Indicator Choice in Pay-for-Performance

Majid Mahzoon[a*], Ali Shourideh[b], Ariel Zetlin-Jones[b]

[a] Department of Economics, University of Central Florida, Orlando, FL, United States

[b] Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, United States

September 19, 2026

Abstract

We study the classical moral-hazard model with limited liability when the party being measured chooses the performance measure. Before the principal offers a contract, the agent commits to an indicator—any garbling of a fixed performance technology—and compensation can depend only on its realization. She withholds information even when a perfectly informative measure is available, and any implementable outcome can be replicated with a binary pass/fail indicator. We recast her design problem as a geometric game in which she chooses a set of likelihood-ratio vectors and the principal chooses a point in it. We show that the problem reduces, without loss, to the choice of a single vector from the set made feasible by the technology. Using this representation, we characterize when she can implement a given effort, including the first-best effort, while capturing the entire surplus, and show that a more informative technology can reduce equilibrium surplus. For single-index technologies, in which effort enters the output distribution through a scalar index, an optimal indicator is a threshold on output. Our results provide a rationale for coarse performance measures that depends on who holds the design right rather than on the cost or precision of measurement.

Keywords: moral hazard; informativeness principle; information design; limited liability; pay for performance; principal–agent model

*Corresponding author. E-mail address: majid.mahzoon@ucf.edu

1 Introduction

Performance pay is often tied to measures far coarser than the data available. Bonuses depend on whether performance crosses a threshold rather than on its precise level, and the measure itself is frequently an object of negotiation. Executives bargain over the vesting conditions attached to their stock grants, and athletes over which statistics trigger their bonuses. In 2023, the Writers Guild of America negotiated a residual bonus tied to a single viewership threshold—of twenty percent of a service's domestic subscribers within ninety days—although streaming platforms observe viewership in fine detail. In each case, the party being measured helped choose a measure coarser than the data behind it.[1]

The classical theory of moral hazard usually takes the performance measure on which compensation is based as given. When the measure is chosen, it is typically chosen either by the principal, who by the informativeness principle prefers more informative measures, or by a designer who is free to select any signal of effort, unconstrained by the available performance data. This paper instead gives the choice to the party being measured and restricts her to signals that can be constructed from those data. We ask what measure she chooses and what her choice does to efficiency. We find that she withholds information even when a perfectly informative measure is available, that a single pass/fail signal is all she ever needs, that, under conditions we characterize, her preferred measure implements the first-best effort while leaving her the entire surplus, and that richer performance data can lower total surplus.

We study this question in the textbook moral-hazard model with a risk-neutral agent protected by limited liability, as in Innes (1990), with one stage added at the beginning. A fixed *performance technology* maps the agent's costly effort e into a distribution over performance outcomes x. Before the principal offers a contract, the agent chooses an *indicator*, which can be any garbling of x, and the principal can condition compensation only on that indicator. Formally, an indicator is an information structure, a stochastic map from output to signals. The agent commits to the indicator but not to her effort. Once the indicator is chosen, the parties play the standard moral-hazard game.[2] We interpret the

[1] See Bebchuk and Fried (2004) on executive compensation and this article in the Daily Mail on athletes. The terms of the writers' bonus are set out in the Guild's summary of the 2023 agreement (Writers Guild of America (2023)).

[2] Throughout, we assume that the agent can commit to the indicator but cannot commit to the effort level. A natural justification is that employment contracts are typically enforceable, while effort choices are not verifiable and cannot be contracted upon directly. This asymmetry is also the source of the agent's rent: if she could commit to effort as well, the principal would not need to provide incentives and would

agent's control over the indicator as bargaining power over which performance measures enter the contract, as in executive compensation or the writers' negotiations described above.

The agent faces a genuine trade-off. Suppose for the moment that the performance technology is perfect, so that x reveals effort. If she fully reveals x, the principal can infer her effort and need pay her only enough to cover its cost. If she reveals nothing, the principal cannot provide incentives, so no costly effort is exerted and the agent again earns no rent. The question is how much information she should withhold. One might expect her to sacrifice some efficiency in order to earn a rent. Under a perfect technology, however, she need not: she can design an indicator that implements the first-best effort and leaves her the entire surplus.

The agent's choice of indicator works by making the principal a residual claimant on total surplus. She discloses only a pass/fail signal whose probability of passing rises with the cost of effort, with the bonus calibrated so that each extra unit of cost buys exactly one unit of expected pay. She is then indifferent among all efforts the indicator makes implementable, so implementing any of them costs the principal that effort's cost plus a fixed rent. The principal therefore chooses the effort that maximizes surplus—the first best. The agent raises the rent until the principal's outside option binds, since he can always implement the zero-cost effort at no wage. With a perfectly revealing measure that the agent could not garble, the principal would implement the same first-best effort but keep the surplus himself. Giving the agent control over the measure thus changes the division of surplus without changing the implemented effort.

Once the performance technology is imperfect, the indicator must be a garbling of a noisy performance outcome, and it is no longer clear whether the agent can preserve the first-best effort or capture the full surplus. Analyzing the problem requires a different approach from standard information design. Each effort induces a different distribution over output, the principal chooses a wage schedule before the signal is realized, and incentives depend on how likely each signal is not only under the target effort but also under off-path efforts. The problem is therefore not posterior-separable, so the standard concavification approach of Kamenica and Gentzkow (2011) does not apply.

Our first contribution recasts the agent's design problem as a geometric game: she chooses a set, and the principal chooses a point in it. Each signal is represented by the vector of likelihood ratios comparing its probability under alternative efforts with its prob-

pay nothing.

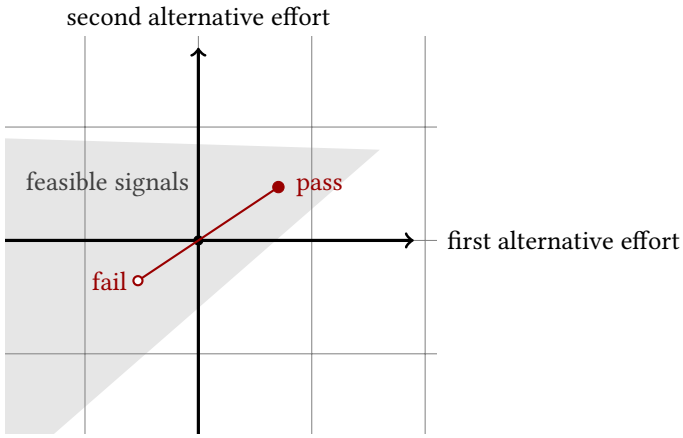

Figure 1: The agent's design problem for a technology with three efforts. The shaded region represents the signals supported by the performance technology in likelihood-ratio coordinates, one for each alternative to the target effort. A pass/fail indicator generates a segment through the origin. The principal's contract selects a point on this segment—here the pass endpoint—and payoffs depend only on that point.

ability under the target. Continuation payoffs depend on the indicator only through the convex hull of these vectors, and the performance technology determines the largest hull available. We then reduce the problem further: without loss, the agent can choose a single vector from that hull, which can be generated by an indicator with only two realizations. Figure 1 illustrates the set chosen by the agent—a segment generated by a pass/fail indicator—inside the largest hull made available by the performance technology. An infinite-dimensional design problem thus becomes a finite-dimensional one. The representation itself does not depend on the agent holding the design right: it characterizes the principal's continuation problem for any given indicator.

Our second contribution characterizes when the agent can implement a target effort and capture the entire surplus; the leading case is the first best. The idea in this case is similar to the one described for a perfect performance technology. There is a particular likelihood-ratio vector that leaves the agent indifferent across all efforts, playing the same role as above. If this vector lies in the feasible set generated by the performance technology, the agent can implement the first-best effort and capture the entire surplus. Under a perfect technology, the feasible set is as large as possible and always contains this vector. An imperfect technology shrinks the set, but full extraction at the first best remains possible as long as the vector is still feasible. The feasibility condition reflects the severity

of the agency problem: it is easier to satisfy when the technology is more informative or when effort costs are closer together. Checking the condition amounts to solving a linear feasibility problem. We also provide sufficient conditions stated directly in the primitives, involving the likelihood ratios of a single output.

We also ask how equilibrium surplus varies with the technology. A more informative technology enlarges the set of indicators available to the agent, so her equilibrium payoff is weakly higher, but total surplus need not be. We give an example in which a Blackwell improvement in the technology strictly lowers equilibrium surplus because it allows the agent to target an effort that gives her a higher payoff but is less efficient. A decline in the largest equilibrium surplus requires at least four effort levels and a technology rich enough that the feasible set is more than one-dimensional; with three or fewer efforts, and in the single-index class we turn to next, the largest equilibrium surplus cannot fall.

In one tractable class—single-index technologies, where effort shifts the output distribution through a scalar index—an optimal indicator is a threshold on output: pass or fail according to whether output clears a cutoff, as in the writers' bonus. The same class also reveals a limit to the efficiency result: the agent's preferred threshold generates weakly less total surplus than the one the principal would choose.

Which indicators determine pay is also an active policy question. Employers increasingly use algorithmic monitoring to evaluate workers, while recent proposals call for greater disclosure and worker voice over the technologies used to measure performance (Bernhardt et al. (2023)). Our model studies the limiting case in which the worker controls the choice of indicator, and our results show that such control need not reduce efficiency: under conditions we characterize, the worker's preferred indicator implements the first-best effort. The framework also accommodates two forms of regulation. Guaranteeing the principal access to baseline metrics places a lower bound on the information available for contracting, which we study in Appendix A. Restrictions on monitoring instead place an upper bound on available information, and our results show that such restrictions need not reduce equilibrium surplus and can increase it, even before accounting for any privacy benefits.

1.1 Related Literature

Our paper relates to several strands of the literature on contracting and information design.

Within the moral-hazard literature, our key innovation is to study the agent's choice of performance indicators and to show that this choice can eliminate all agency inefficiencies. In the classical model, Innes (1990) and Poblete and Spulber (2012) analyze moral hazard with limited liability and a risk-neutral agent and show that simple debt contracts are optimal. Carroll (2015) and Walton and Carroll (2022) consider moral hazard with limited liability when the principal faces non-Bayesian uncertainty about the production technology and seeks to maximize her worst-case payoff. [3]

A related strand studies the choice among imperfect performance measures. In the multitask framework of Holmstrom and Milgrom (1991) and Feltham and Xie (1994), imperfect measures distort the allocation of effort across tasks, while Baker (2002) analyzes the resulting trade-off between distortion and risk in choosing among them. In our model, the measure is a garbling of output itself, so it may be coarse but is never misaligned. Its imperfection determines the surplus generated by the implemented effort and how that surplus is divided.

While information-design incentives are often abstracted away in moral-hazard models, there are notable exceptions. Holmström (1979) and, more recently, Chaigneau et al. (2019) study how changes in the performance technology affect incentives. By the informativeness principle, more informative outputs reduce the wage necessary to induce effort. Gjesdal (1982) and Kim (1995) rank information systems in agency models, with Kim's ranking based on mean-preserving spreads of the likelihood-ratio distribution. Demougin and Fluet (2001) provide a related ranking under risk neutrality and limited liability, which matches our wage environment. These rankings are made from the principal's side over exogenously given information systems. Corollary 1 instead ranks technologies, the feasible sets from which the agent designs, by the equilibrium value of her design problem, with the comparison expressed through inclusion of the convex hulls of likelihood-ratio vectors rather than through dispersion of the likelihood-ratio distribution. Moreover, these papers treat the performance technology as exogenous and do not allow the agent to influence the information that the principal observes.

Two of the closest papers to ours are Garrett et al. (2023) and Babichenko et al. (2025). Garrett et al. (2023) study a setting in which the agent designs the production technology itself, choosing the cost of generating each output distribution. They term this problem "technology design" and show that the agent-optimal technology uses only binary distributions. In contrast, in our model the performance technology is fixed, and the agent

[3] A large related literature also examines models with risk-averse agents and risk-neutral principals.

instead selects an indicator of performance, i.e., an information structure that garbles the principal's observations. Georgiadis et al. (2024) study the polar case of our fixed, finite technology: given the cost of producing each output distribution, the agent can choose any distribution in response to the contract. Whether the technology is designed, as in Garrett et al. (2023), or flexible by assumption, the margin of choice in both papers concerns the output distribution. In our model, the production technology is fixed and the agent chooses only how output is measured for compensation. In addition, our paper contributes technically by reformulating the problem in terms of likelihood ratios, which provides a geometric representation of the agent's information-design problem.

Babichenko et al. (2025) study information design in a principal–agent problem in which a regulator chooses the information structure observed by the principal, and characterize the implementable actions and utility profiles when information structures are unconstrained or exogenously restricted. Their unconstrained environment corresponds to the boundary case of our model in which the performance technology is perfect. In that case, any signal of the action is a feasible indicator, and our characterization recovers their implementability result exactly (Section 5.1). Away from this boundary, the indicator must arise as a garbling of a fixed performance technology, and the feasible indicators are summarized by a technology-dependent convex set of likelihood-ratio vectors, the object at the center of our analysis. Everything we do beyond this nesting concerns that object: how the performance technology, together with the other primitives, determines when the agent can extract the full surplus; how equilibrium payoffs and surplus vary with the informativeness of the technology; and the analysis of a tractable class in which the technology generates a one-dimensional feasible set of likelihood-ratio vectors. These questions have no analogue in the unconstrained environment, where there is no technology to vary and no feasible set to characterize. The identity of the designer is not the essential difference. Our framework applies to any party holding the design right, including a regulator; the essential difference is the technological constraint.

Our support-based characterization is closely related to recent work on worst-case privacy measures in Liu (2025). By support-based, we mean that the relevant object is the set of posterior realizations that can arise, and hence the induced set of likelihood-ratio vectors, rather than the probabilities assigned to those realizations. There, privacy loss admits a max-separable representation over signal-specific likelihood vectors. Since the associated index function is homogeneous of degree zero, scaling a likelihood vector changes the probability of the corresponding signal without changing its privacy index.

In our model, a related homogeneity property follows from wage normalization in Lemma 1. Proportional changes in signal frequencies can be offset by inverse changes in wages, leaving incentive constraints and expected payments unchanged.

Georgiadis and Szentes (2020) study moral hazard with limited liability and a risk-averse agent in a setting where the principal can continuously acquire signals about the agent's effort at a constant marginal cost. They show that the principal-optimal information-acquisition strategy is a two-threshold policy. Barron et al. (2020) analyze a model in which the agent can costlessly add mean-preserving noise to the output. This feature is also present in our setting; however, in our model the noisy output is used only for contracting purposes and does not affect the principal's payoff relative to the original output. Moreover, in our timing the agent first chooses the information structure, then the principal offers a contract, and finally the agent chooses her effort. In Barron et al. (2020), the sequence is reversed: the principal first offers a contract, and the agent then chooses both effort and the information structure.

Another related strand of the moral-hazard literature focuses on career concerns, introduced by Holmström (1999), and on changes in information structures observed by both the principal and the agent. In this context, Holmström (1999) shows that noisy performance signals can strengthen incentives. Similarly, Dewatripont et al. (1999) demonstrate that more informative signals in the sense of Blackwell do not necessarily increase incentives. In contrast, in our model, indicator design, that is, the agent's ability to manipulate the information available to the principal, serves as a tool to reallocate surplus between the agent and the principal while still achieving efficiency.

In the context of team production and moral hazard, Halac et al. (2021) show that the principal can benefit from private contract offers by exploiting rank uncertainty: each agent observes only her own bonus and a distribution over rankings, and her bonus makes working dominant if higher-ranked agents also work. In our setting, by contrast, it is the agent who leverages the principal's uncertainty about performance to improve efficiency.

We also contribute to the growing literature on incentives in Bayesian persuasion. Several papers, including Boleslavsky and Kim (2018), Rosar (2017), Perez-Richet and Skreta (2022), Ball (2019), Saeedi and Shourideh (2026), and Zapechelnyuk (2020), study environments in which a third party designs an information structure and a "sender" exerts costly effort to influence the distribution of the underlying state. Technically, our setting differs from this class of models. A key distinction is that the agent's ex-post incentive to exert effort depends on the signal distributions induced by both the target effort and off-path

effort choices. Whereas Kamenica and Gentzkow (2011) formulate information design in terms of beliefs or posteriors, we formulate the problem in likelihood-ratio space and characterize the solution using a geometric game. Because each effort induces its own prior and the contracting problem links signal realizations through the agent's incentive constraints, the problem is not posterior-separable, and the standard concavification approach does not apply. The likelihood-ratio approach may also apply to other information-design problems in which both on-path and off-path beliefs matter.

The rest of the paper is organized as follows. Section 2 introduces the basic model. Section 3 illustrates how the agent can use indicator design to implement the first-best effort and extract the entire surplus when the underlying performance technology perfectly reveals effort. Section 4 develops the geometric interpretation of the indicator-choice game between the principal and the agent for general performance technologies. Section 5 characterizes when the agent can extract the entire surplus even when the performance technology is noisy, and studies how equilibrium payoffs and total surplus vary with the informativeness of the technology. Section 6 shows that threshold signals are optimal for single-index performance technologies, in which effort enters the output distribution through a scalar index. Section 7 concludes.

2 Model

Our general model builds upon the textbook moral hazard framework. Consider a principal employing an agent to perform a task whose output is represented by $x \in X$, where X is finite. The agent chooses effort $e \in E = \{e_1, \cdots, e_m\}$ to perform the task, where E is finite. The agent's effort choice induces a probability distribution $f(x|e)$ over the outcome space X, where $\sum_x f(x|e) = 1, \forall e \in E$. We refer to $f(\cdot|\cdot)$ as the performance technology. Effort is costly to the agent; the cost of exerting effort e is given by $c(e)$, where $c : E \to \mathbb{R}_+$.

Throughout the analysis, we assume that $e_1 \in E$ represents the unique effort with the lowest cost; for simplicity, let $c(e_1) = 0$.[4] The principal's gross expected payoff from effort e is given by $g(e)$, where $g : E \to \mathbb{R}$. As in the standard moral-hazard model, the principal cannot observe the agent's effort and thus cannot offer a contract contingent on

[4]If several efforts had zero cost, the principal's outside option would be the largest value of g among them, and the results would be unchanged, with condition (i) of Proposition 8 imposed on every zero-cost effort.

effort; he can only offer contracts contingent on observable outcomes.

Our departure from the textbook moral-hazard model is that the agent can influence the principal's information about output by choosing an information structure (S, π). Here, S is a signal space and $\pi(\cdot|x) : X \to \Delta(S)$ is a stochastic mapping from the output space X to the signal space, where $\sum_s \pi(s|x) = 1,\ \forall x \in X$. The principal observes only the signal $s \in S$ generated by this information structure and can thus offer a contract contingent solely on the realized signal. Therefore, the principal's choice of contract can be represented by a function $w : S \to \mathbb{R}_+$, where $w(s)$ denotes the wage paid to the agent when signal s is realized. One interpretation of this contractual restriction is that the task output is not observable to the principal, but the agent can verifiably disclose information about it by committing to a disclosure policy of this form. An alternative interpretation is that the output is observable, but the agent can choose the performance measure on which she will be evaluated and compensated.

We assume that the agent enjoys limited liability, that is, the contract offered by the principal guarantees a non-negative wage regardless of the effort she exerts. The principal's payoff u_P equals the gross expected payoff from the implemented effort minus the wage paid to the agent. The agent's payoff u_A is equal to the wage she receives from the principal minus the cost of her effort. Both the principal and the agent are assumed to maximize expected utility. Notice that the principal can always implement the zero-cost effort e_1, by offering $w(s) = 0,\ \forall s \in S$. Therefore, his outside option is to implement e_1 and obtain $\underline{u}_P = g(e_1)$.[5]

The timing of the game is as follows:

- The agent chooses an information structure (S, π). To ease exposition, we assume that the signal space S is finite and simply represent the information structure by π. [6]

- Upon observing the information structure (S, π) chosen by the agent, the principal offers a contract $w : S \to \mathbb{R}_+$ contingent on the realized signal.

- After observing the contract w offered by the principal (and the information struc-

[5] If the principal also has an exogenous outside option $\bar{u}$ from declining to contract, his effective reservation payoff is $\max\{g(e_1), \bar{u}\}$, since the contract $w \equiv 0$ already guarantees $g(e_1)$. We maintain $\bar{u} \leq g(e_1)$, so the principal's reservation payoff is $g(e_1)$ throughout.

[6] This restriction is without loss of generality: binary information structures suffice by Proposition 3, and Proposition 4 shows that allowing richer signal spaces does not change the value of the agent's design problem.

ture (S, π) she has chosen), the agent chooses how much effort e to exert.

- Given the effort e chosen by the agent, the output x is realized according to $f(x|e)$ and then signal $s \in S$ is realized according to $\pi(s|x)$.

- Payoffs are realized: the agent's payoff is $u_A = w(s) - c(e)$, and the principal's payoff is $u_P = g(e) - w(s)$.

To summarize, the game is played sequentially in three stages. In the first stage, the agent chooses the information structure π, which determines how much the principal learns about the realized output. In the second stage, the principal offers the agent a contract w contingent on the signal realization. In the final stage, the agent chooses her effort e.

The equilibrium concept is the standard subgame perfect equilibrium (SPE). The agent's strategy $(\pi, \sigma_e(\pi, w))$ consists of a choice of information structure π and a choice of effort $\sigma_e(\pi, w)$ as a function of π and the contract w offered by the principal. The principal's strategy $\sigma_w(\pi)$ is the contract he offers to the agent as a function of the information structure π chosen by the agent.

Definition 1. An SPE of the game $\sigma^* = (\pi^*, \sigma_e^*(\pi, w), \sigma_w^*(\pi))$ consists of a strategy for the agent $(\pi^*, \sigma_e^*(\pi, w))$ and a strategy for the principal $\sigma_w^*(\pi)$ such that:

- $\sigma_e^*(\pi, w)$ maximizes the agent's expected utility for every information structure π and contract w chosen by the principal.

- $\sigma_w^*(\pi)$ maximizes the principal's expected utility for every information structure π chosen by the agent, given the agent's equilibrium effort strategy $\sigma_e^*(\pi, w)$.

- π^* maximizes the agent's expected utility given the principal's equilibrium strategy $\sigma_w^*(\pi)$ and her own equilibrium effort strategy $\sigma_e^*(\pi, w)$.

Throughout, we resolve indifference as follows. In the final stage, when the agent is indifferent among efforts, she chooses the effort the contract is intended to implement whenever that effort is among her optimal choices. In the second stage, when the principal is indifferent among contracts implementing different efforts, he chooses a contract implementing the effort that the information structure is intended to implement, whenever such a contract is among his optimal choices.[7] Any remaining indifference is resolved in favor

[7]Formally, intentions can be made explicit by having the agent announce a target effort alongside the information structure and the principal announce a recommended effort alongside the contract. When

of the other player. Since the agent designs the information structure to implement the effort that maximizes her own payoff, these tie-breaking rules select the agent-preferred subgame-perfect equilibrium. The first is the usual compliance convention at binding incentive constraints in contract theory, while the second parallels sender-preferred selection in the persuasion literature.

Given an information structure (S, π), let $p(\cdot|e) \in \Delta(S)$ denote the induced distribution of signals under effort e. For each $e \in E$ and $s \in S$, $p(s|e) = \sum_x f(x|e)\pi(s|x)$. Conditional on a given information structure (S, π), both the principal and the agent evaluate expected payoffs using the stochastic mapping $p(\cdot|\cdot)$.

We now formulate the problems solved by the principal and the agent, beginning with the agent's choice of effort. The agent's expected utility is $U_A = \sum_s p(s|e)w(s) - c(e)$. Therefore, the agent's problem in the last stage of the game (where π and w have already been chosen in the preceding stages) is

$$\max_e \sum_s p(s|e)w(s) - c(e). \tag{1}$$

The principal's expected utility is $U_P = g(e) - \sum_s p(s|e)w(s)$ if he chooses to implement effort e. For a given desired effort level e, the principal chooses a wage schedule that solves

$$\begin{aligned} \min_w \ & \sum_s p(s|e)w(s) \text{ s.t.} \\ & \sum_s p(s|e)w(s) - c(e) \geq \sum_s p(s|\hat{e})w(s) - c(\hat{e}), \ \forall \hat{e} \in E, \\ & w(s) \geq 0, \ \forall s \in S. \end{aligned} \tag{2}$$

The constraints represent the agent's incentive-compatibility condition and a set of limited liability constraints. Notice that we have not imposed a participation constraint for the agent. This is because the assumption $c(e_1) = 0$ together with limited liability implies that setting $e = e_1$ guarantees a non-negative payoff for the agent. Let $W(e, \pi)$ denote the optimal value of the principal's problem in (2). Thus, $W(e, \pi)$ is the minimum expected wage required to implement effort e under information structure π.

the principal is indifferent among optimal contracts, he selects one whose recommended effort coincides with the agent's announced target whenever such a contract is optimal; when the agent is indifferent among optimal efforts, she selects the principal's recommended effort whenever it is among them. This formulation is equivalent to the tie-breaking convention used in the analysis and changes no result.

