Optimal Rating Design under Moral Hazard*

Maryam Saeedi † Ali Shourideh ‡

November 25, 2026

Abstract

We study optimal monotone rating design under moral hazard and strategic manipulation. An intermediary observes a noisy indicator of effort and commits to a rating policy that shapes market beliefs and pay. We characterize optimal ratings via concavification of a gain function. When effort increases tail risk, optimal ratings use lower censorship, pooling poor outcomes to encourage risk-taking; when effort reduces tail risk, upper censorship discourages negligence. In multi-task settings, censorship can discourage window dressing relative to productive effort, even when the indicator satisfies the monotone likelihood ratio property for each action. Redistributive tests can feature mid-censorship.

Keywords: Information Design, Moral Hazard, Window Dressing, Majorization, Censorship. **JEL codes:** D82, D83, D86, L15, L86, D63

*This paper was previously circulated under the title "Optimal Rating Design". We thank Nageeb Ali, Ian Ball, James Best, Aislinn Bohren, Odilon Camara, Joyee Deb, Rahul Deb, Maryam Farboodi, Nima Haghpanah, Hugo Hopenhayn, Emir Kamenica, Navin Kartik, Alexey Kushnir, Stephen Morris, Aniko Öry, Paula Onuchic, Jacopo Perego, Ilya Segal, Vasiliki Skreta, and Alex Wolitzky, as well as numerous seminar and conference participants for their helpful comments. We have used Claude.ai similarly to an RA.

†Carnegie Mellon University, Email: msaeedi@andrew.cmu.edu

‡Carnegie Mellon University, Email: ashourid@andrew.cmu.edu

1 Introduction

Many markets rely on information disclosure or ratings to facilitate trade and incentivize quality provision. Environmental, Social, and Governance (ESG) rating agencies aim to incentivize companies to improve their environmental and social impact. Online platforms such as Amazon, Airbnb, Upwork, and eBay design reputation systems to incentivize and signal providers' quality. Standardized tests communicate student ability to universities. In each case, an intermediary observes signals about agent behavior and must decide how to convey this information to a market that rewards agents based on perceived quality. A fundamental challenge arises: agents strategically respond to rating policies by potentially using window dressing and other manipulation strategies, and the information disclosed shapes their incentives.

Despite the ubiquity of these systems in markets suffering from moral hazard, several core theoretical questions are yet to be answered: What are the fundamental trade-offs in designing rating systems when participants can anticipate and react to them? How does the effort technology shape optimal ratings? How should an intermediary design information structures to account for multi-tasking and window-dressing incentives? This paper answers these questions by incorporating information design in a static variant of the career-concerns model introduced by Holmström (1999).

Our model features an agent (e.g., a company seeking an ESG rating or a seller on eBay) who takes costly actions that create value for a competitive market. These actions also generate a noisy indicator observed by an intermediary (e.g., an ESG rating firm or an online platform), which must then decide how to convey this information via a rating to the market. The market, in turn, pays the agent its expected value based on the signal and its belief about the agent's action. Our primary objective is to find the optimal information structure to maximize a flexible objective function for the intermediary. This function can target a particular action or social welfare, capturing environments where market values do not fully internalize the social value of actions—such as the positive externalities of ESG activities or the welfare losses from window-dressing activities.[1]

Our main results shed light on the optimal provision of incentives for one- and multi-dimensional effort. First, we show how interactions between the distribution of the indicator and effort affect the optimal ratings when effort is a single variable. Namely, we show that when effort increases tail risk, lower-censorship ratings maximize effort, while the opposite is true when effort decreases tail risk. Second, in a multi-task model with window dressing as in Holmström and

[1] Since Holmström (1999), it is well known that the *implicit incentives* provided by career concerns do not necessarily lead to efficient effort levels because they fail to fully internalize the social benefit of the agent's action. This externality is also present in our model, and the intermediary's objective can be thought of as addressing such externalities.

Milgrom (1991) and Dewatripont et al. (1999), we identify a condition on the information technology, the monotone relative informativeness property (MRIP), under which optimal ratings are upper or lower censorship.

Our main technical challenge is the interplay between information structures and the agent's incentive constraints.[2] This is because the agent's incentives are determined by her interim payoff, her expected payoff at the stage where she knows the indicator (state) and does not know the outcome of the possible random signal. We introduce the concept of *interim prices*—the agent's interim expectation of the market's posteriors—as the sufficient statistic that determines incentives from the agent's perspective. This object, which can be interpreted as the agent's second-order belief, allows us to transform the problem of choosing an information structure into a tractable constrained optimization problem with majorization and incentive compatibility as its constraints. In general, the set of implementable interim prices is not well characterized: being a mean-preserving contraction of market values is necessary but not sufficient.[3] Our first result, Proposition 1, shows that when these interim prices are comonotone with market values, a price schedule can be implemented by some rating if and only if it is a mean-preserving contraction of the market values. This implies that we can cast the problem of rating design as a standard moral hazard problem with transfers subject to a majorization constraint. Throughout the paper, we therefore restrict attention to *monotone ratings*—those whose interim prices are comonotone with market values—a restriction we motivate by the agent's ability to manipulate the indicator downward ex post.

With this result in hand, in Theorem 1, we show that optimal ratings can be characterized through concavification of a *gain function* in the quantile space. As a result, when this concavification is achieved by a unique rating, a condition that we can verify using properties of the technology, the optimal information structure is a deterministic monotone partition of the indicator space: the intermediary either fully reveals the indicator on some regions or pools contiguous intervals. This provides a sharp foundation for the prevalence of coarse, threshold-based rating systems.

Our first set of results derives general properties of optimal ratings in classic moral hazard where the agent has a single-dimensional effort and the objective is to maximize effort levels. Under the canonical Monotone Likelihood Ratio Property (MLRP) assumption in the moral hazard literature, we show that the gain function in this case is concave and, as a result, full-information

[2]This is in addition to the usual difficulty of solving models with moral hazard. We maintain the validity of the standard first-order approach to address this and verify it ex post for each model considered. The verification is left to the Online Appendix C.

[3]In a standard sender–receiver setting, Doval and Smolin (2024) are also interested in the interim payoff of a sender. They use a duality approach to characterize the frontier of interim payoff profiles of the sender – as a function of the state. Our comonotonicity restriction allows us to simplify the problem and restrict interim prices to be a mean-preserving contraction of market values.

disclosure is optimal.

Despite the prominence of the MLRP assumption, many economically relevant activities violate MLRP in systematic ways. Innovative activities often increase both upside potential and downside risk, e.g., R&D effort can lead to breakthroughs or failures. Conversely, maintenance activities typically reduce variance through more consistent outcomes. To capture these patterns, we introduce two new distributional properties: the *expanding likelihood ratio property* (ELRP), where increased effort expands the distribution's tails, and the *compressing likelihood ratio property* (CLRP), where increased effort compresses outcomes toward the center. Under ELRP, optimal ratings take the form of lower censorship, providing insurance against downside risk to encourage risk-taking. In contrast, under CLRP, optimal ratings are upper censorship, pooling high realizations while revealing low ones, which punishes poor outcomes and encourages tail-compressing effort.

To see the intuition, first note that the marginal benefit of effort is given by the covariance between the interim price and the marginal percentage change in the probability of different outcomes induced by a marginal increase in effort. In the MLRP case, higher effort shifts the distribution toward higher realizations in a uniform, monotone way. Full revelation is therefore optimal for incentives: because interim prices must be mean-preserving contractions of the full-information market values, no rating can make the price schedule "steeper" than under full revelation, and any pooling only flattens it.

Under ELRP, by contrast, higher effort expands tail events: it increases the probability of both very high and very low outcomes, so full revelation exposes the agent to downside risk that effort itself makes more likely. Pooling low realizations of the indicator provides insurance against these outcomes—not in the risk-sharing sense, since the agent is risk neutral, but by withholding the penalty on realizations that effort itself makes more likely—and thereby strengthens the covariance between interim prices and likelihood changes, increasing incentives to exert effort. Under CLRP, higher effort compresses the distribution toward the center, so extremely high outcomes are relatively more likely under low effort. Pooling the top realizations while revealing low ones similarly improves the alignment between interim prices and likelihood changes, strengthening incentives to exert effort.

Our second set of results, in Section 4, pertains to a multi-task moral hazard model with window dressing à la Holmström and Milgrom (1991), in the career-concerns variant of Dewatripont et al. (1999). In the multi-task model, the agent allocates efforts across productive activity and a window-dressing one (an action that boosts the indicator without creating value for the market). Different values of the indicator can have different informational content about the two actions: an increase in the indicator can be more informative about window dressing or about productive effort. Theorem 2 identifies a condition on the information technology, which we re-

fer to as the *monotone relative informativeness property* (MRIP): high realizations of the indicator are either always more informative of window dressing or of productive activity. With MRIP, welfare-maximizing ratings are either upper or lower censorship. Interestingly, this is true even if the indicator satisfies the standard MLRP condition with respect to each effort.

At an intuitive level, pooling higher (lower) values of the indicator can have a larger effect on window-dressing incentives relative to incentives for productive effort, especially when higher (lower) values are relatively more informative about window dressing. We illustrate this in a class of location–scale technologies, where the direction of censorship is determined by whether window dressing contributes more to the level or to the dispersion of the indicator and how the marginal costs of effort are related to their impact on the mean of the indicator.

Finally, to illustrate the range of problems that can be addressed by our formulation, in Section 5 we apply our framework to the problem of redistributive test design with heterogeneous students. We show that optimal tests may involve "mid-censorship" to balance incentive provision with redistributive motives towards disadvantaged groups.

Our analysis also offers practical guidance for regulators and rating system design. As data collection has intensified, several institutions have formed around using data to incentivize behavior, which in turn has created incentives for manipulation and window dressing (see for example Mayzlin et al. (2014)). Our results provide guidance on how ratings should be designed in such environments and rationalize the heterogeneity observed across rating systems. For example, platforms have adopted a variety of ratings: some platforms (such as Shipt or Instacart) allow for low-rating forgiveness and a fresh start, which could be interpreted as lower censorship; in others, such as Airbnb, ratings are concentrated at the top (see for example Zervas et al. (2021)), which effectively pools high outcomes and can be interpreted as upper censorship. Our results suggest these differences may reflect optimal responses to underlying differences in how effort affects outcome distributions, either in the single-task model (MLRP vs. ELRP/CLRP) or in the multi-task model via the direction of the informativeness of the indicator about window dressing. More broadly, the paper provides a toolkit for evaluating rating policies across domains—from ESG certification to educational testing to online marketplaces—by connecting features of agent technology to the optimal structure of information disclosure.

1.1 Related Literature

Our paper is related to a few strands of the literature in information economics and mechanism design. It is closely related to a recent literature that studies information design when strategic behavior affects the state in response to the choice of the information structure (e.g., Frankel and Kartik (2019), Ball (2025), and Perez-Richet and Skreta (2022)). Our exercise substantively differs

from Ball (2025), where the agent's actions purely manipulate several indicators, preventing a receiver from correctly predicting a correlated quality. In contrast, our setting completely assumes away adverse selection and screening, and pooling of indicators occurs solely in order to motivate the agent. Moreover, relative to Ball (2025) and Frankel and Kartik (2019), our mathematical result on second-order expectations allows us to study a larger class of objectives and state spaces, albeit under the restriction that ratings are monotone in the indicator. In our model, the presence of window-dressing incentives is similar to the falsification model in Perez-Richet and Skreta (2022). The main difference from our setting is the existence of noise in the ability of the agent to manipulate the signal observed by the intermediary.

A related paper to ours is Boleslavsky and Kim (2021). They study a model of Bayesian persuasion with moral hazard, similar to ours, in which an agent chooses an effort level that affects the distribution of the state, and a sender affects a receiver's action using an information structure. The papers differ in terms of focus and technique: we focus on a career-concerns model where the information structure only affects the agent's incentives. Additionally, we use majorization tools, which allow us to work with larger state spaces. Beyond technique, the economic difference is that our characterization ties the shape of the optimal rating—full disclosure, lower or upper censorship—to how effort shifts the distribution of the indicator, a question that is difficult to answer in their setting.

In a related dynamic career-concerns model, Hörner and Lambert (2021) elegantly show that optimal ratings that aggregate the agent's past performance over time are linear in histories. They are interested in the dynamics of ratings, while in our model all incentive provision happens instantaneously and our focus is on how properties of single- and multiple-task technologies translate into properties of optimal ratings. Moreover, in their setting, the process for the indicator is a Brownian process whose trend is controlled by the agent, and, as a result, instantaneous shocks are small. It remains to be seen how large shocks, for example as in jump processes, can impact the dynamics of optimal ratings.[4]

From a technical perspective, our results are related to the new literature in information economics that uses optimization under majorization constraints – see Kleiner et al. (2021). Their solution method uses the characterization of extreme points of the set of monotone functions that majorize a certain function. Similarly, Bergemann et al. (2026) and Bergemann et al. (2022) use the same strategy as our work to cast the problem in terms of quantiles and use concavification to derive optimal mechanisms—see also Rayo (2013). While their focus is on screening models with hidden information, ours is closer to classic moral hazard. Our focus on monotone ratings also

[4]Relatedly, a recent paper by Madsen et al. (2025) studies a moral hazard model with non-monetary incentives. Our paper is related to their work to the extent that our agent is incentivized using ratings, a non-monetary instrument.

connects our work to the literature on monotone persuasion—see Mensch (2021) and Kolotilin et al. (2024)—which provides conditions under which optimal monotone signals are deterministic and take censorship forms. In that literature, the state is exogenous, and the conditions for censorship are placed on the sender's payoff over states. Here, the distribution of the state is chosen by the agent, so the analogous conditions fall on the effort technology—MLRP, ELRP, and CLRP with a single task, MRIP with multiple—rather than on the sender's preferences.

Our paper is also related to the literature concerned with the problem of certification and its interactions with moral hazard: Albano and Lizzeri (2001), Zubrickas (2015), Onuchic and Ray (2023), and Zapechelnyuk (2020). A notable contribution is that of Albano and Lizzeri (2001), where the key assumption that the intermediary can charge an arbitrary fee schedule leads to an indeterminacy between using transfers and ratings to implement desired outcomes. Zubrickas (2015), Zapechelnyuk (2020), and Onuchic and Ray (2023) also study related problems, but they focus on *deterministic* technologies where the agent's effort deterministically translates into values for the market. In contrast, in our model, the presence of noise allows us to disentangle the indicator from market values, which in turn leads to an inefficient level of effort under full information and enables us to study window dressing and manipulation incentives. Other related contributions to this literature include Rodina and Farragut (2020) on inducing effort through grades and Xiao (2025) on incentivizing agents through ratings. Relatedly, in the context of team production, Halac et al. (2021) show that uncertainty about a worker's compensation ranking in a team can remove low-effort equilibria. Our result on how censorship for some technologies can improve incentives can be regarded as the single-agent version of their result.[5]

Finally, our paper complements the empirical literature on certification and disclosure in markets with asymmetric information, such as online platforms (e.g., Hui et al. (2023) and Nosko and Tadelis (2015)), health insurance markets (Vatter (2025)), food labeling (Barahona et al. (2023)), and ESG investing (Berg et al. (2022)). We contribute to this literature by developing theoretical methods and general lessons for the optimal design of rating systems.

The remainder of the paper proceeds as follows. Section 2 presents the model and provides our key technical results. Section 3 establishes general properties of optimal ratings with a single task under alternative distributional assumptions, including MLRP, ELRP, and CLRP. Section 4 focuses on multi-task moral hazard and window dressing. Section 5 applies our results to redistributive test design. Section 6 concludes. Proofs are relegated to the Appendix, and additional results, including the verification of the first-order approach, are in the Online Appendix.

[5] Ali et al. (2022) study a model with adverse selection (i.e., exogenous state), where optimal disclosure involves uncertainty, but it is a way of uniquely implementing an intermediary's desirable outcome.

2 A Model of Moral Hazard and Optimal Ratings

In this section, we describe our basic model of rating design and provide some preliminary analysis of the restrictions implied by the fact that incentives are provided through ratings.

We are interested in settings in which an intermediary observes some information about an agent's chosen actions and decides how to convey this information to a competitive market, henceforth "the market," which then pays its posterior mean as a price to the agent.

More specifically, the agent exerts an effort vector $a \in A \subset \mathbb{R}^N$ at a cost $c(a)$, where A is compact. This effort generates a random outcome $(v, y) \in \mathbb{R}^2$, where v represents the value of the output to the market and y is a noisy *indicator* observed by the intermediary. We denote the cumulative distribution function of the indicator y conditional on action a by $G(y|a)$ and assume that it is absolutely continuous with density $g(y|a)$.

