_{EFK} is indeed minimized—and so (40) maximized—by taking $(\widehat{\mathcal{A}}, c_P)$ to be an optimal input.)

An eligible monotone EFK contract exists if and only if

$$V_P\big(w\big[\alpha^*, \beta_{\text{EFK}}\big(\alpha^*, (\widehat{\mathcal{A}}^*, c_P^*)\big)\big]|\widehat{\mathcal{A}}^*, c_P^*\big) > 0.$$

References

Barron, Daniel, George Georgiadis, and Jeroen Swinkels (2020), "Optimal contracts with a risk-taking agent." *Theoretical Economics*, 15, 715–761. [1627]

Bhattacharyya, Sugato and Francine Lafontaine (1995), "Double-sided moral hazard and the nature of share contracts." *The RAND Journal of Economics*, 26, 761–781. [1624, 1647]

Carroll, Gabriel (2015), "Robustness and linear contracts." *American Economic Review*, 105, 536–563. [1625, 1627, 1630, 1632, 1646]

Chassang, Sylvain (2013), "Calibrated incentive contracts." *Econometrica*, 81, 1935–1971. [1627]

Cooper, Russell and Thomas W. Ross (1985), "Product warranties and double moral hazard." *The RAND Journal of Economics*, 16, 103–113. [1624]

Corbett, Charles J. and Gregory A. DeCroix (2001), "Shared-savings contracts for indirect materials in supply chains: Channel profits and environmental impacts." *Management Science*, 47, 881–893. [1624]

Corbett, Charles J., Gregory A. DeCroix, and Albert Y. Ha (2005), "Optimal shared-savings contracts in supply chains: Linear contracts and double moral hazard." *European Journal of Operational Research*, 163, 653–667. [1624]

Dai, Tianjiao and Juuso Toikka (2022), "Robust incentives for teams." *Econometrica*, 90, 1583–1613. [1626, 1631, 1646]

Demski, Joel S. and David E. M. Sappington (1991), "Resolving double moral hazard problems with buyout agreements." *The RAND Journal of Economics*, 22, 232–240. [1626]

Diamond, Peter (1998), "Managerial incentives: On the near linearity of optimal compensation." *Journal of Political Economy*, 106, 931–957. [1624, 1627]

Eswaran, Mukesh and Ashok Kotwal (1985), "A theory of contractual structure in agriculture." *American Economic Review*, 75, 352–367. [1624]

Georgiadis, George and Balazs Szentes (2020), "Optimal monitoring design." *Econometrica*, 88, 2075–2107. [1624]

Grossman, Sanford J. and Oliver D. Hart (1983), "An analysis of the principal-agent problem." *Econometrica*, 51, 7–45. [1623]

Hébert, Benjamin (2018), "Moral hazard and the optimality of debt." *The Review of Economic Studies*, 85, 2214–2252. [1624]

Holmström, Bengt (1979), "Moral hazard and observability." *The Bell Journal of Economics*, 10, 74–91. [1623]

Holmström, Bengt and Paul Milgrom (1987), "Aggregation and linearity in the provision of intertemporal incentives." *Econometrica*, 55, 303–328. [1624, 1627]

Innes, Robert D. (1990), "Limited liability and incentive contracting with ex-ante action choices." *Journal of Economic Theory*, 52, 45–67. [1623, 1624]

Kim, Son Ku and Susheng Wang (1998), "Linear contracts and the double moral-hazard." *Journal of Economic Theory*, 82, 342–378. [1624]

Lafontaine, Francine (1992), "Agency theory and franchising: Some empirical results." *The RAND Journal of Economics*, 23, 263–283. [1635]

Lopomo, Giuseppe, Luca Rigotti, and Chris Shannon (2011), "Knightian uncertainty and moral hazard." *Journal of Economic Theory*, 146, 1148–1172. [1624]

Roels, Guillaume, Uday S. Karmarkar, and Scott Carr (2010), "Contracting for collaborative services." *Management Science*, 56, 849–863. [1624]

Co-editor Simon Board handled this manuscript.

Manuscript received 11 June, 2021; final version accepted 23 October, 2022; available online 25 October, 2022.