Given $W(e, \pi)$, the principal's problem is to choose an effort level e to maximize his expected utility, that is,

$$\max_{e} \; g(e) - W(e, \pi). \tag{3}$$

It is possible to express the above as an incentive-compatibility constraint and thus write the agent's problem in the first stage of the game as

$$\max_{e,\pi} \; W(e, \pi) - c(e)$$

subject to the principal's incentive-compatibility constraint

$$g(e) - W(e, \pi) \geq g(\hat{e}) - W(\hat{e}, \pi), \; \forall \hat{e} \in E. \tag{4}$$

2.1 Remarks on the Environment

It is useful to discuss various interpretations of the model as well as our key assumptions.

Performance Measure as Information Structure

In the textbook model of moral hazard, the principal cannot observe the agent's effort. He therefore relies on an imperfect signal of effort to incentivize the agent. If the principal can observe output, he offers an output-contingent contract, making output the performance measure for the agent. In our model, the principal does not observe output but does observe a signal that may be correlated with effort, and he offers a signal-contingent contract. As a result, the signal becomes the relevant performance measure for the agent. By choosing the information structure, the agent influences the observables that determine her compensation. We therefore interpret this choice as selecting her performance measure.

We assume that the principal cannot offer output-contingent contracts. One interpretation is that the principal cannot observe output directly. Another is that the restriction arises from negotiations over the performance measures included in the contract.

Commitment

We assume that the agent commits to an information structure in the first stage of the game. Our interpretation of the information structure as a contractual performance measure provides a natural justification of the commitment assumption. When signing a con-

tract, both parties are aware of and agree on the probabilistic relationship between the performance measure and the underlying output. Because the performance measure is written into the contract, commitment is a natural feature of the environment rather than an extra behavioral assumption.

On-Path and Off-Path Efforts

The agent commits to the indicator but not to her effort, and nothing requires the principal to implement the effort she has in mind. Given the indicator, he offers whichever contract maximizes his own payoff. This asymmetry in commitment is what gives the agent her rent. If she could also commit to an effort, the principal would have no need to provide incentives and would offer no wage, so she would commit to the zero-cost effort and obtain nothing. The indicator must therefore make the target attractive to the principal and every alternative unattractive, which is why the analysis turns on the likelihood ratios of signals under efforts that are never taken in equilibrium. This is the pattern in the writers' negotiation described in the introduction: the parties bargained over an enforceable performance measure, while the effort underlying it remained unverifiable.

Comparison to the Literature on Bayesian Persuasion

As in Bayesian persuasion, the problem can be expressed in terms of distributions of posterior beliefs. Two features, however, distinguish our environment from the standard persuasion framework and explain why the usual concavification argument does not apply. First, there is no single prior governing the problem. The distribution of output depends on effort, and effort is chosen in the final stage, after the information structure is designed and the contract is offered. Each effort e therefore induces its own prior $f(\cdot \mid e)$, so a given information structure generates a family of posterior distributions, one for each effort, with Bayes plausibility holding separately for each prior. These distributions are nevertheless linked through their supports: once the support under one effort is known, the supports under all other efforts are determined.

Second, the principal's action is not chosen as a function of the realized posterior. The principal commits to a wage schedule before the signal is realized, and the wage associated with a signal is determined by the entire information structure. In particular, incentives depend on the likelihood of that signal under both the target effort and off-path effort choices. Moreover, the contracting problem has a homogeneity property: a proportional

reduction in a signal's probability, holding its likelihood ratios fixed, can be offset by an inverse increase in its wage, leaving both incentive constraints and expected payments unchanged. As a result, absolute signal probabilities drop out of continuation payoffs. Information structures that generate the same support of posteriors, and hence the same likelihood-ratio vectors, are payoff-equivalent even if they assign different probabilities to those posteriors. This differs from the posterior-separable formulation underlying standard Bayesian persuasion, in which the value of an information structure is an expectation of a fixed value function over posterior beliefs and therefore depends on the probabilities assigned to those beliefs. The concavification approach of Kamenica and Gentzkow (2011) is thus not directly applicable. Section 4 develops the alternative: continuation payoffs depend on the information structure through the likelihood-ratio vectors generated by the support of its posterior distributions, and the analysis proceeds directly in this likelihood-ratio space.

3 Perfect Performance Technologies

In this section, we study equilibrium outcomes when the performance technology fully reveals the agent's effort. In this benchmark, the indicator can be any garbling of effort, so the agent's problem reduces to unconstrained information design over her own action. This is the simplest setting in which the main forces of the model appear: the role of off-path likelihood ratios and an indicator that leaves the agent indifferent across effort levels, thereby making the principal a residual claimant on total surplus. The remainder of the paper asks how these results change when the technology is imperfect and the indicator must be a garbling of a fixed f.

Recall that the performance technology specifies the mapping from the agent's effort to the probability distribution over the outcome space X. Here, we assume this mapping fully reveals the agent's effort.

Assumption 1. *The performance technology is* perfect*: $X = E$ and $f(x|e) = 1$ if and only if $x = e$.*

Under a perfect performance technology, for any information structure (S, π) chosen by the agent, the induced mapping from effort to signals given by $p(s|e)$ satisfies $p(s|e) = \pi(s|e)$. Even though the underlying performance technology is perfect, the agent may choose information structures that are not perfect, and her wage will depend

on the chosen signal. Proposition 1 shows that the agent prefers her wage to depend on an imperfect signal of effort rather than on the perfectly revealing performance measure, without sacrificing total surplus.

Let e_{FB} denote the surplus-maximizing, first-best effort level that solves

$$e_{FB} \in \arg\max_{e \in E} \{g(e) - c(e)\}$$

We then have the following proposition:

Proposition 1. *Suppose the performance technology is perfect as in Assumption 1. Then there exists an information structure (S, π) that implements e_{FB}, and the agent obtains all the surplus so that $u_A = g(e_{FB}) - c(e_{FB}) - g(e_1)$.*

Why can the agent design an indicator that transfers the entire surplus without distorting effort? The key property is indifference. Suppose the indicator is such that the cheapest contract implementing any positive-cost effort leaves the agent indifferent among all efforts implementable under it. Her payoff under such a contract is then a constant K, regardless of which effort she chooses, so the expected wage required to implement effort e is $c(e) + K$. The principal's payoff from implementing e is therefore $g(e) - c(e) - K$, that is, total surplus minus a fixed rent. This makes the principal a residual claimant on surplus. When the performance measure is exogenous, the information rent required to implement an effort generally varies with the effort level. The principal's preferred effort therefore need not maximize total surplus. Here, by contrast, the rent is the same constant K regardless of the effort implemented. Maximizing the principal's payoff is therefore equivalent to maximizing total surplus, so he implements e_{FB}.

The only limit on the agent's rent is the principal's outside option. He can always implement the zero-cost effort e_1 at zero wage, so K cannot exceed $\big[g(e_{FB}) - c(e_{FB})\big] - g(e_1)$. The agent-optimal indicator sets K exactly at this upper bound. At this point, the principal is indifferent between implementing e_{FB} and taking his fallback, while the agent is indifferent among all efforts implementable under the indicator. The tie-breaking conventions of Section 2 resolve both indifferences: the principal chooses a contract implementing the effort the indicator is intended to implement, here e_{FB}, and the agent complies with that intended effort. Thus, e_{FB} is implemented and the agent captures the entire surplus.

The proof of Proposition 1 constructs precisely such an indicator, namely the pass/fail device described in the introduction. To ease notation, normalize $g(e_1) = 0$ and consider an information structure with $S = \{L, H\}$, where $(\pi(H \mid e))$ is defined as

[8]

$$\pi(H|e) = \begin{cases} 1 - \frac{c(e_{FB}) - c(e)}{g(e_{FB})} & \text{if } c(e) < c(e_{FB}) \\ 1 & \text{if } c(e) \geq c(e_{FB}) \end{cases} . \tag{5}$$

The construction delivers indifference through two features of (5). First, for efforts satisfying $c(e) \leq c(e_{FB})$, the probability of the high signal rises linearly with cost at slope $1/g(e_{FB})$. Thus, a bonus of $g(e_{FB})$ paid following H compensates each additional unit of effort cost with exactly one unit of expected pay. The agent is therefore indifferent among all such efforts, while any smaller bonus compensates effort cost less than one-for-one and induces only the zero-cost effort. The principal consequently cannot implement any positive-cost effort with a wage spread below $g(e_{FB})$, and it is optimal to pay zero following L. Second, for efforts with $c(e) > c(e_{FB})$, the high signal already occurs with probability one, so these efforts generate the same signal distribution as e_{FB} and cannot be implemented at any wage. With the wage spread equal to $g(e_{FB})$, the expected wage required to implement an effort $\hat{e}$ is $c(\hat{e}) + \big[g(e_{FB}) - c(e_{FB})\big]$, that is, its cost plus a rent equal to the first-best surplus. The principal's payoff is therefore $\big[g(\hat{e}) - c(\hat{e})\big] - \big[g(e_{FB}) - c(e_{FB})\big]$. He obtains his outside-option payoff of zero from e_{FB}, and no other implementable effort gives him a higher payoff. The tie-breaking convention of Section 2 therefore selects e_{FB}. The principal earns zero, while the agent captures the entire surplus. The formal argument is provided in Appendix B.

The constructed indicator is deliberately coarse: it is less informative about effort than the underlying performance outcome X. This loss of information tightens the principal's incentive-compatibility constraints and is the source of the agent's rent. Yet because the indicator makes the principal a residual claimant on surplus, this rent comes at no efficiency cost: rent extraction and surplus maximization coincide.

Since information structures in our model are garblings of the underlying performance metric, the perfect-technology benchmark maximizes the set of indicators available to the agent: there is no underlying information friction. In most cases of interest, performance metrics are imperfect signals of hidden effort, and the requirement that the indicator be a garbling of a fixed technology may rule out the indifference construction.

[8] Note that $\pi(H|e) \in [0, 1]$: it is increasing in $c(e)$, equals 1 at $e = e_{FB}$, and at its lowest, $e = e_1$, equals $1 - \dfrac{c(e_{FB})}{g(e_{FB})} \geq 0$, where the inequality is $g(e_{FB}) - c(e_{FB}) \geq g(e_1) - c(e_1) = 0$ (first-best surplus weakly exceeds e_1's).

4 A Geometric Analysis of the Game

We now characterize the equilibrium outcomes of the game. We first describe the set of effort levels that are implementable under some information structure. We then establish a "coarseness" result: it is without loss of generality to restrict the agent's choice of information structures to binary ones, in which the signal space contains only two points. In the next section, we use this binary-signal result to characterize when the agent can implement a target effort while extracting the entire surplus, and to derive sufficient conditions stated in terms of the primitives.

While it is possible to work with zero-probability events and define likelihood ratios by specifying how division by zero is treated, we impose the following assumption to avoid these complications:

Assumption 2. *The performance technology has full support; that is,* $\forall x \in X, \forall e \in E, f(x|e) > 0$.

This assumption ensures that all likelihood ratios introduced below are well defined.

Implementable Effort. To characterize implementable effort levels, we use backward induction and recast the principal's problem geometrically. For any target effort e^* the likelihood ratios of each signal under alternative efforts relative to e^* are sufficient statistics for the principal's contracting problem. The information structure determines the feasible set of likelihood-ratio vectors, and the principal's compensation choice corresponds to selecting a point from this set

More formally, we define an implementable effort level e^* as follows:

Definition 2. An effort level e^* is *implementable* if there exists an information structure (S, π) such that e^* solves the principal's problem (3).

Given this definition, an implementable effort level e^* must satisfy the incentive-compatibility constraints in (4) and (1) for both the principal and the agent. In particular, the agent's incentive-compatibility constraint must hold for all possible histories, including those following any deviation by the principal to a contract that implements an alternative effort level.

Let e^* denote the effort level that the agent seeks to implement in the first stage of the game. To motivate the relevance of likelihood ratios, consider the agent's interim incentive-compatibility constraint when the principal only pays the agent following a

single signal realization s:

$$p\left(s|e^*\right)w\left(s\right)-c\left(e^*\right)\geq p\left(s|\hat{e}\right)w\left(s\right)-c\left(\hat{e}\right),\forall\hat{e}\in E.$$

These constraints imply that for any effort level $\hat{e}$ for which signal s is less likely than under e^*, i.e., $p\left(s|\hat{e}\right)<p\left(s|e^*\right)$, the wage must satisfy

$$w\left(s\right)\geq\frac{c\left(e^*\right)-c\left(\hat{e}\right)}{p\left(s|e^*\right)-p\left(s|\hat{e}\right)},\tag{6}$$

Conversely, if signal s is more likely under $\hat{e}$ than under e^*, then the wage must satisfy

$$\frac{c\left(\hat{e}\right)-c\left(e^*\right)}{p\left(s|\hat{e}\right)-p\left(s|e^*\right)}\geq w\left(s\right).\tag{7}$$

The first set of constraints (6) then implies that the expected wage must satisfy

$$p\left(s|e^*\right)w\left(s\right)\geq\max_{\hat{e}:p(s|e^*)>p(s|\hat{e})}\frac{c\left(e^*\right)-c\left(\hat{e}\right)}{1-\frac{p(s|\hat{e})}{p(s|e^*)}}.$$

Thus, the expected cost of implementing e^* depends on the likelihood ratio of signal s under the alternative effort relative to e^*. The second set of constraints (7) places an upper bound on the wages the principal may offer while respecting the agent's incentives. As we show below, this restriction can also be expressed in terms of likelihood ratios.

The preceding illustration assumes that the principal compensates the agent following only one single signal realization. We now consider an arbitrary wage schedule $w\left(s\right)$ chosen to implement any effort level $e_i\in E$, whether on or off the equilibrium path. We may write the expected wage paid to the agent as

$$\begin{aligned}\sum_s w\left(s\right)p\left(s|e_i\right)&=\sum_s w\left(s\right)p\left(s|e^*\right)\frac{p\left(s|e_i\right)}{p\left(s|e^*\right)}\\&=\sum_s w\left(s\right)p\left(s|e^*\right)\cdot\sum_s\frac{w\left(s\right)p\left(s|e^*\right)}{\sum_{s'}w\left(s'\right)p\left(s'|e^*\right)}\frac{p\left(s|e_i\right)}{p\left(s|e^*\right)}\\&=\sum_s w\left(s\right)p\left(s|e^*\right)\cdot\sum_s\alpha_s\frac{p\left(s|e_i\right)}{p\left(s|e^*\right)},\end{aligned}$$

where $\alpha_s=\frac{w(s)p(s|e^*)}{\sum_{s'}w(s')p(s'|e^*)}\geq 0$ and $\sum_s\alpha_s=1$. Since the weights α_s do not depend on

e_i, we may write the agent's interim incentive-compatibility constraint as

$$\sum_s w(s)\, p(s|e^*) \cdot \sum_s \alpha_s \frac{p(s|e_i)}{p(s|e^*)} - c(e_i) \geq \sum_s w(s)\, p(s|e^*) \cdot \sum_s \alpha_s \frac{p(s|e_j)}{p(s|e^*)} - c(e_j).$$

Therefore, if we define $\ell_i = 1 - \sum_s \alpha_s p(s|e_i)/p(s|e^*)$, this incentive-compatibility constraint may be written as

$$\sum_s w(s)\, p(s|e^*) \cdot [\ell_j - \ell_i] \geq c(e_i) - c(e_j), \forall j = 1, \cdots, |E|. \tag{8}$$

The constraints depend on payments only through the products $w(s)p(s \mid e^*)$. Rescaling a signal's probability by $\lambda > 0$, while holding its likelihood ratios fixed and dividing its wage by λ, therefore leaves every incentive constraint and every expected payment unchanged. Absolute signal frequencies are thus irrelevant, and only the likelihood ratios, which capture how informative each signal is about deviations, matter. We use this freedom in Section 4.1, when choosing how often the paid signal occurs.

The principal's choice of contract $w(s)$ can be decomposed into a choice of the scalar $\sum_s w(s)\, p(s|e^*)$ and a choice of the weights $\{\alpha_s\}$, or equivalently, the likelihood ratios $\{\ell_j\}$. In other words, if we represent the likelihood-ratio profile by the vector

$$\mathbf{l} = \left(\ell_1, \cdots, \ell_{|E|}\right),$$

then $\mathbf{l} \in \text{conv}\left(\left\{\left(1 - \frac{p(s|e_1)}{p(s|e^*)}, \cdots, 1 - \frac{p\left(s|e_{|E|}\right)}{p(s|e^*)}\right)\right\}_{s \in S}\right) = \text{co}(\mathbf{p})$. Thus, the principal's choice can be summarized by the overall level of compensation $\sum_s p(s|e^*)\, w(s) = \overline{w}$, together with an element of the set $\text{co}(\mathbf{p})$. We thus have the following lemma:

Lemma 1. *Consider an implementable effort e^* and its associated information structure $\mathbf{p} = \{p(\cdot|\cdot)\}$. Then the cost to the principal from implementing any effort e_i is given by*

$$W(e_i, \mathbf{p}) = \min_{\mathbf{l} \in co(\mathbf{p}) \cap \Omega_i} (1 - \ell_i) \cdot \max\left\{0, \max_{j:\ell_j \geq \ell_i} \frac{c(e_i) - c(e_j)}{\ell_j - \ell_i}\right\}, \tag{9}$$

where

$$\Omega_i = \left\{\mathbf{l} \in \mathbb{R}^{|E|} : \max\left\{0, \max_{j:\ell_j \geq \ell_i} \frac{c(e_i) - c(e_j)}{\ell_j - \ell_i}\right\} \leq \min_{j:\ell_j < \ell_i} \frac{c(e_i) - c(e_j)}{\ell_j - \ell_i}\right\}. \tag{10}$$

The cone Ω_i collects the likelihood-ratio profiles at which a single nonnegative wage

can satisfy all of the agent's incentive constraints at once[9]: large enough to deter every effort that makes the paid signal less likely than e_i (the lower bounds in (8)), yet small enough not to tempt any effort that makes it more likely (the upper bounds). Implementability of e_i is precisely the requirement that these lower and upper bounds be compatible.

Lemma 1 reduces the principal's problem to choosing an effort level and a point in the convex hull of likelihood ratio vectors, specifically an element of $\mathrm{co}\,(\mathbf{p})$ that lies in the convex cone defined in (10). Note that not all likelihood-ratio profiles $\mathbf{l}$ are feasible in the sense that there may be no compensation scheme that satisfies the agent's interim incentive-compatibility constraint. As in the motivating example above, the likelihood ratios must permit a wage $w(s)$ that lies between the lower and upper bounds in (10). The convex cone Ω_i defines precisely the set of likelihood-ratio profiles that admit incentive-compatible compensation schemes.

By choosing an information structure, the agent determines the convex hull $\mathrm{co}\,(\mathbf{p})$. The principal then selects a point in the intersection of $\mathrm{co}\,(\mathbf{p})$ and the convex cone Ω_i. In particular, continuation play depends on the information structure only through $\mathrm{co}\,(\mathbf{p})$: any two structures inducing the same convex hull are equivalent from the second stage onward. This observation by itself does not make the analysis more tractable, since it does not directly characterize the set of feasible convex hulls $\mathrm{co}(\mathbf{p})$. The following proposition provides such a characterization. Note that, analogous to $\mathrm{co}\,(\mathbf{p})$, the set $\mathrm{co}\,(\mathbf{f})$ is the convex hull of the points

$$a_x = \left(1 - \frac{f(x \mid e_1)}{f(x \mid e^*)}, \ldots, 1 - \frac{f(x \mid e_{|E|})}{f(x \mid e^*)}\right), \qquad x \in X,$$

where a_x is the likelihood-ratio vector generated by output x. Thus, $\mathrm{co}\,(\mathbf{f}) = \mathrm{conv}\{a_x\}_{x \in X}$.

Proposition 2. *For any information structure (S, π) with $|S| < \infty$, its associated $\mathrm{co}\,(\mathbf{p})$ is a subset of $\mathrm{co}\,(\mathbf{f})$ and contains the origin $\mathbf{0} = (0, \cdots, 0)$ in its relative interior. Additionally, for any closed convex subset $C \subseteq \mathrm{co}\,(\mathbf{f})$ that contains the origin in its relative interior and has a finite set of extreme points Z, there exists an information structure (S, π) with $|S| \leq |Z|+1$ such that $\mathrm{co}\,(\mathbf{p}) = C$.*

[9]We adopt the extended-real convention $\frac{a}{0} = +\infty$ for $a > 0$ and $\frac{a}{0} = -\infty$ for $a < 0$. Thus, if some e_j has $\ell_j = \ell_i$ and $c(e_j) < c(e_i)$, the maximand is $+\infty$ and e_i is not implementable at that $\mathbf{l}$, since no finite wage can deter a lower-cost deviation with the same likelihood coordinate. By contrast, if $c(e_j) > c(e_i)$, the corresponding term is $-\infty$ and therefore drops out. The case $\ell_j = \ell_i$ with $c(e_j) = c(e_i)$ is excluded by convention because (8) then holds trivially.

Each extreme point of co $(\mathbf{p})$ must be interpreted as the likelihood-ratio vector generated by a signal realization. If we restrict attention to information structures with as many realizations as extreme points, the latter must therefore determine a distribution of posteriors that satisfies Bayes plausibility. Each extreme point z corresponds to a posterior $\mu_z(\cdot)$ under e^*, which specifies the coefficients in a convex representation of that point in terms of the likelihood-ratio vectors associated with $x \in X$. At the same time, since the origin lies in the relative interior of co $(\mathbf{p})$, it can be written as a convex combination of these extreme points, with coefficients given by the signal (equivalently, posterior) probabilities $p(s|e^*) > 0$. However, Bayes plausibility is not automatic; it requires that the mixture of extreme-point posteriors reproduce the prior,

$$\sum_z p(z|e^*)\mu_z(\cdot) = f(\cdot|e^*).$$

For a given collection of extreme points, there need not exist probabilities satisfying this identity. The induced mixture may place too much or too little probability on some states relative to the prior. In either case, the extreme points alone cannot generate a feasible information structure. Introducing an additional signal realization corresponding to an interior point of co $(\mathbf{p})$ provides the flexibility needed to absorb these discrepancies and restore Bayes plausibility.

Example 1. Suppose $X = \{x_1, x_2, x_3, x_4\}$ and $E = \{e_1, e_2, e_3\}$, with $e^* = e_3$. The performance technology is

$$f = \begin{bmatrix} 0.02 & 0.54 & 0.08 & 0.36 \\ 0.18 & 0.06 & 0.04 & 0.72 \\ 0.2 & 0.6 & 0.02 & 0.18 \end{bmatrix},$$

where $f_{ij} = f(x_j|e_i)$. The induced likelihood-ratio points $\{a_1, a_2, a_3, a_4\}$ are illustrated in Figure 2, where co $(\mathbf{f})$ = conv (a_1, a_2, a_3, a_4) is the grey shaded area. Consider $\mathbf{l}^H = (0.3, 0.3)$ and $\mathbf{l}^L = -4\mathbf{l}^H = (-1.2, -1.2)$, and let co $(\mathbf{p})$ = conv $\left(\left\{\mathbf{l}^H, \mathbf{l}^L\right\}\right)$. With a binary signal $S = \{H, L\}$, any implementable pair must satisfy $\mathbf{l}^L = -\frac{p(H|e^*)}{1-p(H|e^*)}\mathbf{l}^H$, which requires $p(H|e^*) = 0.8$. However, it can be shown that any posterior μ_H generating $\mathbf{l}^H$ implies $p(H|e^*)\mu_H(x) = 0.8\mu_H(x) \geq f(x|e^*)$ for some $x \in X$, violating Bayes plausibility. Hence, no binary information structure can implement co $(\mathbf{p})$. In fact, fixing $\mathbf{l}^H$, the furthest point implementable with a binary information structure is

$\mathbf{l}^{\max} = -1.25\mathbf{l}^H = (-0.375, -0.375)$. Equivalently, $\mathbf{l}^{\max}$ corresponds to the largest signal probability $p(H \mid e^*)$ consistent with the fixed $\mathbf{l}^H$: raising $p(H|e^*)$ pushes $\mathbf{l}^L$ outward until the feasibility constraint $p(H \mid e^*)\mu_H(x) \leq f(x|e^*)$ first binds at x_1; beyond $\mathbf{l}^{\max}$ no binary structure is feasible.[10]

Now consider an additional signal, so that $S = \{H, L, 0\}$. Fix posteriors $\mu_H = (0.36, 0.56, 0, 0.08)$ and $\mu_L = (0, 0.32, 0.276, 0.404)$, which generate $\mathbf{l}^H$ and $\mathbf{l}^L$, respectively. Choose signal probabilities under e^* as $p(H|e^*) = 0.2$, $p(L|e^*) = 0.05$, and $p(0|e^*) = 0.75$. Define the residual posterior by

$$\mu_0 = \frac{f(\cdot|e^*) - p(H|e^*)\mu_H - p(L|e^*)\mu_L}{p(0|e^*)} \approx (0.1707,\ 0.6293,\ 0.0083,\ 0.1917).$$

Then Bayes plausibility holds and μ_0 generates the origin, so $\mathbf{l}^0 = \mathbf{0}$. Therefore, $\text{co}\,(\mathbf{p}) = \text{conv}\left\{\mathbf{l}^H, \mathbf{l}^L, \mathbf{l}^0\right\}$. This example illustrates that, with two signals, feasibility bounds on $p(s|e^*)$ prevent the implementation of distant endpoints, while a third signal allows the extreme signals to be made sufficiently rare and the residual mass to be absorbed by an interior posterior. This does not conflict with the coarseness results below. By Proposition 3, two signals always suffice to implement a given target effort at a given expected wage. What two signals cannot generally do is generate an arbitrary convex hull $\text{co}\,(\mathbf{p})$. This is why the construction in the proof of Proposition 2 uses a third, residual signal, as in this example.