The market consists of competitive buyers who value the agent's output at v. However, the market observes neither the true value v nor the agent's action a directly. Instead, it forms expectations based on information provided by the intermediary. If the market observes the indicator y and believes the agent's action to be $\hat{a}$, the expected value of the output is given by:

$$\overline{v}(y; \hat{a}) = \mathbb{E}[v \mid y, \hat{a}]$$

We refer to $\overline{v}(y; \hat{a})$ as *market values*, i.e., the most informative assessment of the valuation of the market. Throughout the paper, we impose the following assumptions:

Assumption 1. *For all (deterministic) market beliefs $\hat{a}$ and effort a, the distribution of the indicator $g(y|a)$, market values $\overline{v}(y; \hat{a})$, and the cost function $c(a)$ satisfy:*

1. *the support of y is a convex (possibly unbounded) subset of $\mathbb{R}$,*
2. *the market value $\overline{v}(y; \hat{a})$ is continuous in $(y, \hat{a})$ and increasing in y,*
3. *the gradients of the distribution $\nabla_a g(y|a)$, $\nabla_a G(y|a)$ are continuous in (y, a),*
4. *market values and likelihood ratios have finite second moments that are locally uniformly bounded, i.e., for every a there exists an open neighborhood $B(a) \subset \mathbb{R}^N$ such that*

$$\sup_{a' \in B(a) \cap A} \mathbb{E}\left[v^2 | a'\right], \sup_{a' \in B(a) \cap A} \mathbb{E}\left[\frac{\|\nabla_a g(y|a')\|^2}{g(y|a')^2} | a'\right] < \infty$$

5. *the cost function $c(a)$ is continuously differentiable in a, i.e., $\nabla_a c(a)$ is continuous in a.*

The second part of Assumption 1 states that market values are ranked based on the values of the indicator. It implies that changes in market belief do not affect the ranking of market values. The third and fourth parts of Assumption 1 are technical conditions that allow us to work with

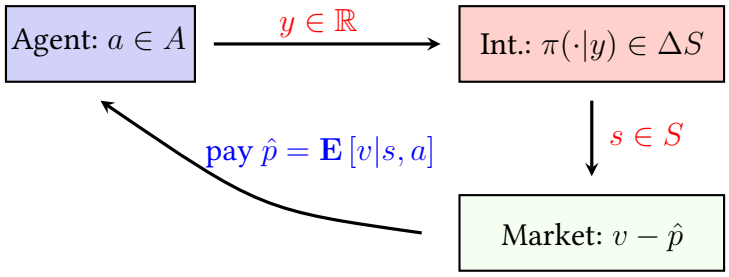


Figure 1: General structure of the model

unbounded values for the indicator and market values. They can be further relaxed if compactness of the support of (v, y) is added as an assumption.

The intermediary observes the indicator y (at no cost) and controls the information observed by the market. Specifically, the intermediary commits to an information structure $(S, \pi(\cdot|y))$, where S is a set of signal realizations and $\pi(\cdot|y) \in \Delta S$ is the distribution over signals conditional on the realization of y. Having observed s, the market pays its expected value $\mathbb{E}[v|s, \hat{a}]$ to the agent when it believes the agent chooses $\hat{a}$.[6] This expectation is calculated using the information available, s, and the common belief about equilibrium play.[7] The timing is summarized in Figure 1: the intermediary commits to an information structure, the agent chooses her effort, the indicator realizes, a rating is drawn and sent to the market, and the market pays the agent its posterior expectation of the value.

Given an information structure $(S, \pi(\cdot|y))$ and market belief $\hat{a}$, the agent's expected payoff is:

$$\int_Y \int_S \mathbb{E}[v|s, \hat{a}] \, d\pi(s|y) \, dG(y|a) - c(a). \tag{1}$$

In equilibrium, the agent chooses a to maximize (1).

The ex-post market price $\mathbb{E}[v|s, a]$ depends on the information structure (S, π) and also on the market's prior about the distribution of (a, y), which depends on the agent's equilibrium strategy. More specifically, the market uses its beliefs about the equilibrium strategy of the agent a to form a prior $G(y|a)$ and uses Bayes' rule to form the posterior expectation $\mathbb{E}[v|s, a]$ satisfying

$$\int_Y \int_{S'} \mathbb{E}[v|s, a] \, d\pi(s|y) \, dG(y|a) = \int_Y \overline{v}(y; a) \, \pi(S'|y) \, dG(y|a), \forall S' \subset S. \tag{2}$$

The above defines a Bayesian Nash equilibrium given the information structure (S, π). More specifically, given an information structure (S, π), an equilibrium is an effort a together with the

[6] We assume that the buyers are on the long side of the market, thus willing to pay their expected value. Our analysis remains unchanged if the market keeps a constant fraction of their expected value.

[7] An information structure is a family of probability spaces $\{(S, \mathscr{S}, \pi(\cdot|y))\}_{y \in Y}$, where S is the space of signal realizations and $\mathscr{S}$ is a σ-algebra. Throughout the paper, we work with S a standard Borel space, and $\mathscr{S}$ its Borel σ-algebra. Henceforth, we drop references to σ-algebras in our analysis. Additionally, when describing subsets, we refer to Borel subsets.

market's belief $\hat{a} = a$ and $\mathbb{E}[v|s, \hat{a}]$ such that a maximizes expression (1), and given a, the market beliefs satisfy Bayesian updating as defined in equation (2).

It should be noted that all of our results implicitly include an equilibrium selection assumption where we allow the intermediary to choose the equilibrium – as in classic contract theory and mechanism design. In other words, under our optimal ratings, there could be other equilibria, but the analysis of unique implementation, or of the full set of equilibria induced by an optimal rating, is beyond the scope of this paper.

The intermediary chooses an information structure to maximize an objective that may differ from total surplus. In the majority of the paper (Sections 2 through 4), we assume that the intermediary has $W(a)$ as its objective. The function $W(a)$ captures externalities or direct preferences over effort. For example, if the intermediary wishes to correct an externality not internalized by market prices – as in the case of ESG – then $W(a) \neq V(a)$, with $V(a) = \mathbb{E}[v|a] - c(a)$ being the total surplus. An alternative is to assume that $W(a) = V(a)$, which can be relevant when full revelation of the indicator leads to inefficient market outcomes. Finally, in Section 5, we consider a model with distributive concerns and describe how to recast the problem under such preferences.

The environment fits a number of applications. Online platforms observe rich performance data about their providers and convey it to the market through certification policies such as Airbnb's Superhost or eBay's Top Rated Seller; such policies correspond to the information structure in our model, see Hui et al. (2016). Ratings also invite manipulation and window dressing: the agent may take costly actions that inflate the indicator without enhancing fundamental value, as when sellers pay for positive reviews, see He et al. (2022); an illustration of this is described in Example 1, which we develop further in Section 4.

The following example illustrates how one can use our model to think about manipulation, and how ratings can potentially help:

Example 1. Suppose that market values and the indicator are given by

$$v = a_1 + \varepsilon_1, y = v + a_2(1 + \varepsilon_2)$$

where $\varepsilon_1, \varepsilon_2 \sim \mathcal{N}(0, 1)$ are independent standard normals. Here, a_1 increases the market value but a_2 can manipulate the observed indicator.

The market values $\overline{v}(y; a)$ are given by

$$\overline{v}(y; a) = a_1 + \frac{y - a_1 - a_2}{1 + a_2^2}$$

while $y|a \sim \mathcal{N}(a_1 + a_2, 1 + a_2^2)$. Suppose further that the cost satisfies $c(a) = \frac{\kappa_1 a_1^2 + \kappa_2 a_2^2}{2}$.

In this example, total welfare is given by $V(a) = a_1 - c(a)$. In other words, both efforts can affect the indicator but effort a_1 creates value since $\mathbb{E}v = a_1$ while a_2 is purely costly and

creates no value. Thus, we can view a_1 as productive effort while a_2 is a window-dressing effort. As we establish in Section 4, when the intermediary has an objective of $W(a) = V(a)$, a rating mechanism with censoring can discourage window dressing more than productive effort and lead to higher welfare. ■

The remainder of this section reduces optimal rating design to a tractable problem in two steps: we first characterize the set of interim price schedules that some rating can implement, and we then use this characterization to solve for the optimal rating.

2.1 Interim Prices: Definition and Characterization

In this section, we introduce *interim prices*, a mathematical object that determines the agent's incentives in the choice of effort and allows us to simplify the problem of rating design. Specifically, we define *interim prices* as

$$p(y) = \int \mathbb{E}[v|s] \, d\pi(s|y). \tag{3}$$

In words, p is the expected payment to the agent conditional on the indicator y, integrating over possible signals s given the rating system. Additionally, it is an equilibrium object as it depends on $\mathbb{E}[v|s]$, which depends on the market's beliefs about the agent's action profile.[8] It can also be interpreted as the agent's "second-order belief": her belief about the belief of the market on values.

Critically, interim prices are a sufficient statistic for the information structure from the agent's perspective. Specifically, for any choice of a, the agent's payoff is given by

$$\int p(y) \, dG(y|a) - c(a).$$

As a result, the problem of designing an optimal rating system is isomorphic to the problem of choosing an interim price schedule p, subject to the constraint that p must be implementable via some information structure (S, π). Thus, we need to characterize the set of feasible interim prices. Generally, there are no simple conditions to characterize the set of interim price profiles that result from a particular information structure and action profiles. However, as we will show next, under some restriction on information structures, a simple characterization exists.

To see the restrictions that ratings place on interim prices, note that market values are the interim prices associated with the fully revealing information structure, and that any interim price schedule is an iterated conditional expectation of market values. Viewed as random variables, we must therefore have $p \succeq_{\text{c.v.}} \overline{v}$, i.e., p is a mean-preserving contraction of $\overline{v}$.[9] Intuitively, a rating can

[8] Throughout this section and the next one, we suppress the market belief $\hat{a}$ in the notation in order to avoid clutter. This is without loss since the payoff of the agent depends on market beliefs only through the interim prices.

[9] The relation $p \succeq_{\text{c.v.}} \overline{v}$ represents the concave order, which implies that for all concave functions $\phi : \mathbb{R} \to \mathbb{R}$,

only average market values across realizations of the indicator, and averaging reduces dispersion.

Given the results in the literature – see for example Rothschild and Stiglitz (1970) or Gentzkow and Kamenica (2016) – it is tempting to suggest that the reverse of the above observation is also true: that mean-preserving contraction is a sufficient condition for the existence of a rating that implements p. However, the reverse is not obvious, since the same information structure determines both the market's posterior $\mathbb{E}[v|s]$ and the agent's expectation of it, $\int \mathbb{E}[v|s]\, d\pi(s|y)$. Below we show that the reverse is indeed true when interim prices and market values are comonotone.[10] Formally, we say that p and $\overline{v}$ are comonotone if:

$$p(y) > p(y') \Rightarrow \overline{v}(y;\hat{a}) > \overline{v}(y';\hat{a}).$$

In words, higher prices are associated with higher market values, so the two random variables never move in opposite directions.

Proposition 1. *Suppose that p is a function that maps values of y into $\mathbb{R}$ such that*

1. *p is comonotone with $\overline{v}$, and*
2. *$p \succeq_{c.v.} \overline{v}$.*

Then, there exists an information structure (S, π) such that $p(y) = \int \mathbb{E}[v|s]\, d\pi(s|y)$.

This proposition implies that the comonotone equilibria of the game are characterized by three conditions: the effort is incentive compatible given the interim prices,

$$a \in \arg\max_{\hat{a}\in A} \int p(y)\, dG(y|\hat{a}) - c(\hat{a}) \tag{4}$$

the interim prices dominate market values in the concave order, and interim prices and market values are comonotone.

This reduction allows us to transform the optimal rating design problem into a standard mechanism design problem with transfers, where the "transfers" are the interim prices constrained by the concave order.

Proposition 1 is the main technical tool that allows us to significantly simplify the problem of rating design and to reduce it to the constrained concavification in Theorem 1. The main challenge is the fact that interim prices are second-order expectations – expectation of the market's expectation conditional on the realization of the state. In our proof, we use the characterization result of Kleiner et al. (2021) to construct an appropriate signal structure.

$\mathbb{E}[\phi(p)] \geq \mathbb{E}[\phi(\overline{v})]$. Since Bayes plausibility implies $\mathbb{E}[p] = \mathbb{E}[\overline{v}]$, this definition is equivalent to majorization, second-order stochastic dominance, and increasing concave order – see for example Shaked and Shanthikumar (2007) Section 4.A.

[10] In the Online Appendix D, we provide an example that illustrates that, without comonotonicity, mean-preserving contraction is no longer sufficient and additional conditions are needed. We also discuss its relationship with similar results in the literature.

Remark on Comonotonicity In the rest of the paper, we focus on ratings that satisfy the monotonicity constraint. One can justify this decision using ex-post manipulation: if the agent can manipulate the indicator downwards but not upwards (hide some positive outcomes), then the effective interim prices are monotone.[11]

2.2 The Main Characterization

Given our result above, the problem of optimal rating design can be stated as maximizing the objective $W(a)$ subject to incentive compatibility (4), comonotonicity, and majorization. We refer to an effort $a \in A$ as *implementable* when there exists an interim price function for which it satisfies incentive compatibility, comonotonicity, and majorization.

We solve this problem in two steps. First, we transform prices and values into quantile space (for the indicator), where the concave order becomes a majorization constraint and the value of any objective can be computed by concavification. Second, we incorporate the first-order version of the incentive constraints through a Lagrangian argument, which leads to our main characterization.

Let $v_Q(i)$ denote the market value associated with the i-th quantile of the indicator distribution. Formally,

$$v_Q(i) = \overline{v}(G^{-1}(i|a); a). \tag{5}$$

For ease of notation, we have dropped a from the expression of the quantile value. Similarly, let $p_Q(i)$ be the quantile representation of the interim price. Given comonotonicity of p and $\overline{v}$, $p_Q(i)$ is the interim price associated with market value $v_Q(i)$.

The characterization of implementable interim prices allows us to evaluate any objective by concavification: as we show in Appendix A.1, maximizing an objective that is linear in interim prices, ignoring the incentive constraints, over implementable interim prices amounts to concavifying an integrated version of its weighting function in the quantile space. Intuitively, pooling an interval of realizations replaces the objective's value on that interval with its average, and the best monotone pooling is described by a concave envelope.

In order to incorporate incentive compatibility, we first define

$$F(i|\hat{a}; a) = G\left(G^{-1}(i|a)\,|\hat{a}\right).$$

In words, F is the distribution over quantiles when the agent chooses $\hat{a}$ but quantiles are defined according to a. In quantile space, incentive compatibility (4) requires that a maximizes $-\int_0^1 F(i|\hat{a}; a)\, dp_Q(i) - c(\hat{a})$ over $\hat{a} \in A$.

[11] The classic moral hazard literature has imposed such constraints. The seminal paper by Innes (1990) is an example.

The First-Order Approach In order to sidestep many of the complications that typically arise in moral hazard problems and in line with much of the moral hazard literature,[12] we will use the first-order approach (FOA) throughout the paper. That is, we replace the incentive constraint (4) with its first-order condition. Since our characterization identifies a candidate rating—in the leading cases, an upper- or lower-censorship of the indicator—it is sufficient to verify that the agent's payoff is quasi-concave in effort given this candidate, so that the local incentive constraint implies global incentive compatibility. In the Online Appendix C, we will use an approach similar to Chade and Swinkels (2020) to provide sufficient conditions on the distribution functions for the validity of the first-order approach. The conditions often reduce to comparing the curvature of the c.d.f. $G(y|a)$ with respect to a to that of the cost function $c(a)$ and are typically satisfied provided that $c(a)$ is sufficiently convex. We should also emphasize that this is an equilibrium condition since the interim prices and market values are calculated under the market belief about effort, $\hat{a}$, and the agent's choice a needs to coincide with the market belief.[13]

The following provides a characterization of optimal ratings under FOA:

Theorem 1. *Suppose that the highest value of $W(a)$ among implementable efforts is achieved at $a^* \in intA$, and let v_Q be the quantile representation of $\overline{v}(y;a^*)$. Suppose that a^* is not a local maximizer of $W(a)$ among all $a \in A$. Then, under the validity of FOA, there exists $\lambda^* \in \mathbb{R}^N \setminus \{0\}$ such that for every increasing $p^*(y)$ with quantile representation p^*_Q such that:*

(i) *majorization holds: $p^* \succeq_{c.v.} \overline{v}$,*

(ii) *FOA incentive compatibility is satisfied:*

$$-\int_0^1 \frac{\partial F}{\partial \hat{a}_n}(i|a^*;a^*)\, dp^*_Q(i) = \frac{\partial c}{\partial a_n}(a^*), \forall n = 1, \cdots, N,$$

then the following holds:

$$\int \Gamma(i;\lambda^*,a^*)\, dp^*_Q(i) = \int cav\Gamma(i;\lambda^*,a^*)\, dv_Q(i) \tag{D}$$

where

$$\Gamma(i;\lambda,a) = -\sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\Big|_{\hat{a}=a}$$

Theorem 1 provides a partial characterization of optimal ratings for optimal efforts a^*. It states that for any optimal rating that is implementable (comonotone, less dispersed than market values, FOA incentive compatible), the value of a *gain function* under the optimal rating should be the

[12] An exhaustive list of papers that resort to FOA would be too long. See Azevedo and Wolff (2025) and the list of papers therein for a recent discussion of the literature.

[13] The literature on verification of the first-order approach in moral hazard is quite extensive. A few examples are Mirrlees (1999), Rogerson (1985), Jewitt (1988), and more recently Conlon (2009) and Jung and Kim (2015).

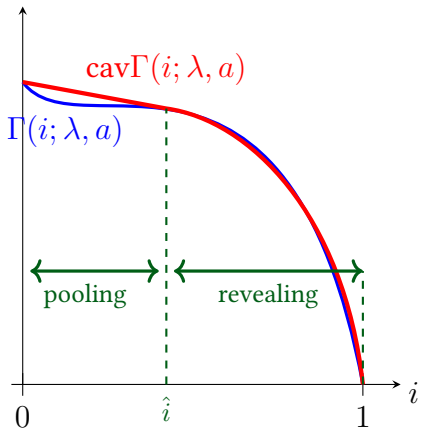


Figure 2: An illustration of the concavification of the gain function Γ and construction of optimal ratings

same as its concavification under market values. As a result, when the gain function is such that this concavification is achieved by a unique rating, the concavification determines the optimal rating. An example of this gain function is demonstrated in Figure 2. In this figure, the gain function and its concavification are depicted. Since there is a unique rating that achieves this concavification – see Proposition 2 below – the optimal rating is pooling in the lower interval and fully revealing on the upper one. It should be noted that $\Gamma(i)$ captures the value, in terms of the incentive constraints, of raising the interim price for all realizations of the indicator above quantile i. Thus, the optimal rating pools where Γ is below its concavification and reveals where the two coincide.

We should remark on the proof of Theorem 1. There are two complications that arise in our proof: first, the supports of the values and the indicator are possibly unbounded; second, incentive compatibility really captures an equilibrium concept – as in Holmström (1999)'s original career-concerns model. The additional continuity and boundedness assumptions in Assumption 1 guarantee that we can use a separating hyperplane theorem. At the heart of it is proving that if, at some effort, the first-order conditions can be perturbed in any direction by changing the interim price, then the same is true at all nearby efforts; see the Online Appendix E.

Optimality of Threshold-Based Ratings While Theorem 1 implies that concavification of the gain function is a determinant of optimal ratings, it does not yet imply that optimal ratings are simple. In fact, since the gain function to be *concavified* is only a function of the quantile i, by Carathéodory's theorem, the concave envelope of $\Gamma(i;\lambda,a)$ at any i can be achieved by using at most two points. This logic establishes that the value in (D) is always achieved by a deterministic monotone partition: the intermediary either fully reveals the indicator on some regions or pools

contiguous intervals into coarse categories. The complication is that the optimal rating must also satisfy the incentive constraints in Theorem 1; when there are multiple optimal ratings that achieve the value in (D), satisfying these constraints might require randomization. The following proposition identifies cases where monotone partitional signals are uniquely optimal:

Proposition 2. *Under the conditions of Theorem 1:*

1. *There is an optimal rating that randomizes over at most $N+1$ deterministic monotone partitions of the indicator.*

2. *Suppose that $\Gamma(\cdot;\lambda^*,a^*)$ is such that for any $i\in(0,1)$ with $\Gamma(i;\lambda^*,a^*)=\text{cav}\Gamma(i;\lambda^*,a^*)$, the function $\text{cav}\Gamma(\cdot;\lambda^*,a^*)$ is not affine on any neighborhood of i. Then the optimal rating is unique and a monotone partition. It is revealing for values of i where $\Gamma(i;\lambda^*,a^*)=\text{cav}\Gamma(i;\lambda^*,a^*)$ and pooling for maximal intervals with $\Gamma(i;\lambda^*,a^*)\neq\text{cav}\Gamma(i;\lambda^*,a^*)$.*

3. *Specifically, if $\Gamma(\cdot;\lambda^*,a^*)$ is convex–concave (resp. concave–convex) on $[0,1]$ and strictly concave on its concave part, the optimal rating is unique and deterministic: a lower (resp. upper) censorship.*

Proposition 2 clarifies the implications of Theorem 1 for the design of optimal ratings. Specifically, we do not need to mix over more than $N+1$ monotone partitional signals, which is an application of Carathéodory's theorem and the fact that there are at most N binding and linear incentive constraints under FOA. The second and third parts illustrate conditions under which the optimal rating implied by concavification of the gain function is unique. While these are conditions on the gain function, which is itself endogenous to the multipliers attached to incentive compatibility, as we show in the applications in Sections 3 through 5, they can often be checked directly using properties of the indicator technology $g(y|a)$.

Theorem 1 and Proposition 2 together provide a characterization of optimal ratings. In what follows, we use them to provide general properties of optimal ratings with single and multiple efforts or tasks.

3 Optimal Single-Task Ratings

In this section, we provide general properties of optimal ratings and how they depend on the joint distribution of the indicator and market values. As we show, whether optimal ratings involve full disclosure or censorship—and whether censorship occurs at the bottom or the top of the distribution—is determined by how effort shifts the distribution of the indicator. In this section, we focus on problems in which effort is one-dimensional. In Section 4, we study multiple tasks.

With a single action, an elementary observation is that the problem of optimal rating design boils down to a characterization of the highest implementable effort, since any lower value of effort can also be targeted by rating systems. Thus, we can always choose $W(a)$ as an increasing function in a and apply Theorem 1. The analysis in this section is particularly useful in identifying values of effort that are higher than those achieved by a fully revealing rating system. Especially in markets with positive externalities, where fully informative ratings lead to inefficiently low levels of effort, one can use ratings (absent MLRP) to improve market efficiency.

To make the model tractable, let us assume the following:

Assumption 2. *The action space A and the distribution function $g(y|a)$ satisfy the following:*

1. *The action space is $A = [0, \overline{a}] \subset \mathbb{R}$.*
2. *For all $a > 0$, the support of $g(y|a)$ is a (potentially unbounded) interval $I = [\underline{y}, \overline{y}]$ and $g(y|a)$ is twice differentiable.*

Both parts of Assumption 2 are fairly common in the moral hazard literature. Note that since we need optimal effort for the agent to be interior to A, we have to assume that $\overline{a}$ is large enough so that it is not implementable.

By Theorem 1, under FOA, the optimal rating is found by concavification of the function $\Gamma(i; \lambda, a) = -\lambda \frac{\partial F(i|\hat{a};a)}{\partial \hat{a}}\Big|_{\hat{a}=a}$, where λ is the Lagrange multiplier on the local incentive compatibility constraint, and $F(i|\hat{a}; a)$ is the induced distribution of the quantiles of the indicator when the agent chooses effort $\hat{a}$ while the market believes it to be a. The following calculation ties the object to be concavified to properties of the distribution function $G(y|a)$:

$$-\lambda \frac{\partial^2}{\partial i^2} \frac{\partial F(i|\hat{a};a)}{\partial \hat{a}}\Big|_{\hat{a}=a} = -\lambda \frac{1}{g(y|a)} \frac{\partial^2}{\partial y \partial a} \log g(y|a)\Big|_{y=G^{-1}(i|a)}$$

In other words, the concavity of $\Gamma(i; \lambda, a)$ at a particular quantile i is determined by the sign of the cross partial of the log-likelihood function $\log g(y|a)$. This implies that optimal ratings are directly tied to the supermodularity of the log-likelihood function $\log g(y|a)$. In what follows, we discuss a few cases and their economic interpretation and implications for optimal ratings.

Let us start with the canonical assumption made in the moral hazard literature, the so-called *MLRP* assumption:

Definition 1. A distribution function $g(y|a)$ is said to satisfy the Monotone Likelihood Ratio Property (**MLRP**) when $g(y|a)$ is log-supermodular. That is, $\frac{\partial^2}{\partial a \partial y} \log g(y|a) = \frac{\partial}{\partial y} \frac{g_a(y|a)}{g(y|a)} \geq 0, \forall y \in I, a \in A$.

MLRP implies that an increase in effort leads to a rightward shift of the distribution of indicator realizations. Moreover, it also implies that the same is true for the conditional distribution of the

indicator when restricted to an interval of values of y.[14]

Our first result establishes that in the presence of MLRP, the highest implementable effort is indeed associated with full information:

Proposition 3 (Full Disclosure under MLRP)**.** *Suppose the conditions of Theorem 1 and Assumption 2 hold and that $g(y|a)$ satisfies MLRP. Then the highest implementable effort is associated with interim price $p(y) = \overline{v}(y; a_{FI})$. That is, it is the highest effort level that satisfies*

$$a_{FI} \in \arg\max_{a \in A} \int \overline{v}(y; a_{FI})\, g(y|a)\, dy - c(a).$$

The above result states that the often-assumed MLRP has strong implications for what can be achieved via ratings. Specifically, it states that using ratings, it is not possible to increase the level of effort beyond what the market can achieve by fully observing the indicator.

We should note that the definition of the highest implementable effort a_{FI} involves the calculation of a fixed point. This is because market values $\overline{v}(y; a)$ should be calculated under the belief of the market that the action taken is a_{FI} while the agent is able to deviate from it. In Proposition 3, a_{FI} is defined as the highest such fixed point.

The proof of Proposition 3 follows directly from Theorem 1. Specifically, under MLRP, the function $\Gamma(i; \lambda, a)$ is either concave or convex for all values of i, depending on the sign of λ. This means that when $\lambda > 0$, Γ is concave and coincides with its concavification. Thus, given our construction of optimal ratings in Section 2, the optimal rating becomes fully revealing. In turn, if $\lambda < 0$, Γ is convex in i and thus its concavification is simply the 0 function – since $\Gamma(0; \lambda, a) = \Gamma(1; \lambda, a) = 0$. In other words, the optimal rating involves providing no information, which results in $a = 0$, which cannot be optimal, so λ cannot be negative.

3.1 Expanding and Compressing Likelihood Ratios

Many economic activities violate MLRP in systematic ways. For example, consider activities where greater effort affects not just the mean outcome but also its variance. Innovative activities often increase both upside potential and downside risk—greater R&D effort can lead to breakthroughs or failures. On the other hand, activities such as maintenance typically reduce variance—more careful attention produces more consistent outcomes. These patterns correspond to distributions where the cross-derivative of the log-likelihood changes sign.

In what follows, we define two classes of distributions and characterize the optimal ratings.

Definition 2. A distribution function $g(y|a)$ is said to satisfy:

[14] Formally, MLRP is equivalent to the statement that an increase in a increases the distributions over the indicator y according to the likelihood ratio order. See Shaked and Shanthikumar (2007), Section 1.C.

1. Expanding likelihood ratio property **(ELRP)** if for any $a \in A$, there exists $\hat{y}$ such that $\frac{\partial^2}{\partial a \partial y} \log g(y|a) > 0$ when $y > \hat{y}$ and $\frac{\partial^2}{\partial a \partial y} \log g(y|a) < 0$ when $y < \hat{y}$,

2. Compressing likelihood ratio property **(CLRP)** if for any $a \in A$, there exists $\hat{y}$ such that $\frac{\partial^2}{\partial a \partial y} \log g(y|a) > 0$ when $y < \hat{y}$ and $\frac{\partial^2}{\partial a \partial y} \log g(y|a) < 0$ when $y > \hat{y}$.

The terminology reflects how effort affects the signal distribution's tails. Under ELRP, increased effort expands the tails, while under CLRP, increased effort compresses the distribution toward the center.

For example, consider $\log y \sim \mathcal{N}(\log a, a^\gamma)$, that is, $\log y$ has a normal distribution with mean $\log a$ and variance a^γ. In this case, we can use the definition of the density of the normal distribution to show that

$$\frac{\partial^2}{\partial a \partial y} \log g(y|a) = \frac{1 + \gamma \log y/a}{a^{1+\gamma} y}.$$

When $\gamma > 0$, g satisfies ELRP since the above is positive if and only if $y/a \geq e^{-1/\gamma}$. In contrast, when $\gamma < 0$, g satisfies CLRP since the above is positive if and only if $y/a \leq e^{-1/\gamma}$.

The notions of ELRP and CLRP are tied to whether an increase in effort a leads to a higher or lower variance of the indicator y. Specifically, suppose that $y = f(\sigma(a)\varepsilon + m(a))$ with m and f increasing, and ε has density $e^{h(\varepsilon)}$ such that $h(\varepsilon)$ is concave and $\varepsilon + h'(\varepsilon)/h''(\varepsilon)$ is increasing and unbounded in ε. In this case, y exhibits ELRP (CLRP) if $\sigma(a)$ is increasing (decreasing) in a. Several classes of distributions satisfy these properties for h: Normal, Gumbel, Generalized Normal, Logistic, etc. For this class of distributions, a constant $\sigma(a)$ leads to MLRP.

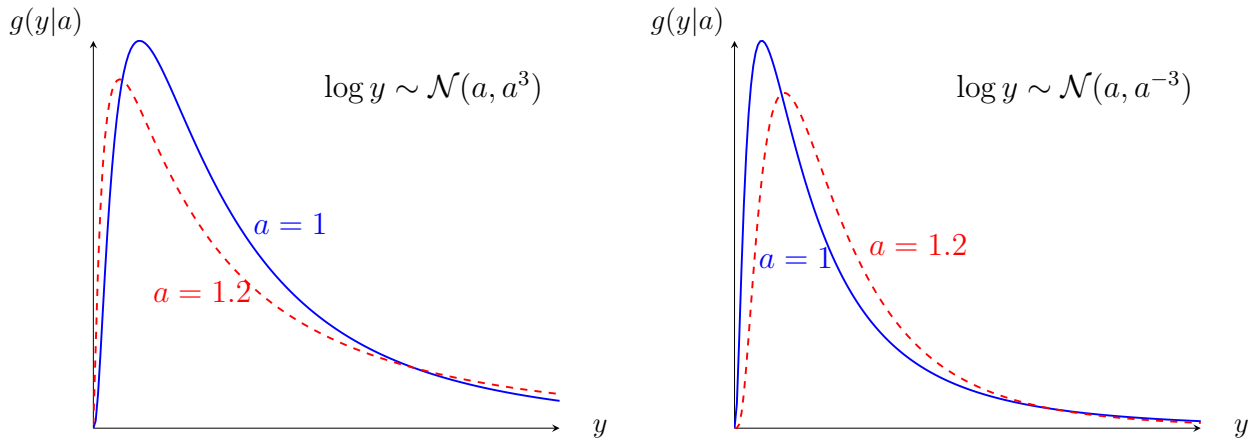
A version of these examples, with mean parameter a in place of $\log a$, is depicted in Figure 3. As can be seen, in the case of ELRP, the tail densities increase as effort a increases, while the densities for mid-realizations decline. In contrast, under CLRP, tail densities decline, while the densities for mid-realizations increase. In both cases, the two densities intersect exactly twice, which is in contrast with the single crossing of MLRP.

Given these definitions, we can state our result on optimal ratings for these classes of distributions:

Proposition 4 (Censorship under ELRP and CLRP)**.** *Suppose Assumptions 1 and 2 hold and FOA is valid.*

1. *If $g(y|a)$ satisfies ELRP, then the highest implementable effort a_{LC} is the highest value of effort that is incentive compatible for an interim price associated with lower-censorship ratings, i.e., ratings that pool values of y below a threshold and reveal higher values.*

2. *If $g(y|a)$ satisfies CLRP, then the highest implementable effort a_{UC} is the highest value of effort that is incentive compatible for an interim price associated with upper-censorship ratings, i.e., ratings that pool values of y above a threshold and reveal lower values.*



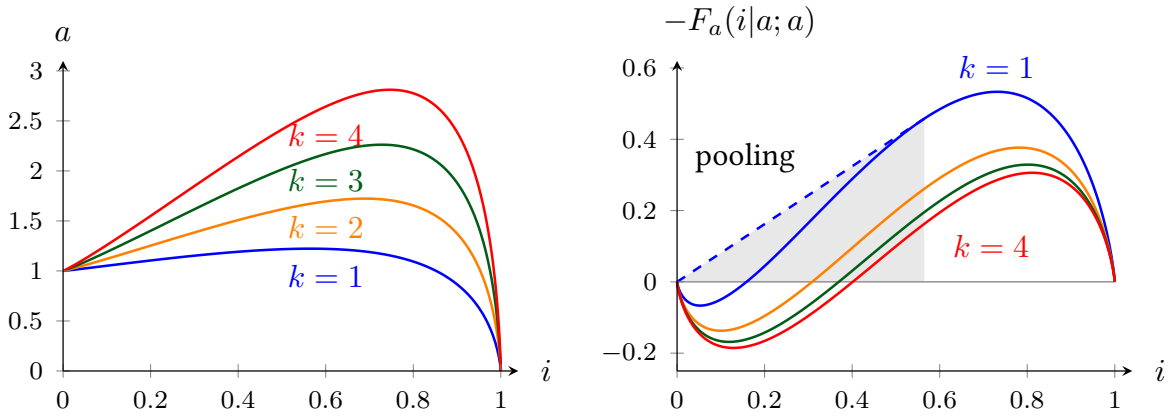
(a) ELRP (b) CLRP

Figure 3: Distributions with ELRP (left) and CLRP (right)

As the above proposition establishes, optimal ratings for ELRP and CLRP distributions are fairly simple. They involve either lower censorship (in the case of ELRP) or upper censorship (in the case of CLRP). In what follows, we provide an example and discuss its implications for various tasks and technologies.

Suppose that $y \sim \mathcal{N}\left(a, (ka)^2\right)$, that y determines market values, i.e., $\overline{v}\left(y; a\right) = y$, and that the cost is $c\left(a\right) = a^2/2$. Under a full-information rating, $p\left(y\right) = y$, the profit of the agent is $a - a^2/2$, which is maximized at $a_{FI} = 1$. It can be easily checked that in this case $G\left(y|a\right)$ satisfies ELRP. For any value of $i \in [0, 1]$, we can find the highest level of effort that is a best response to pooling of the i lowest realizations of the indicator y (lower censorship). This is depicted in Figure 4 (left panel) for values of $k = 1, 2, 3, 4$. At the lowest value, $i = 0$ (full information), optimal effort is $a_{FI} = 1$. Optimal effort peaks at some threshold, 0.56, 0.69, 0.73, and 0.75, respectively, and falls to zero as i tends to 1 (complete pooling). It should be noted that as the variance of y becomes more responsive to a, the highest possible value of effort increases. Figure 4 (right panel) depicts the marginal change in the quantile as a result of an increase in a, $-F_a\left(i|a; a\right)$, and its concavification (in the case of $k = 1$). The threshold for pooling on the right coincides with the peak of the left plot.

The results of this section can be summarized by how effort moves the distribution of the indicator. When effort shifts the indicator in the sense of MLRP, any pooling of realizations only weakens the link between the indicator and the market's assessment, and full disclosure maximizes effort (Proposition 3). Censorship becomes optimal when effort changes the shape of



(a) Optimal effort for pooling the lowest i realizations (b) Concavification of the marginal change in quantiles of y

Figure 4: Optimal efforts for lower-censorship policies

the distribution rather than simply shifting it. If higher effort leads to a riskier outcome (ELRP), such as innovative activities, very low realizations become relatively more likely under high effort while moderately low ones become less likely. Pooling of low realizations via lower-censorship ratings then transfers reward from the latter to the former, provides insurance against possible downsides, and encourages risk-taking. On the other hand, for an activity where more effort leads to a more precise outcome (CLRP), such as maintenance activities, very high realizations are relatively more likely under low effort. Full revelation at low values and pooling at higher values then encourage higher effort to avoid punishments for low outcomes (Proposition 4). In each case, the rating pools exactly the region of the indicator where the likelihood ratio moves against effort, and reveals the rest.