This proof sketch and Example 1 also clarify a key distinction between our framework and standard Bayesian persuasion. In persuasion models, the entire distribution of posteriors is payoff-relevant. In contrast, in our setting, only the support of the induced distribution matters. The likelihood ratios across effort levels are the only objects that affect incentives in the subsequent moral-hazard stage. In particular, for any signal realization $s \in S$ and effort level $e \in E$,

$$\frac{p(s|e)}{p(s|e^*)} = \mathbb{E}_{\mu_s}\left[\frac{f(x|e)}{f(x|e^*)}\right].$$

Thus, information structures that generate the same support of posteriors induce the same likelihood-ratio vectors and are equivalent from an incentive perspective, even if they

[10] Letting $\pi(H|x) = p(H|e^*)\mu_H(x)/f(x|e^*)$ denote the probability that output x generates signal H, this constraint is equivalently $\pi(H|x) \leq 1$; at $\mathbf{l}^{\max}$ it binds at x_1, so output x_1 generates signal H with probability one.

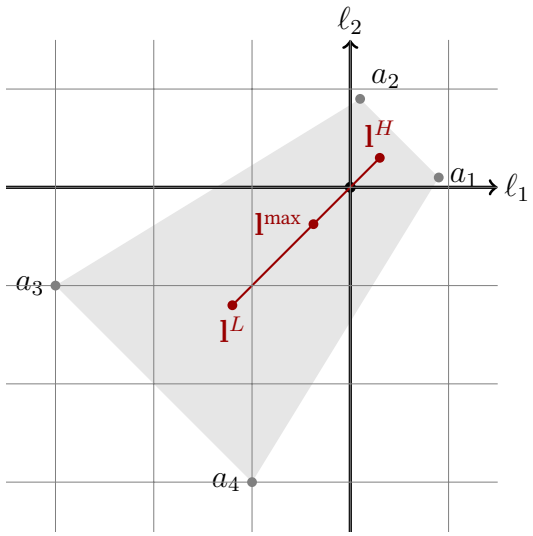

Figure 2: Geometric representation of Proposition 2. The points $\mathbf{l}^H$, $\mathbf{l}^L$ correspond to the signals H and L, respectively. Fixing $\mathbf{l}^H$, the point $\mathbf{l}^{\max}$ is the furthest endpoint implementable with a binary signal.

assign different probabilities to those posteriors.

Proposition 2 implies that any convex subset of $\text{co}(\mathbf{f})$ that contains the origin in its relative interior can be selected by the agent as $\text{co}(\mathbf{p})$. We may therefore represent the agent's choice of information structure as the choice of a convex subset of $\text{co}(\mathbf{f})$ containing the origin. In this sense, the agent-optimal information structure that implements e^* can be represented as the equilibrium of the following game:

- **Stage 1.** The agent chooses a finite set of points $L \subseteq \text{co}(\mathbf{f})$ such that the convex hull of these points $\text{conv}(L)$ contains the origin.
- **Stage 2.** The principal chooses an effort level $e_i \in E$ and a point $\ell \in \text{conv}(L) \cap \Omega_i$ to maximize $g(e_i) - (1 - \ell_i) \cdot \max_{j:\ell_j > \ell_i} \frac{c(e_i) - c(e_j)}{\ell_j - \ell_i}$.
- **Stage 3.** The principal's optimal choice of effort in stage 2 coincides with e^*.

As Section 4.1 shows, Stage 1 reduces further: without loss, the agent chooses a single point in $\text{co}(\mathbf{f})$, implemented by a binary information structure whose likelihood-ratio set is a line segment through the origin.

The following example illustrates the construction of co $(\mathbf{f})$ and the principal's equilibrium.

Example 2. Suppose $X = \{x_1, x_2, x_3\}$ and $E = \{e_1, e_2, e_3\}$, where $c(e_1) = 0$, $c(e_2) = 0.1$, and $c(e_3) = 0.3$. The performance technology is

$$f = \begin{bmatrix} 0.35 & 0.50 & 0.15 \\ 0.05 & 0.50 & 0.45 \\ 0.10 & 0.15 & 0.75 \end{bmatrix},$$

where $f_{ij} = f(x_j|e_i)$. Moreover, the principal's gross expected payoffs are $g(e_1) = 0.8$, $g(e_2) = 1.4$, and $g(e_3) = 1.65$. Therefore, the first-best effort is e_3. Setting $e^* = e_3$, note that $1 - f(x|e_3)/f(x|e^*) = 0$, so the set co $(\mathbf{f})$ can be embedded in $\mathbb{R}^2$. The gray region in Figure 3 depicts this set, where each point a_i corresponds to $x_i \in X$.

To see the geometry of the game, consider a signal structure with four realizations $S = \{s_1, s_2, s_3, s_4\}$ and

$$\pi = \begin{bmatrix} 0.5 & 0.2 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.3 & 0.2 \\ 0.05 & 0.05 & 0.1 & 0.8 \end{bmatrix},$$

where $\pi_{ij} = \pi(s_j|x_i)$. Using $p(s|e) = \sum_{x \in X} \pi(s|x) f(x|e)$, we can construct the likelihood-ratio vectors. The corresponding convex hull co$(\mathbf{p})$ is shown as the green region in the left panel of Figure 3.

To illustrate the principal's incentives, suppose the agent fully reveals x to the principal. Then, as shown earlier, choosing a contract by the principal is equivalent to choosing a point $\mathbf{l} \in$ co $(\mathbf{f})$. If the principal wants to implement e_3, he chooses $\mathbf{l} \in$ co $(\mathbf{f})$ to minimize its cost given by

$$\max\left\{\frac{c(e_3) - c(e_2)}{\ell_2}, \frac{c(e_3) - c(e_1)}{\ell_1}\right\} = \max\left\{\frac{0.2}{\ell_2}, \frac{0.3}{\ell_1}\right\}$$

The lower contour sets of this function are translated convex cones (positive orthants), illustrated by the blue shaded region in the right panel of Figure 3. In this example, the minimum is achieved at the point $\hat{\mathbf{l}}$. If instead the principal wants to implement e_2, he chooses $\mathbf{l} \in$ co $(\mathbf{f})$ to minimize $(1 - \ell_2)\frac{0.1}{\ell_1 - \ell_2}$ when $\ell_1 \geq \ell_2$. This minimum is achieved at a_3. For e_3 to be implementable, the principal's payoff from choosing e_2, achieved at a_3, must be lower than his payoff from choosing e_3, achieved at $\hat{\mathbf{l}}$. The set of likelihood-

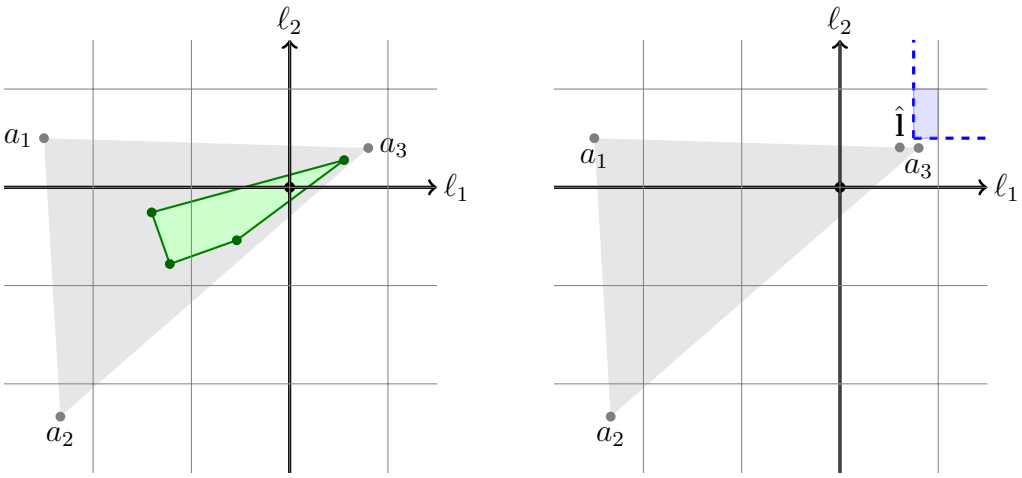

Figure 3: Geometric representation of information structures in the space of likelihood ratios. The left panel depicts an example with four signals. The right panel shows the principal's incentives under full revelation of x by the agent.

ratio points that satisfy this requirement is shown as the blue shaded region in the right panel of Figure 3. Importantly, $\hat{\mathbf{l}}$ is not contained in this set. Thus, under full revelation of x, the agent cannot induce the principal to implement e_3. One can verify that e_2 is implementable under full revelation.

4.1 Binary Information Structures

We use the geometric interpretation to show that, without loss of generality, the agent may restrict attention to information structures with at most two signals.

Proposition 3. *If e^* is implementable by some information structure (S, π) and requires an expected wage $W(e^*, \pi)$, then e^* is also implementable by a binary information structure $\left(\hat{S}, \hat{\pi}\right)$ with $|\hat{S}| = 2$ and $W(e^*, \hat{\pi}) = W(e^*, \pi)$.*

Figure 4 illustrates the intuition behind the proof. Consider an implementable effort level e^*, and suppose that its associated information structure is represented in Figure 4. The green region depicts the convex hull $\text{co}(\mathbf{p})$ corresponding to this information structure. Let point b denote the principal's optimal likelihood-ratio vector in $\text{co}(\mathbf{p})$ when implementing e^*. The red line represents the convex hull generated by the binary information structure $\hat{\pi}$; with only two signal realizations, $\text{co}(\hat{\mathbf{p}})$ must be a line segment. Since

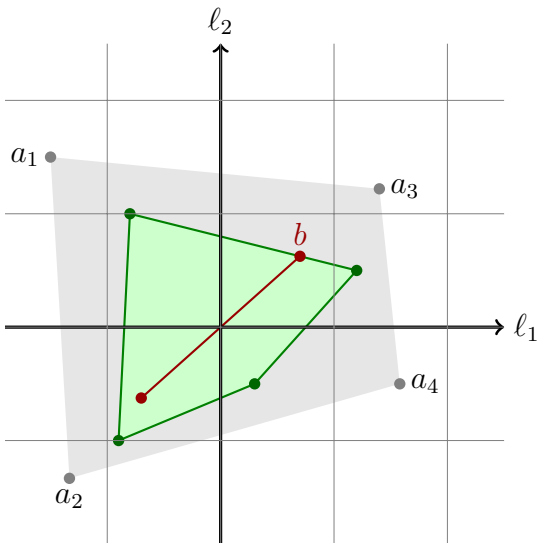

Figure 4: Geometric intuition for the construction of a binary information structure

$b \in \text{co}\,(\hat{\mathbf{p}})$ and b is also optimal under the original information structure π, the principal continues to choose b when implementing e^* under $\hat{\pi}$. Moreover, because $\text{co}\,(\hat{\mathbf{p}}) \subseteq \text{co}\,(\mathbf{p})$, the principal's minimized cost for any deviation $e_i \neq e^*$ under $\hat{\pi}$ must be weakly higher than under π. Consequently, e^* remains implementable under $\hat{\pi}$, and it yields the same outcome as under the original information structure.

Proposition 3 therefore allows us to restrict attention to binary information structures when characterizing agent-optimal designs. We now apply this result to characterize the agent-optimal information structure in Example 2.

Example 2 (Continued). Recall Example 2, in which we argued that under full revelation the principal chooses to implement e_2, and this choice is inefficient since total surplus under e_2 is lower than under e_3.

It remains to determine whether e_3 is implementable and how well the agent can do by choosing an information structure. By Proposition 3, we may restrict attention to information structures with only two signal realizations. Because $c(e_1) = 0$, the principal can implement e_1 at zero wage and guarantee a payoff of $g(e_1) = 0.8$.

Now consider the point $\mathbf{l}^*$ satisfying

$$\frac{0.2}{\hat{\ell}_2} = \frac{0.3}{\hat{\ell}_1} = 0.85$$

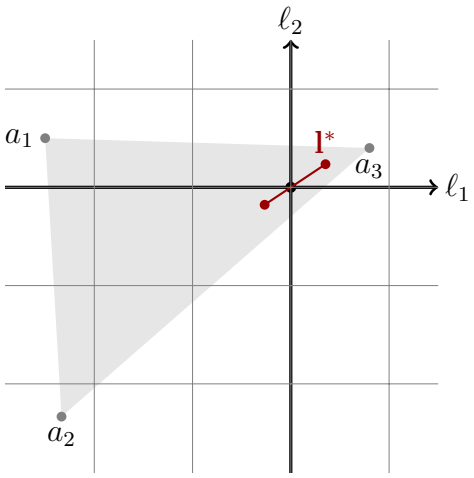

Figure 5: Agent-optimal information structure in Example 2

At this point, the principal's payoff from implementing each effort level is given by

$$\begin{aligned} e_3 : 1.65 - \max\left\{\frac{0.2}{\ell_2^*}, \frac{0.3}{\ell_1^*}\right\} &= 0.8 \\ e_2 : 1.40 - (1-\ell_2^*)\cdot\frac{0.1}{\ell_1^*-\ell_2^*} &= 1.40 - 0.65 = 0.75 \\ e_1 : 0.8 - 0 &= 0.8 \end{aligned}$$

At $\mathbf{l}^*$, the principal's payoff is maximized by choosing e_3. If $\mathbf{l}^* \in \text{co}(\mathbf{f})$, there exists a binary information structure whose convex hull contains $\mathbf{l}^*$, so e_3 may be implementable. In this example, $\mathbf{l}^*$ indeed lies in $\text{co}(\mathbf{f})$. Figure 5 illustrates the point $\mathbf{l}^*$ together with a two-point information structure that implements it (the red line passing through $\mathbf{l}^*$). Moreover, one can verify that under this information structure, the principal's cost of implementing either e_2 or e_3 is minimized at $\mathbf{l}^*$. Specifically, let $S = \{L, H\}$ and define

$$\pi(H|x) = \begin{cases} \frac{27}{83}, & \text{if } x = x_1 \\ \frac{15}{83}, & \text{if } x = x_2 \\ \frac{41}{83}, & \text{if } x = x_3 \end{cases}, \pi(L|x) = 1 - \pi(H|x).$$

One can verify directly that the information structure (S, π) generates the likelihood-ratio point depicted in Figure 5.

By selecting this information structure, the agent implements the efficient effort e_3 and captures the entire surplus. Under this structure, the bonus is paid with probability $p(H \mid e_3) = \frac{35.7}{83} \approx 0.43$ and equals $\frac{0.85}{0.43} \approx 1.98$. Section 5 provides conditions on the performance technology under which this outcome is achievable.

By Proposition 3, the agent's choice of a binary information structure can be summarized by two objects: the likelihood-ratio point $\mathbf{l} \in \operatorname{co}(\mathbf{f})$ associated with the paid signal and its probability $\lambda \in (0, \lambda_{\max}]$ under the target effort. Lemma 3 in Appendix B shows that every such pair can be generated by a binary information structure, with $\lambda_{\max}$ given by the Bayes-plausibility constraint. The probability λ matters only through the location of the unpaid signal's likelihood-ratio point, $-\frac{\lambda}{1-\lambda}\,\mathbf{l}$.

This second object turns out to be essentially free. Making the paid signal rarer moves the unpaid endpoint toward the origin without changing the expected wage associated with the paid endpoint. By the homogeneity property discussed after (8), the bonus increases in inverse proportion to the probability of the paid signal, leaving every incentive constraint and expected payment unchanged. Once λ is sufficiently small, implementing any positive-cost effort on the negative portion of the segment requires a wage large enough that the principal prefers his outside option, so continuation play is governed by the paid endpoint alone. The agent's design problem therefore reduces, without loss, to the choice of a single likelihood-ratio vector in $\operatorname{co}(\mathbf{f})$.

We now combine the preceding results into a single statement. Recall from Lemma 1 that the principal's contract choice reduces to the choice of a point in $\operatorname{co}(\mathbf{p})$. For $\mathbf{l} \in \operatorname{co}(\mathbf{f})$, let $W_i(\mathbf{l})$ denote the minimum expected wage at which the principal can implement effort e_i at $\mathbf{l}$, and let $\Pi_i(\mathbf{l})$ denote his payoff from doing so. When e_i is implementable at $\mathbf{l}$, Lemma 1 gives

$$W_i(\mathbf{l}) = (1 - \ell_i) \max_{j:\, \ell_j \geq \ell_i} \frac{c(e_i) - c(e_j)}{\ell_j - \ell_i}, \qquad \Pi_i(\mathbf{l}) = g(e_i) - W_i(\mathbf{l}),$$

with the extended-real conventions of footnote 9. When e_i is not implementable at $\mathbf{l}$, we set $W_i(\mathbf{l}) = +\infty$ and hence $\Pi_i(\mathbf{l}) = -\infty$. In particular, $W_1 \equiv 0$ and $\Pi_1 \equiv g(e_1)$. Define

the agent's rent at $\mathbf{l}$ by[11]

$$R(\mathbf{l}) = \max\Big\{W_i(\mathbf{l}) - c(e_i) : \ i \in \arg\max_k \Pi_k(\mathbf{l})\Big\}.$$

This is the agent's payoff when the principal chooses his preferred effort at $\mathbf{l}$; the maximum reflects the convention of Section 2, under which the agent can obtain the best of the efforts that are optimal for the principal.

Proposition 4. *The value of the agent's design problem is* $\sup_{\mathbf{l}\in co(\mathbf{f})} R(\mathbf{l})$.

Proposition 4 reduces the agent's optimization over information structures to the finite-dimensional problem $\sup_{\mathbf{l}\in \text{co}(\mathbf{f})} R(\mathbf{l})$.[12] Thus, the value of the design problem is determined by the position of a single likelihood-ratio vector in co $(\mathbf{f})$.

The proof does not rely on Proposition 3. Instead, it constructs a binary information structure directly from the point $\mathbf{l}$, rather than by garbling an existing structure. Proposition 3 also follows from the argument: applying the construction to the point chosen by the principal under any given information structure yields a binary structure that implements the same effort at the same expected wage.

The agent-optimal structure is coarse in signals but need not involve a moderate bonus. The expected wage is determined by the paid signal's likelihood-ratio vector, while the probability of the paid signal must be small enough that no rival effort is profitably implementable at the unpaid signal; the proof of Proposition 4 gives the corresponding bound. Since the bonus equals the expected wage divided by this probability, it can be large when costlier rival efforts remain serious threats at the unpaid signal. Risk neutrality and limited liability allow this rescaling without changing the expected wage, while a

[11] The coordinates are defined relative to a fixed reference effort, but the program below and its value do not depend on this choice. To see this, consider the paid signal s and take e^* as the reference effort. Let $\varphi_i = \frac{p(s|e_i)}{p(s|e^*)}$, where $\ell_i = 1 - \varphi_i$. The implementation cost in Lemma 1 can then be written as $W_i = \varphi_i \max\left\{0, \max_{j:\varphi_j \le \varphi_i} \frac{c(e_i)-c(e_j)}{\varphi_i - \varphi_j}\right\}$. This expression is homogeneous of degree zero in φ. If the reference effort is changed to $\tilde{e}$, then $\frac{p(s|e_i)}{p(s|\tilde{e})} = \frac{\varphi_i}{\varphi_{\tilde{e}}}$, so the entire vector φ is rescaled by the same positive factor. It follows that W_i, and hence Π_i and R, are unchanged.

[12] A subgame-perfect equilibrium exists if and only if this supremum is attained. Attainment can fail because R need not be upper semicontinuous: the principal's preferred effort may change discontinuously when an alternative effort becomes implementable. Attainment holds under the full-extraction conditions of Section 5.1, where a maximizer is exhibited explicitly, while in the single-index class of Section 6 it depends on a boundary condition. Even when the supremum is not attained, it remains the relevant prediction. For every $\varepsilon > 0$, choosing $\mathbf{l}$ such that $R(\mathbf{l}) > \sup R - \varepsilon$ and applying the construction in the proof of Proposition 4 yields an ε-equilibrium in which the agent's payoff is within ε of the supremum. Continuation play consists of exact best responses after every history, and only the agent's first-stage choice is ε-optimal, since her gain from deviating is at most $\sup R - R(\mathbf{l}) < \varepsilon$.

wage cap or risk aversion would limit the same construction. The bound need not force a small probability: in Example 2 the bonus is paid with probability 0.43. In the single-index class of Section 6 the probability of the paid signal is bounded away from zero except at the attainment boundary discussed there.

5 Full Extraction and the Performance Technology

When can the agent capture the entire surplus, and how do equilibrium outcomes vary with the performance technology? Section 5.1 provides an exact characterization of full surplus extraction and uses it to derive a sequence of sufficient conditions stated in increasingly primitive terms. Section 5.2 then studies how equilibrium outcomes vary with the informativeness of the performance technology.

5.1 Characterizing Full Extraction

When can the agent implement a target effort e^* and extract the entire surplus? By Proposition 4, this question can be stated in terms of a single likelihood-ratio profile, which must satisfy two properties. First, implementing e^* at that profile must leave the principal exactly at his outside option. The set of profiles satisfying this requirement is $L(e^*) = \{\mathbf{l} : \Pi_{e^*}(\mathbf{l}) = g(e_1)\}$. Second, no rival effort can be strictly more attractive to the principal at the same profile. For each rival effort e_j, define $D_j = \{\mathbf{l} : \Pi_j(\mathbf{l}) > g(e_1)\}$. Thus, D_j collects the profiles at which implementing e_j gives the principal strictly more than $g(e_1)$, and therefore, on the locus ($L(e^*)$), strictly more than the target effort.[13] Full extraction at e^* is possible exactly when some feasible profile lies in $L(e^*)$ and outside every D_j:

Proposition 5. *There exists an information structure (S, π) implementing e^* with the agent's payoff $U_A = g(e^*) - g(e_1) - c(e^*)$ and the principal's payoff $g(e_1)$ if and only if $(L(e^*) \cap co\,(\mathbf{f})) \not\subseteq \bigcup_{e_j \neq e^*} D_j$.*

Full extraction therefore fails exactly when the sets D_j jointly cover $L(e^*) \cap \text{co}\,(\mathbf{f})$.

[13] Substituting the wage from (9) and rearranging, using $1 - \ell_j > 0$, which follows from full support, the condition $\Pi_j(\mathbf{l}) > g(e_1)$ can be written as a finite collection of affine inequalities, one for each effort cheaper than e_j. Implementability of e_j at $\mathbf{l}$ adds the affine inequality $\big(c(e_j) - c(e_i)\big)(\ell_j - \ell_k) \leq \big(c(e_k) - c(e_j)\big)(\ell_i - \ell_j)$ for each cheaper effort e_i and costlier effort e_k, together with $\ell_i \geq \ell_j$ for each equal-cost rival. Every inequality binds at $\mathbf{1} = (1, \ldots, 1)$, so D_j is a convex polyhedral cone with vertex at $\mathbf{1}$. If $g(e_j) \leq g(e_1)$, then $D_j = \varnothing$.

Example 2 (Continued). Proposition 5 can be used to determine whether the agent can implement e_3 and extract the entire surplus. For this target, the full-extraction locus consists of two rays:

$$L(e_3) = \left\{ \mathbf{l} | \ell_1 = \frac{c(e_3) - c(e_1)}{g(e_3) - g(e_1)}, \ell_2 \geq \frac{c(e_3) - c(e_2)}{g(e_3) - g(e_1)} \right\} \cup \left\{ \mathbf{l} | \ell_2 = \frac{c(e_3) - c(e_2)}{g(e_3) - g(e_1)}, \ell_1 \geq \frac{c(e_3) - c(e_1)}{g(e_3) - g(e_1)} \right\}.$$

The only rival effort that can give the principal strictly more than his outside option on $L(e_3)$ is e_2. Two conditions determine D_2. First, e_2 is implementable if and only if

$$\mathbf{l} \in \Omega_2 = \left\{ \mathbf{l} | \ell_2 \leq 0, \ell_1 > \ell_2 \right\} \cup \left\{ \mathbf{l} | \ell_2 > 0, \ell_1 \geq \frac{c(e_3) - c(e_1)}{c(e_3) - c(e_2)} \ell_2 \right\}.$$

Second, at such a profile the principal's payoff from e_2 exceeds $g(e_1)$ if and only if

$$\frac{g(e_2) - g(e_1)}{c(e_2) - c(e_1)} \ell_1 + \left(1 - \frac{g(e_2) - g(e_1)}{c(e_2) - c(e_1)} \right) \ell_2 > 1.$$

D_2 is the set of profiles satisfying both. As Figure 6 shows, $L(e_3) \cap \text{co}(\mathbf{f}) \not\subseteq D_2$, so the agent can implement e_3 and extract the full surplus $g(e_3) - g(e_1) - c(e_3) = 0.55$. Applying the same argument to e_2, the agent can also implement e_2 and extract the full surplus, $g(e_2) - g(e_1) - c(e_2) = 0.5$. The right panel of Figure 6 displays the corresponding sets in the same coordinates. Since $0.55 > 0.5$, the agent prefers e_3 to e_2.

The characterization is exact, but applying it requires computing $L(e^*)$ and the sets D_j for the technology at hand, as in Example 2. The remainder of this subsection instead identifies particular profiles that satisfy the characterization, reducing the condition to the feasibility of a single likelihood-ratio vector. We begin with the first-best target and the profile at which the agent is indifferent across all efforts.