4 Multi-Task Moral Hazard and Window Dressing

In this section, we illustrate the value of our characterization results above by applying them to the classic multi-task moral hazard model. Since the seminal work of Holmström and Milgrom (1991), multi-task principal-agent models have become the workhorse for analyzing incentives in settings where agents can use several actions to affect the observed outcomes – see also Baker (1992) and Dewatripont et al. (1999).[15] Despite this prevalence of multi-task moral hazard in ap-

[15]Since Dewatripont et al. (1999) analyze a variant of Holmström (1999)'s career-concerns model, their model is closer to ours, where the agent's compensation is determined by market expectations as opposed to an endogenous contract chosen by the principal, as in Holmström and Milgrom (1991) and Baker (1992).

Several empirical studies have looked at variants of the multi-task moral hazard model. A partial list includes Dumont et al. (2008) and Alexander (2020) for compensation of doctors, Acemoglu et al. (2020) for incentives in military and security forces, De Janvry et al. (2023) for incentives of government employees, Andrabi and Brown

plied settings and the importance of disclosure/rating policies in it, a unified theoretical analysis of rating policies has received little attention. We consider a variant of the model in Dewatripont et al. (1999) to understand how rating design can be used to mitigate multi-task incentive problems. As we will see, censorship can discourage window dressing relative to productive effort, and the direction of censorship has a natural analogue to the results of Section 3.

As before, the agent chooses a vector of efforts $a = (a_1, a_2) \in [0, \overline{a}]^2$, with a_1 being the *productive* effort and a_2 being a *window dressing* effort. The difference between productive and window-dressing effort is defined in Assumption 3 below:

Assumption 3 (Productive Effort vs. Window Dressing)**.** *The distribution of* (v, y) *conditional on* a *satisfies:*

1. *market value* $\overline{v}(y; a) = \mathbb{E}[v|y; a]$ *is increasing in* a_1*,*
2. *both efforts increase the indicator, i.e.,* $\mathbb{E}[y|a]$ *is increasing in* a_n *for* $n = 1, 2$*,*
3. *welfare is decreasing in* a_2*, i.e.,* $V(a) = \mathbb{E}[v|a] - c(a)$ *is decreasing in* a_2 *and quasi-concave in* a_1 *given* a_2*.*

The above are standard properties of multi-tasking models. They imply that efficiency leads to $a_2^* = 0$ and $a_1^* \in \arg\max_{a_1} V(a_1, 0)$.

An example of this is the standard model of Dewatripont et al. (1999) where

$$\begin{pmatrix} v \\ y \end{pmatrix} = \begin{pmatrix} b_v \cdot a \\ b_y \cdot a \end{pmatrix} + \varepsilon, \qquad \varepsilon \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma_v^2 & \sigma_{vy} \\ \sigma_{vy} & \sigma_y^2 \end{pmatrix}\right), \sigma_{vy} > 0$$

where $b_v, b_y \in \mathbb{R}_+^2$ capture the effect of a on the market value and the indicator. Using properties of the normal distribution, we can show that market values conditional on y and belief $\hat{a}$ are:

$$\overline{v}(y; \hat{a}) = \mathbb{E}[v|y; \hat{a}] = \beta y + (b_v - \beta b_y) \cdot \hat{a}$$

where $\beta = \sigma_{vy}/\sigma_y^2 > 0$. Since a_2 must be a window-dressing effort, we should have $b_{v,2} = 0, b_{y,2} > 0$. Moreover, for market values to be increasing in a_1, we need $b_{v,1} > \beta b_{y,1}$. This model of multi-tasking can be reduced to a single effort. Hence, the results from Section 3 apply to it. In what follows, we are concerned with technologies that are not reducible to a single task.

Throughout the section, our focus will be on welfare-maximizing rating design, i.e., we set $W(a) = V(a)$. The key trade-off is that, since window dressing is a completely wasteful activity but has private benefits for the agent, an intermediary that cares about welfare might want to hide some information to differentially affect the incentives for the two activities. As we will show, the key criterion is how various values of the indicator are relatively more informative about a_1 or a_2.

(2022) for incentives of teachers, and Mayzlin et al. (2014) and Hui et al. (2025) for incentives on online platforms.

Applying Theorem 1 to this setting, the gain function is given by

$$\Gamma(i;\lambda,a) = -\lambda_1 \left.\frac{\partial G\left(G^{-1}(i|a)\,|\hat{a}\right)}{\partial \hat{a}_1}\right|_{\hat{a}=a} - \lambda_2 \left.\frac{\partial G\left(G^{-1}(i|a)\,|\hat{a}\right)}{\partial \hat{a}_2}\right|_{\hat{a}=a}.$$

As in Section 3, we can write

$$\Gamma'(i;\lambda,a) = -\lambda_1 \left.\frac{\partial \log g(y|a)}{\partial a_1}\right|_{y=G^{-1}(i|a)} - \lambda_2 \left.\frac{\partial \log g(y|a)}{\partial a_2}\right|_{y=G^{-1}(i|a)}.$$

We will call $\frac{\partial \log g(y|a)}{\partial a_n}$ the *score* of task n and refer to it as $\ell_n(y|a)$.

To understand our conditions on the multi-task technology, it is useful to use an information-theoretic intuition to understand the score better. To see this, imagine that the vector $a = (a_1, a_2)$ is unknown and distributed according to an improper uniform prior density. Consider observing the indicator y and having a posterior density $\varphi(a|y)$. Then Bayesian updating implies that

$$\log \varphi(a|y) = \log g(y|a) + \text{function of } y$$

This implies that

$$\log \frac{\varphi(a + \varepsilon e_n|y)}{\varphi(a|y)} \approx \varepsilon \ell_n(y|a)$$

with e_n being a vector with its n-th element equal to 1 and 0 otherwise. The above says that, for a fixed observation of the indicator y, $\ell_n(y|a)$ measures the change in likelihood of having a slightly higher a_n.

The key property that we use concerns the relative change of the score for different values of y, as shown in the following definition:

Definition 3. A multi-tasking technology $G(y|a)$ exhibits the *Monotone Relative Informativeness Property (MRIP)* if for all $y_2 > y_1$ and $a \in [0, \overline{a}]^2$:

$$D(y_1, y_2|a) = \frac{\partial \ell_1(y_1|a)}{\partial y}\frac{\partial \ell_2(y_2|a)}{\partial y} - \frac{\partial \ell_1(y_2|a)}{\partial y}\frac{\partial \ell_2(y_1|a)}{\partial y}$$

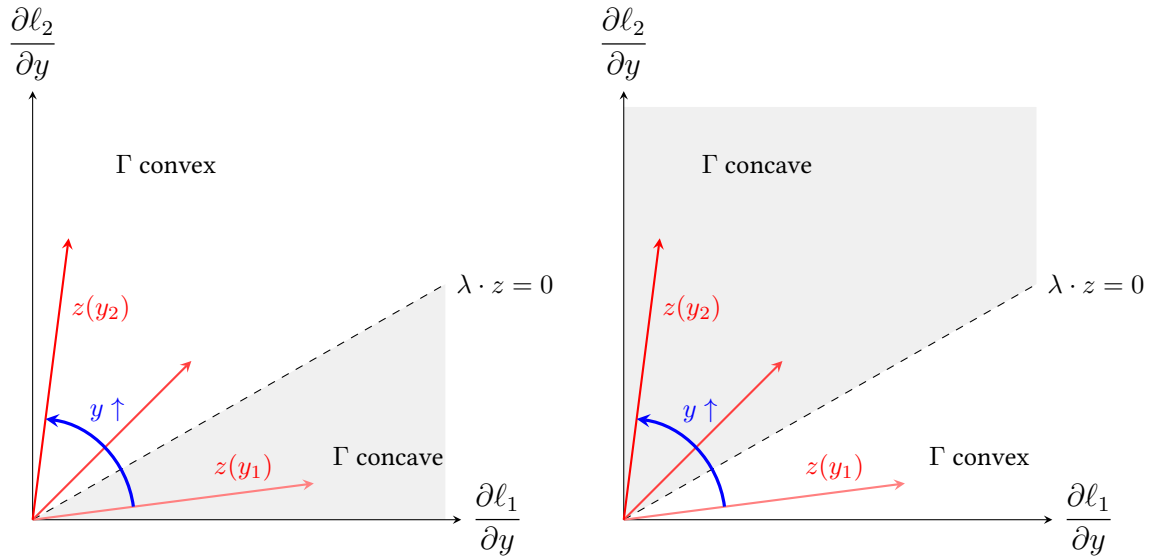
is never zero and always has the same sign, i.e., it is negative or positive for all such y_1, y_2 and a.

As its name suggests, MRIP captures how informative higher values of the indicator are about productive and window-dressing effort respectively. To see this, consider the vector $z(y|a) = \left(\frac{\partial \ell_1}{\partial y}(y|a), \frac{\partial \ell_2}{\partial y}(y|a)\right)$ in the plane. Basic arithmetic of trigonometric functions allows us to write

$$D(y_1, y_2|a) = \|z(y_1|a)\|_2 \|z(y_2|a)\|_2 \sin\left(\theta(z(y_2|a)) - \theta(z(y_1|a))\right)$$

where in the above $\|\cdot\|_2$ is a vector's length in $\mathbb{R}^2$ and $\theta(z)$ is the angle of the vector z with the x-axis. Hence, the statement that D is always negative (positive) means that as y increases, the angle with the x-axis decreases (increases). This is depicted in Figure 5.

But beyond this geometric and mathematical property, it can also be seen that MRIP is capturing how informative higher values of the indicator are about action a_1, productive action, relative to



(a) $\lambda_1 > 0 > \lambda_2$, upper censorship

(b) $\lambda_1 < 0 < \lambda_2$, lower censorship

Figure 5: Monotone relative informativeness (MRIP) in Example 2

a_2, window-dressing action. To see this, suppose that y satisfies MLRP with respect to a_2, i.e., $\frac{\partial \ell_2(y|a)}{\partial y} > 0$. Then, we can write D as

$$D\left(y_1, y_2|a\right) = \frac{\partial \ell_2\left(y_2|a\right)}{\partial y} \frac{\partial \ell_2\left(y_1|a\right)}{\partial y} \left[\frac{\frac{\partial \ell_1(y_1|a)}{\partial y}}{\frac{\partial \ell_2(y_1|a)}{\partial y}} - \frac{\frac{\partial \ell_1(y_2|a)}{\partial y}}{\frac{\partial \ell_2(y_2|a)}{\partial y}} \right]$$

In the above, the term in the bracket determines the sign of D. Hence, $D > 0$ suggests that lower values of the indicator are relatively more informative of the productive effort than of the window-dressing effort, and vice versa. This is the sense in which MRIP captures relative informativeness. The formulation in Definition 3 allows us to generalize this to arbitrary technologies without requiring the MLRP assumption. Let us use an example to clarify this:

Example 2. Assume:

$$v = a_1 + \varepsilon_1, y = v + a_2\varepsilon_2, \varepsilon_1 \sim \mathcal{N}\left(0, 1\right), \varepsilon_2 \sim \text{Exp}\left(1\right)$$

with $\varepsilon_1, \varepsilon_2$ independent. In this case, we can solve for the distribution of y, which is given by

$$g\left(y|a\right) = \frac{1}{a_2} e^{-\frac{u}{a_2} - \frac{1}{2a_2^2}} \Phi\left(u\right), u = y - a_1 - \frac{1}{a_2},$$

with Φ the c.d.f. of the standard normal. Moreover, the expected value of v given y and a is given by

$$\overline{v}\left(y; a\right) = a_1 + \frac{1}{a_2} - \frac{\phi\left(u\right)}{\Phi\left(u\right)}$$

with ϕ being the density of the standard normal. Since ϕ/Φ is decreasing in u and has a derivative

higher than -1, the market value is strictly increasing in y and a_1. Finally, as long as the cost function is convex in a_1 and increasing in a_2, Assumption 3 is satisfied. Note that in this case, both scores $\ell_n(y|a)$ are monotone increasing in y, which means that MLRP is satisfied with respect to each effort. This implies that the sign of D is determined by the monotonicity properties of

$$\frac{\frac{\partial \ell_1(y|a)}{\partial y}}{\frac{\partial \ell_2(y|a)}{\partial y}} = a_2^2 \frac{-\frac{d}{du}\frac{\phi(u)}{\Phi(u)}}{1+\frac{d}{du}\frac{\phi(u)}{\Phi(u)}}$$

which is decreasing in y and, as a result, $D > 0$. ■

The following theorem states our main result on optimal ratings in the multi-task model:

Theorem 2. *Suppose that the conditions of Theorem 1 hold, $G(y|a)$ is a technology that exhibits MRIP, and FOA is valid. Then welfare-maximizing ratings are either upper or lower censorship. In other words, the gain function is either concave–convex or convex–concave.*

The significance of Theorem 2 is that it ties a natural notion of informativeness, relative informativeness between the productive and the window-dressing efforts, to general properties of welfare-maximizing ratings. Its implication can be seen in Figure 5. In the figure, the dashed line represents the value of y for which the second derivative of the gain function is 0. In Figure 5a, higher values of y are associated with convexity of the gain function which leads to pooling above a threshold and thus upper-censorship ratings. In Figure 5b, the opposite happens, and thus the optimal rating is lower censorship. It should be noted that not all values of y for which the gain is concave should be separated, and the dashed line only represents the inflection point of the gain function and not the censorship threshold.

Theorem 2 does not determine the direction of the censorship. While one expects the sign of $D(y_1, y_2|a)$ to be indicative of that direction, this ultimately depends on one important factor: how changes in incentives affect market beliefs, which in turn affect the incentives through their effect on $\overline{v}(y; a)$. This is in addition to the complementarity between the two efforts in terms of cost. To this end, we make two assumptions: first, we assume that $\overline{v}(y; a)$ is affine in y; second, we assume that $G(y|a)$ belongs to the location–scale class of distributions. Namely, we consider the location–scale family $G(y|a) = H\left(\frac{y-\mu(a)}{\sigma(a)}\right)$ such that a controls the mean and variance of y, with H the c.d.f. of a distribution with a log-concave density. Several classes of distributions satisfy MRIP within this family. These include Normal, Logistic, Gumbel, Generalized Normal, Gamma (with shape parameter ≥ 1), and Weibull (with shape parameter ≥ 1). We will use μ_n, σ_n, and c_n for the partial derivatives of μ, σ, and c, respectively, with respect to a_n. We have the following proposition:[16]

[16]MRIP fails for H being Cauchy or Fréchet but neither of these has log-concave densities. Evidently, MRIP and log-concavity of the density of H are not equivalent and it is possible to come up with examples that are log-concave but not MRIP. However, among the textbook classes of distributions they tend to be equivalent.

Proposition 5. *Suppose that G is a location–scale technology that exhibits MRIP with a log-concave density and that $\overline{v}(y;a)$ is a linear function of y. Moreover, suppose that FOA is valid and $\mu_1\sigma_2 - \sigma_1\mu_2 \neq 0$. Then the welfare-maximizing rating is upper censorship if and only if*

$$\frac{\mu_2 c_1 - \mu_1 c_2}{\mu_1\sigma_2 - \sigma_1\mu_2} \geq 0 \tag{6}$$

and it is lower censorship otherwise.

In words, if window dressing is the cheaper way to raise the mean of the indicator – so that the numerator of (6) is positive – then when window dressing has a larger impact on the dispersion of the indicator relative to its mean, the optimal rating is upper censorship; lower censorship arises when the impact of window dressing is more pronounced on the level of the indicator relative to its dispersion. When instead productive effort is the cheaper source of the mean of the indicator, the numerator of (6) is negative, and the direction of censorship is reversed. We provide the derivations behind Proposition 5 in the Online Appendix B.

The logic behind this result is one of implementation. Under welfare maximization, the objective depends on the rating only through the action that it implements, and thus the role of the rating is to make the implemented action incentive compatible. Within the location–scale class, the rating affects incentives only through two statistics: the marginal reward that it attaches to the level of the indicator and the marginal reward that it attaches to its dispersion. The agent's first-order conditions determine both at the implemented action. When window dressing is the cheaper source of the level of the indicator, a rating that rewards the level alone induces too much window dressing. The only remaining margin is dispersion, and since window dressing is the more dispersive effort, incentive compatibility requires that dispersion be punished. Since pooling high realizations of the indicator removes the reward for upside risk, upper censorship is the rating that punishes dispersion. When either of the two comparisons in (6) is reversed, the same logic delivers lower censorship.

Proposition 5 is the analogue of the results of Section 3. While there, the direction of the censorship was determined by whether more effort expanded (ELRP) or compressed (CLRP) tail events, here something similar holds. The difference, however, is that in this multi-task model, what matters is the relative impact of window dressing and productive efforts on tail events as illustrated by (6), weighed against the relative cost of the two efforts in moving the level of the indicator. When window dressing has a larger impact on tail events, high realizations of the indicator should be censored, and when productive effort has a higher impact on tail events, low realizations should be censored. One can thus view MRIP as the multi-task counterpart of the MLRP, ELRP, and CLRP conditions: it plays the same role for the question of which task the indicator speaks about that those conditions play for the question of how much it speaks. This is also the sense in which our results go beyond the logic of low-powered incentives in Holmström

and Milgrom (1991): since prices are determined by the market, the power of incentives cannot be lowered uniformly, but a rating can remove incentives locally by pooling indicators, and it is optimal to do so exactly where the indicator mainly reflects manipulation.