Section 3 showed that an indicator leaving the agent indifferent across efforts turns the principal into a residual claimant on surplus. In likelihood-ratio coordinates, the corresponding profile is the vector $\mathbf{l}^*$ defined by

$$\ell_i^* = \frac{c(e_{FB}) - c(e_i)}{g(e_{FB}) - g(e_1)}, \qquad \forall i = 1, \ldots, |E|.$$

Equivalently, at $\mathbf{l}^*$, every incentive constraint binds at the same expected wage, $g(e_{FB}) -$

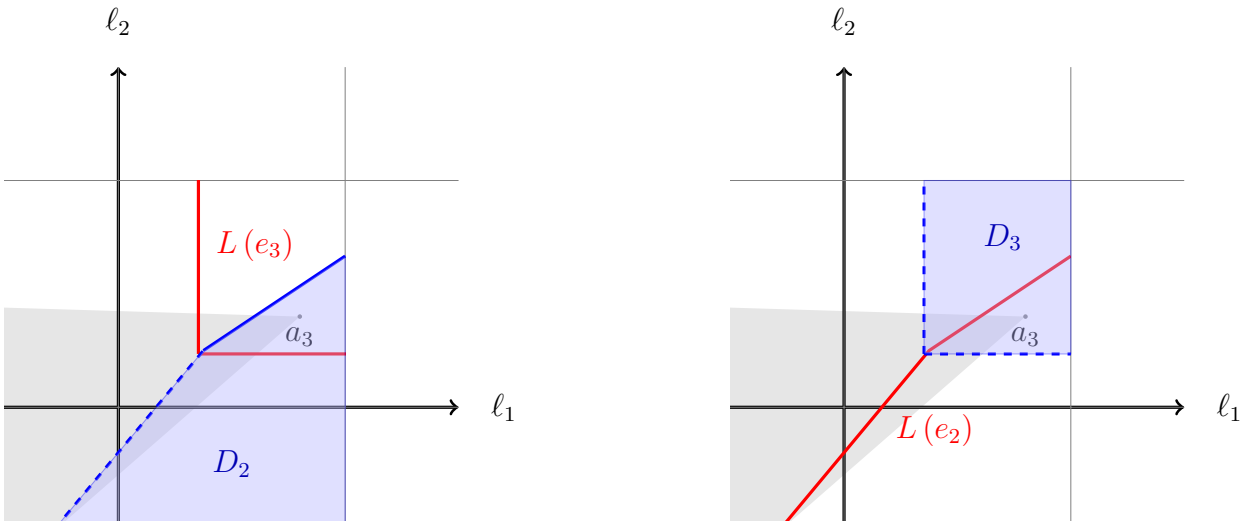

Figure 6: The full-extraction condition of Proposition 5 for the technology in Example 2. Both panels plot likelihood-ratio profiles relative to e_3. The left panel shows the relevant sets for implementing e_3, while the right panel shows the corresponding sets for implementing e_2.

$g(e_1)$: efforts cheaper than e_{FB} require at least this wage to be deterred, while costlier efforts require at most this wage, and the two bounds coincide. The agent is therefore indifferent across all effort levels; see the reformulation of (1) in (8). Relative to the construction in Section 3, indifference here also extends to efforts costlier than e_{FB}, which that construction instead made unimplementable.

The two conditions of the characterization hold at $\mathbf{l}^*$. Every chord $\frac{c(e_j)-c(e_i)}{l_i^*-l_j^*}$ equals $g(e_{FB})-g(e_1)$. Thus, implementing e_{FB} requires an expected wage of exactly this amount and leaves the principal at his outside option, so $\mathbf{l}^* \in L(e_{FB})$. Implementing any rival effort e_j requires its own cost plus the same fixed rent $g(e_{FB}) - g(e_1) - c(e_{FB})$, implying $\Pi_j(\mathbf{l}^*) = g(e_1) + \big[g(e_j) - c(e_j)\big] - \big[g(e_{FB}) - c(e_{FB})\big]$. Since e_{FB} maximizes total surplus, this expression is at most $g(e_1)$, and hence $\mathbf{l}^*$ lies outside every D_j. Full extraction at the first best therefore obtains whenever $\mathbf{l}^*$ is feasible.

Proposition 6. *Suppose that $g\,(e_{FB}) > g\,(e_1)$ and that $\mathbf{l}^* \in co\,(\mathbf{f})$. Then the first-best effort e_{FB} is implementable, and there exists an information structure (S, π) such that the agent's*

payoff is $U_A = g(e_{FB}) - g(e_1) - c(e_{FB})$, *and the principal's payoff is* $g(e_1)$, *so that the agent captures the entire surplus.*

Proposition 6 reduces full extraction at the first best to a single feasibility check. Since $\text{co}(\mathbf{f})$ is the convex hull of the $|X|$ vectors a_x in a space of dimension $|E| - 1$, testing whether $\mathbf{l}^* \in \text{co}(\mathbf{f})$ amounts to asking whether there exists some $\mu \in \Delta(X)$ such that $\sum_x \mu(x) a_x = \mathbf{l}^*$, a linear program that can be solved in polynomial time. The test is constructive: a feasible μ is precisely the posterior that the extracting indicator assigns to the paid signal, so verifying the condition and constructing the indicator are part of the same computation.

Two forces govern the condition, both reflecting the severity of the agency problem: the incentives that must be provided and the information available to provide them. Since $\mathbf{l}^*$ depends on the primitives only through the ratios $\frac{c(e_{FB})-c(e_i)}{g(e_{FB})-g(e_1)}$, the condition is easier to satisfy when effort costs are closer together relative to the gross payoff gain, and when the technology is more informative, so that $\text{co}(\mathbf{f})$ is larger. The two margins offset one another. The following family illustrates the informational margin.

Almost-Perfect Performance Technology. Suppose $X = E \subset \mathbb{R}_+, g(e) = e$, and for some $0 < \varepsilon < 1/(m-1)$,

$$f(e_j|e_j) = 1 - (m-1)\varepsilon, \forall j = 1, \cdots, |E|, \qquad f(e_i|e_j) = \varepsilon, \forall i \neq j.$$

As $\varepsilon \to 0$, the technology converges to a perfect one, in which observing $x \in X$ fully reveals effort. For each ε, $\text{co}(\mathbf{f})$ is the convex hull of the likelihood-ratio vectors $\mathbf{l}(e_k)$, whose coordinates in the corresponding $(|E|-1)$-dimensional space are

$$\mathbf{l}_i(e_k) = 1 - \frac{f(e_k|e_i)}{f(e_k|e_{FB})} = \begin{cases} 1 - \frac{\varepsilon}{1-(m-1)\varepsilon}, & \text{if } e_k = e_{FB}, i \neq k, \\ 0, & \text{if } e_k \neq e_{FB}, i \neq k, \\ m - \frac{1}{\varepsilon}, & \text{if } e_k \neq e_{FB}, i = k. \end{cases}$$

As $\varepsilon \to 0$, this set expands. The vector $\mathbf{l}(e_{FB})$ converges to $(1, \cdots, 1)$, while for $e_k \neq e_{FB}$, $\mathbf{l}(e_k)$ converge to $(0, \cdots, 0, -\infty, 0, \cdots, 0)$, where $-\infty$ appears in the coordinate corresponding to e_k. Hence, $\text{co}(\mathbf{f}) \to (-\infty, 1]^{|E|-1}$, as illustrated in Figure 7. This is the largest feasible likelihood-ratio set and corresponds to the limiting perfect performance technology. Therefore, unless the zero-cost effort is itself first best, $\mathbf{l}^* \in \text{co}(\mathbf{f})$ for ε sufficiently small.

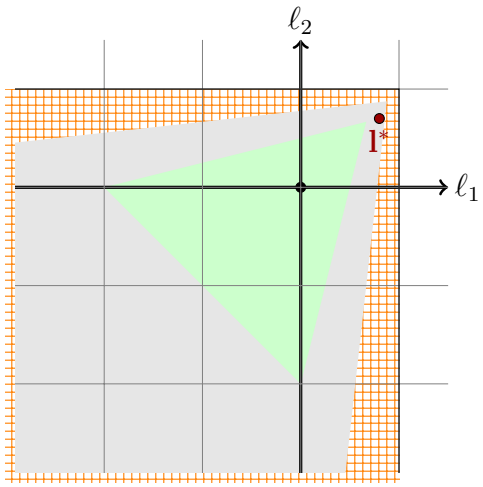

Figure 7: Almost-Perfect Performance Technology for $|E| = 3$. The green and gray shaded regions depict co $(\mathbf{f})$ for progressively smaller values of ε. As $\varepsilon \to 0$, co $(\mathbf{f})$ converges to the shaded quarter-space southwest of $(1, 1)$.

The point $\mathbf{l}^*$ is not the only profile that delivers full extraction. Since the expected wage must equal $g(e_{FB}) - g(e_1)$, at least one incentive constraint must bind. It is enough for the binding constraint to be the one against the zero-cost effort e_1, since keeping that constraint tight is sufficient to make every rival effort unattractive to the principal. These profiles form the branch $B = \{\mathbf{l} : \ell_1 = \ell_1^*,\ \ell_i \geq \ell_i^* \text{ for all } i \neq 1\}$, whose vertex is $\mathbf{l}^*$. Every point in B implements e_{FB} and leaves the principal at his outside option (see the end of the proof of Proposition 6), so it is sufficient that co $(\mathbf{f})$ intersect B. This is a weaker requirement and can be checked by a linear program of the same form. The branch also contains the likelihood-ratio profile generated by the construction in Section 3: the point that assigns coordinate zero to every effort costlier than e_{FB}.

That last profile extends beyond the first best. The first-best target is special because no rival generates more surplus, so indifference across efforts is compatible with full extraction. For an arbitrary target e^*, however, a costlier effort may generate more surplus, so making the agent indifferent between it and the target need not deter the principal from implementing it; such an effort must instead be made unimplementable. The construction of Section 3 does exactly this and delivers full extraction provided that cheaper efforts generate less surplus than the target: assign each cheaper effort its break-even

coordinate, $\ell_j(e^*) = \frac{c(e^*)-c(e_j)}{g(e^*)-g(e_1)}$, and each costlier effort $\ell_j(e^*) = 0$. At this profile, implementing a cheaper effort e_j gives the principal $g(e_1) + \big[g(e_j) - c(e_j)\big] - \big[g(e^*) - c(e^*)\big]$, which, under surplus monotonicity, is at most $g(e_1)$. A costlier effort is not implementable, since it is no more likely to be rewarded than e^* while costing more. Hence, $\mathbf{l}(e^*)$ lies in $L(e^*)$ and outside every D_j.

Proposition 7. *Suppose that $g(e_{FB}) > g(e_1)$ and that total surplus is weakly increasing in the cost of effort. Define the likelihood vector $\mathbf{l}(e^*) = \big(\ell_1(e^*), \ldots, \ell_{|E|}(e^*)\big)$ by*

$$\ell_j(e^*) = \begin{cases} \frac{c(e^*)-c(e_j)}{g(e^*)-g(e_1)}, & \text{if } c(e^*) > c(e_j), \\ 0, & \text{if } c(e^*) \leq c(e_j). \end{cases}$$

If $\mathbf{l}(e^*) \in co(\mathbf{f})$, then e^* is implementable, and there exists an information structure (S, π) such that the agent's payoff is $U_A = g(e^*) - g(e_1) - c(e^*)$, and the principal's payoff is $g(e_1)$, so that the agent captures the entire surplus.

Under surplus monotonicity, e_{FB} is the highest-cost effort. Thus, when $e^* = e_{FB}$, every rival is assigned its break-even coordinate and $\mathbf{l}(e_{FB})$ coincides with $\mathbf{l}^*$. Proposition 7 therefore reduces to Proposition 6. Without surplus monotonicity, $\mathbf{l}(e_{FB})$ and $\mathbf{l}^*$ differ, but $\mathbf{l}(e_{FB})$ still lies in B. Its feasibility therefore remains sufficient for full extraction at the first best.

The conditions so far ask whether a particular likelihood-ratio vector is feasible. The simplest conditions avoid this feasibility check altogether: they are inequalities involving the likelihood ratios of a single output, stated directly in terms of the primitives.

Proposition 8. *Suppose $g(e^*) > g(e_1)$, and suppose there exists an output $x \in X$ such that (i) $\frac{f(x|e_1)}{f(x|e^*)} \leq \left(1 - \frac{c(e^*)}{g(e^*)-g(e_1)}\right)$, (ii) $\frac{f(x|e_j)}{f(x|e^*)} \leq 1$, $\forall j$ such that $c(e_j) > c(e^*)$, and (iii) $\frac{f(x|e_j)}{f(x|e^*)} < \frac{c(e^*)-c(e_j)}{c(e^*)}\frac{f(x|e_1)}{f(x|e^*)} + \frac{c(e_j)}{c(e^*)}$, $\forall e_j \neq e^*$ such that $0 < c(e_j) \leq c(e^*)$. Then e^* is implementable, and there exists an information structure (S, π) such that the agent's payoff is $U_A = g(e^*) - g(e_1) - c(e^*)$, and the principal's payoff is $g(e_1)$, so that the agent captures the entire surplus.*

Like the earlier results, Proposition 8 works by exhibiting a profile in $L(e^*)$ that lies outside every D_j. Here, that profile is constructed from a single output rather than pinned down by the incentive conditions: it is a scaling toward the origin of the likelihood-ratio vector generated by x, which lies in co $(\mathbf{f})$ by convexity, so feasibility is automatic.

The conditions have a direct interpretation in terms of indicators. Consider the indicator that pays exactly when output x is realized. Conditions (ii) and (iii) rule out every other positive-cost effort at this profile: costlier efforts are no more likely to be rewarded, and among cheaper efforts the deviation to e_1 is the most attractive. Thus, the profile lies outside every D_j. Since neither condition is affected by scaling, the same is true along the entire ray. Condition (i) implies that this indicator implements e^* at an expected wage no greater than $g(e^*) - g(e_1)$. The agent can then garble the indicator by paying with probability less than one after x and with positive probability after other outputs. This makes the paid signal less discriminating between e^* and e_1, raising the wage required to deter e_1 until it equals $g(e^*) - g(e_1)$, at which point the profile lies in $L(e^*)$.

The conditions span two extremes. The zero-cost effort must be deterred entirely through information, which is condition (i). Efforts costlier than e^* are already deterred by their cost, so condition (ii) requires only that x be no more likely under them than under e^*. Condition (iii) interpolates between these cases: as the cost of a cheaper effort rises toward $c(e^*)$, the permitted likelihood ratio rises from $f(x \mid e_1)/f(x \mid e^*)$ toward 1, so efforts closer to e^* in cost require less informational separation. As with Proposition 6, the conditions are easier to satisfy when effort costs are close together relative to the gross payoff gain, that is, when the agency problem is not too severe.

The same argument delivers any division of the surplus, not only full extraction. Replacing $g(e^*) - g(e_1)$ by an arbitrary $\omega \in [c(e^*),\ g(e^*) - g(e_1)]$ in condition (i) gives conditions under which an indicator implements e^* at expected wage ω, leaving the agent with payoff $\omega - c(e^*)$ and the principal with payoff $g(e^*) - \omega$. In the boundary case of a perfect technology, the conditions hold for every $\omega \in [c(e^*),\ g(e^*) - g(e_1)]$. Thus, every split of the surplus generated by e^* that gives the agent nonnegative utility and the principal at least his outside-option payoff is implementable. Our characterization therefore recovers, as a limiting statement, the implementability characterization in Theorem 3.1 of Babichenko et al. (2025). [14]

[14] Assumption 2 excludes a perfect technology, but the conditions remain well defined in that case. Taking $x = e^*$, we have $f(x \mid e^*) = 1$ and $f(x \mid e_j) = 0$ for every $e_j \neq e^*$, so the conditions hold for every ω in this interval. The same conclusion is obtained under Assumption 2 from the almost-perfect family introduced above. For $\varepsilon < 1/m$, the output $x = e^*$ satisfies $f(x \mid e_j)/f(x \mid e^*) = r(\varepsilon) := \varepsilon/\big(1 - (m-1)\varepsilon\big)$ for every $e_j \neq e^*$. The conditions of Proposition 8 then hold for every $\omega \in \big[c(e^*)/(1 - r(\varepsilon)),\ g(e^*) - g(e_1)\big]$. As $\varepsilon \to 0$, this interval converges to $[c(e^*),\ g(e^*) - g(e_1)]$.

5.2 Comparative Statics in the Performance Technology

Section 5.1 asked when a given technology permits full extraction. We now vary the technology and ask how equilibrium payoffs and the resulting surplus respond.

Corollary 1. *Suppose a subgame-perfect equilibrium exists under both f and f'.*[15] *Then the agent's equilibrium payoff U_A is weakly increasing in the informativeness of the performance technology: if $\text{co}(\mathbf{f}') \subseteq \text{co}(\mathbf{f})$ (in particular, whenever f is Blackwell more informative than f'), then $U_A(f) \geq U_A(f')$.*

More generally, the holder of the design right maximizes over convex subsets of $\text{co}(\mathbf{f})$, so the designer's payoff is weakly increasing in the informativeness of the technology. When the principal holds the design right, that is the classical informativeness principle.

Corollary 1 concerns only the agent's payoff. Neither the cost of the implemented effort nor the total surplus it generates is generally monotone in the technology: a Blackwell improvement can strictly reduce equilibrium surplus, even under standard regularity conditions.

Example 3. Let $X = \{x_1, x_2, x_3\}$ and $E = \{e_1, e_2, e_3, e_4\}$, with $c = (0,\ 0.1,\ 0.4,\ 0.5)$ and $g = (0,\ 0.8,\ 1.4,\ 1.8)$. The corresponding surpluses are $g(e) - c(e) = (0,\ 0.7,\ 1,\ 1.3)$, which are strictly increasing in cost, so e_4 is first-best. Consider the performance technologies

$$f = \begin{bmatrix} 0.40 & 0.50 & 0.10 \\ 0.30 & 0.55 & 0.15 \\ 0.10 & 0.70 & 0.20 \\ 0.08 & 0.70 & 0.22 \end{bmatrix}, \qquad f' = \begin{bmatrix} 0.12 & 0.78 & 0.10 \\ 0.09 & 0.76 & 0.15 \\ 0.03 & 0.77 & 0.20 \\ 0.024 & 0.756 & 0.22 \end{bmatrix},$$

where $f_{ij} = f(x_j \mid e_i)$, and similarly for f'. The technology f' is obtained from f through the garbling matrix

$$G = \begin{bmatrix} 0.3 & 0.7 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

so that $f' = fG$. Thus, x_1 is misrecorded as x_2 with probability 0.7, while x_2 and x_3 are recorded correctly. Both technologies have full support and satisfy strict MLRP under the

[15] The existence hypothesis is not restrictive for the comparison. By Proposition 4, the value of the agent's design problem is $\sup_{\mathbf{l} \in \text{co}(\mathbf{f})} R(\mathbf{l})$, which is weakly increasing under $\text{co}(\mathbf{f}') \subseteq \text{co}(\mathbf{f})$ whether or not the supremum is attained. The same argument applies to Corollary 2: the supremum of the agent's payoff over admissible information structures weakly falls as the constraint tightens, since her feasible set shrinks.

ordering x_1, x_2, x_3. Moreover, $\text{co}(\mathbf{f}') \subseteq \text{co}(\mathbf{f})$ for every target effort.

Output x_2 generates exactly the likelihood-ratio vector $\mathbf{l}(e_3)$ from Proposition 7. In particular, $f(x_2 \mid \cdot) = (0.50,\ 0.55,\ 0.70,\ 0.70)$, so $1 - \frac{f(x_2|e_1)}{f(x_2|e_3)} = \frac{2}{7} = \frac{c(e_3)}{g(e_3)}$,$1 - \frac{f(x_2|e_2)}{f(x_2|e_3)} = \frac{3}{14} = \frac{c(e_3)-c(e_2)}{g(e_3)}$, and $f(x_2 \mid e_4) = f(x_2 \mid e_3)$. Proposition 7 therefore implies that the agent extracts the full surplus at e_3: $U_A = g(e_3) - c(e_3) = 1$. The extracting indicator pays only on x_2. The bonus is paid with probability $f(x_2 \mid e_3) = 0.70$ and equals $\frac{1.4}{0.70} = 2$, consistent with the discussion at the end of Section 4.1.

By contrast, under f, targeting e_4 gives the agent a maximum payoff of approximately 0.86. Hence, the equilibrium under f implements e_3 and generates surplus 1. After the garbling, $\mathbf{l}(e_3) \notin \text{co}(\mathbf{f}')$; in fact, e_3 is not implementable under f'. Under f', the agent can obtain a payoff of 0.826 at e_4. This exceeds the maximum payoff available from targeting e_2, $g(e_2) - c(e_2) = 0.7$, while e_1 yields zero. The agent therefore optimally targets e_4. The equilibrium under f' implements the first-best effort, the agent's payoff is at least 0.826, and equilibrium surplus equals 1.3. Thus, expanding $\text{co}(\mathbf{f}')$ to $\text{co}(\mathbf{f})$ changes the implemented effort from e_4 to e_3 and reduces equilibrium surplus from 1.3 to 1, even though the agent's equilibrium payoff weakly increases, as Corollary 1 requires. The coarser technology therefore implements the first-best effort in this example, whereas the more informative technology does not.[16]

There are two limits to this phenomenon. First, it requires at least four efforts.

Proposition 9. *Let $|E| \leq 3$ and $\text{co}(\mathbf{f}') \subseteq \text{co}(\mathbf{f})$. Suppose a subgame-perfect equilibrium exists under both f and f'. Then the largest equilibrium surplus under f is at least the largest equilibrium surplus under f'.*

Second, the reversal cannot occur within the single-index technologies of Section 6, where greater informativeness weakly raises both parties' equilibrium payoffs and therefore equilibrium surplus. Thus, a decline in the largest equilibrium surplus requires at least four efforts and a multidimensional feasible likelihood-ratio set. What matters is not whether $\text{co}(\mathbf{f})$ expands, but the directions in which it expands.

The reversal also has a policy implication. Restrictions on workplace monitoring reduce the information available for contracting and therefore constrain $\text{co}(\mathbf{f})$. The example shows that such restrictions need not reduce the surplus generated by the implemented

[16]The reversal is not specific to the reduced-form specification of g. We have also constructed a variant with monetary output, $g(e) = \mathbb{E}[x|e]$, and a mean-preserving garbling, so that the surplus schedule is held fixed across technologies. The same reversal occurs: the coarser technology implements the first-best effort, while the Blackwell improvement does not.

effort and can instead increase it. More informative performance measurement therefore need not improve efficiency when the measured party influences the choice of indicator.

6 Single-Index Technologies

The results so far hold for an arbitrary performance technology. We now study a class in which the feasible set of likelihood-ratio vectors collapses to a single line through the origin. The agent's design problem, which in general involves choosing a vector in co $(\mathbf{f})$, then reduces, under the monotone-likelihood-ratio assumption introduced below, to the choice of a scalar, and the analysis can be carried out directly in terms of that scalar. This allows us to state the results for an arbitrary compact effort set E, either finite as in Sections 2--5 or an interval, and an arbitrary outcome space X.[17]

We say that the performance technology is single-index if effort enters the output distribution linearly through a scalar index:

$$f(x \mid e) \;=\; f_0(x) + \Gamma(e)\, D(x), \qquad \textstyle\sum_x D(x) = 0,$$

where f_0 is a probability distribution over X and $f(x \mid e) > 0$, or equivalently $1 + \Gamma(e)\rho(x) > 0$, where $\rho(x) \equiv D(x)/f_0(x)$.

Rescaling Γ to take values in $[0, 1]$, the technology can equivalently be written as $f(x \mid e) = (1 - \Gamma(e))\, g_0(x) + \Gamma(e)\, g_1(x)$, where g_1/g_0 is an increasing function of ρ.[18] Costlier efforts therefore place greater weight on outputs with higher ρ. For $\Gamma(e') > \Gamma(e)$, the likelihood ratio

$$\frac{f(x \mid e')}{f(x \mid e)} = \frac{1 + \Gamma(e')\rho(x)}{1 + \Gamma(e)\rho(x)}$$

is increasing in $\rho(x)$, so the family satisfies MLRP when outputs are ordered by $\rho(x)$ and efforts by $\Gamma(e)$. A familiar instance is the Gaussian mixture $(1 - \Gamma)\mathcal{N}(\mu_0, \sigma^2) + \Gamma\,\mathcal{N}(\mu_1, \sigma^2)$ with $\mu_1 > \mu_0$. For the remainder of the section, we order outputs so that ρ is nondecreasing.[19]

[17] Sections 2–5 take X and E to be finite. Here, when X is a continuum, $f(\cdot \mid e)$ is a density and integrals over X replace sums; when E is an interval, we additionally assume that c and g are continuous. The proof of Proposition 3 covers these cases, so the agent may again restrict attention to binary indicators.

[18] This class is often referred to as a mixture or "spanning" technology and is widely used in the moral-hazard literature; see Grossman and Hart (1983) and Kirkegaard (2017).