While Proposition 5 is useful in terms of intuition for the direction of censorship, ultimately the condition is in terms of the effort a, which itself is endogenous. The following example is meant to shed light on this:

Example 1 (continued)**.** Consider the setup in Example 1, which is a special case of the location–scale family in Proposition 5. In this case,

$$\mu_1\sigma_2 - \sigma_1\mu_2 = \frac{a_2}{\sqrt{1+a_2^2}} > 0$$

This, together with Proposition 5, implies that the optimal rating is upper censorship if and only if

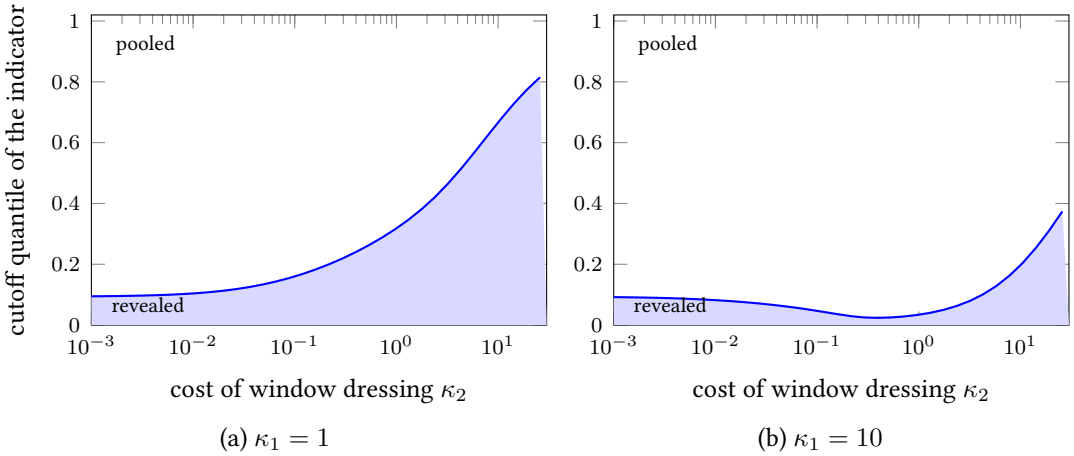
$$\mu_2 c_1 - \mu_1 c_2 = \kappa_1 a_1 - \kappa_2 a_2 \geq 0.$$

In particular, we establish in the Online Appendix B that the welfare-maximizing rating is upper censorship and is never full disclosure. In this location–scale family, ratings affect incentives through their effect on the location (mean of y) and the scale (standard deviation of y) of the indicator. Both efforts increase the mean of the indicator, so the reward to location affects both efforts, while the reward to scale affects only window dressing. A negative reward to scale thus discourages window dressing alone, and this is achieved via upper censorship.

To see this trade-off better, suppose that $\kappa_1 = 1$ and $\kappa_2 = 0.25$, so that window dressing is four times cheaper than productive effort. Under full disclosure, both efforts raise the mean of the indicator by the same amount. Since the variance of y is $1 + a_2^2$, window dressing also makes the indicator more noisy. Thus, the market puts a low weight on the indicator, which further weakens the incentives for productive effort. In equilibrium, the agent chooses $a_1 = 0.35$ and $a_2 = 1.38$ and the welfare of full disclosure is 0.05. The optimal rating is upper censorship: it reveals realizations of the indicator below its 21st percentile and pools all realizations above it. Pooling high realizations punishes dispersion, and since only window dressing affects the dispersion of the indicator, it discourages window dressing alone: a_2 falls from 1.38 to 0.66. With less window dressing, the indicator becomes more precise, and the weight that the market puts on it rises as well. Even though pooling lowers the reward to the level of the indicator for a given response of the market, this stronger response raises productive effort from 0.35 to 0.38, and welfare increases from 0.05 to 0.25.

■

Following Holmström and Milgrom (1991), a question that arises in our multi-task environment is how incentives react to a decline in the cost of window dressing. In a linear contracting



(a) $\kappa_1 = 1$ (b) $\kappa_1 = 10$

Figure 6: Cutoff of the optimal rating for Example 1: realizations of the indicator below the cutoff are revealed, and those above it are pooled

environment, they show that declines in window-dressing cost lead to lower-powered incentives – they reduce the dependence of wages on output (equivalent to our indicator). One can ask a similar question in our environment, with the caveat that low-powered incentives are not as well defined. We can use the above example and see how the cutoff is going to change as the cost of window dressing declines. As shown in the example, the optimal rating is upper censorship. This is depicted in Figure 6. When productive effort is relatively cheap (left panel), censorship is used to reduce window dressing and shift incentives to productive effort, and the optimal rating becomes less informative as window dressing becomes cheaper. When productive effort is expensive (right panel), there is little productive effort to protect, and the benefits of pooling come mainly from reducing the cost of window dressing itself. These benefits are small both when window dressing is very costly, since the agent does little of it, and when it is very cheap, since its cost is then small. Thus, there is more pooling at intermediate levels of window-dressing cost, and the informativeness of the optimal rating is non-monotone in the cost of window dressing.

5 Application: Redistributive Test Design

Recent public discourse in the education realm has highlighted the biases of standardized testing (such as the SAT) and testing of difficult subjects (such as math) against students with socioeconomic disadvantages.[17] These observations have prompted a movement to relax the requirement

[17] In July 2023, the California Board of Education adopted the controversial California Mathematics Framework, which sets guidelines for mathematics education in California public schools. Citing the students' socioeconomic disadvantages, the framework calls for some relaxations in testing standards and education of mathematics. For more information, see the article in the New Yorker on the California Mathematics Framework.

that students take such tests. This includes several universities' policies to make the SAT optional (see Dessein et al. (2025)) and attempts at making mathematics education more accessible and easier. More recently, however, several highly selective colleges, including MIT, Dartmouth, and Harvard, have reinstated their testing requirements, arguing that test scores help predict academic success and identify well-prepared students from less-advantaged backgrounds (Wren, 2022; Dartmouth College, 2024; Hoekstra, 2024); see Friedman et al. (2025) for evidence on the predictive power of test scores at such colleges. Inspired by this, in this section, we provide an alternative answer to this debate in the form of optimal test design.

To see this, consider a student who could be of type $\theta \in \{R, P\}$ with probabilities f_R, f_P. Suppose that both types can exert effort a_θ, which leads to a distribution of an indicator y that is distributed according to $g(y|a_\theta)$ with support given by $I = \left[\underline{y}, \overline{y}\right]$ – with the possibility that $\underline{y} = -\infty$ and $\overline{y} = \infty$. The cost of effort for each type is $k_\theta c(a)$, where $c(a)$ is a convex and increasing function, and $0 < k_R < k_P$. Finally, let $v = y$ so that the market values are simply the value of the indicator, and let $p(y)$ be the interim price function that is increasing. Let us assume that an increase in a leads to an increase in $G(y|a)$ in the sense of first-order stochastic dominance. Here, the market can be thought of as colleges that value students according to their test scores.

Consider a rating designer who wishes to maximize the following objective

$$\alpha_P f_P \left[\int_I p(y)\, dG(y|a_P) - k_P c(a_P)\right] + \alpha_R f_R \left[\int_I p(y)\, dG(y|a_R) - k_R c(a_R)\right] \tag{7}$$

where $\alpha_P f_P + \alpha_R f_R = 1$ and $\alpha_P > 1 > \alpha_R$. The distributional weights α_θ are attached to the student's type—that is, to the socioeconomic disadvantage captured by the cost parameter k_θ—and not to the realization of the test score itself.

The problem of optimal rating design is then to find $p(y)$ and a_R, a_P to maximize the objective in (7) subject to incentive compatibility for both types and to the constraint that $p(y)$ is a mean-preserving contraction of y, which is distributed according to $f_P G(y|a_P) + f_R G(y|a_R) = \overline{G}(y|a_R, a_P)$. In this case, a similar proof to that of Theorem 1 implies that, under the validity of FOA, the optimal rating can be found by concavification of the gain function

$$\begin{aligned}\Gamma(i) = &-\alpha_P f_P G\left(\overline{G}^{-1}(i)\,|a_P\right) - \alpha_R f_R G\left(\overline{G}^{-1}(i)\,|a_R\right) \\ &- \lambda_P G_a\left(\overline{G}^{-1}(i)\,|a_P\right) - \lambda_R G_a\left(\overline{G}^{-1}(i)\,|a_R\right)\end{aligned}$$

where in the above i is the quantile of y according to $\overline{G}$ and λ_θ's are the multipliers on the associated incentive compatibility constraints.

The following proposition guarantees that, for a class of distribution functions, the second derivative of the above object switches sign at most two times. This would imply that the optimal rating is always switching between at most three regions of pooling and separation:

Proposition 6. *Suppose that* $\log g(y|a) = \zeta(y) + r(y)\, m(a) - b(a)$ *where* $r(y), m(a), b(a)$ *are increasing functions and* $m(a) > 0$. *Then the optimal rating that maximizes (7) always has at most three alternating intervals of pooling and revelation.*

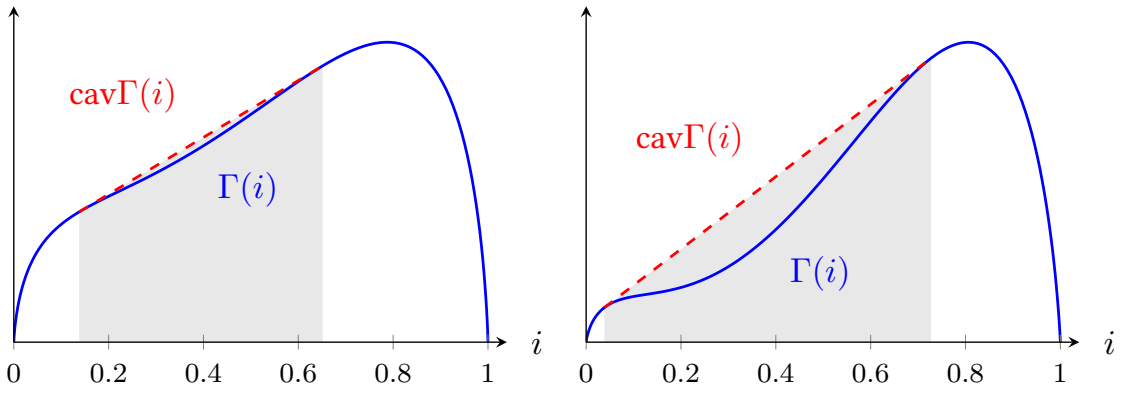
The class of distributions considered in Proposition 6 includes some of the fairly common ones that are used in applied work, including 1. a normal distribution such that the mean is controlled by a, 2. a log-normal distribution such that the mean of $\log y$ is controlled by a, 3. a Gumbel distribution such that the location is controlled by a, 4. a Fréchet distribution such that the scale parameter is controlled by a, among others. Note that the assumption implies that $\log g$ is supermodular in (y, a), i.e., it satisfies MLRP.

One might expect redistributive motives to always favor pooling low realizations, since pooling at the bottom shields the disadvantaged type. However, pooling low realizations weakens both incentive constraints. Under certain conditions on the likelihood ratio g_a/g for low realizations of y – specifically, when it becomes arbitrarily large in absolute value as $y \to \underline{y}$ – low realizations become very informative about effort, and the incentive effect dominates the redistributive motives for low realizations, and optimal ratings become mid-censorship. To see this, let us consider the following example.

Example 3. Suppose that $y = a + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, 1)$ and that $c(a) = a^2/2$. Let us also assume that $\alpha_P = 1/f_P, \alpha_R = 0$, so that the objective is to maximize the payoff of the high-cost type. As mentioned before, this example satisfies the requirement of Proposition 6, and hence optimal ratings switch at most twice between pooling and revelation. Our calculations illustrate that optimal ratings are indeed mid-censorship: those that pool middle observations of y and separate the extreme realizations. We further assume that $k_R = 1/2$ and that $f_P = f_R = 1/2$. Figure 7 depicts the optimal rating and the concavification of the gain function described above for different values of the cost for type P.[18] In the left panel, the cost of type P is closer to that of type R, $k_P = 3/5$. In this case, the optimal rating pools observations approximately between the 14th and 65th percentiles of the y distribution. When the cost for type P increases to 4/5, the optimal rating pools observations between the 4th and the 73rd percentiles. As can be seen, as the difference between the two cost types increases, the rating policy becomes less informative in order to redistribute more across the types.

Relative to full disclosure, the objective of the designer increases from 0.833 to 0.841 when $k_P = 3/5$ and from 0.625 to 0.675 when $k_P = 4/5$. It should be noted that this comes at the cost of lower effort by both types: when $k_P = 4/5$, a_P falls from 1.25 to 0.77 and a_R from 2 to 1.78. ■

[18] In order for the difference between the concavification and the function to be more visible, we subtract $1 - i$ from the gain function Γ in the plot. Since this subtraction is linear, it does not affect the resulting concavification in terms of the pooling and separating intervals.



(a) Low cost for type P, $k_P = 3/5$

(b) High cost for type P, $k_P = 4/5$

Figure 7: Determining the optimal rating for Example 3

Viewed through the lens of the current debate on standardized testing, our analysis suggests an alternative to making tests optional: rather than letting students withhold low scores, which effectively pools the bottom of the score distribution, a designer with redistributive motives can pool intermediate scores while preserving the incentive effects of the extremes.

6 Conclusion

In this paper, we have developed a general framework for the design of rating systems in the presence of moral hazard and strategic manipulation. Our approach makes this problem tractable by identifying "interim prices" as the sufficient statistic for the agent's incentives. Under a natural monotonicity restriction, implementability is equivalent to a majorization restriction: feasible interim prices are mean-preserving contractions of the full-information market values. This converts rating design into a classic moral-hazard problem with a majorization constraint.

Building on this characterization, we provide a general solution method. Under a first-order approach to incentive compatibility, optimal rating design reduces to concavifying a gain function in quantile space. This formulation yields two broad takeaways. First, it delivers a set of sufficient statistics for optimal transparency: the technology matters through how effort shifts the quantile distribution of the indicator and, in turn, the distribution of market values. Second, it implies structure on optimal information policies. Under mild conditions, optimal ratings are simple deterministic monotone partitions of the indicator space: the intermediary either fully reveals the indicator on some regions or pools contiguous regions into coarse categories.

The economic insights that emerge from our analysis connect the statistical properties of the

task to the structure of optimal disclosure. In the canonical benchmark with monotone likelihood ratios, full-information disclosure achieves the highest implementable effort. Departures from this benchmark generate systematic and testable patterns of censorship. For innovative activities where greater effort expands outcome variance (ELRP), *lower-censorship* ratings that pool poor realizations provide insurance against downside risk, encouraging risk-taking. For maintenance activities where effort compresses variance (CLRP), *upper-censorship* ratings that pool high realizations punish poor outcomes, deterring negligence and encouraging consistency. Strong redistributive motives create a tension between incentive provision and protecting disadvantaged agents, which our test-design application resolves with mid-censorship. More broadly, the framework clarifies why "more transparency" is not a universal remedy: whether additional information strengthens incentives depends on which parts of the outcome distribution effort affects.

Our second set of results concerns multi-task environments with window dressing, where more informative ratings can intensify incentives for manipulable activities that improve measured performance without an increase in underlying value. Under the monotone relative informativeness property (MRIP)—the multi-task counterpart of the single-task taxonomy—welfare-maximizing ratings are again upper or lower censorship, even when the technology satisfies MLRP and the single-task logic would call for full disclosure. Finally, in our application to redistributive test design, we show how the same concavification logic rationalizes *mid-censorship* rules that pool intermediate outcomes while separating extremes. These results formalize a common policy intuition: the optimal granularity of evaluation depends jointly on incentive provision, manipulability, and distributional objectives.

A Proofs

A.1 Proof of Proposition 1

Proof. As in Section 2.2, we work with the quantile representations p_Q, v_Q. By Assumption 1, both p_Q, v_Q are integrable monotone functions that map $[0,1]$ to $\mathbb{R}$ – p_Q is integrable because $p_Q \succeq_{\text{c.v.}} v_Q$. Hence, they satisfy the hypotheses of Theorem 1 in Kleiner et al. (2021).[19]

Since $p_Q \succeq_{\text{c.v.}} v_Q$, if we define P to be the set of such monotone functions, Theorem 1 in Kleiner et al. (2021) implies that the extreme points of P consist of monotone functions for each of which there is a partition of $[0,1]$ into a collection of intervals $\mathcal{I} = \{I_\alpha\}_{\alpha \in K} \cup \{J_\beta\}_{\beta \in L}$ for which either $p_Q(i) = v_Q(i), \forall i \in I_\alpha$ or $p_Q(i) = \mathbb{E}[v_Q | i \in J_\beta, a], \forall i \in J_\beta$. Let ext$P$ be the set of such

[19] A possible issue can arise because our indicator and market values are unbounded – Kleiner et al. (2021)'s Proposition 1 only shows compactness of P for bounded functions. In Lemma 9 in the Online Appendix, we show that their result also works for integrable functions.

functions, each of which is identified by $\mathcal{I}_p$, and for each $i \in [0,1]$, let $\mathcal{I}_p(i)$ be the member of $\mathcal{I}_p$ that contains i. When i is fully revealed, i.e., $i \in I_\alpha$ for some $\alpha \in K$, we use the convention that $\mathcal{I}_p(i) = \{i\}$.