[19] This is without loss of generality: outputs enter payoffs only through f, and outputs with the same value of ρ generate the same likelihood-ratio vector, so they can be indexed by their value of ρ. Interpreting a threshold rule as a cutoff in the underlying output variable, however, requires ρ to be increasing in that

Proposition 10. *Suppose X and E are finite. The performance technology is single-index if and only if the likelihood-ratio vectors $\{a_x\}_{x\in X}$ lie on a single line through the origin, or equivalently, if and only if* $\operatorname{co}(\mathbf{f})$ *is at most one-dimensional.*[20]

Concretely, for any target e^*, the likelihood-ratio vector generated by output x is

$$a_x = s(x)\,\mathbf{v}, \qquad v_j = \Gamma(e^*) - \Gamma(e_j), \qquad s(x) = \frac{D(x)}{f(x \mid e^*)} = \frac{\rho(x)}{1 + \Gamma(e^*)\rho(x)}.$$

Hence,

$$\operatorname{co}(\mathbf{f}) = \left\{ t\mathbf{v} : \; t \in \left[\min_x s(x), \; \max_x s(x) \right] \right\}.$$

The direction $\mathbf{v}$ depends only on the effort levels, while outputs matter only through $s(x)$, which is increasing in $\rho(x)$. Proposition 10 gives the geometric characterization for the finite model; the results that follow allow for arbitrary E and X, and maintain the following assumption throughout the remainder of the section.

Assumption 3. *Γ and c are comonotone, that is, $(\Gamma(e) - \Gamma(e'))(c(e) - c(e')) \geq 0$ for all $e, e' \in E$.*

Within the single-index family, likelihood ratios are monotone in ρ. Thus, with outputs ordered by ρ, Assumption 3 is equivalent to the monotone likelihood ratio property when efforts are ranked by cost: higher-cost effort shifts the output distribution upward in the MLRP sense.[21]

A binary indicator is described by $\pi(x) = \pi(H \mid x)$, the probability that output x generates the paid signal. Since only $c \circ \Gamma^{-1}$ and $g \circ \Gamma^{-1}$ enter payoffs, we index efforts by their index value and normalize $\Gamma(e) = e$ for the remainder of the section, with the minimum index, corresponding to the zero-cost effort, normalized to $e_1 = 0$.[22] For $e > 0$,

variable; that is, the technology must satisfy MLRP with respect to the natural ordering of outputs.

[20] For general E and X, the technology is single-index if and only if the family of densities $\{f(\cdot \mid e) : e \in E\}$ lies on a line, in the sense that its affine hull in the space of signed measures on X has dimension at most one. The proof in Appendix B establishes this version and derives the finite-case statement above from it.

[21] Assumption 3 is without loss when g is nondecreasing in the index. Under any contract, the expected wage can be written as $W_0 + W_D\Gamma(e)$, where $W_0 = \sum_s w(s) \sum_x \pi(s|x) f_0(x)$ and $W_D = \sum_s w(s) \sum_x \pi(s|x) D(x)$. If $W_D \geq 0$, an effort with a weakly lower index and weakly higher cost than another effort, with at least one inequality strict, is never incentive compatible. If $W_D < 0$, incentive compatibility against e_1 implies that the implemented effort cannot have an index above $\Gamma(e_1)$, so the principal cannot obtain more than his fallback payoff. Such dominated efforts are therefore never implemented in equilibrium. Dropping them leaves equilibrium payoffs unchanged and yields a comonotone family.

[22] As with outputs, efforts affect f only through $\Gamma(e)$, so efforts with the same index are informationally

let

$$d(e) = \sup_{e' < e} \frac{c(e) - c(e')}{e - e'}$$

denote the steepest slope of a chord joining a lower effort to e, and set $d(0) = 0$. Under Assumption 3, c is nondecreasing, and hence $d(e) \geq 0$. Not every effort is implementable by a wage contract: effort e is implementable if and only if $d(e) \leq \inf_{e' > e} \frac{c(e') - c(e)}{e' - e}$, a condition that depends only on the cost function. When c is convex, chord slopes are nondecreasing, so this condition holds at every effort. When, in addition, E is an interval and c is differentiable, the chord slopes increase as $e' \uparrow e$, so $d(e) = c'(e)$ for every $e > 0$, with c' interpreted as the left derivative at the upper endpoint. As the next result shows, $d(e)$ determines the required wage.

Proposition 11. *Let the technology be single-index. (i) For any binary indicator, all payoffs depend on π only through the price shift $A(\pi) = \sum_x \pi(x) f_0(x) \,/\, \sum_x \pi(x) D(x)$. Its attainable range is an interval unbounded above, with infimum $A_{min} = 1/\sup_x \rho(x)$.*[23] *For every effort e implementable by a wage contract, the expected wage required to implement it is $(A(\pi) + e)\; d(e)$, and a lower $A(\pi)$ corresponds to a more informative indicator. (ii) In equilibrium, an optimal indicator can be taken to be a threshold rule: for some x^* and $q \in [0, 1]$,*[24]

$$\pi(x) = \begin{cases} 1, & x > x^*, \\ q, & x = x^*, \\ 0, & x < x^*. \end{cases}$$

Given an indicator with price shift A, the principal chooses $e_B(A) \in \arg\max_e \; g(e) - (A + e) d(e)$, where the maximum is taken over efforts implementable by a wage contract. Write $V(A)$ for the agent's payoff when the principal best responds in this way. The value of her design problem is the supremum of $V(A)$ over the achievable price shifts.[25] Let A^*

identical, and retaining only the cheapest leaves payoffs unchanged. The relabeled effort set is finite when E is finite, and is an interval when E is an interval and Γ is continuous.

[23] The infimum is attained if ρ reaches its supremum on a set of outputs with positive probability; in particular, it is always attained when X is finite.

[24] When X has atoms, including when X is finite, randomization at the threshold may be required to attain every feasible value of A. The indicator assigns the low signal below a threshold, the high signal above it, and randomizes between the two signals at the threshold.

[25] The function V can be discontinuous only where the implemented effort changes. At such points, the principal is indifferent between the efforts implemented on either side, and the tie-breaking convention of Section 2 selects the one the agent prefers. Hence, V is upper semicontinuous. Moreover, $V(A) \to 0$ as $A \to \infty$. For any $\varepsilon > 0$, implementing an effort with $c(e) \geq \varepsilon$ requires an expected wage of at least $A\varepsilon/\bar{e}$,

denote the price shift generated by the agent's equilibrium indicator. When the agent's optimal payoff is positive and several price shifts are optimal, we select the largest. When her optimal payoff is zero, only the zero-cost effort is implemented at any achievable price shift, so the choice of indicator affects neither party's payoff, and any optimal indicator may be selected.[26]

Proposition 12. *Let c be strictly convex. (i) e_B is weakly decreasing in A, as is equilibrium total surplus. (ii) The value of the agent's design problem is weakly decreasing in A_{min}. (iii) For every A, $e_B(A)$ is weakly below e_{FB}, the largest maximizer of $g - c$.*

The class admits an informativeness principle: since the wage required to implement any effort rises with A, the principal's payoff is decreasing in A. If he controlled the indicator, he would therefore choose $A = A_{\min}$, the most informative indicator the technology permits. Part (i) says more: effort and total surplus are also weakly decreasing in A, so greater informativeness benefits not only the principal but also total surplus. Since the agent must choose $A^* \geq A_{\min}$, agent-controlled design implements weakly less effort and generates weakly less total surplus than principal-controlled design. This monotonicity is a one-dimensional phenomenon: a single price shift, which captures all of the indicator's payoff-relevant content, moves every effort's implementation cost together without changing which efforts are implementable, ruling out the reversal of Section 5.2. Part (ii) gives the comparative static in informativeness: a lower $A_{\min}$ expands the agent's feasible set and weakly raises her payoff. It also weakly lowers A^*, since only price shifts below the current optimum are added; by the monotonicity above, effort, total surplus, and the principal's payoff therefore all weakly increase. Whether further improvements matter once the bound is slack depends on the shape of V; under the single-peakedness condition of Proposition 13, they do not.[27] Part (iii) reflects the agent's rent, $(A+e)d(e) - c(e)$,

since $d(e) \geq c(e)/e$. For sufficiently large A, all such efforts are therefore dominated by the zero-cost effort, so $c(e_B(A)) \to 0$ and hence $e_B(A) \to e_1$; when E is finite, $e_B(A) = e_1$ for all sufficiently large A. Since the principal never pays more than $g(e_B(A)) - g(e_1)$, the agent's rent satisfies $V(A) \leq g(e_B(A)) - g(e_1) \to 0$. The supremum over $[A_{\min}, \infty)$ is therefore attained whenever $A_{\min}$ is feasible. If A_{min} is not attained, the feasible set is open at its lower boundary, and the supremum is attained if and only if the agent's optimal price shift is interior.

[26] Every achievable price shift is strictly positive: either $A > A_{\min} \geq 0$, or $A = A_{\min}$ is attained, which requires $\sup_x \rho(x) < \infty$ and hence $A_{\min} = \frac{1}{\sup_x \rho(x)} > 0$. If a positive-cost effort e were implemented at price shift A, the agent's rent, $(A+e)d(e) - c(e)$, would be at least $\frac{A}{e}c(e) > 0$, since $d(e) \geq c(e)/e$. Hence, a zero optimal payoff implies that only the zero-cost effort is implemented, at zero wage.

[27] With finitely many efforts, V rises between the principal's switch points and jumps downward at each switch. If V has multiple peaks, a reduction in $A_{\min}$ can make a higher peak feasible even when the current bound is slack.

which increases with effort. The principal therefore maximizes total surplus net of a rent that rises with effort and may stop short of the first best.

Proposition 13 characterizes the equilibrium under one additional condition.

Proposition 13. *Suppose V is single-peaked with peak A°: it is strictly increasing below A° and, wherever V is positive, strictly decreasing above A°. Suppose also that V is positive at some achievable price shift. If $A^\circ > A_{min}$ or $A_{\min}$ is attained, then $A^* = \max\{A_{min}, A^\circ\}$. Otherwise, no optimal price shift exists. When A_{min} is attained, the agent and principal prefer the same price shift if and only if $A_{min} \geq A^\circ$.*

The example below takes E to be an interval, so the implemented effort varies smoothly with A rather than jumping as it would on a finite grid. Under the linear-quadratic specification, the division of surplus then has a closed-form expression.

Example 4. To illustrate the equilibrium regimes and the division of surplus, let $E = [0, \bar{e}]$, $c(e) = e^2/2$, $g(e) = \kappa e$ with $\kappa > 0$, and assume $\bar{e} \geq \kappa/2$. The economy is summarized by (κ, A_{min}), where κ measures the marginal value of effort and A_{min} is the lower bound on the price shift imposed by the available performance technology.[28] For this example, suppose $A_{\min}$ is attained. The principal's best response is $e_B(A) = \max\{0, (\kappa - A)/2\}$. Since $A \geq 0$, the unconstrained maximizer $(\kappa - A)/2$ never exceeds $\kappa/2 \leq \bar{e}$, so the upper boundary never binds. The agent's payoff along this best response is concave on the positive-effort region $A < \kappa$, attains its maximum at $A = \kappa/3$, and is zero for $A \geq \kappa$. Hence, it is single-peaked with peak $\kappa/3$. When $\kappa \leq A_{\min}$, no positive effort is implemented and the agent's payoff is zero at every feasible price shift, so she is indifferent among all indicators. When $\kappa > A_{\min}$, Proposition 13 implies $A^* = \max\{A_{\min}, \kappa/3\}$. When $A_{\min} < \kappa \leq 3A_{\min}$, the agent's preferred price shift $\kappa/3$ is infeasible, so she chooses $A^* = A_{\min}$, which is also the principal's preferred price shift. The implemented effort is $e = (\kappa - A_{\min})/2$. When $\kappa > 3A_{\min}$, the lower bound no longer binds. The agent chooses $A^* = \kappa/3 > A_{\min}$ and induces $e_A = \kappa/3$, while a principal who controlled the indicator would choose A_{min} and induce $e_B = (\kappa - A_{\min})/2 > e_A$.

The example also gives closed-form expressions for the division of surplus. Let $s = U_A/(U_A + U_P)$ denote the agent's share of total surplus. In the coincidence regime, where

[28] A single-index technology consistent with any $\bar{e} > 0$ and $A_{\min} > 0$ exists. Nonnegativity of the affine densities $f_0(x)\big(1 + \rho(x)e\big)$ for $e \in [0, \bar{e}]$, restricts only the negative values of ρ, requiring $\rho(x) \geq -1/\bar{e}$, while $A_{\min} = \frac{1}{\sup_x \rho(x)}$ depends only on its positive values. The positive and negative values of ρ can therefore be chosen independently, subject to $\int \rho(x) f_0(x)\, dx = 0$.

$A^* = A_{min}$,

$$U_A = \tfrac{(\kappa - A_{\min})(\kappa + 3A_{\min})}{8}, \qquad U_P = \tfrac{(\kappa - A_{\min})^2}{4}, \qquad s = \frac{\kappa + 3A_{\min}}{3\kappa + A_{\min}}.$$

Holding κ fixed, a decrease in $A_{\min}$ raises both payoffs and lowers the agent's share of surplus. Thus, better information expands total surplus while shifting its division toward the principal. At the opposite extreme, $s \to 1$ as $A_{\min} \uparrow \kappa$, although both payoffs and total surplus vanish in this limit. In the distinct-designs regime, where $A^* > A_{\min}$,

$$U_A = \tfrac{\kappa^2}{6}, \qquad U_P = \tfrac{\kappa^2}{9}, \qquad s = \tfrac{3}{5}.$$

Here the technological lower bound on A is slack, so both equilibrium payoffs and their division are independent of further improvements in the technology. Thus, in this example, informativeness affects the division of surplus only when $A_{\min}$ binds; once it is slack, the quadratic cost specification yields the constant split $s = 3/5$.

The example also quantifies the value to the agent of controlling the indicator. This value is zero in the coincidence regime, where the two parties choose the same price shift. In the distinct-designs regime, it equals $(\kappa - 3A_{\min})^2/24$, which is strictly positive, increases as $A_{\min}$ falls, and converges to $\kappa^2/24$ as $A_{\min} \to 0$.

The comparative statics for payoff levels extend beyond the linear-quadratic specification, whereas the closed-form division of surplus does not.

Remark. In the single-index family, expected output is affine in the index:

$$\mathbb{E}[x|e] = \mu_0 + \delta\Gamma(e), \qquad \delta = \sum_x xD(x) > 0,$$

where the inequality holds when ρ is increasing in output, which here is a property of the technology rather than a labeling convention. Under the normalization $\Gamma(e) = e$, the specification $g(e) = \kappa e$ in Example 4 is therefore equivalent, up to the additive constant μ_0, to taking the principal's gross payoff to be expected output, with $\kappa = \delta$. The constant μ_0 does not affect effort choices or payoff differences.

The finite and continuous models differ in their implications for full extraction. With finitely many efforts, the principal's best alternative to a target may be a discrete step away, possibly the zero-cost effort. The agent can therefore raise the incentive price while holding the target fixed until that alternative binds. With a continuum of efforts, nearby

alternatives are always available, and further coarsening changes the implemented effort rather than merely raising the wage required to sustain a fixed target. Under the additional conditions $c'(0) = 0$ and single-peakedness of the principal's objective in effort,[29] his payoff in the continuous single-index model is therefore strictly positive whenever the implemented effort is positive, and can reach zero only as the implemented effort, and hence the surplus, vanishes.

The 2023 writers' strike provides a useful example in which the performance measure itself became an object of bargaining. Streaming platforms observe viewership at a fine level, but writers' residual compensation had not been tied directly to those data. The settlement introduced a negotiated performance metric for a bonus: a show qualifies if its domestic viewership reaches at least twenty percent of the service's subscribers within ninety days, while the underlying data are shared with the union only confidentially and in aggregate. The episode parallels several features of the model. The metric was negotiated by the agent side through collective bargaining, the contracted measure is a coarse binary statistic constructed from a much richer underlying variable, and the coarsening takes the form of a threshold rule. More broadly, the dispute illustrates how control over performance measurement can itself become part of compensation bargaining. The model also offers a possible interpretation of the timing. When success could be measured using relatively coarse market outcomes, there may have been less reason to bargain explicitly over the performance metric. As streaming replaced those market signals with proprietary and highly granular viewership data, the floor on how demanding a metric could be effectively vanished, and the model suggests this is when metric design turns contested: with coarse measurement the parties' preferred metrics coincide, while with fine measurement they diverge and where the cutoff sits becomes an object of bargaining. The 2023 dispute is consistent with the industry crossing that boundary.

7 Conclusion

This paper studies contract design when the party being measured chooses the performance measure. Our results have three implications. First, the geometric methodology may be useful in other environments in which off-equilibrium-path beliefs shape payoffs.

[29] A sufficient condition for single-peakedness is that g be concave and $c''' \geq 0$. These conditions hold in Example 4.

The information structure that supports the equilibrium effort also governs what the principal would infer from efforts never taken, and these off-path likelihoods affect both his contracting problem and the division of surplus.

Second, the paper isolates a margin distinct from how much information exists: who chooses the measure. A more informative technology benefits whoever holds the design right, but it need not increase total surplus. A Blackwell improvement can reduce equilibrium surplus by changing which effort the principal implements.

Third, the efficiency consequences of assigning that right to the measured party depend on the environment. When the agency problem is mild, with effort costs close together relative to the gains from effort and the technology sufficiently informative, her preferred indicator can implement the first-best effort while allowing her to capture the entire surplus. In the single-index class, where the technology is far from perfect with more than two effort levels, her preferred indicator generates weakly less total surplus than the one the principal would choose.

Taken together, these results suggest that the coarseness of contracted performance measures may reflect the allocation of the design right rather than the limits of measurement. Testing this would require variation in bargaining power across contracts written on comparable underlying performance data, which we leave to future work.

A Minimum-Information Constraints

In the analysis above, the performance technology f plays the role of an upper bound on the information that can be made available to the principal for contracting purposes. Interpreted through the geometric representation developed in Section 4, the full technology f determines the maximal feasible set of likelihood–ratio vectors $\mathrm{co}\,(\mathbf{f})$. Proposition 2 establishes that any information structure chosen by the agent induces a convex set $\mathrm{co}\,(\mathbf{p}) \subseteq \mathrm{co}\,(\mathbf{f})$ containing the origin, and conversely that any convex subset of $\mathrm{co}\,(\mathbf{f})$ whose relative interior contains the origin can be generated by some finite information structure. In this sense, the agent's indicator choice amounts to selecting how much of the available information encoded in f is revealed to the principal. Throughout the main body of the paper, this perspective allows us to treat f as the maximal information benchmark, with all feasible contracts arising from garblings of the underlying performance measure.

In many applications, however, the principal may already possess the right to condition contracts on some baseline information that cannot be removed, or may be able to

secure that right through bargaining. We model such constraints in reduced form. Let q denote such an exogenously given information structure, and let co $(\mathbf{q})$ be its associated convex hull of likelihood-ratio vectors. Given the constraint q, we call an information structure *admissible* if co $(\mathbf{q}) \subseteq$ co $(\mathbf{p})$.[30] Throughout this section, the agent's choice of information structure is restricted to admissible ones; the game is otherwise unchanged. Since payoffs depend on an information structure only through its likelihood-ratio hull, admissibility requires the principal to retain at least the contracting possibilities available under q. Any information structure that reveals q is admissible, since refinement implies co $(\mathbf{q}) \subseteq$ co $(\mathbf{p})$ by the argument of Proposition 2. The converse need not hold, so admissibility is weaker than requiring the agent's indicator itself to reveal q. By Proposition 2, every admissible hull lies within co $(\mathbf{f})$, so the agent chooses a convex set between co $(\mathbf{q})$ and co $(\mathbf{f})$. In this section, we study how such minimum-information constraints affect implementability, surplus extraction, and, for single-index technologies, the agent's optimal price shift.

A.1 Payoff Monotonicity and Implementability

Because the constraint enters the agent's problem only by shrinking her feasible set, Corollary 1 has an immediate analogue.

Corollary 2. *Suppose a subgame-perfect equilibrium exists under both constraints q and q'. Tightening the minimum-information constraint weakly lowers the agent's equilibrium payoff: if* co $(\mathbf{q}) \subseteq$ co $(\mathbf{q}')$, then $U_A(q) \geq U_A(q')$.

Corollaries 1 and 2 bound the agent's design problem from opposite sides. The technology f determines the maximum information available for contracting, so enlarging it can only benefit the agent. The constraint q determines the minimum information that must remain available for contracting, so enlarging it can only reduce her payoff. As with Corollary 1, this monotonicity concerns only the agent's payoff. Section A.3 shows that, within the single-index class, implemented effort and total surplus move in the opposite direction. Outside that class, this opposite monotonicity need not hold.

The signal-count reduction of Proposition 3 also extends to this setting: implementability and the expected wage are unchanged if the agent uses at most $|supp(\mathbf{q})| + 3$ signals.

[30] Results establishing that the agent cannot achieve a given outcome under admissibility extend a fortiori to the stricter institution in which the principal observes q directly, since every information structure feasible there is admissible.

Proposition 14. *If e^* is implementable by some admissible information structure (S, π) with associated stochastic mapping p such that co $(\mathbf{q}) \subseteq$ co$(\mathbf{p}) \subseteq$ co $(\mathbf{f})$ and requires an expected wage $W(e^*, \pi)$, then e^* is also implementable by an admissible information structure $\left(\hat{S}, \hat{\pi}\right)$ with $\left|\hat{S}\right| \leq |supp(\mathbf{q})| + 3$ and $W(e^*, \hat{\pi}) = W(e^*, \pi)$.*

A.2 Surplus Extraction

Full extraction is characterized by the same two objects as in Section 5.1: the locus $L(e^*)$ of profiles at which implementing e^* leaves the principal exactly his outside option, and the sets D_j of profiles at which he strictly prefers a rival effort. The minimum-information constraint changes the problem in two ways. First, the principal now optimizes over a set that must contain co $(\mathbf{q})$. Since co $(\mathbf{p})$ is convex and must contain both co $(\mathbf{q})$ and the extraction profile, it must also contain their convex hull. The relevant incentive conditions must therefore hold not only at the extraction profile, but throughout its convex hull with co $(\mathbf{q})$.

Second, the constraint creates a deviation with no counterpart in Section 5.1. Because co $(\mathbf{p})$ must contain co $(\mathbf{q})$, the principal may retain access to profiles at which implementing e^* itself requires an expected wage below $g(e^*) - g(e_1)$. At such a profile, he implements the target effort while obtaining more than his outside option, and no admissible choice of information structure can remove that profile. In parallel with the sets D_j, define $D_{e^*} = \{\mathbf{l} : \Pi_{e^*}(\mathbf{l}) > g(e_1)\}$ as the set of such profiles. The wage bounds in (8) give a coordinate characterization: $D_{e^*} = O(e^*) \cap \Omega_{e^*}$, where

$$O(e^*) = \left\{\mathbf{l} : \ell_j > \ell_j^* \text{ for every } e_j \text{ with } c(e_j) < c(e^*)\right\}, \qquad \ell_j^* = \frac{c(e^*) - c(e_j)}{g(e^*) - g(e_1)}.$$

Only efforts cheaper than e^* enter the definition of $O(e^*)$: the expected wage required to implement e^* is determined by the incentive constraints against cheaper efforts, while costlier efforts affect only whether e^* is implementable. The sufficient condition below is stated in terms of $O(e^*)$ rather than D_{e^*}; the reason becomes clear in the proof.

Proposition 15. *(i) If there exists $\mathbf{l} \in L(e^*) \cap$ co $(\mathbf{f})$ such that conv$\big($co $(\mathbf{q}) \cup \{\mathbf{l}\}\big)$ is disjoint from the closure of D_j for every $e_j \neq e^*$ and disjoint from $O(e^*)$, then there is an admissible information structure that implements e^*, gives the agent $U_A = g(e^*) - g(e_1) - c(e^*)$, and leaves the principal with $g(e_1)$. (ii) If such an admissible information structure exists, then there is some $\mathbf{l} \in L(e^*) \cap$ co $(\mathbf{f})$ such that conv$\big($co $(\mathbf{q}) \cup \{\mathbf{l}\}\big)$ is disjoint from D_j for every*

$e_j \neq e^*$ *and from* $O(e^*) \cap \Omega_{e^*}$.

The sufficient and necessary conditions differ in two respects. First, they differ when the hull $\operatorname{conv}\big(\operatorname{co}(\mathbf{q}) \cup \{\mathbf{l}\}\big)$ meets the boundary of some D_j, where the principal is exactly indifferent between e^* and e_j. Second, they differ on $O(e^*) \setminus \Omega_{e^*}$, the set of profiles at which the incentive constraints against cheaper efforts would permit an expected wage below $g(e^*) - g(e_1)$, but e^* itself is not implementable. The principal cannot deviate to such a profile, so the necessary condition need not exclude it. The sufficient condition does exclude it because the construction in Appendix B perturbs the hull slightly, and a small perturbation can restore implementability at such a profile while the strict wage comparisons continue to hold.

Both differences disappear in two polar cases. When $\operatorname{co}(\mathbf{q})$ has full dimension, or more generally when $\mathbf{l}$ lies in the span of $\operatorname{co}(\mathbf{q})$, the construction in Appendix B requires no profile beyond the hull itself, and the characterization is exact using the open sets. Full extraction under the constraint then fails exactly when, for every profile in $L(e^*) \cap \operatorname{co}(\mathbf{f})$, its convex hull with $\operatorname{co}(\mathbf{q})$ intersects some D_j or D_{e^*}. When q is uninformative, so that $\operatorname{co}(\mathbf{q}) = \{\mathbf{0}\}$, the hull is the segment from the origin to $\mathbf{l}$. Along this segment, Lemma 4 rules out every deviation: scaling toward the origin raises the wage required to implement every effort, so no point on the segment lies in any D_j or in D_{e^*}. The characterization is therefore exact in this case as well, and Proposition 5 is recovered, with the hull condition reducing to the proposition's pointwise condition on $\mathbf{l}$.

The conditions are stated on the convex hull rather than separately on $\operatorname{co}(\mathbf{q})$ and $\mathbf{l}$ because the hull may intersect some D_j or $O(e^*)$ even when neither $\operatorname{co}(\mathbf{q})$ nor $\mathbf{l}$ does. Thus, a minimum-information constraint can prevent full extraction through its interaction with the extraction profile.

A.3 Single-Index Technologies

On the single-index class of Section 6, the minimum-information constraint also reduces to a restriction on the price shift. The set $\operatorname{co}(\mathbf{q})$ is a subsegment of the feasible line; let $A_q \geq A_{\min}$ denote the price shift associated with its endpoint on the reward side. We assume q is informative, so that $A_q < \infty$; otherwise the constraint is vacuous and the analysis of Section 6 applies unchanged. Admissibility requires the reward-side endpoint of $\operatorname{co}(\mathbf{p})$ to extend at least as far as that of $\operatorname{co}(\mathbf{q})$. The required extension on the opposite side is irrelevant for incentives: a paid signal on that side is less likely under every

positive-cost effort than under e_1, and therefore cannot implement a positive-cost effort under nonnegative wages. The agent's admissible choices can thus be represented by $A \in [A_{\min}, A_q]$. Hence, the minimum-information constraint places an upper bound on the price shift, while the technology imposes the lower bound $A_{\min}$.

Proposition 16. *Let c be strictly convex. (i) The agent's equilibrium payoff is weakly increasing in A_q. (ii) The implemented effort and total surplus are weakly decreasing in A_q. Thus, a stricter minimum-information constraint weakly raises both.*

Proposition 17. *Suppose, as in Proposition 13, that the agent's payoff V, evaluated at the principal's best response $e_B(A)$, is single-peaked in A with peak A°, and that V is positive at some achievable price shift. If $A^\circ > A_{\min}$ or $A_{\min}$ is attained, then $A^* = \min\{\max\{A_{\min}, A^\circ\}, A_q\}$. Otherwise, no optimal price shift exists. Thus, the agent's unconstrained choice in Proposition 13 is truncated at the upper bound A_q. The constraint is irrelevant when $A_q > \max\{A_{\min}, A^\circ\}$ and binds otherwise.*

Three observations follow. First, the comparative statics with respect to the constraint run in the opposite direction from those with respect to the technology. A more informative technology weakly benefits the agent (Proposition 12), whereas a tighter minimum-information constraint weakly reduces her payoff (Proposition 16). When $A^* = A_{\min}$, a reduction in $A_{\min}$ can also lower A^*, while when $A^* = A_q$, an increase in A_q can raise A^*. Under the single-peakedness condition of Proposition 17, once either bound becomes slack, relaxing it further has no effect. Second, tightening a binding minimum-information constraint weakly raises the implemented effort and total surplus while reducing the agent's rent. The equilibrium therefore moves toward the principal-design outcome and coincides with it when $A_q = A_{\min}$. Third, this welfare result is specific to the single-index class. Outside this class, tightening the constraint can strictly reduce equilibrium surplus.

B Proofs

Lemma 2. *Let T denote the lowest-dimensional linear subspace of $\mathbb{R}^{|E|}$ that contains $co(\mathbf{f})$. If $f(\cdot|e^*)$ has full support, then the origin is an interior point of $co(\mathbf{f})$ relative to T.*

Proof. Notice that the origin can be written as a convex combination of the points defining

co $(\mathbf{f})$ using the weights $f(x|e^*)$:

$$\sum_x f(x|e^*)\left(1-\frac{f(x|e_i)}{f(x|e^*)}\right)=\sum_x f(x|e^*)-\sum_x f(x|e_i)=1-1=0,\ \forall i,$$

where $f(x|e^*)>0,\ \forall x\in X$ by the full-support assumption. Thus the origin always belongs to co $(\mathbf{f})$. Suppose, toward a contradiction, that the origin is not an interior point of co $(\mathbf{f})$ relative to T. By the supporting hyperplane theorem, there would exist a hyperplane in the linear subspace T passing through the origin that contains co $(\mathbf{f})$ entirely on one side. Because $f(x|e^*)>0,\ \forall x\in X$, this is possible only if co $(\mathbf{f})$ lies entirely within that hyperplane. But this contradicts the definition of T as the lowest-dimensional linear subspace of $\mathbb{R}^{|E|}$ containing $co\,(\mathbf{f})$. Hence the origin must be an interior point of $co\,(\mathbf{f})$ relative to T. □

Lemma 3. *Fix $\mathbf{l}^*\in co\,(\mathbf{f})$ and fix a representation $\ell_i^*=1-\sum_{x\in X}\alpha_x\frac{f(x|e_i)}{f(x|e^*)},\forall i$, where $\alpha=(\alpha_x)_{x\in X}\in\Delta(X)$. Then for every $\lambda\in(0,\lambda_{max}(\alpha)]$, where $\lambda_{max}(\alpha):=\min_x\frac{f(x|e^*)}{\alpha_x}$, there exists an information structure (S,π) with $|S|=2$ such that (i) co $(\mathbf{p})$ is a line segment having $\mathbf{l}^*$ as one endpoint, and (ii) the other endpoint equals $\hat{\mathbf{l}}=-\frac{\lambda}{1-\lambda}\mathbf{l}^*$, which lies on the line through $\mathbf{l}^*$ and the origin, on the side opposite to $\mathbf{l}^*$. In particular, as $\lambda\to 0$, the endpoint $\hat{\mathbf{l}}$ converges to the origin, while at $\lambda=\lambda_{max}(\alpha)$ it reaches the maximal feasible point.*

Proof. Fix $\mathbf{l}^*$ and the associated weights $(\alpha_x)_{x\in X}$. For any $\lambda\in(0,\lambda_{\max}(\alpha)]$, define the information structure (S,π) with $S=\{L,H\}$ by

$$\pi\left(H|x\right)=\lambda\frac{\alpha_x}{f\left(x|e^*\right)},\pi\left(L|x\right)=1-\pi\left(H|x\right).$$

The bound $\lambda\leq\lambda_{\max}(\alpha)$ ensures that $\pi(H|x)\leq 1$. Then

$$\begin{aligned}1-\frac{p\left(H|e_i\right)}{p\left(H|e^*\right)}&=1-\frac{\sum_x\pi\left(H|x\right)f\left(x|e_i\right)}{\sum_x\pi\left(H|x\right)f\left(x|e^*\right)},\\&=1-\frac{\sum_x\frac{\alpha_x}{f(x|e^*)}f\left(x|e_i\right)}{\sum_x\frac{\alpha_x}{f(x|e^*)}f\left(x|e^*\right)}\\&=1-\frac{\sum_x\frac{\alpha_x}{f(x|e^*)}f\left(x|e_i\right)}{\sum_x\alpha_x}\\&=1-\sum_x\frac{\alpha_x}{f\left(x|e^*\right)}f\left(x|e_i\right)=\ell_i^*,\quad\forall i,\end{aligned}$$

so the likelihood-ratio point induced by signal H equals $\mathbf{l}^*$. Since $p(L|e^*) = 1 - \lambda > 0$, the likelihood-ratio point induced by signal L is well defined and equals

$$\hat{\ell}_i = 1 - \frac{p\,(L|e_i)}{p\,(L|e^*)} = -\frac{\lambda}{1-\lambda}\ell_i^*, \quad \forall i.$$

Hence $\mathrm{co}\,(\mathbf{p})$ is a line segment with one endpoint at $\mathbf{l}^*$ and another at $\hat{\mathbf{l}} = -t\mathbf{l}^*$ with $t = \lambda/(1-\lambda)$, as claimed. □

Lemma 4. *For any effort e_i, any $\mathbf{l} \in \Omega_i$, and any $s \in (0,1]$, $s\mathbf{l} \in \Omega_i$ and $W_i(s\mathbf{l}) \geq W_i(\mathbf{l})$.*

Proof. All coordinate differences scale by s, so every bound in (8) scales by $1/s$, while membership in Ω_i is unchanged. Writing M for the maximum term in $W_i(\mathbf{l})$, we have $M \geq 0$: if $c(e_i) > 0$, implementability requires $\ell_1 > \ell_i$, so the constraint against e_1 gives a positive chord; if $c(e_i) = 0$, then $W_i \equiv 0$ and the claim is immediate. Thus, for $s \in (0,1]$,

$$W_i(s\mathbf{l}) = (1 - s\ell_i)\frac{M}{s} \geq (1-\ell_i)M = W_i(\mathbf{l}),$$

since $1 - s\ell_i \geq s(1-\ell_i)$. □

Proof of Proposition 1. Let $\Delta \equiv g(e_{FB}) - g(e_1)$. If $\Delta = 0$, then $g(e_{FB}) - c(e_{FB}) \geq g(e_1) - c(e_1)$ and $c(e_1) = 0$ imply $c(e_{FB}) = 0$. Hence, under the tie-breaking conventions of Section 2, the zero contract $w \equiv 0$ implements e_{FB}, with $u_A = 0 = g(e_{FB}) - c(e_{FB}) - g(e_1)$, while the principal obtains $g(e_1)$. Suppose, therefore, that $\Delta > 0$. Consider the binary information structure $S = L, H$ defined by

$$\pi(H|e) = \begin{cases} 1 - \frac{c(e_{FB}) - c(e)}{\Delta} & \text{if } c(e) < c(e_{FB}) \\ 1 & \text{if } c(e) \geq c(e_{FB}) \end{cases}.$$

This is the construction in 5 with Δ in place of $g(e_{FB})$; the two coincide under the normalization $g(e_1) = 0$ used in the text. Note that $\pi(H \mid e) \in [0,1]$. It is minimized at e_1, where $\pi(H \mid e_1) = 1 - \frac{c(e_{FB})}{\Delta} \geq 0$, since $g(e_{FB}) - c(e_{FB}) \geq g(e_1) - c(e_1)$. For any effort level $\hat{e}$ that the principal wishes to implement, wages must satisfy

$$p(H \mid \hat{e})w(H) + \big(1 - p(H \mid \hat{e})\big)w(L) - c(\hat{e}) \geq p(H \mid e)w(H) + \big(1 - p(H \mid e)\big)w(L) - c(e),$$

for every $e \in E$. These constraints immediately imply that any effort satisfying $c(\hat{e}) > c(e_{FB})$ is not implementable, since the agent can deviate to e_{FB} without changing the

distribution of signals and strictly reduce her effort cost. Hence, the principal is restricted to efforts satisfying $c(\hat{e}) \leq c(e_{FB})$. Consider now an arbitrary $\hat{e}$ with $c(\hat{e}) \leq c(e_{FB})$. Using $p(H \mid e) = \pi(H \mid e)$, substituting the definition of π , and simplifying gives

$$\frac{w(H) - w(L) - \Delta}{\Delta} c(\hat{e}) \geq \frac{w(H) - w(L) - \Delta}{\Delta} c(e)$$

for every e satisfying $c(e) \leq c(e_{FB})$. If $c(\hat{e}) > 0$, comparison with e_1 requires $w(H) - w(L) \geq \Delta$. If, in addition, $c(\hat{e}) < c(e_{FB})$, comparison with e_{FB} requires the reverse inequality, so the wage spread must equal Δ. If $c(\hat{e}) = c(e_{FB})$, the minimum feasible spread is again Δ. Thus, the cheapest contract implementing any positive-cost effort has $w(H) - w(L) = \Delta$. At this spread, the agent is indifferent among all efforts satisfying $c(e) \leq c(e_{FB})$. Given this wage spread, the principal's expected payoff from implementing $\hat{e}$ is $g(\hat{e}) - w(L) - p(H \mid \hat{e})\big[w(H) - w(L)\big]$. Limited liability therefore makes it optimal to set $w(L) = 0$, and hence $w(H) = \Delta$. Substituting for $p(H \mid \hat{e})$, the principal's payoff becomes $\big[g(\hat{e}) - c(\hat{e})\big] - \big[g(e_{FB}) - c(e_{FB})\big] + g(e_1)$. Since e_{FB} maximizes total surplus, this expression is no greater than $g(e_1)$, with equality for any surplus-maximizing effort. The principal can also obtain $g(e_1)$ by implementing e_1 under the zero contract. By the tie-breaking conventions of Section 2, he selects the contract intended to implement e_{FB}, and the agent complies with that intended effort. The agent's payoff is therefore $\pi(H \mid e_{FB})w(H) - c(e_{FB}) = \Delta - c(e_{FB}) = g(e_{FB}) - g(e_1) - c(e_{FB})$, while the principal obtains $g(e_1)$. □

Proof of Proposition 2. Let (S, π) be an information structure. For any $s \in S$ and any $i = 1, \ldots, |E|$,

$$\frac{p(s|e_i)}{p(s|e^*)} = \frac{\sum_x f(x|e_i)\pi(s|x)}{\sum_x f(x|e^*)\pi(s|x)} = \sum_x \frac{f(x|e^*)\pi(s|x)}{\sum_{x'} f(x'|e^*)\pi(s|x')} \frac{f(x|e_i)}{f(x|e^*)} = \sum_x \mu_s(x) \frac{f(x|e_i)}{f(x|e^*)},$$

where $\sum_x \mu_s(x) = 1$ and $\mu_s(x)$ is the principal's posterior probability of x after observing s (under e^*). Hence,

$$\left(1 - \frac{p(s|e_1)}{p(s|e^*)}, \cdots, 1 - \frac{p(s|e_{|E|})}{p(s|e^*)}\right) = \sum_x \mu_s(x) \left(1 - \frac{f(x|e_1)}{f(x|e^*)}, \cdots, 1 - \frac{f(x|e_{|E|})}{f(x|e^*)}\right).$$

Therefore, the left hand side is a member of $\text{co}(\mathbf{f})$ for all $s \in S$, so $\text{co}(\mathbf{p}) \subset \text{co}(\mathbf{f})$.

Moreover, since $f(x|e^*) > 0$ for all x, we have

$$p(s|e^*) = \sum_x f(x|e^*)\pi(s|x) > 0$$

for all non-redundant signals $s \in S$. Without loss of generality, we restrict attention to such signals. Now note that, for each i,

$$\sum_s p(s|e^*)\left(1 - \frac{p(s|e_i)}{p(s|e^*)}\right) = 1 - \sum_s p(s|e_i) = 0.$$

Thus $\mathbf{0} \in \text{co}\,(\mathbf{p})$. Since $p(s|e^*) > 0$ for all non-redundant signals $s \in S$, by an argument analogous to that in Lemma 2, the origin lies in the relative interior of co $(\mathbf{p})$.

Let $C \subseteq \text{co}\,(\mathbf{f})$ be a closed convex set with the origin in its relative interior and extreme points $Z = \{\mathbf{l}^1, \ldots, \mathbf{l}^{|Z|}\}$. Since $C \subseteq \text{co}\,(\mathbf{f})$, for each z there exists $\mu_z \in \Delta(X)$ such that

$$\mathbf{l}^z = \sum_{x \in X} \mu_z(x)\left(1 - \frac{f(x|e_i)}{f(x|e^*)}\right), \qquad i = 1, \ldots, |E|\,.$$

Because the origin is in the relative interior of C, there exist $\lambda_z > 0$ with $\sum_z \lambda_z = 1$ and $\sum_z \lambda_z l^z = \mathbf{0}$.

Define $\delta = \min_{x \in X} \frac{f(x|e^*)}{\sum_{z=1}^{|Z|} \lambda_z \mu_z(x)} > 0$, and choose $\varepsilon > 0$ such that $\varepsilon < \min\{\delta, 1\}$. Let $S = Z \cup \{0\}$ and define

$$\pi(z \mid x) = \varepsilon \frac{\lambda_z \mu_z(x)}{f(x \mid e^*)}, \quad \text{for } z = 1, \ldots, |Z|\,, \qquad \pi(0 \mid x) = 1 - \sum_{z=1}^{|Z|} \pi(z \mid x).$$

Then $\pi(\cdot \mid x)$ is a probability distribution.

Let $p(\cdot|e)$ be the induced signal distribution. For $z = 1, \ldots, |Z|$,

$$p(z|e^*) = \sum_{x \in X} \pi(z \mid x)\, f(x \mid e^*) = \varepsilon \lambda_z > 0, \qquad \Pr(x|z, e^*) = \frac{\pi(z|x)\, f(x|e^*)}{p(z|e^*)} = \mu_z(x).$$

Hence,

$$1 - \frac{p(z \mid e_i)}{p(z \mid e^*)} = \sum_{x \in X} \mu_z(x)\left(1 - \frac{f(x \mid e_i)}{f(x \mid e^*)}\right) = \mathbf{l}_i^z,$$

and $p(0|e^*) = 1 - \varepsilon$. Let $\mathbf{l}^0$ be the likelihood–ratio vector for signal 0. We have

$$\sum_{s \in S} p(s|e^*)\, \mathbf{l}^s = \mathbf{0} \Rightarrow (1-\varepsilon)\mathbf{l}^0 + \sum_{z=1}^{|Z|} \varepsilon \lambda_z \mathbf{l}^z = \mathbf{0}.$$

Using $\sum_{z=1}^{|Z|} \lambda_z \mathbf{l}^z = \mathbf{0}$,

$$(1-\varepsilon)\mathbf{l}^0 = \sum_{z=1}^{|Z|} (1-\varepsilon)\lambda_z\, \mathbf{l}^z \Rightarrow \mathbf{l}^0 = \sum_{z=1}^{|Z|} \lambda_z \mathbf{l}^z = \mathbf{0}.$$

Hence, $\mathbf{l}^0 \in \text{co}\left(\{\mathbf{l}^z : z = 1, \ldots, |Z|\}\right) = C$. Hence, $\text{co}\left(\mathbf{p}\right) = \text{conv}\{\mathbf{l}^1, \ldots, \mathbf{l}^{|Z|}, \mathbf{l}^0\} = \text{conv}(Z)$. Since C is closed, bounded, and convex, it equals the convex hull of its extreme points. Therefore, $\text{co}\left(\mathbf{p}\right) = C$. □

Proof of Proposition 3. Suppose the effort level e^* is implementable by (S, π), and consider the optimization problem in (9). Let $\mathbf{l}^* \in \text{co}\left(\mathbf{p}\right)$ be the principal's optimal choice of likelihood-ratio profile when implementing e^*. By the definition of $\text{co}\left(\mathbf{p}\right)$, there exist weights $\{\alpha_s\}_{s \in S}$ with $\alpha_s \geq 0$ and $\sum_{s \in S} \alpha_s = 1$ such that

$$\ell_i^* = 1 - \sum_{s \in S} \alpha_s \frac{p\left(s|e_i\right)}{p\left(s|e^*\right)}, \forall i = 1, \cdots, |E|\,.$$

Define the information structure $\left(\hat{S}, \hat{\pi}\right)$ by $\hat{S} = \{L, H\}$ and

$$\hat{\pi}\left(H|x\right) = \sum_{s} \beta_s \pi\left(s|x\right), \quad \hat{\pi}\left(L|x\right) = 1 - \hat{\pi}\left(H|x\right),$$

where $\beta_s = \frac{\alpha_s/p(s|e^*)}{\sum_s \alpha_s/p(s|e^*)}$. Then

$$\begin{aligned}
1 - \frac{\hat{p}\left(H|e_i\right)}{\hat{p}\left(H|e^*\right)} &= 1 - \frac{\sum_x \hat{\pi}\left(H|x\right) f\left(x|e_i\right)}{\sum_x \hat{\pi}\left(H|x\right) f\left(x|e^*\right)}, \\
&= 1 - \frac{\sum_{s\in S} \frac{\alpha_s}{p(s|e^*)} \sum_x \pi\left(s|x\right) f\left(x|e_i\right)}{\sum_{s\in S} \frac{\alpha_s}{p(s|e^*)} \sum_x \pi\left(s|x\right) f\left(x|e^*\right)} \\
&= 1 - \frac{\sum_{s\in S} \frac{\alpha_s}{p(s|e^*)} p\left(s|e_i\right)}{\sum_{s\in S} \frac{\alpha_s}{p(s|e^*)} p\left(s|e^*\right)} \\
&= 1 - \sum_{s\in S} \frac{\alpha_s}{p\left(s|e^*\right)} p\left(s|e_i\right) = \ell_i^*, \forall i = 1, \cdots, |E|.
\end{aligned}$$

Thus $\mathbf{l}^* \in \text{co}\left(\hat{\mathbf{p}}\right)$. In particular, if the principal chooses e^*, the likelihood-ratio profile that minimizes his cost under (S, π) remains feasible under the binary information structure $\left(\hat{S}, \hat{\pi}\right)$. To complete the argument, we must show that for any alternative $e_i \neq e^*$, $W\left(e_i, \hat{\pi}\right) \geq W\left(e_i, \pi\right)$. It is sufficient to show that the likelihood-ratio point associated with signal L,

$$\left(1 - \frac{\hat{p}\left(L|e_1\right)}{\hat{p}\left(L|e^*\right)}, \cdots, 1 - \frac{\hat{p}\left(L|e_{|E|}\right)}{\hat{p}\left(L|e^*\right)}\right),$$

lies in co $(\mathbf{p})$. If this holds, then co $\left(\hat{\mathbf{p}}\right)$, which is the line segment between this point and $\mathbf{l}^*$, is a subset of co $(\mathbf{p})$, ensuring that the principal cannot achieve a lower cost under $\left(\hat{S}, \hat{\pi}\right)$. We compute:

$$\begin{aligned}
1 - \frac{\hat{p}\left(L|e_i\right)}{\hat{p}\left(L|e^*\right)} &= 1 - \frac{\sum_x \left(1 - \hat{\pi}\left(H|x\right)\right) f\left(x|e_i\right)}{\sum_x \left(1 - \hat{\pi}\left(H|x\right)\right) f\left(x|e^*\right)} \\
&= 1 - \frac{\sum_x \sum_s \left(1 - \beta_s\right) \pi\left(s|x\right) f\left(x|e_i\right)}{\sum_x \sum_s \left(1 - \beta_s\right) \pi\left(s|x\right) f\left(x|e^*\right)} \\
&= 1 - \frac{\sum_s \left(1 - \beta_s\right) p\left(s|e_i\right)}{\sum_s \left(1 - \beta_s\right) p\left(s|e^*\right)} \\
&= 1 - \sum_s \frac{\left(1 - \beta_s\right) p\left(s|e^*\right)}{\sum_s \left(1 - \beta_s\right) p\left(s|e^*\right)} \frac{p\left(s|e_i\right)}{p\left(s|e^*\right)}.
\end{aligned}$$

The expression is thus a convex combination of the likelihood-ratio vectors defining co $(\mathbf{p})$, so it lies in co $(\mathbf{p})$. Therefore, co $\left(\hat{\mathbf{p}}\right) \subseteq$ co $(\mathbf{p})$, which implies $W\left(e_i, \hat{\pi}\right) \geq W\left(e_i, \pi\right)$ for every $e_i \neq e^*$ and proves the result.

The construction extends beyond finite E and finite X. Suppose e^* is implementable

by (S, π) and let w be a contract solving the principal's problem. If $w \equiv 0$, an uninformative binary structure with zero wage yields the same outcome. Otherwise, define the information structure $(\hat{S}, \hat{\pi})$ by $\hat{S} = \{L, H\}$ and

$$\hat{\pi}(H \mid x) = \sum_{s \in S} \beta_s \, \pi(s \mid x), \quad \hat{\pi}(L \mid x) = 1 - \hat{\pi}(H \mid x),$$

where now $\beta_s = w(s) / \sum_{s' \in S} w(s')$, and set $\hat{w}(H) = \sum_{s \in S} w(s)$ and $\hat{w}(L) = 0$. Then, for every $e \in E$, $\hat{w}(H)\, \hat{p}(H \mid e) = \sum_{s \in S} w(s)\, p(s \mid e)$, so the expected wage at every effort is unchanged. The agent therefore faces the same choice problem, with the same best responses and the same ties, and e^* is induced at the same expected wage. Moreover, any contract (w_H, w_L) under $(\hat{S}, \hat{\pi})$ generates the same expected-wage schedule as the contract $v(s) = w_L + (w_H - w_L)\, \beta_s \geq 0$ under (S, π). Thus, every expected-wage schedule available to the principal under $(\hat{S}, \hat{\pi})$ is also available under (S, π). Since payoffs depend on a contract only through its expected-wage schedule, the contract constructed above remains optimal for the principal. Hence e^* solves his problem under $(\hat{S}, \hat{\pi})$, with the same payoffs for both parties. The argument uses no structure on E. When X is a continuum, the sums defining $p(\cdot \mid e)$ are replaced by integrals.