By Choquet's theorem — see for example Phelps (2001), page 14 — $p_Q(i)$ must be a convex combination of the extreme points identified by Kleiner et al. (2021). Formally, there exists a $\tau \in \Delta \text{ext}P$ such that

$$p_Q(i) = \int p(i)\, d\tau(p), \forall i$$

Let the set of signals be $S = \{(p, J) \,|p \in \text{ext}P, J = \mathcal{I}_p(i), i \in [0,1]\}$ and

$$\pi(C \times \{J\} \,|i) = \tau(\{p \in C, \mathcal{I}_p(i) = J\}), \forall C \subset \text{ext}P, \exists i' \in [0,1], J = \mathcal{I}_{\tilde{p}}(i'), \tilde{p} \in \text{ext}P.$$

In words, π is associated with a signal that reveals which partition in $\mathcal{I}_p$ is used with probability $d\tau(p)$ and then reveals i if the partition in $\mathcal{I}_p$ reveals i, and otherwise the interval J_β that i belongs to. Moreover, π is well defined since $p \mapsto \mathcal{I}_p(i)$ is a measurable mapping.

Fix a realization $s = (p, J) \in S$. If $J = \{i\}$ is a singleton, under the signal associated with p, the market learns the quantile i and its posterior mean is $v_Q(i) = \bar{v}(y; a)$ – p is drawn independently of y. If $J = J_\beta$ is a pooling interval, the market learns only that the quantile lies in J_β, and since the quantile is uniform, its posterior mean is $\frac{1}{|J_\beta|}\int_{J_\beta} v_Q(j)\, dj = p(i)$ for every $i \in J_\beta$. In both cases, the posterior mean equals $p(i)$, where p is the extreme point drawn. Therefore, for $i = G(y|a)$,

$$\int \mathbb{E}[v \mid s, a]\; d\pi(s \mid i) = \int_{\text{ext}P} p(i)\, d\tau(p) = p_Q(i),$$

so that p_Q is the interim price quantile associated with the signal structure π. This proves the claim. □

A.2 Proof of Theorem 1

To prove this theorem, we first need to prove the following lemma.

Lemma 1. *Fix an effort level $a \in A$ and its associated market value quantile function v_Q. Let $H : [0,1] \to \mathbb{R}$ be a continuous function such that $H(1) = H(0) = 0$. Let cav$H$ be the concave envelope of H and let $\mathcal{P}_a$ be the set of monotone quantile functions p_Q that satisfy $p_Q \succeq_{c.v.} v_Q$. Then:*

1. *Any $p_Q \in \mathcal{P}_a$ for which $\int H(i)\, dp_Q(i)$ is well defined satisfies:*

$$\int_0^1 H(i)\, dp_Q(i) \leq \int_0^1 \text{cav}H(i)\, dv_Q(i). \tag{8}$$

This holds even when $\int \text{cav}H\, dv_Q = \infty$.

2. *A $\hat{p}_Q \in \mathcal{P}_a$ can be constructed such that $\int_0^1 H(i)\, d\hat{p}_Q(i) = \int_0^1 \text{cav}H(i)\, dv_Q(i)$. The interim price $\hat{p}_Q(i)$ satisfies $\hat{p}_Q(i) = v_Q(i)$ when $H(i) = \text{cav}H(i)$. Moreover, if for some maximal interval $I \subset [0,1]$, $\text{cav}H(i) > H(i)$, then $\hat{p}_Q(i) = \mathbb{E}[v_Q | i \in I]$.*

Proof. Before stating the proof, let us state some preliminary results that we use throughout:

1. Since H is continuous and $H(1) = H(0) = 0$, we must have that $\text{cav}H(1) = \text{cav}H(0) = 0$ and $\text{cav}H \geq 0$. This is because $\text{cav}H$ can always be lowered at the limits while keeping concavity in an interval around it. The non-negativity of $\text{cav}H$ then follows from its concavity and its end values being 0. Moreover, $\text{cav}H$ is continuous at 0 and 1.

2. Integrability of p_Q implies that $|jp_Q(j)| \to 0$ as $j \to 0$ and $|(1-j)\, p_Q(j)| \to 0$ as $j \to 1$. This is because p_Q is integrable and monotone, and if $p_Q(0) = -\infty$, we can write for j small enough
$$|jp_Q(j)| \leq \int_0^j |p_Q(i)|\, di \to 0 \text{ as } j \to 0$$
which follows from the integrability of p_Q. A similar argument holds for the upper limit and v_Q.

3. Since $\text{cav}H$ is concave, it has a right derivative given by $(\text{cav}H(t))'$, and concavity of $\text{cav}H$ implies that $d(-\text{cav}H)' \geq 0$. We can then write the following
$$\begin{aligned} \text{cav}H(i) &= \int_0^1 \min\{i(1-t), t(1-i)\}\, d(-\text{cav}H(t))' \\ &= (1-i)\int_0^i t d(-\text{cav}H(t))' + i\int_i^1 (1-t)\, d(-\text{cav}H(t))' \end{aligned} \tag{9}$$
This is because we can use integration by parts and write
$$\begin{aligned} (1-i)\int_0^i t d(-\text{cav}H(t))' &= t(-\text{cav}H(t))'\big|_{t=0}^{i} \times (1-i) + (1-i)\int_0^i (\text{cav}H(t))'\, dt \\ &= t(-\text{cav}H(t))'\big|_{t=0}^{i} \times (1-i) + (1-i)\,\text{cav}H(i) \\ i\int_i^1 (1-t)\, d(-\text{cav}H(t))' &= (1-t)(-\text{cav}H(t))'\big|_{t=i}^{1} \times i + i\text{cav}H(i) \end{aligned}$$
Since $\text{cav}H$ is concave, $t\text{cav}H(t)' \leq \text{cav}H(t) - \text{cav}H(0)$. We also know that $\text{cav}H$ is non-negative and thus its derivative near 0 should be non-negative, and as we send t to 0, $t(-\text{cav}H(t))'$ vanishes. The same holds for $(1-t)\left(-\text{cav}H(t)'\right)$ as t tends to 1. Adding the above two expressions then gives (9) since the boundary terms at $t = 0, 1$ vanish and the terms for $t = i$ cancel.

We now use these to prove the main claim of the lemma. Note that since p_Q is monotone, i.e.,

$dp_Q \geq 0$, and that $H(i) \leq \text{cav}H(i)$ for all $i \in [0,1]$, we must have that

$$\int_0^1 H(i)\, dp_Q(i) \leq \int_0^1 \text{cav}H(i)\, dp_Q(i)$$

for $p_Q \in \mathcal{P}_a$ for which the LHS is well defined. Now we can use the equality (9) and write

$$\begin{aligned}
\int_0^1 \text{cav}H(i)\, dp_Q(i) &= \int_0^1 \int_0^1 \min\{i(1-t), t(1-i)\}\, d(-\text{cav}H(t))'\, dp_Q(i) \\
&= \int_0^1 \int_0^1 \min\{i(1-t), t(1-i)\}\, dp_Q(i)\, d(-\text{cav}H(t))' \\
&= -\int_0^1 \left[(1-t)\int_0^t p_Q(i)\, di - t\int_t^1 p_Q(i)\, di\right] d(-\text{cav}H(t))' \qquad (10)
\end{aligned}$$

where the second step uses switching of the integrals (by Tonelli's theorem, this is valid as $dp_Q, d(-\text{cav}H)', \min\{\cdots\}$ are all non-negative) and the last step uses integration by parts for the inside integral and the fact that $p_Q(i)\, i$ and $p_Q(i)(1-i)$, respectively, vanish at 0 and 1. Since $p_Q \succeq_{\text{c.v.}} v_Q$ and the concave order is equivalent to majorization, we must have that

$$(1-t)\int_0^t p_Q(i)\, di - t\int_t^1 p_Q(i)\, di \geq (1-t)\int_0^t v_Q(i)\, di - t\int_t^1 v_Q(i)\, di$$

This is because p_Q and v_Q have the same integral over $[0,1]$, and by majorization, $\int_0^t p_Q di \geq \int_0^t v_Q di$. Using the above inequality, we can write the earlier expression and have

$$\begin{aligned}
\int_0^1 \text{cav}H(i)\, dp_Q(i) &\leq -\int_0^1 \left[(1-t)\int_0^t v_Q(i)\, di - t\int_t^1 v_Q(i)\, di\right] d(-\text{cav}H(t))' \\
&= \int_0^1 \text{cav}H(i)\, dv_Q(i)
\end{aligned}$$

where in the above we have used the fact that $d(-\text{cav}H(t))' \geq 0$. This proves the first claim. Note that in the above we have only used Tonelli's theorem, which imposes no restriction on the finiteness of the integrals.

To prove the second part, consider the set $\{i : \text{cav}H(i) > H(i)\}$. Continuity of H and cav$H$ implies that this set is open and thus a union of open intervals $\{J_\beta\}_{\beta \in K}$ such that cav$H(i)$ is affine over each $J_\beta = (i_{1,\beta}, i_{2,\beta})$. We thus have

$$\begin{aligned}
\int_{i_{1,\beta}}^{i_{2,\beta}} \text{cav}H(i)\, dv_Q(i) &= \int_{i_{1,\beta}}^{i_{2,\beta}} \left[\frac{i_{2,\beta} - i}{i_{2,\beta} - i_{1,\beta}} H(i_{1,\beta}) + \frac{i - i_{1,\beta}}{i_{2,\beta} - i_{1,\beta}} H(i_{2,\beta})\right] dv_Q(i) \\
&= H(i)\, v_Q(i)\big|_{i_{1,\beta}}^{i_{2,\beta}} - \int_{i_{1,\beta}}^{i_{2,\beta}} v_Q(i)\, \frac{H(i_{2,\beta}) - H(i_{1,\beta})}{i_{2,\beta} - i_{1,\beta}} di
\end{aligned}$$

using integration by parts. Now consider $\hat{p}_Q(i)$ given by

$$\hat{p}_Q(i) = \begin{cases} v_Q(i_{1,\beta}) & i = i_{1,\beta} \\ v_Q(i_{2,\beta}) & i = i_{2,\beta} \\ \frac{\int_{i_{1,\beta}}^{i_{2,\beta}} v_Q(j)dj}{i_{2,\beta}-i_{1,\beta}} & i \in J_\beta \end{cases}.$$

We have

$$\int_{i_{1,\beta}}^{i_{2,\beta}} H(i)\, d\hat{p}_Q(i) = H(i_{1,\beta}) \left[\frac{\int_{i_{1,\beta}}^{i_{2,\beta}} v_Q(j)\, dj}{i_{2,\beta} - i_{1,\beta}} - v_Q(i_{1,\beta}) \right] + H(i_{2,\beta}) \left[v_Q(i_{2,\beta}) - \frac{\int_{i_{1,\beta}}^{i_{2,\beta}} v_Q(j)\, dj}{i_{2,\beta} - i_{1,\beta}} \right]$$

which is the same as $\int_{J_\beta} \text{cav}H dv_Q$. Outside of the set $\cup_\beta J_\beta$, we have $\text{cav}H = H$. Hence, if we construct $\hat{p}_Q$ as being equal to the average values of v_Q over J_β and otherwise equal to $v_Q(i)$, $\hat{p}_Q$ is obviously a member of $\mathcal{P}_a$, and $\int H d\hat{p}_Q$ is the same as the RHS in (8). This proves the second part. That the equality holds even when the integrals are infinite follows from the fact that all integrals over J_β's and their complements are non-negative. □

A.2.1 Proof of the Theorem

Proof. Fix a^* to be the implementable effort that maximizes $W(a)$. Let $\mathcal{P}_a$ be the set of quantile interim prices p_Q that are monotone and satisfy $p_Q \succeq_{\text{c.v.}} v_Q$ – with v_Q being the market value quantile associated with a. We will think of $\mathcal{P}_a$ as a subset of $L_2([0,1])$ and use the topology induced by $\|\cdot\|_2$. Assumption 1 together with $p_Q \succeq_{\text{c.v.}} v_Q$ guarantees that each p_Q has a finite L_2 norm. Let $T_a : \mathcal{P}_a \to \mathbb{R}^N$ be the affine operator associated with incentive compatibility:

$$T_{a,n}p = -\int_0^1 \frac{\partial F}{\partial \hat{a}_n}(i|a;a)\, dp_Q(i) - \frac{\partial c(a)}{\partial a_n}$$

Our proof consists of four steps.

Step 1: *T_a is a well-defined affine operator over $\mathcal{P}_a$.* We first show that the above is finite and thus well-defined. From the definition of $F(\cdot|\cdot;\cdot)$, we must have

$$\frac{\partial}{\partial \hat{a}_n} F(i|a;a) = \frac{\partial G}{\partial \hat{a}_n}\left(G^{-1}(i|a)\,|\hat{a}\right)\Big|_{\hat{a}=a},$$

and let us refer to $\frac{\partial \log g\left(G^{-1}(i|a)|\hat{a}\right)}{\partial \hat{a}_n}\Big|_{\hat{a}=a}$ as $s_n(i;a)$. Using integration by parts, we can write

$$T_{a,n}p = -\frac{\partial c}{\partial a_n}(a) + \int_0^1 p_Q(i)\, s_n(i;a)\, di - \frac{\partial G}{\partial \hat{a}_n}\left(G^{-1}(i|a)\,|\hat{a}\right)\Big|_{\hat{a}=a} p_Q(i)\Big|_{i=0}^1 \tag{11}$$

The integral satisfies

$$\left| \int_0^1 p_Q(i)\, s_n(i;a)\, di \right|^2 \leq \int_0^1 p_Q(i)^2\, di \int_0^1 s_n(i;a)^2\, di$$

by an application of the Cauchy–Schwarz inequality. Since the quadratic function is convex and $p_Q \succeq_{\text{c.v.}} v_Q$, we must have that

$$\int_0^1 p_Q(i)^2 di \leq \int_0^1 v_Q(i;a)^2 di < \infty$$

Both integrals $\int v_Q^2 di$, $\int s_n^2 di$ are finite by Assumption 1. Hence, for the existence of T_a, it suffices to show that the boundary term in (11) is finite. To see that, note that if $p_Q(i)$ is bounded below and above, the fact that at quantiles $i = 0, 1$ the c.d.f. remains unchanged in response to a change in effort implies that the boundary term in (11) is 0. Suppose that $p_Q(i) \to -\infty$ as $i \to 0$. Then for i small enough, $p_Q(i)$ is negative, and we have

$$\left| \frac{\partial G}{\partial \hat{a}_n} \left(G^{-1}(i|a) | \hat{a} \right) \Big|_{\hat{a}=a} \right| = \left| \int_0^i s_n(j;a) dj \right| \leq \sqrt{i \int_0^1 s_n(j;a)^2 dj}$$

by applying the Cauchy–Schwarz inequality. We also have that, since $p_Q(i)$ is negative and monotone,

$$i p_Q(i)^2 \leq \int_0^i p_Q(j)^2 dj$$

Since p_Q^2 is integrable, the RHS of the above converges to 0 as $i \to 0$. Moreover,

$$\left| \frac{\partial G}{\partial \hat{a}_n} \left(G^{-1}(i|a) | \hat{a} \right) \Big|_{\hat{a}=a} \right| |p_Q(i)| \leq \sqrt{i \int_0^1 s_n(j;a)^2 dj} \, |p_Q(i)|$$

Since we have shown that $i p_Q(i)^2$ converges to 0 as $i \to 0$, so does the above. Using a similar argument for $i \to 1$, we can establish that the boundary term in (11) is indeed 0. This proves the claim that $T_a p$ is well-defined and finite.

Step 2: $T_a \mathcal{P}_a$ is convex. Since T_a is affine and $\mathcal{P}_a$ is convex, its image $T_a \mathcal{P}_a$ in $\mathbb{R}^N$ is also a convex set.

Step 3: 0 is on the boundary of $T_{a^*} \mathcal{P}_{a^*}$ i.e., $0 \in \partial T_{a^*} \mathcal{P}_{a^*}$. To see this, suppose to the contrary that 0 is in the interior of $T_{a^*} \mathcal{P}_{a^*}$. We show that every a in a neighborhood of a^* becomes incentive compatible. This, together with the fact that a^* is not a local maximizer of $W(a)$ in A, delivers a contradiction.

To show this, note that if $0 \in \text{int} T_{a^*} \mathcal{P}_{a^*}$, then there exists an open ball in the image $T_{a^*} \mathcal{P}_{a^*}$ around 0.

In the Online Appendix E, Lemma 8, we show that the set $\{a : 0 \in \text{int}(T_a \mathcal{P}_a)\}$ is open. Since this set contains a^*, it should include an open neighborhood around a^*. Since a^* is not a local maximizer of $W(a)$ in A, an open neighborhood of a^* should contain an effort level $\hat{a} \in A$ such that $W(\hat{a}) > W(a^*)$. Since $0 \in T_{\hat{a}} \mathcal{P}_{\hat{a}}$, $\hat{a}$ is implementable by an interim price vector, which is in contradiction with the optimality of a^*. Thus, it has to be that $0 \in \partial T_{a^*} \mathcal{P}_{a^*}$.

Step 4: (D) has to hold. Since 0 is on the boundary of the convex set $T_{a^*} \mathcal{P}_{a^*}$, there exists $\lambda \in$

$\mathbb{R}^N \setminus \{0\}$ such that for all $x \in T_{a^*}\mathcal{P}_{a^*}$, $\lambda \cdot x \leq 0$. Hence, if $\hat{p}_Q \in \mathcal{P}_{a^*}$ is such that $T_{a^*}\hat{p}_Q = 0$, we have

$$-\sum_{n=1}^{N} \lambda_n \left[\int_0^1 \frac{\partial F}{\partial \hat{a}_n} (i|a^*; a^*) \, dp_Q(i) + \frac{\partial c(a^*)}{\partial a_n} \right] \leq 0 = -\sum_{n=1}^{N} \lambda_n \left[\int_0^1 \frac{\partial F}{\partial \hat{a}_n} (i|a^*; a^*) \, d\hat{p}_Q(i) + \frac{\partial c(a^*)}{\partial a_n} \right], \forall p_Q \in$$

Thus, if we define $\Gamma(i; \lambda, a^*) = -\sum_n \lambda_n \frac{\partial F}{\partial \hat{a}_n}(i|a^*; a^*)$, the above inequality implies that

$$\int_0^1 \Gamma(i; \lambda, a^*) \, dp_Q(i) \leq \int_0^1 \Gamma(i; \lambda, a^*) \, d\hat{p}_Q(i) = \lambda \cdot \nabla_a c(a^*) < \infty.$$

Equation (D) then follows from the above and Lemma 1. □

A.3 Proof of Proposition 2

Proof. We prove Parts 1 and 2. Part 3 is an evident corollary of Part 2.