Finally, the restriction to finite S plays no role beyond guaranteeing that the wage is bounded. If S is infinite and the wage w is bounded, the same construction applies by setting $\beta_s = w(s)/M$ for any constant $M \geq \sup_{s \in S} w(s)$ and paying $\hat{w}(H) = M$. The sums over S are then replaced by integrals. □

Proof of Proposition 4. Fix $\mathbf{l} \in \operatorname{co}(\mathbf{f})$ and a representation $\alpha \in \Delta(X)$. By Lemma 3, there exists a binary information structure whose likelihood-ratio set is the segment from $\mathbf{l}$ to $-\frac{\lambda}{1-\lambda}\mathbf{l}$. For λ sufficiently small, no positive-cost effort implemented on the negative portion of the segment gives the principal as much as $g(e_1)$.[31] On the positive portion,

[31] The principal's gain from implementing any effort is bounded by $\max_k [g(e_k) - g(e_1)]$, which is finite. On the negative portion of the ray, each positive-cost effort is either not implementable or requires an expected wage that diverges as the endpoint approaches the origin. Hence, the endpoint can be chosen sufficiently close to the origin that every implementable positive-cost effort requires a wage exceeding this bound. No effort implemented on that portion can then give the principal as much as $g(e_1)$. Explicitly, write the unpaid endpoint as $-s\mathbf{l}$, with $s > 0$. For each positive-cost effort e_k, Lemma 1 implies that the chords at $-s\mathbf{l}$ scale by $1/s$, so the maximum in (9) is B_k/s, where $B_k \equiv \max_{j:\, \ell_j \leq \ell_k} \frac{c(e_k) - c(e_j)}{\ell_k - \ell_j}$. If $B_k \leq 0$, then e_k is not implementable at $-s\mathbf{l}$ at any nonnegative wage. If $B_k > 0$, the required expected wage is $(1 + s\ell_k)\frac{B_k}{s}$. Let $M \equiv \max_j \big[g(e_j) - g(e_1)\big]$. This wage exceeds M whenever $s < \frac{B_k}{M + |\ell_k| B_k}$. Taking the minimum over all k with $B_k > 0$ gives the required bound on s; if no such k exists, no restriction on s is needed. The paid signal then occurs with probability $\lambda = \frac{s}{1+s}$.

Lemma 4 implies that $\mathbf{l}$ is the cheapest point at which to implement each effort. The principal therefore solves $\max_k \Pi_k(\mathbf{l})$, and, by the tie-breaking convention of Section 2, the agent receives $R(\mathbf{l})$. Since this construction is available for every $\mathbf{l} \in \text{co}(\mathbf{f})$, the value of the agent's problem is at least $\sup_{\mathbf{l}\in\text{co}(\mathbf{f})} R(\mathbf{l})$.

For the reverse inequality, fix any information structure and a continuation equilibrium. If the principal implements e_1, the agent's payoff is zero, which is at most $\sup R$. Indeed, $R \geq 0$, since for any implementable effort e_i, the incentive constraint against e_1 implies$W_i(\mathbf{l}) - c(e_i) \geq 0$. So suppose he implements some e^* with $c(e^*) > 0$. By Lemma 1, the principal's contract corresponds to some $\mathbf{l}' \in \text{co}(\mathbf{p}) \cap \Omega_{e^*}$, and Proposition 2 gives $\text{co}(\mathbf{p}) \subseteq \text{co}(\mathbf{f})$. Optimality of e^* implies, for every k,

$$\Pi_{e^*}(\mathbf{l}') \geq g(e_k) - \min_{\mathbf{l}\in\text{co}(\mathbf{p})\cap\Omega_k} W_k(\mathbf{l}) \geq \Pi_k(\mathbf{l}').$$

Hence $e^* \in \arg\max_k \Pi_k(\mathbf{l}')$, and the agent's payoff, $W_{e^*}(\mathbf{l}') - c(e^*)$, is at most $R(\mathbf{l}')$. Since $\mathbf{l}' \in \text{co}(\mathbf{f})$, it is at most $\sup_{\mathbf{l}\in\text{co}(\mathbf{f})} R(\mathbf{l})$. □

Proof of Proposition 5. Suppose an information structure (S, π) implements e^* with the stated payoffs. By Lemma 1, the principal's cost-minimizing contract corresponds to a likelihood-ratio point $\mathbf{l} \in \text{co}(\mathbf{p}) \cap \Omega_{e^*}$, and by Proposition 2, $\text{co}(\mathbf{p}) \subseteq \text{co}(\mathbf{f})$. Since the principal's payoff is $g(e_1)$, the expected wage required to implement e^* is $W_{e^*}(\mathbf{l}) = g(e^*) - g(e_1)$, which is equivalent to $\Pi_{e^*}(\mathbf{l}) = g(e_1)$, and hence $\mathbf{l} \in L(e^*)$. Moreover, since the principal's optimal payoff is $g(e_1)$ and $\mathbf{l} \in \text{co}(\mathbf{p})$, implementing any rival effort e_j at $\mathbf{l}$ can yield at most $g(e_1)$. Thus, $\Pi_j(\mathbf{l}) \leq g(e_1)$, with $\Pi_j(\mathbf{l}) = -\infty$ if e_j is not implementable at $\mathbf{l}$. Hence, $\mathbf{l} \notin D_j$. Conversely, fix $\mathbf{l} \in L(e^*) \cap \text{co}(\mathbf{f})$ such that $\mathbf{l} \notin D_j$ for every $e_j \neq e^*$. By Lemma 3, there exists a binary information structure whose likelihood-ratio set is a line segment with endpoint $\mathbf{l}$ and opposite endpoint $-\frac{\lambda}{1-\lambda}\mathbf{l}$. Efforts potentially implemented on the negative portion of the segment can be ignored. By Lemma 3, the negative endpoint can be chosen arbitrarily close to the origin, making the wage required to implement any positive-cost effort there arbitrarily large and therefore never optimal for the principal (see footnote 31 in the proof of Proposition 4). On the positive portion, Lemma 4 implies that the cost of implementing any effort is minimized at the endpoint $\mathbf{l}$. The principal's problem therefore reduces to choosing which effort to implement at $\mathbf{l}$. Since $\mathbf{l} \in L(e^*)$, implementing e^* yields exactly $g(e_1)$. Since $\mathbf{l} \notin D_j$, no rival effort gives the principal strictly more than $g(e_1)$ at $\mathbf{l}$, while implementing e_1 at zero wage yields $g(e_1)$. The principal's optimal set therefore contains both e^* and e_1, each yielding $g(e_1)$.

By the tie-breaking convention of Section 2, the principal offers the contract implementing the effort the information structure is intended to implement, namely e^*, and the agent complies. The agent's payoff is $W_{e^*}(\mathbf{l}) - c(e^*) = g(e^*) - g(e_1) - c(e^*)$, while the principal's payoff is $g(e_1)$. Hence, an information structure with the stated payoffs exists if and only if some point in $L(e^*) \cap \text{co}(\mathbf{f})$ lies outside every D_j. □

Proof of Proposition 6. Suppose $\mathbf{l}^* \in \text{co}(\mathbf{f})$, and let $\Delta \equiv g(e_{FB}) - g(e_1) > 0$. Since likelihood-ratio coordinates are defined relative to the target e_{FB}, we have $\ell^*_{FB} = 0$. By definition of $\mathbf{l}^*$, $\ell^*_i = \frac{c(e_{FB}) - c(e_i)}{\Delta}$ for every i. Hence, $\ell^*_i - \ell^*_j = \frac{c(e_j) - c(e_i)}{\Delta}$, so every nontrivial chord at $\mathbf{l}^*$ equals Δ. Consider first the target effort e_{FB}. For every effort with $\ell^*_i > 0$, and there is at least one since $\Delta > 0$ implies $c(e_{FB}) > 0 = c(e_1)$, the corresponding lower bound on the expected wage is $\frac{c(e_{FB}) - c(e_i)}{\ell^*_i} = \Delta$. For every effort with $\ell^*_i < 0$, the corresponding upper bound is also Δ. Thus, the lower and upper incentive bounds coincide, so e_{FB} is implementable at $\mathbf{l}^*$ and $W_{FB}(\mathbf{l}^*) = \Delta$. It follows that $\Pi_{FB}(\mathbf{l}^*) = g(e_{FB}) - \Delta = g(e_1)$, and therefore $\mathbf{l}^* \in L(e_{FB})$. Now consider any rival effort e_j with $c(e_j) > 0$. Since some effort is then strictly cheaper than e_j, and since every nontrivial chord at $\mathbf{l}^*$ equals Δ, the lower and upper incentive bounds for implementing e_j again coincide. Hence e_j is implementable at $\mathbf{l}^*$ and $W_j(\mathbf{l}^*) = (1 - \ell^*_j)\Delta = \Delta + c(e_j) - c(e_{FB})$. Therefore,

$$\Pi_j(\mathbf{l}^*) = g(e_j) - W_j(\mathbf{l}^*) = g(e_1) + \big[g(e_j) - c(e_j)\big] - \big[g(e_{FB}) - c(e_{FB})\big] \leq g(e_1),$$

where the inequality follows because e_{FB} maximizes total surplus. The only remaining rival is e_1, for which $\Pi_1(\mathbf{l}^*) = g(e_1)$, so $D_1 = \varnothing$. Thus, $\mathbf{l}^* \notin D_j$ for every $e_j \neq e_{FB}$. Proposition 5 therefore applies with $e^* = e_{FB}$ and yields an information structure implementing e_{FB} with agent payoff $g(e_{FB}) - g(e_1) - c(e_{FB})$ and principal payoff $g(e_1)$.

More generally, let $\mathbf{l} \in B \cap \text{co}(\mathbf{f})$. By definition of B, $\ell_1 = \ell^*_1$ and $\ell_i \geq \ell^*_i$ for every i. Consider first the target effort e_{FB}. For every effort with $\ell_i > 0$, the corresponding lower bound is $\frac{c(e_{FB}) - c(e_i)}{\ell_i} = \frac{\ell^*_i \Delta}{\ell_i} \leq \Delta$, with equality at $i = 1$. Similarly, for every effort with $\ell_i < 0$, the corresponding upper bound is $\frac{c(e_{FB}) - c(e_i)}{\ell_i} = \frac{\ell^*_i \Delta}{\ell_i} \geq \Delta$, since $\ell^*_i \leq \ell_i < 0$. Thus, the lower and upper incentive bounds are compatible, and the constraint against e_1 binds, so e_{FB} is implementable at $\mathbf{l}$ and $W_{FB}(\mathbf{l}) = \Delta$. It follows that $\Pi_{FB}(\mathbf{l}) = g(e_{FB}) - \Delta = g(e_1)$, and therefore $\mathbf{l} \in L(e_{FB})$. Now consider any rival effort e_j with $c(e_j) > 0$. If $\ell_j = \ell_1$, then e_1 has the same coordinate and strictly lower cost, so by the conventions of footnote 9, e_j is not implementable at $\mathbf{l}$. If $\ell_j > \ell_1$, the constraint against e_1 imposes

the upper bound $\frac{c(e_j)}{\ell_1-\ell_j} < 0$, so e_j is again not implementable. In either case, $\mathbf{l} \notin D_j$. It remains to consider $\ell_j < \ell_1$. The constraint against e_1 then gives $W_j(\mathbf{l}) \geq \frac{(1-\ell_j)c(e_j)}{\ell_1-\ell_j}$. Since $\ell_1 = \ell_1^*$, $\ell_j \geq \ell_j^*$, and $x \mapsto (1-x)/(\ell_1 - x)$ is increasing on $(-\infty, \ell_1)$, we obtain

$$W_j(\mathbf{l}) \geq \frac{(1-\ell_j)c(e_j)}{\ell_1 - \ell_j} \geq \frac{(1-\ell_j^*)c(e_j)}{\ell_1^* - \ell_j^*} = (1-\ell_j^*)\Delta.$$

Therefore,

$$\Pi_j(\mathbf{l}) \leq g(e_1) + \big[g(e_j) - c(e_j)\big] - \big[g(e_{FB}) - c(e_{FB})\big] \leq g(e_1),$$

where the final inequality follows because e_{FB} maximizes total surplus. The only remaining rival is e_1, for which $\Pi_1(\mathbf{l}) = g(e_1)$, so $D_1 = \varnothing$. Thus, $\mathbf{l} \notin D_j$ for every $e_j \neq e_{FB}$. The conclusion of the proposition therefore holds at every point of $B \cap \operatorname{co}(\mathbf{f})$. □

Proof of Proposition 7. Suppose $\mathbf{l}(e^*) \in \operatorname{co}(\mathbf{f})$, and let $\Delta^* \equiv g(e^*) - g(e_1) > 0$. Since likelihood-ratio coordinates are defined relative to the target e^*, we have $\ell_{e^*}(e^*) = 0$. By definition of $\mathbf{l}(e^*)$, $\ell_j(e^*) = \frac{c(e^*)-c(e_j)}{\Delta^*}$ for every effort cheaper than e^*, while $\ell_j(e^*) = 0$ for every effort at least as costly as e^*. Consider first the target effort e^*. For every cheaper effort, and there is at least one since $\Delta^* > 0$ implies $c(e^*) > 0 = c(e_1)$, the corresponding lower bound on the expected wage is $\frac{c(e^*)-c(e_j)}{\ell_j(e^*)} = \Delta^*$. Every effort at least as costly as e^* satisfies $\ell_j(e^*) = 0 = \ell_{e^*}(e^*)$ and imposes no constraint, since (8) then holds trivially. Hence, e^* is implementable at $\mathbf{l}(e^*)$ and $W_{e^*}(\mathbf{l}(e^*)) = \Delta^*$. It follows that $\Pi_{e^*}(\mathbf{l}(e^*)) = g(e^*) - \Delta^* = g(e_1)$, and therefore $\mathbf{l}(e^*) \in L(e^*)$. Now consider any rival effort e_j with $0 < c(e_j) \leq c(e^*)$. Some effort is then strictly cheaper than e_j, every nontrivial lower bound equals Δ^*, and every upper bound is at least Δ^*, with equality for deviations costing at most $c(e^*)$. Hence, e_j is implementable at $\mathbf{l}(e^*)$ and $W_j(\mathbf{l}(e^*)) = (1-\ell_j(e^*))\Delta^*$. Therefore,

$$\Pi_j(\mathbf{l}(e^*)) = g(e_j) - W_j(\mathbf{l}(e^*)) = g(e_1) + \big[g(e_j) - c(e_j)\big] - \big[g(e^*) - c(e^*)\big] \leq g(e_1),$$

where the inequality follows by surplus monotonicity. Any rival effort e_j strictly costlier than e^* is not implementable at $\mathbf{l}(e^*)$, since the constraint against e^* in (8) becomes $0 \geq c(e_j) - c(e^*) > 0$. Hence, $\Pi_j(\mathbf{l}(e^*)) = -\infty$. The only remaining rival is e_1, for which $\Pi_1(\mathbf{l}(e^*)) = g(e_1)$, so $D_1 = \varnothing$. Thus, $\mathbf{l}(e^*) \notin D_j$ for every $e_j \neq e^*$. Proposition 5 therefore applies and yields an information structure implementing e^* with agent payoff

$g(e^*) - g(e_1) - c(e^*)$ and principal payoff $g(e_1)$. □

Proof of Proposition 8. Write $\Delta^* \equiv g(e^*) - g(e_1) > 0$, and let $\mathbf{l} \in \text{co}(\mathbf{f})$ denote the likelihood-ratio vector generated by the specified output. In these coordinates, the three inequalities state that (i) $\ell_1 \geq c(e^*)/\Delta^*$, (ii) $\ell_j \geq 0$ for every effort strictly costlier than e^*, and (iii) $\big(c(e^*) - c(e_j)\big)\ell_1 < c(e^*)\,\ell_j$ for every rival effort with $0 < c(e_j) \leq c(e^*)$. Condition (iii) implies $\ell_j > 0$ for every such effort, since its left-hand side is non-negative. Conditions (ii) and (iii) are invariant to positive scaling, and by Lemma 2 and convexity, $t\mathbf{l} \in \text{co}(\mathbf{f})$ for every $t \in (0, 1]$. By condition (i), we can choose t such that $t\ell_1 = c(e^*)/\Delta^*$. Set $\mathbf{l}' = t\mathbf{l}$. Consider first the target effort e^*. Since likelihood-ratio coordinates are defined relative to e^*, we have $\ell'_{e^*} = 0$. The deviation to e_1 gives the lower bound $c(e^*)/\ell'_1 = \Delta^*$, while by (iii), every rival effort with $0 < c(e_j) \leq c(e^*)$ gives the strictly smaller lower bound $\big(c(e^*) - c(e_j)\big)/\ell'_j < \Delta^*$. Efforts strictly costlier than e^* satisfy $\ell'_j \geq 0 = \ell'_{e^*}$ while costing more, so they impose no lower bound. Moreover, no effort has a negative coordinate, so there are no upper bounds. Hence e^* is implementable at $\mathbf{l}'$ and $W_{e^*}(\mathbf{l}') = \Delta^*$. It follows that $\Pi_{e^*}(\mathbf{l}') = g(e^*) - \Delta^* = g(e_1)$, and therefore $\mathbf{l}' \in L(e^*)$. Now consider any rival effort e_j with $0 < c(e_j) \leq c(e^*)$. If $\ell'_j \geq \ell'_1$, then e_1 has a weakly lower coordinate and strictly lower cost, so e_j is not implementable at $\mathbf{l}'$. Otherwise, the deviation to e_1 gives the lower bound $c(e_j)/(\ell'_1 - \ell'_j)$ while the deviation to e^* gives the upper bound $\big(c(e^*) - c(e_j)\big)/\ell'_j$. Condition (iii) can be written as $c(e_j)\,\ell'_j > \big(c(e^*) - c(e_j)\big)\big(\ell'_1 - \ell'_j\big)$, which is equivalent to the lower bound exceeding the upper bound. Hence, e_j is not implementable at $\mathbf{l}'$. Any rival effort e_j strictly costlier than e^* is also not implementable at $\mathbf{l}'$, since the constraint against e^* in (8) implies $0 \geq c(e_j) - c(e^*) > 0$. Thus, $\Pi_j(\mathbf{l}') = -\infty$ in all these cases. The only remaining rival is e_1, for which $\Pi_1(\mathbf{l}') = g(e_1)$, so $D_1 = \varnothing$. Thus, $\mathbf{l}' \notin D_j$ for every $e_j \neq e^*$. Proposition 5 therefore applies and yields an information structure implementing e^* with agent payoff $g(e^*) - g(e_1) - c(e^*)$ and principal payoff $g(e_1)$. □

Proof of Corollary 1. By Proposition 2, the agent's feasible sets are the convex subsets of $\text{co}(\mathbf{f})$ that contain the origin in their relative interior, and each such set is induced by an information structure. Since $\text{co}(\mathbf{f}') \subseteq \text{co}(\mathbf{f})$, every set feasible under f' is also feasible under f. By Lemma 1, continuation play depends on an information structure only through the induced set $\text{co}(\mathbf{p})$, so the same set generates the same continuation payoff under either technology. Hence, every payoff attainable under f' is also attainable under f, and the agent's equilibrium payoff is weakly higher under f. □

Proof of Proposition 9. For $|E| \leq 2$, the result is immediate. Suppose therefore that $E = \{e_1, e_2, e_3\}$, with $c(e_1) = 0$. For a technology with likelihood-ratio set $\text{co}(\mathbf{f})$, define

$$R_i = \sup \{W_i(\mathbf{l}) - c(e_i) : \mathbf{l} \in \text{co}(\mathbf{f}),\ e_i \text{ is implemented at } \mathbf{l}\}.$$

By Lemma 1 and Proposition 2, the equilibrium payoff is the rent at some point in $\text{co}(\mathbf{f})$. Conversely, by Lemmas 3 and 4, together with footnote 31 in the proof of Proposition 6, the agent can secure the rent at any point of $\text{co}(\mathbf{f})$. Since her equilibrium choice is optimal, her payoff is therefore $\sup_i R_i$, and this supremum is attained by the implemented effort. Moreover, each R_i is weakly increasing as $\text{co}(\mathbf{f})$ expands, and $R_i \leq g(e_i) - c(e_i) - g(e_1)$, because the principal can always obtain $g(e_1)$.

Fix $\mathbf{l}^\star \in \text{co}(\mathbf{f})$ at which some effort e_k is implemented, and consider the points $\mathbf{l}(t) = t\mathbf{l}^\star$, $t \in (0, 1]$. Along this ray, all chord terms $\frac{c(e_i)-c(e_j)}{\ell_j-\ell_i}$ in (9) scale by $1/t$. Hence, by Lemma 4 and its proof, $W_i(t\mathbf{l}^\star) = \alpha_i + \frac{\beta_i}{t}$, where $\beta_i > 0$, for every $i \neq 1$ for which e_i is implementable, while implementability is unchanged with t. It follows that $\Pi_i(t\mathbf{l}^\star)$ is continuous and strictly increasing in t for $i \neq 1$, whereas $\Pi_1 = g(e_1)$. Moreover, $\Pi_i(t\mathbf{l}^\star) \to -\infty$ as $t \downarrow 0$. Since $\Pi_k(\mathbf{l}^\star) \geq g(e_1)$, define

$$t_0 = \max \left\{ t \in (0, 1] : \max_{i \neq 1} \Pi_i(t\mathbf{l}^\star) \leq g(e_1) \right\}.$$

Then $t_0 \in (0, 1]$, and at t_0 some effort e_m satisfies $\Pi_m(t_0\mathbf{l}^\star) = g(e_1) \geq \Pi_i(t_0\mathbf{l}^\star)$ for all i. Thus, e_m can be implemented with full extraction at $t_0\mathbf{l}^\star \in \text{co}(\mathbf{f})$, and therefore $R_m \geq g(e_m) - c(e_m) - g(e_1)$. Since e_k is the effort chosen by the agent, $R_k \geq R_m \geq g(e_m) - c(e_m) - g(e_1)$.

Now suppose the agent implements e_k, and consider an effort e_j with $g(e_j) - c(e_j) < g(e_k) - c(e_k)$. If $j = 1$, then $R_k \geq 0$. Otherwise, e_j is the only positive-cost effort other than e_k. Applying the preceding argument at a point where e_k is implemented gives $R_k \geq g(e_m) - c(e_m) - g(e_1)$ for some $e_m \in \{e_k, e_j\}$. If $m = k$, then $R_k \geq g(e_k) - c(e_k) - g(e_1) > g(e_j) - c(e_j) - g(e_1)$. If $m = j$, then $R_j \geq g(e_j) - c(e_j) - g(e_1)$, and since the agent chooses e_k, $R_k \geq R_j \geq g(e_j) - c(e_j) - g(e_1)$. Hence, in either case, $R_k \geq g(e_j) - c(e_j) - g(e_1)$.

Finally, let e_a attain the highest equilibrium surplus under f', and suppose, contrary to the claim, that the highest equilibrium surplus under f is strictly lower. Then every effort e_b implemented in an equilibrium under f satisfies $g(e_b) - c(e_b) < g(e_a) - c(e_a)$. Applying

the preceding result under $\mathrm{co}(\mathbf{f}')$ gives $R_a\big(\mathrm{co}(\mathbf{f}')\big) \geq g(e_b) - c(e_b) - g(e_1)$. Therefore,

$$R_a\big(\mathrm{co}(\mathbf{f})\big) \geq R_a\big(\mathrm{co}(\mathbf{f}')\big) \geq g(e_b) - c(e_b) - g(e_1) \geq R_b\big(\mathrm{co}(\mathbf{f})\big),$$

where the first inequality follows from $\mathrm{co}(\mathbf{f}') \subseteq \mathrm{co}(\mathbf{f})$ and the last from $R_b \leq g(e_b) - c(e_b) - g(e_1)$. Since e_b is implemented in an equilibrium under f, $R_b\big(\mathrm{co}(\mathbf{f})\big)$ is the agent's maximal rent under f. Hence all three inequalities must hold with equality. In particular, $R_a\big(\mathrm{co}(\mathbf{f})\big) = R_a\big(\mathrm{co}(\mathbf{f}')\big)$. The latter is attained by the equilibrium under f'. By Proposition 2, the same likelihood-ratio hull is feasible under f, and since continuation payoffs depend on the information structure only through its hull, the same outcome is an equilibrium under f. It implements e_a and generates surplus $g(e_a) - c(e_a)$, contradicting the supposition. □

Proof of Proposition 10. If $f(x \mid e) = f_0(x) + \Gamma(e)D(x)$, then every density $f(\cdot \mid e)$ lies on the affine line $\{f_0 + tD : t \in \mathbb{R}\}$ in the space of signed measures, so the affine hull of $\{f(\cdot \mid e) : e \in E\}$ has dimension at most one. Conversely, suppose the affine hull of $\{f(\cdot \mid e) : e \in E\}$ has dimension at most one. If it is a singleton, set $f_0 = f(\cdot \mid e)$ for any e, $D \equiv 0$, and $\Gamma \equiv 0$. Otherwise, choose e', e'' such that $f(\cdot \mid e') \neq f(\cdot \mid e'')$. Then, for every e, there is a unique scalar $t(e)$ such that $f(\cdot \mid e) = f(\cdot \mid e') + t(e)\big(f(\cdot \mid e'') - f(\cdot \mid e')\big)$. Set $f_0 = f(\cdot \mid e')$, $D = f(\cdot \mid e'') - f(\cdot \mid e')$, and $\Gamma(e) = t(e)$. Then f_0 is a density and $\int D(x)\,dx = 0$, since D is the difference of two densities.

When X and E are finite, the affine-hull condition is equivalent to the statement of the proposition. If $f(x \mid e) = f_0(x) + \Gamma(e)D(x)$, then, for any target effort e^*, the likelihood-ratio vector generated by output x has coordinates

$$a_x(e_j) = \frac{f(x \mid e^*) - f(x \mid e_j)}{f(x \mid e^*)} = s(x)\big(\Gamma(e^*) - \Gamma(e_j)\big),$$

where $s(x) = \frac{D(x)}{f(x|e^*)}$. Thus, every a_x lies on the line through the origin spanned by $\mathbf{v} = \big(\Gamma(e^*) - \Gamma(e_j)\big)_j$. Conversely, suppose that $a_x = s(x)\mathbf{v}$ for every x. Normalize $\Gamma(e^*) = 0$, set $D(x) = s(x)f(x \mid e^*)$, $\Gamma(e_j) = -v_j$, and let $f_0(x) = f(x \mid e^*)$. Then

$$f(x \mid e_j) = f(x \mid e^*) - a_x(e_j)f(x \mid e^*) = f_0(x) + \Gamma(e_j)D(x).$$

Moreover, $\sum_x D(x) = 0$ whenever the technology is nonconstant, since the difference between any two probability distributions integrates to zero; the constant case is imme-

diate with $D \equiv 0$. Hence, the technology is single-index. Therefore, $\mathrm{co}(\mathbf{f})$, the convex hull of the vectors a_x, is at most one-dimensional if and only if the technology is single-index. □

Proof of Proposition 11. (i) For a binary indicator with paid signal H and $\pi(x) = \pi(H \mid x)$,

$$p(H \mid e) = \int_X \pi(x) f_0(x)\, dx + e \int_X \pi(x) D(x)\, dx = P_0 + e P_D,$$

where $P_0 = \int_X \pi(x) f_0(x)\, dx$ and $P_D = \int_X \pi(x) D(x)\, dx$. If $P_D = 0$, the indicator is uninformative. Relabeling the signals if necessary, suppose $P_D > 0$ and define $A = P_0/P_D$, so that $p(H \mid e) = P_D\big(A + e\big)$. A wage w paid on H implements effort e if and only if $wP_D\big(e - e'\big) \geq c(e) - c(e')$ for all e'. By Assumption 3, c is nondecreasing, so efforts $e' < e$ are cheaper and their constraints give lower bounds on wP_D. The constraints from costlier efforts give the corresponding upper bounds. Hence, e is implementable if and only if $d(e)$ does not exceed the smallest of these upper bounds, a condition independent of A. When e is implementable, the minimum wage is $w = \frac{d(e)}{P_D}$, and the corresponding expected wage is $w\, p(H \mid e) = \big(A + e\big) d(e)$. For $e = 0$, there are no lower-bound constraints, so the minimum wage is zero, and the expected-wage formula continues to hold with $d(0) = 0$. Thus, all payoffs depend on π only through A.[32]

When E is an interval and c is differentiable and convex, for every $e > 0$ the chords from below have supremum $c'(e)$, while those from above have infimum $c'(e)$. Hence every effort is implementable and, for $e > 0$, $d(e) = c'(e)$, so the expected wage is $(A + e)c'(e)$. When E is finite, $d(e)$ is the maximum of finitely many chord slopes at each positive effort.

Reporting H' whenever the original signal is H, and with probability β when it is L, yields a binary garbling with $P_D' = (1 - \beta)P_D$ and $A' = A + \frac{\beta}{(1-\beta)P_D}$. As β varies over $(0, 1)$, A' ranges over (A, ∞). Since payoffs depend on the indicator only through A, any indicator with a higher price shift is payoff-equivalent to a garbling of the original indicator. Hence, a lower A corresponds to a more informative indicator.[33] Finally,

[32] For finite E, the same conclusion also follows directly from the likelihood-ratio representation. Relative to a target effort e^*, the paid signal has coordinates $\ell_j = \frac{e^* - e_j}{A + e^*}$, which depend on π only through A. Since the unpaid signal carries no wage, payoffs are determined by the paid signal's likelihood-ratio vector, and Lemma 1 yields the same conclusion.

[33] For finite E, this also has a geometric interpretation. Relative to a target effort e^*, the paid signal's likelihood-ratio vector is $\frac{\mathbf{v}}{A + e^*}$, so a lower A moves the vector farther from the origin along the fixed ray generated by the single-index technology. Consistent with Proposition 2, the garbling constructed above then produces a segment $\mathrm{co}(\mathbf{p})$ contained in that of the original indicator.

$\frac{1}{A} = \frac{\int_X \pi(x)D(x)\,dx}{\int_X \pi(x)f_0(x)\,dx}$ is an average of $\rho(x) = D(x)/f_0(x)$ under weights proportional to $\pi(x)f_0(x)$. Therefore, $A \geq \frac{1}{\sup_x \rho(x)} \equiv A_{\min}$. The infimum is approached by concentrating π on outputs where ρ approaches its supremum, and is attained when the supremum is achieved on a set of positive f_0-measure.

(ii) As noted in part (i), for a threshold indicator, $1/A$ is the average of ρ over the paid region. As the threshold moves down from the top of the support, allowing for randomization at atoms, the paid region adds outputs with weakly lower values of ρ. Hence this average decreases continuously from $\sup_x \rho(x)$ to $\int_X \rho(x)f_0(x)dx = \int_X D(x)dx = 0$.[34] It follows that $1/A$ takes every value in $(0, \sup_x \rho(x))$, or equivalently that A takes every value in $(A_{\min}, \infty)$. By part (i), payoffs depend on the indicator only through A, and by Proposition 3 the agent may restrict attention to binary indicators. Thus, replacing any equilibrium binary indicator by a threshold indicator with the same price shift leaves all payoffs unchanged. □

Proof of Proposition 12. Write $\Pi(e, A) = g(e) - \big(A + e\big)d(e)$ for the principal's payoff in the single-index parameterization, where d is defined as in the proof of Proposition 11. Since c is strictly convex, chord slopes are strictly increasing. Hence, d is strictly increasing. Let $r(e, A) = (A + e)\,d(e) - c(e)$ denote the agent's rent. For $e > e'$, incentive compatibility of the contract implementing e gives $r(e, A) \geq (A + e')\,d(e) - c(e')$. Since $d(e) \geq d(e')$ and $A + e' \geq 0$,

$$(A + e')\,d(e) - c(e') \geq (A + e')\,d(e') - c(e') = r(e', A).$$

The inequality $A + e' \geq 0$ follows from $p(H|e') = P_D(A + e') \geq 0$ with $P_D > 0$. Thus, $r(e, A)$ is increasing in e.

(i) Take $A < A'$ and let e and e' maximize $\Pi(\cdot, A)$ and $\Pi(\cdot, A')$, respectively. Adding the optimality inequalities $\Pi(e, A) \geq \Pi(e', A)$ and $\Pi(e', A') \geq \Pi(e, A')$ gives $\big(d(e') - d(e)\big)(A' - A) \leq 0$. Since $A' - A > 0$, we have $d(e') \leq d(e)$. Because d is strictly increasing, $e' \leq e$. Hence, $e_B(A)$ is weakly decreasing in A. Total surplus decomposes as $g(e) - c(e) = \Pi(e, A) + r(e, A)$. Optimality of e at A gives $\Pi(e, A) \geq \Pi(e', A)$ and $r(e, A) \geq r(e', A)$ since $e \geq e'$ and $r(\cdot, A)$ is increasing. Adding the two inequalities yields $g(e) - c(e) \geq \Pi(e', A) + r(e', A) = g(e') - c(e')$.

[34] Off atoms, P_D is a tail integral of D. The function D is integrable, since for any $e \neq 0$, $\int_X |D(x)|\,dx \leq \frac{2}{|e|}$, because $\Gamma(e)D = f(\cdot \mid e) - f_0$ is the difference of two densities. Hence the tail integral, and therefore the corresponding average, is continuous in the threshold. At an atom, randomization at the threshold interpolates the jump linearly.

(ii) The value of the agent's design problem is the supremum of $V(A)$, her payoff when the principal best responds to A, over the achievable price shifts: all $A > A_{\min}$, together with $A_{\min}$ itself when it is attained. Lowering $A_{\min}$ enlarges this set, since every shift achievable under a higher $A_{\min}$ remains achievable under a lower one. Hence, the agent's value is weakly decreasing in $A_{\min}$.

(iii) Since $\Pi(e, A) = \big[g(e) - c(e)\big] - r(e, A)$, the principal maximizes total surplus net of a rent that is increasing in effort. Let e_{FB} be the largest maximizer of $g - c$. For any $e > e_{FB}$, $g(e) - c(e) < g(e_{FB}) - c(e_{FB})$ and $r(e, A) \geq r(e_{FB}, A)$, so $\Pi(e, A) < \Pi(e_{FB}, A)$. Hence, no effort above e_{FB} can be a best response, and $e_B(A) \leq e_{FB}$. □

Proof of Proposition 13. Since the agent's payoff V, evaluated at the principal's best response $e_B(A)$, is single-peaked with peak A°, its maximum over the set of achievable price shifts is attained at A° when $A^\circ > A_{\min}$, and at $A_{\min}$ when $A^\circ \leq A_{\min}$ and $A_{\min}$ is attained; hence, $A^* = \max\{A_{\min}, A^\circ\}$. If $A^\circ \leq A_{\min}$ and $A_{\min}$ is not attained, the achievable set is $(A_{\min}, \infty)$. Since V is positive at some achievable price shift and nonincreasing above A°, it is positive near $A_{\min}$. Because it is strictly decreasing wherever positive, its supremum is approached only as $A \downarrow A_{\min}$ and is never attained. Hence, no optimal price shift exists. By Proposition 11(i), the wage is increasing in A, and the principal's payoff is therefore decreasing in A. Thus, when $A_{\min}$ is attained, the principal prefers $A_{\min}$, and the two parties prefer the same price shift if and only if $A_{\min} \geq A^\circ$: when $A_{\min} < A^\circ$, the agent strictly prefers A° to $A_{\min}$ because V is strictly increasing below A°. □

Proof of Corollary 2. Under the constraint q, the agent chooses among information structures satisfying $\text{co}(\mathbf{q}) \subseteq \text{co}(\mathbf{p}) \subseteq \text{co}(\mathbf{f})$. By Lemma 1, continuation play depends on the information structure only through $\text{co}(\mathbf{p})$. As $\text{co}(\mathbf{q})$ expands, the agent's feasible set shrinks. Her equilibrium payoff is therefore weakly lower under the tighter constraint. □

Proof of Proposition 14. Since q is an information structure, Proposition 2 implies that $\text{co}(\mathbf{q})$ is a subset of $\text{co}(\mathbf{f})$ and contains the origin in its relative interior. Moreover, $\text{co}(\mathbf{q})$ has at most $|supp(\mathbf{q})|$ extreme points. Suppose the effort level e^* is implementable by an information structure (S, π). Let $\mathbf{l} \in \text{co}(\mathbf{p})$ be the optimal choice for the principal when inducing effort e^*. Let C denote the convex hull of $\text{co}(\mathbf{q}) \cup \{\mathbf{l}, -\varepsilon\mathbf{l}\}$, where $\varepsilon > 0$ is small enough that $-\varepsilon\mathbf{l} \in \text{co}(\mathbf{p})$. Such an ε exists because $\mathbf{0} \in \text{relint}(\text{co}(\mathbf{p}))$ by Proposition 2 and $\mathbf{l} \in \text{co}(\mathbf{p})$. This convex set has at most $|supp(\mathbf{q})| + 2$ extreme points and contains the origin in its relative interior: since the origin lies in the relative interior of $\text{co}(\mathbf{q})$ and

strictly between $\mathbf{l}$ and $-\varepsilon\mathbf{l}$, it is a positive convex combination of the extreme points of $\text{co}(\mathbf{q})$ together with $\mathbf{l}$ and $-\varepsilon\mathbf{l}$.[35] Therefore, by Proposition 2, there exists an information structure $\left(\hat{S},\hat{\pi}\right)$ with $\left|\hat{S}\right| \leq |supp\left(\mathbf{q}\right)| + 3$ signals such that the associated stochastic mapping $\hat{p}$ satisfies $\text{co}\left(\hat{\mathbf{p}}\right) = C$. In particular $\text{co}\left(\mathbf{q}\right) \subseteq \text{co}(\hat{\mathbf{p}})$, so $\left(\hat{S},\hat{\pi}\right)$ is admissible. Since $\mathbf{l} \in \text{co}\left(\hat{\mathbf{p}}\right)$, $\mathbf{l}$ is feasible for the principal under the new information structure $\left(\hat{S},\hat{\pi}\right)$. Hence, if the principal chooses effort e^*, he can achieve the same objective value under $\left(\hat{S},\hat{\pi}\right)$ as under (S,π). To complete the argument, it suffices to show that for any alternative $e_i \neq e^*$, $W\left(e_i,\hat{\pi}\right) \geq W\left(e_i,\pi\right)$. This inequality holds if $\text{co}\left(\hat{\mathbf{p}}\right) \subseteq \text{co}\left(\mathbf{p}\right)$. Admissibility gives $\text{co}\left(\mathbf{q}\right) \subseteq \text{co}\left(\mathbf{p}\right)$. Together with $\mathbf{l}, -\varepsilon\mathbf{l} \in \text{co}\left(\mathbf{p}\right)$ and the convexity of $\text{co}\left(\mathbf{p}\right)$, this yields $\text{co}\left(\hat{\mathbf{p}}\right) \subseteq \text{co}\left(\mathbf{p}\right)$. Therefore, no alternative effort becomes strictly more attractive under $\left(\hat{S},\hat{\pi}\right)$, and e^* remains implementable with the same expected wage. This completes the proof. □

Proof of Proposition 15. (i) Let $\mathbf{l} \in L(e^*) \cap \text{co}\left(\mathbf{f}\right)$ satisfy the stated disjointness condition. For sufficiently small $\varepsilon > 0$ define $C_\varepsilon = \text{conv}\left(\text{co}\left(\mathbf{q}\right) \cup \{\mathbf{l}, -\varepsilon\mathbf{l}\}\right)$. In particular, $-\varepsilon\mathbf{l} \in \text{co}\left(\mathbf{f}\right)$ for sufficiently small ε, since the origin lies in the relative interior of $\text{co}\left(\mathbf{f}\right)$. This convex set contains the origin in its relative interior: since, by Proposition 2, the origin lies in the relative interior of $\text{co}(\mathbf{q})$ and strictly between $\mathbf{l}$ and $-\varepsilon\mathbf{l}$, it is a positive convex combination of the extreme points of $\text{co}(\mathbf{q})$ together with $\mathbf{l}$ and $-\varepsilon\mathbf{l}$. By Proposition 2, together with the signal-count bound of Proposition 14, there exists an information structure such that the associated stochastic mapping p satisfies $\text{co}\left(\mathbf{p}\right) = C_\varepsilon$. In particular, $\text{co}\left(\mathbf{q}\right) \subseteq C_\varepsilon = \text{co}\left(\mathbf{p}\right)$ by construction, so the structure is admissible. Every point of C_ε lies within $\varepsilon\|\mathbf{l}\|$ of the compact set $\text{conv}\big(\text{co}\left(\mathbf{q}\right) \cup \{\mathbf{l}\}\big)$, which is at positive distance from the closure of each set D_j. Hence, for ε sufficiently small, $C_\varepsilon \cap D_j = \emptyset$ for every $e_j \neq e^*$, so no point in $\text{co}\left(\mathbf{p}\right)$ gives the principal more than $g(e_1)$ from any rival effort. No point of C_ε gives him more than $g(e_1)$ from e^* either. Any $\mathbf{x} \in C_\varepsilon$ can be written as $\mathbf{x} = (1-t)\mathbf{h} + t(-\varepsilon\mathbf{l})$, for some $\mathbf{h} \in \text{conv}\big(\text{co}(\mathbf{q}) \cup \{\mathbf{l}\}\big)$ and $t \in [0,1]$. By hypothesis, $\mathbf{h} \notin O(e^*)$, so there is some effort e_j with $c(e_j) < c(e^*)$ such that $\ell_j(\mathbf{h}) \leq \ell_j^*$. Since e^* is implementable at $\mathbf{l}$ with expected wage $g(e^*) - g(e_1)$, every such effort satisfies

[35] If $\mathbf{l}$ lies in the span of $\text{co}\left(\mathbf{q}\right)$, in particular whenever $\text{co}\left(\mathbf{q}\right)$ has full dimension, the point $-\varepsilon\mathbf{l}$ is unnecessary. In that case, $C = \text{conv}\big(\text{co}\left(\mathbf{q}\right) \cup \{\mathbf{l}\}\big)$ has the same affine hull as $\text{co}\left(\mathbf{q}\right)$, so the fact that the origin lies in the relative interior of $\text{co}\left(\mathbf{q}\right)$ implies that it also lies in the relative interior of C. The signal bound then improves to $|\,\text{supp}(q)| + 2$.

$\ell_j(\mathbf{l}) \geq \ell_j^* \geq 0$. Therefore,

$$\ell_j(\mathbf{x}) = (1-t)\ell_j(\mathbf{h}) - t\varepsilon\ell_j(\mathbf{l}) \leq (1-t)\ell_j^* \leq \ell_j^*,$$

so $\mathbf{x} \notin O(e^*)$ as well. If $\ell_j(\mathbf{x}) \leq 0$, then e^* is not implementable at $\mathbf{x}$. If $\ell_j(\mathbf{x}) > 0$, the incentive constraint against e_j alone requires an expected wage of at least $\frac{c(e^*)-c(e_j)}{\ell_j(\mathbf{x})} \geq \frac{c(e^*)-c(e_j)}{\ell_j^*} = g(e^*) - g(e_1)$. In either case, $\Pi_{e^*}(\mathbf{x}) \leq g(e_1)$. The principal's payoff at every point of $\text{co}(\mathbf{p})$, from every effort, is therefore at most $g(e_1)$. At $\mathbf{l} \in L(e^*)$, implementing e^* gives him exactly $g(e_1)$, and ties at $g(e_1)$ are resolved in favor of e^* by the convention of Section 2.

(ii) In any equilibrium with the stated payoffs, the principal implements e^* at expected wage $g(e^*) - g(e_1)$, so his chosen point $\mathbf{l}$ lies in $L(e^*) \cap \text{co}(\mathbf{f})$. Admissibility gives $\text{co}(\mathbf{q}) \subseteq \text{co}(\mathbf{p})$, and since $\text{co}(\mathbf{p})$ is convex and contains $\mathbf{l}$, it also contains $\text{conv}\big(\text{co}(\mathbf{q}) \cup \{\mathbf{l}\}\big)$. If any point in this hull belonged to some D_j, the principal could obtain strictly more than $g(e_1)$ by implementing e_j there, contradicting his choice of e^*. Likewise, if any point in the hull belonged to D_{e^*}, the principal could obtain strictly more than $g(e_1)$ by implementing e^* there, yielding the same contradiction. This proves necessity. □

Proof of Proposition 16. (i) The feasible interval $[A_{\min}, A_q]$ expands with A_q, so the agent's equilibrium payoff is weakly increasing in A_q.

(ii) As A_q increases, the feasible interval $[A_{\min}, A_q]$ expands to the right, so A^*, defined as the largest optimal price shift, weakly increases. By Proposition 12(i), both the implemented effort $e_B(A)$ and equilibrium total surplus are weakly decreasing in A, and hence weakly decreasing in A_q. □

Proof of Proposition 17. The price shifts achievable under the constraint are those available in the unconstrained problem that satisfy $A \leq A_q$. The endpoint A_q is itself achievable, since it is attained by q. Since the agent's payoff V, evaluated at the principal's best response $e_B(A)$, is single-peaked with peak A°, its maximum is attained at A° when $A_{\min} < A^\circ \leq A_q$, at A_q when $A^\circ > A_q$, and at $A_{\min}$ when $A^\circ \leq A_{\min}$ and $A_{\min}$ is attained; hence, $A^* = \min\{\max\{A_{\min}, A^\circ\}, A_q\}$. If $A^\circ \leq A_{\min}$ and $A_{\min}$ is not attained, the achievable set is $(A_{\min}, A_q]$. Since V is positive at some achievable price shift and nonincreasing above A°, it is positive near $A_{\min}$. Because it is strictly decreasing wherever positive, its supremum is approached only as $A \downarrow A_{\min}$ and is never attained. Hence, no optimal price shift exists. □

Acknowledgements

We thank James Best, Marina Halac, Aniko Öry, and Maryam Saeedi, as well as conference and seminar participants at SITE, WATE, the Stony Brook Game Theory Conference, the Canadian Economic Theory Conference, the Society for Economic Dynamics, Carnegie Mellon University, and Washington University in St. Louis.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude (Anthropic) in order to assist with editing the manuscript and with checking proofs and references, in the manner of a research assistant. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Babichenko, Y., I. Talgam-Cohen, H. Xu, and K. Zabarnyi (2025): "Information design in the principal-agent problem," *Games and Economic Behavior*. 6, 7, 37

Baker, G. (2002): "Distortion and risk in optimal incentive contracts," *Journal of human resources*, 728–751. 6

Ball, I. (2019): "Scoring strategic agents," *arXiv preprint arXiv:1909.01888*. 8

Barron, D., G. Georgiadis, and J. Swinkels (2020): "Optimal contracts with a risk-taking agent," *Theoretical Economics*, 15, 715–761. 8

Bebchuk, L. A. and J. M. Fried (2004): *Pay without performance: The unfulfilled promise of executive compensation*, Harvard University Press. 2

Bernhardt, A., L. Kresge, and R. Suleiman (2023): "The data-driven workplace and the case for worker technology rights," *ILR Review*, 76, 3–29. 5

Boleslavsky, R. and K. Kim (2018): "Bayesian persuasion and moral hazard," *Available at SSRN 2913669*. 8

Carroll, G. (2015): "Robustness and linear contracts," *American Economic Review*, 105, 536–563. 6

Chaigneau, P., A. Edmans, and D. Gottlieb (2019): "The informativeness principle without the first-order approach," *Games and Economic Behavior*, 113, 743–755. 6

Demougin, D. and C. Fluet (2001): "Ranking of information systems in agency models: an integral condition," *Economic Theory*, 17, 489–496. 6

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The economics of career concerns, part I: Comparing information structures," *The Review of Economic Studies*, 66, 183–198. 8

Feltham, G. A. and J. Xie (1994): "Performance measure congruity and diversity in multi-task principal/agent relations," *Accounting review*, 429–453. 6

Garrett, D. F., G. Georgiadis, A. Smolin, and B. Szentes (2023): "Optimal technology design," *Journal of Economic Theory*, 209, 105621. 6, 7

Georgiadis, G., D. Ravid, and B. Szentes (2024): "Flexible moral hazard problems," *Econometrica*, 92, 387–409. 7

Georgiadis, G. and B. Szentes (2020): "Optimal monitoring design," *Econometrica*, 88, 2075–2107. 8

Gjesdal, F. (1982): "Information and incentives: The agency information problem," *The Review of Economic Studies*, 49, 373–390. 6

Grossman, S. J. and O. D. Hart (1983): "Implicit contracts under asymmetric information," *The quarterly journal of economics*, 123–156. 40

Halac, M., E. Lipnowski, and D. Rappoport (2021): "Rank uncertainty in organizations," *American Economic Review*, 111, 757–86. 8

Holmström, B. (1979): "Moral hazard and observability," *The Bell journal of economics*, 74–91. 6

——— (1999): "Managerial incentive problems: A dynamic perspective," *The review of Economic studies*, 66, 169–182. 8

Holmstrom, B. and P. Milgrom (1991): "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, 7, 24–52. 6

Innes, R. D. (1990): "Limited liability and incentive contracting with ex-ante action choices," *Journal of economic theory*, 52, 45–67. 2, 6

Kamenica, E. and M. Gentzkow (2011): "Bayesian persuasion," *American Economic Review*, 101, 2590–2615. 3, 9, 15

Kim, S. K. (1995): "Efficiency of an information system in an agency model," *Econometrica: Journal of the Econometric Society*, 89–102. 6

Kirkegaard, R. (2017): "Moral hazard and the spanning condition without the first-order approach," *Games and Economic Behavior*, 102, 373–387. 40

Liu, T. (2025): "Worst-Case Privacy," *November, https://drive. google. com/file/d/1PGy0HN1RkpM4k3Il_lWkv49czidpKqAt/view, Working paper*. 7

Perez-Richet, E. and V. Skreta (2022): "Test design under falsification," *Econometrica*, 90, 1109–1142. 8

Poblete, J. and D. Spulber (2012): "The form of incentive contracts: agency with moral hazard, risk neutrality, and limited liability," *The RAND Journal of Economics*, 40, 215–234. 6

Rosar, F. (2017): "Test design under voluntary participation," *Games and Economic Behavior*, 104, 632–655. 8

Saeedi, M. and A. Shourideh (2026): "Optimal rating design under moral hazard," *arXiv preprint arXiv:2008.09529*. 8

Walton, D. and G. Carroll (2022): "A General Framework for Robust Contracting Models," *Econometrica*. 6

Writers Guild of America (2023): "Summary of the 2023 WGA MBA," https://www.wgacontract2023.org/the-campaign/summary-of-the-2023-wga-mba, accessed September 2026. 2

Zapechelnyuk, A. (2020): "Optimal quality certification," *American Economic Review: Insights*, 2, 161–176. 8