Proof of Part 1: Let $\mathcal{Q} = \{p_Q \in \mathcal{P}_{a^*} : T_{a^*} p_Q = 0\}$; it is non-empty since a^* is implementable. The set $\mathcal{P}_{a^*}$ is convex and weakly compact in $L_2([0,1])$, and by Step 1 of the proof of Theorem 1, T_{a^*} is affine and weakly continuous. Hence, $\mathcal{Q}$ is convex and weakly compact and, by the Krein–Milman theorem, has an extreme point. Since $\mathcal{Q}$ is the intersection of $\mathcal{P}_{a^*}$ with N hyperplanes, by a result of Dubins (1962), this extreme point is a convex combination of at most $N+1$ extreme points of $\mathcal{P}_{a^*}$, which by Kleiner et al. (2021) are monotone partitions.

Proof of Part 2: We show the claim by proving that, under the hypothesis, a unique interim price achieves the value in (D). From the proof of Lemma 1, we know that

$$\forall p_Q \in \mathcal{P}_{a^*}, \int_0^1 \Gamma dp_Q \leq \int_0^1 \text{cav}\Gamma dp_Q \leq \int_0^1 \text{cav}\Gamma dv_Q$$

Suppose $\hat{p}_Q$ achieves the value in (D), i.e., both inequalities above hold with equality. For the first inequality to hold, it has to be that in any open interval with $\Gamma < \text{cav}\Gamma$, $d\hat{p}_Q = 0$.

In the case of the second inequality, we can use (10) to write

$$\int_0^1 \text{cav}\Gamma d\hat{p}_Q = -\int_0^1 \left[(1-t) \int_0^t \hat{p}_Q(i) \, di - t \int_t^1 \hat{p}_Q(i) \, di \right] d\left(-\text{cav}\Gamma(t)\right)'$$

Writing the same equality for v_Q, the second inequality holds with equality only if

$$\int_0^t \hat{p}_Q(i) \, di = \int_0^t v_Q(i) \, di$$

for all t in the support of the measure $d(-\text{cav}\Gamma)'$. By assumption, every $i \in (0,1)$ with $\Gamma(i) = \text{cav}\Gamma(i)$ is in this support. Hence, $\hat{p}_Q$ is constant on each maximal interval with $\Gamma < \text{cav}\Gamma$ and, since the above equality holds at its endpoints, equal to the average of v_Q over it; elsewhere $\hat{p}_Q = v_Q$ almost everywhere. This pins down $\hat{p}_Q$ uniquely. □

A.4 Proof of Proposition 3

Proof. From Theorem 1, we know that $\lambda \neq 0$. The function $\Gamma(i;\lambda,a)$ satisfies

$$\Gamma'(i;\lambda,a) = -\lambda \frac{\partial g\left(G^{-1}(i|a)|a\right)}{\partial a} \frac{1}{g\left(G^{-1}(i|a)|a\right)}$$

Since g exhibits MLRP, if $\lambda > 0$, then Γ' is decreasing in i and so Γ is concave. If $\lambda < 0$, then Γ' is increasing in i and so Γ is convex. Clearly, λ cannot be negative since a convex Γ leads to a fully uninformative rating. □

A.5 Proof of Proposition 4

Proof. From Theorem 1, we know that $\lambda \neq 0$ exists such that $\Gamma(i;\lambda,a) = -\lambda \left.\frac{\partial F(i|\hat{a};a)}{\partial \hat{a}}\right|_{\hat{a}=a}$. As we have shown in Section 3,

$$\Gamma''(i;\lambda,a) = -\lambda \left.\frac{1}{g(y|a)} \frac{\partial^2 \log g(y|a)}{\partial a \partial y}\right|_{y=G^{-1}(i|a)}$$

Given our definition of ELRP, when λ is negative, Γ is concave–convex, and when λ is positive, Γ is convex–concave. We wish to show that under ELRP, λ is positive, and thus the optimal rating has to be lower censorship.

Suppose to the contrary that λ is negative. In this case, since $\Gamma(0;\lambda,a) = \Gamma(1;\lambda,a) = 0$, there are two possibilities: 1. $\Gamma'(0;\lambda,a) < 0$, in which case $\Gamma(i;\lambda,a)$ is non-positive for all values of i and its concave envelope is the zero function associated with no information; 2. $\Gamma'(0;\lambda,a) > 0$, in which case for an interval $[0,i_1]$ the concave envelope coincides with Γ and for higher values cavΓ is linear. This is associated with an upper-censorship optimal rating. In the first case, the marginal return to effort is zero, and effort with positive marginal cost cannot be supported.

In the second case, let y_1 be the value of the indicator associated with i_1. In this case, the marginal benefit of effort is given by

$$\int p(y) \frac{\partial g(y|a)}{\partial a} dy =$$
$$\int_{\underline{y}}^{y_1} \overline{v}(y;a) \frac{\partial \log g(y|a)}{\partial a} dG + \overline{p} \int_{y_1}^{\overline{y}} \frac{\partial g(y|a)}{\partial a} dy$$

where in the above $\overline{p}$ is the average value of $\overline{v}$ when $y \geq y_1$. Since Γ is concave over $[0,i_1]$, $\frac{\partial \log g(y|a)}{\partial a}$ is decreasing in y over the interval $\left[\underline{y}, y_1\right]$, and thus, using the fact that $\overline{v}$ is increasing in y, we can use Chebyshev's sum inequality to write the above as

$$\int_{\underline{y}}^{y_1} \overline{v}(y;a) \frac{\partial \log g(y|a)}{\partial a} dG + \overline{p} \int_{y_1}^{\overline{y}} \frac{\partial g(y|a)}{\partial a} dy \leq$$
$$\frac{\int_{\underline{y}}^{y_1} \overline{v}(y;a)\, dG}{G(y_1|a)} \int_{\underline{y}}^{y_1} \frac{\partial \log g(y|a)}{\partial a} dG + \overline{p} \int_{y_1}^{\overline{y}} \frac{\partial g(y|a)}{\partial a} dy =$$

$$\frac{\int_{\underline{y}}^{y_1} \overline{v}(y;a)\, dG}{G(y_1|a)} \frac{\partial G(y_1|a)}{\partial a} - \overline{p} \frac{\partial G(y_1|a)}{\partial a}$$

where in the above we have used the fact that $\frac{\partial G(\overline{y}|a)}{\partial a} = 0$. Since $\Gamma(i_1) = -\lambda \frac{\partial G(y_1|a)}{\partial a} > 0$, the above expression satisfies

$$\left(\mathbb{E}\left[\overline{v}|y \leq y_1\right] - \mathbb{E}\left[\overline{v}|y \geq y_1\right]\right) \frac{\partial G(y_1|a)}{\partial a} \leq 0$$

which cannot be the case since the cost of effort is increasing. Hence, $\lambda > 0$, and thus the optimal rating is lower censorship. When g exhibits CLRP, the argument is the mirror of the current argument. □

A.6 Proof of Theorem 2

Proof. From the text, we know that in this case the second derivative of the gain function satisfies

$$g(y|a)\, \Gamma''(i|a)|_{i=G(y|a)} = -\lambda_1 \frac{\partial \ell_1(y|a)}{\partial y} - \lambda_2 \frac{\partial \ell_2(y|a)}{\partial y} = -\boldsymbol{\lambda} \cdot z(y|a),$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \neq 0$ by Theorem 1. In order to prove the result, we show that, under MRIP, the above changes sign at most once. This means that Γ is either concave–convex or convex–concave, and therefore, the optimal rating is either lower or upper censorship. Note that we can assume without loss of generality that $D > 0$, as when $D < 0$, the reverse argument would work. Since $D \neq 0$, $z(y|a) \neq 0$ for all y, and thus its angle $\theta(z(y|a))$ is well-defined; we choose it to be continuous in y. Since $D > 0$, we have $\sin(\theta(z(y_2|a)) - \theta(z(y_1|a))) > 0$ for all $y_1 < y_2$, and by continuity, $\theta(z(y_2|a)) - \theta(z(y_1|a)) \in (0, \pi)$. Hence, $\theta(z(y|a))$ is increasing in y and never takes two values that are π apart. Letting $\theta(\boldsymbol{\lambda})$ be the angle of $\boldsymbol{\lambda}$, we have

$$\boldsymbol{\lambda} \cdot z(y|a) = \|\boldsymbol{\lambda}\|_2 \|z(y|a)\|_2 \cos(\theta(z(y|a)) - \theta(\boldsymbol{\lambda}))$$

which changes sign only when $\theta(z(y|a))$ crosses $\theta(\boldsymbol{\lambda}) \pm \frac{\pi}{2}$. Since these two values are π apart, $\theta(z(y|a))$ crosses at most one of them, which establishes the claim. Finally, since $\Gamma'' \neq 0$ away from this crossing, Γ is strictly concave on its concave part, and by Proposition 2, the optimal rating is unique. This is depicted in Figure 8. □

A.7 Proof of Proposition 6

Proof. It is immediate by using an argument similar to that of Theorem 1 that optimal ratings can be found by concavification of the function

$$\Gamma(i) = -\alpha_P f_P G\left(\overline{G}^{-1}(i)|a_P\right) - \alpha_R f_R G\left(\overline{G}^{-1}(i)|a_R\right) - \lambda_P G_a\left(\overline{G}^{-1}(i)|a_P\right) - \lambda_R G_a\left(\overline{G}^{-1}(i)|a_R\right)$$

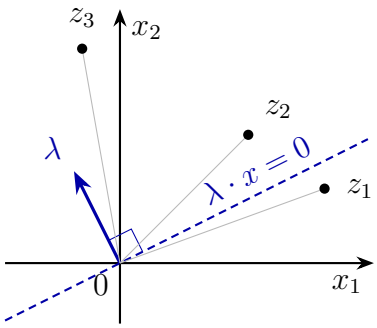


Figure 8: The vectors of likelihood changes under MRIP

for some Lagrange multipliers λ_P, λ_R. We have

$$\Gamma'(i) = -\alpha_P f_P \frac{g\left(\overline{G}^{-1}(i)\,|a_P\right)}{\overline{g}\left(\overline{G}^{-1}(i)\right)} - \alpha_R f_R \frac{g\left(\overline{G}^{-1}(i)\,|a_R\right)}{\overline{g}\left(\overline{G}^{-1}(i)\right)} - \lambda_P \frac{g_a\left(\overline{G}^{-1}(i)\,|a_P\right)}{\overline{g}\left(\overline{G}^{-1}(i)\right)} - \lambda_R \frac{g_a\left(\overline{G}^{-1}(i)\,|a_R\right)}{\overline{g}\left(\overline{G}^{-1}(i)\right)}$$

where in the above $\overline{g}(y) = f_P g(y|a_P) + f_R g(y|a_R)$. Given the functional form of g, we have

$$\begin{aligned} g(y|a) &= e^{\zeta(y)+r(y)m(a)-b(a)} \\ \frac{g(y|a_P)}{g(y|a_R)} &= e^{r(y)(m(a_P)-m(a_R))+b(a_R)-b(a_P)} \\ \frac{g_a(y|a)}{g(y|a)} &= m'(a)\,r(y) - b'(a) \end{aligned}$$

Substituting into the formula for Γ' implies

$$\Gamma'\left(\overline{G}(y)\right) = -\alpha_P f_P \frac{1}{f_P + f_R \frac{g(y|a_R)}{g(y|a_P)}} - \alpha_R f_R \frac{\frac{g(y|a_R)}{g(y|a_P)}}{f_P + f_R \frac{g(y|a_R)}{g(y|a_P)}} - \lambda_P \frac{\frac{g_a(y|a_P)}{g(y|a_P)}}{f_P + f_R \frac{g(y|a_R)}{g(y|a_P)}} - \lambda_R \frac{\frac{g(y|a_R)}{g(y|a_P)} \frac{g_a(y|a_R)}{g(y|a_R)}}{f_P + f_R \frac{g(y|a_R)}{g(y|a_P)}}$$

If we refer to $m(a_R) - m(a_P)$ as Δm and similarly for $b(a_R) - b(a_P)$, we can write the above as

$$\Gamma'\left(\overline{G}(y)\right) = -\frac{\alpha_P f_P + \alpha_R f_R e^{r\Delta m - \Delta b} + \lambda_P\left(m'_P r - b'_P\right) + \lambda_R\left(m'_R r - b'_R\right) e^{r\Delta m - \Delta b}}{f_P + f_R e^{r\Delta m - \Delta b}}$$

If we define $x = e^{r\Delta m}$, then the above has the form

$$-\frac{A_1 + A_2 x + A_3 \log x + A_4 x \log x}{B_1 + B_2 x}$$

We then have

$$(B_1 + B_2 x)^2 \frac{\Gamma''\left(\overline{G}(y)\right)\overline{g}(y)}{\frac{dx}{dy}} = (B_2 A_3 - B_1 A_4)\log x - B_1 A_3/x - B_2 A_4 x + B_2(A_1 - A_3) - B_1(A_2 + A_4)$$

If $a_R = a_P$, Γ'' has a constant sign, and the claim is immediate. Otherwise, $\frac{dx}{dy} = \Delta m\, r'(y)\, x$ has a constant sign, and it suffices to show that the above changes sign at most two times in x. Its

derivative with respect to x is given by

$$\frac{-B_2A_4x^2 + (B_2A_3 - B_1A_4)\,x + B_1A_3}{x^2} = \frac{(B_1 + B_2x)\,(A_3 - A_4x)}{x^2}.$$

Note that in the above $B_1 = f_P, B_2 = f_R e^{-\Delta b}$ are both positive as well as x. This implies that the above changes sign at most once for all positive values of x. In other words, $(B_1 + B_2x)^2 \frac{\Gamma''\left(\overline{G}(y)\right)\overline{g}(y)}{\frac{dx}{dy}}$ is either increasing–decreasing or decreasing–increasing. This means that Γ'' switches sign at most twice. Hence, Γ consists of at most three alternating convex and concave pieces, and by Proposition 2, the optimal rating reveals where $\Gamma = \text{cav}\Gamma$ and pools on the maximal intervals where $\Gamma < \text{cav}\Gamma$. This establishes the claim. □

References

Acemoglu, D., L. Fergusson, J. Robinson, D. Romero, and J. F. Vargas (2020): "The Perils of High-Powered Incentives: Evidence from Colombia's False Positives," *American Economic Journal: Economic Policy*, 12, 1–43. 21

Albano, G. L. and A. Lizzeri (2001): "Strategic Certification and Provision of Quality," *International economic review*, 42, 267–283. 7

Alexander, D. (2020): "How Do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs," *Journal of Political Economy*, 128, 4046–4096. 21

Ali, S. N., N. Haghpanah, X. Lin, and R. Siegel (2022): "How to Sell Hard Information," *The Quarterly Journal of Economics*, 137, 619–678. 7

Andrabi, T. and C. Brown (2022): "Subjective Versus Objective Incentives and Teacher Productivity," *Berkeley Working Paper*. 21

Azevedo, E. M. and I. Wolff (2025): "Broad Validity of the First-Order Approach in Moral Hazard," *arXiv preprint arXiv:2506.18873*. 14

Baker, G. P. (1992): "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 100, 598–614. 21

Ball, I. (2025): "Scoring Strategic Agents," *American Economic Journal: Microeconomics*, 17, 97–129. 5, 6

Barahona, N., C. Otero, and S. Otero (2023): "Equilibrium Effects of Food Labeling Policies," *Econometrica*, 91, 839–868. 7

Berg, F., J. F. Koelbel, and R. Rigobon (2022): "Aggregate Confusion: The Divergence of ESG Ratings," *Review of Finance*, 26, 1315–1344. 7

Bergemann, D., T. Heumann, and S. Morris (2026): "Screening with Persuasion," *Journal of Political Economy*, 134. 6

Bergemann, D., T. Heumann, S. Morris, C. Sorokin, and E. Winter (2022): "Optimal Information Disclosure in Classic Auctions," *American Economic Review: Insights*, 4, 371–388. 6

Billingsley, P. (1995): *Probability and Measure*, New York: Wiley, 3 ed. 63

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24, 265–272. 59

Boleslavsky, R. and K. Kim (2021): "Bayesian Persuasion and Moral Hazard," *Working Paper, Emory University*. 6

Chade, H. and J. Swinkels (2020): "The No-Upward-Crossing Condition, Comparative Statics, and the Moral-Hazard Problem," *Theoretical Economics*, 15, 445–476. 14, 50

Conlon, J. R. (2009): "Two new conditions supporting the first-order approach to multisignal principal–agent problems," *Econometrica*, 77, 249–278. 14

Dartmouth College (2024): "Reactivating Dartmouth's Testing Policy," Dartmouth Admissions, February 5, https://admissions.dartmouth.edu/node/45786. 29

De Janvry, A., G. He, E. Sadoulet, S. Wang, and Q. Zhang (2023): "Subjective Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China," *American Economic Review*, 113, 766–799. 21

Dessein, W., A. Frankel, and N. Kartik (2025): "Test-Optional Admissions," *American Economic Review*, 115, 3130–3170. 29

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The Economics of Career Concerns, part II: Application to Missions and Accountability of Government Agencies," *The Review of Economic Studies*, 66, 199–217. 3, 4, 21, 22

Doval, L. and A. Smolin (2024): "Persuasion and Welfare," *Journal of Political Economy*, 132, 2451–2487. 3

Dubins, L. E. (1962): "On Extreme Points of Convex Sets," *Journal of Mathematical Analysis and Applications*, 5, 237–244. 38

Dumont, E., B. Fortin, N. Jacquemet, and B. Shearer (2008): "Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment," *Journal of Health Economics*, 27, 1436–1450. 21

Frankel, A. and N. Kartik (2019): "Muddled Information," *Journal of Political Economy*, 127, 1739–1776. 5, 6

Friedman, J. N., B. Sacerdote, D. O. Staiger, and M. Tine (2025): "Standardized Test Scores and Academic Performance at Ivy-Plus Colleges," *AEA Papers and Proceedings*, 115, 676–681. 29

Gentzkow, M. and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review*, 106, 597–601. 12

Halac, M., E. Lipnowski, and D. Rappoport (2021): "Rank Uncertainty in Organizations," *American Economic Review*, 111, 757–786. 7

He, S., B. Hollenbeck, and D. Proserpio (2022): "The Market for Fake Reviews," *Marketing Science*, 41, 896–921. 10

Hoekstra, H. (2024): "Return to Required Testing," Letter to the Harvard Faculty of Arts and Sciences, April 11, https://www.fas.harvard.edu/2024/04/11/return-to-required-testing. 29

Holmström, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *The Review of Economic Studies*, 66, 169–182. 2, 15, 21

Holmström, 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. 2, 4, 21, 26, 27

Hörner, J. and N. S. Lambert (2021): "Motivational Ratings," *Review of Economic Studies*, 88, 1892–1935. 6

Hui, X., G. Z. Jin, and M. Liu (2025): "Designing Quality Certificates: Insights from eBay," *Journal of Marketing Research*, 62, 40–60. 22

Hui, X., M. Saeedi, Z. Shen, and N. Sundaresan (2016): "Reputation and Regulations: Evidence from eBay," *Management Science*, 62, 3604–3616. 10

Hui, X., M. Saeedi, G. Spagnolo, and S. Tadelis (2023): "Raising the Bar: Certification Thresholds and Market Outcomes," *American Economic Journal: Microeconomics*, 15, 599–626. 7

Innes, R. D. (1990): "Limited liability and incentive contracting with ex-ante action choices," *Journal of economic theory*, 52, 45–67. 13

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56, 1177–1190. 14, 50

Jung, J. Y. and S. K. Kim (2015): "Information space conditions for the first-order approach in agency problems," *Journal of Economic Theory*, 160, 243–279. 14

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89, 1557–1593. 6, 12, 32, 33, 38, 62, 63

Kolotilin, A., H. Li, and A. Zapechelnyuk (2024): "On Monotone Persuasion," Working paper. 7

Madsen, E., B. Williams, and A. Skrzypacz (2025): "Performance Incentives with Multiple Instruments," *Available at SSRN 5262226*. 6

Mayzlin, D., Y. Dover, and J. Chevalier (2014): "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, 104, 2421–2455. 5, 22

Mensch, J. (2021): "Monotone Persuasion," *Games and Economic Behavior*, 130, 521–542. 7

Mirrlees, J. A. (1999): "The Theory of Moral Hazard and Unobservable Behaviour: Part I," *The Review of Economic Studies*, 66, 3–21. 14

Nosko, C. and S. Tadelis (2015): "The Limits of Reputation in Platform Markets: An Empirical

Analysis and Field Experiment," Tech. rep., National Bureau of Economic Research. 7

Onuchic, P. and D. Ray (2023): "Conveying Value via Categories," *Theoretical Economics*, 18, 1407–1439. 7

Perez-Richet, E. and V. Skreta (2022): "Test Design under Falsification," *Econometrica*, 90, 1109–1142. 5, 6

Phelps, R. R. (2001): *Lectures on Choquet's Theorem*, vol. 1757 of *Lecture Notes in Mathematics*, Springer, 2 ed. 33

Rayo, L. (2013): "Monopolistic Signal Provision," *The B.E. Journal of Theoretical Economics*, 13, 27–58. 6

Rodina, D. and J. Farragut (2020): "Inducing Effort Through Grades," Working paper. 7

Rogerson, W. P. (1985): "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 1357–1367. 14

Rothschild, M. and J. E. Stiglitz (1970): "Increasing Risk: I. A Definition," *Journal of Economic theory*, 2, 225–243. 12, 59

Royden, H. L. and P. M. Fitzpatrick (2010): *Real Analysis*, Boston: Prentice Hall, 4 ed. 63

Shaked, M. and J. G. Shanthikumar (2007): *Stochastic Orders*, Springer Science & Business Media. 12, 18

Strassen, V. (1965): "The Existence of Probability Measures with Given Marginals," *The Annals of Mathematical Statistics*, 36, 423–439. 59

Vatter, B. (2025): "Quality Disclosure and Regulation: Scoring Design in Medicare Advantage," *Econometrica*, 93, 959–1001. 7

Wren, K. (2022): "Q&A: Stuart Schmill on MIT's Decision to Reinstate the SAT/ACT Requirement," MIT News, March 28, https://news.mit.edu/2022/stuart-schmill-sat-act-requirement-0328. 29

Xiao, P. (2025): "Incentivizing Agents through Ratings," Working paper, Boston University. 7

Zapechelnyuk, A. (2020): "Optimal Quality Certification," *American Economic Review: Insights*, 2, 161–76. 7

Zervas, G., D. Proserpio, and J. W. Byers (2021): "A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average," *Marketing Letters*, 32, 1–16. 5

Zubrickas, R. (2015): "Optimal Grading," *International Economic Review*, 56, 751–776. 7

Online Appendix

B Multi-Tasking with Location–Scale Technologies

In this section, we describe optimal ratings when $G(y|a) = H\left(\frac{y-\mu(a)}{\sigma(a)}\right)$, i.e., a controls the location and scale of the indicator y, and provide a proof of Proposition 5. Throughout, we write $\mu_n = \partial\mu(a)/\partial a_n$, $\sigma_n = \partial\sigma(a)/\partial a_n$, $c_n = \partial c(a)/\partial a_n$, $\varepsilon = (y-\mu(a))/\sigma(a)$ and $\Psi(a) = \mu_1\sigma_2 - \mu_2\sigma_1$, all evaluated at the implemented action.

Before providing the analysis, let us describe what MRIP implies within the location–scale class. Consider the family $G(y|a) = H\left(\frac{y-\mu(a)}{\sigma(a)}\right) = H(\varepsilon)$ such that a controls the mean and variance of y, with H the c.d.f. of a distribution with a log-concave density $H'(\varepsilon) = h(\varepsilon) = e^{\eta(\varepsilon)}$. In this case,

$$D(y_1, y_2|a) = \eta''(\varepsilon_1)\eta''(\varepsilon_2)\Psi(a)\left[\varepsilon_2 + \frac{\eta'(\varepsilon_2)}{\eta''(\varepsilon_2)} - \varepsilon_1 - \frac{\eta'(\varepsilon_1)}{\eta''(\varepsilon_1)}\right] \tag{12}$$

where $\varepsilon_n = \frac{y_n-\mu(a)}{\sigma(a)}$. Since h is log-concave, i.e., $\eta'' \leq 0$, the leading product is non-negative, and MRIP for this class is equivalent to monotonicity of $\varepsilon + \frac{\eta'(\varepsilon)}{\eta''(\varepsilon)}$. Monotonicity of $\varepsilon + \eta'(\varepsilon)/\eta''(\varepsilon)$ holds for the Normal (where it equals 2ε), Logistic, Gumbel, Generalized Normal, Gamma (with shape parameter ≥ 1), and Weibull (with shape parameter ≥ 1) families, so all of these exhibit MRIP with a log-concave density.

Fix the implemented action a and let $\pi(\varepsilon) = p(\mu(a) + \sigma(a)\varepsilon)$ be the interim price expressed in units of the standardized indicator. The payoff of the agent from effort a' is $\int p(y)\,dG(y|a') - c(a')$, and differentiating with respect to a'_n at $a' = a$ yields

$$B_n = \mu_n P_1 + \sigma_n P_2, \qquad P_1 = -\frac{1}{\sigma(a)}\int \pi(\varepsilon)\,h'(\varepsilon)\,d\varepsilon, \quad P_2 = -\frac{1}{\sigma(a)}\int \pi(\varepsilon)\,(\varepsilon h(\varepsilon))'\,d\varepsilon$$

Integrating by parts, $P_1 = \frac{1}{\sigma(a)}\int h(\varepsilon)\,d\pi(\varepsilon)$ and $P_2 = \frac{1}{\sigma(a)}\int \varepsilon h(\varepsilon)\,d\pi(\varepsilon)$, where $d\pi$ is the non-negative measure generated by the non-decreasing schedule π. Thus, the rating affects the incentives of the agent only through two statistics: P_1, the marginal reward that the rating attaches to the level of the indicator, and P_2, the marginal reward that it attaches to its dispersion. Moreover, $P_1 > 0$ unless the rating is uninformative.

Lemma 2. *Suppose that the welfare-maximizing rating implements an interior action a and that $\Psi(a) \neq 0$. Then*

$$P_1 = \frac{\sigma_2 c_1 - \sigma_1 c_2}{\Psi}, \qquad P_2 = \frac{\mu_1 c_2 - \mu_2 c_1}{\Psi}$$

Proof. The first-order conditions of the agent are $\mu_n P_1 + \sigma_n P_2 = c_n$ for $n = 1, 2$, two linear equations in (P_1, P_2) whose determinant is $\Psi(a)$. Solving establishes the claim. □

Note that $P_1 > 0$ requires that $\sigma_2 c_1 - \sigma_1 c_2$ and $\Psi(a)$ share a sign, which holds automatically in the leading case in which productive effort compresses the dispersion of the indicator while window dressing expands it, i.e., $\sigma_1 \leq 0 \leq \sigma_2$.

Lemma 3. *Suppose that $\overline{v}(y;a)$ is a linear function of y with a positive slope and that h is log-concave. Then any upper-censorship rating satisfies $P_2 < 0$, any lower-censorship rating satisfies $P_2 > 0$, and the fully revealing rating satisfies $P_2 = 0$.*

Proof. Let $\beta(a) > 0$ be the slope of $\overline{v}$, and consider an upper-censorship rating with threshold $\hat{y} = \mu(a) + \sigma(a)\hat{\varepsilon}$. The measure $d\pi$ has density $\beta(a)\sigma(a)$ below $\hat{\varepsilon}$ and an atom of size $\beta(a)\sigma(a)m(\hat{\varepsilon})$ at $\hat{\varepsilon}$, where $m(x) = \mathbb{E}[\varepsilon - x | \varepsilon \geq x]$ is the mean residual life of ε. Therefore,

$$P_2 = \beta(a)\left[\Theta(\hat{\varepsilon}) + m(\hat{\varepsilon})\hat{\varepsilon}h(\hat{\varepsilon})\right], \qquad \Theta(x) = \int_{-\infty}^{x} \varepsilon h(\varepsilon)\, d\varepsilon < 0$$

When $\hat{\varepsilon} \leq 0$, both terms are non-positive, and the first is strictly negative. When $\hat{\varepsilon} > 0$, since ε has mean zero, $-\Theta(\hat{\varepsilon}) = (m(\hat{\varepsilon}) + \hat{\varepsilon})(1 - H(\hat{\varepsilon}))$, and thus $P_2 < 0$ is equivalent to

$$\hat{\varepsilon} m(\hat{\varepsilon}) h(\hat{\varepsilon}) < (m(\hat{\varepsilon}) + \hat{\varepsilon})(1 - H(\hat{\varepsilon}))$$

Log-concavity of h implies log-concavity of $1 - H$, which implies $m(x) \leq (1 - H(x))/h(x)$; the left-hand side is thus at most $\hat{\varepsilon}(1 - H(\hat{\varepsilon}))$, which is strictly below the right-hand side since $m(\hat{\varepsilon}) > 0$. The argument for lower-censorship ratings is the mirror of this argument and uses $\mathbb{E}[x - \varepsilon | \varepsilon \leq x] \leq H(x)/h(x)$, implied by log-concavity of H. Finally, under full disclosure, $d\pi = \beta\sigma(a)\, d\varepsilon$ and $P_2 = \beta \int \varepsilon h(\varepsilon)\, d\varepsilon = 0$. □

We can now prove Proposition 5.

Proof. An uninformative rating satisfies $B_n = 0$, and thus the first-order conditions of the agent imply $c_n = 0$, which contradicts the interiority of the implemented action. Hence, by Theorem 2, the welfare-maximizing rating is an upper- or a lower-censorship rating. By Lemma 2, $P_2 = (\mu_1 c_2 - \mu_2 c_1)/\Psi$ at the implemented action. When (6) holds, $P_2 \leq 0$, and thus by Lemma 3, the optimal rating cannot be lower censorship; it is therefore upper censorship, degenerating to full disclosure when (6) holds with equality. When (6) fails, $P_2 > 0$, and the optimal rating cannot be upper censorship; it is therefore lower censorship. This establishes the claim. □

B.1 Optimality of Upper Censorship in Example 1

In this section, we show that the optimal rating in Example 1 is indeed upper censorship and is not full disclosure. Recall that in the example

$$\overline{v}(y;a) = a_1 + \frac{y - a_1 - a_2}{1 + a_2^2} \Rightarrow v_Q(i;a) = a_1 + \frac{\Phi^{-1}(i)}{\sqrt{1 + a_2^2}}$$

with $\Phi(y)$ being the c.d.f. of the standard normal.

Note that the FOA incentive compatibility in Lemma 2 can be written as

$$\kappa_1 a_1 = \frac{1}{1+a_2^2}\frac{P_1}{\beta}, \quad \kappa_2 a_2 = \frac{1}{1+a_2^2}\left[\frac{P_1}{\beta} + \frac{a_2}{\sqrt{1+a_2^2}}\frac{P_2}{\beta}\right]$$

where in the above $\beta = 1/(1+a_2^2)$ and P_1 and P_2 are defined above. As Lemma 3 states, $P_2 > 0$ is associated with lower censorship. Here, we show that if $P_2 > 0$, then there exists a perturbation of a that increases a_1 and decreases a_2, and thus increases welfare – this is because $P_1 \leq \beta < 1$ and $P_1 = c_1$. To see this, perturb a to $a' = a + da$. Let $p_Q(i)$ be the initial price vector that satisfies $P_2 > 0$, and consider the following change in the interim prices:

$$\hat{p}_Q(i) = (1-\delta)\left[\mathbb{E}[v|a'] + \frac{\beta(a')\sigma(a')}{\beta(a)\sigma(a)}(p_Q(i) - \mathbb{E}[v|a])\right] + \delta\frac{P_1}{\beta(a)}v_Q(i;a') + \delta\left(1 - \frac{P_1}{\beta(a)}\right)\mathbb{E}[v|a']$$

Note that the above is valid since $P_1 \leq \beta$, with equality for a fully informative interim price. For δ positive and close to 0, the above perturbation guarantees the return to location to be equal to $P_1' = \beta(a')P_1/\beta(a)$. This is because $h = \frac{e^{-\varepsilon^2/2}}{\sqrt{2\pi}}$, $H(\varepsilon) = \Phi(\varepsilon)$, and thus we can rewrite

$$\begin{aligned}
P_1' &= \frac{1}{\sigma(a')}\int_0^1 \hat{p}_Q(i)\Phi^{-1}(i)\,di \\
&= \frac{1}{\sigma(a')}\int_0^1 \left[(1-\delta)\frac{\beta(a')\sigma(a')}{\beta(a)\sigma(a)}p_Q(i) + \delta\frac{P_1}{\beta(a)}v_Q(i;a')\right]\Phi^{-1}(i)\,di \\
&= (1-\delta)\frac{\beta(a')}{\beta(a)}P_1 + \frac{\delta P_1}{\beta(a)}\beta(a') = \frac{P_1\beta(a')}{\beta(a)}
\end{aligned}$$

The new value of P_2' is given by

$$\begin{aligned}
P_2' &= \frac{1}{\sigma(a')}\int_0^1 \hat{p}_Q(i)\left(\Phi^{-1}(i)^2 - 1\right)di \\
&= \frac{1}{\sigma(a')}\int_0^1 \left[(1-\delta)\frac{\beta(a')\sigma(a')}{\beta(a)\sigma(a)}p_Q(i) + \delta\frac{P_1}{\beta(a)}v_Q(i;a')\right]\left(\Phi^{-1}(i)^2 - 1\right)di \\
&= (1-\delta)\frac{\beta(a')}{\beta(a)}P_2
\end{aligned}$$

This means that $P_2'/\beta(a') = (1-\delta)\frac{P_2}{\beta}$. Now lowering P_2/β and keeping P_1/β fixed implies that we can use an implicit derivative to calculate the resulting change in a_2 and a_1

$$\kappa_2 da_2 = -\frac{2a_2}{1+a_2^2}\kappa_2 a_2 da_2 + \frac{1}{(1+a_2^2)^{5/2}}\frac{P_2}{\beta}da_2 - \delta\frac{a_2}{(1+a_2^2)^{3/2}}\frac{P_2}{\beta}$$

$$\Rightarrow \left[\frac{1+3a_2^2}{1+a_2^2}\kappa_2 - \frac{1}{(1+a_2^2)^{5/2}}\frac{P_2}{\beta}\right]da_2 = -\delta\frac{a_2}{(1+a_2^2)^{3/2}}\frac{P_2}{\beta}$$

From the FOA IC constraint above, we have

$$\frac{1+3a_2^2}{1+a_2^2}\kappa_2 - \frac{1}{(1+a_2^2)^{5/2}}\frac{P_2}{\beta} = \frac{1+3a_2^2}{1+a_2^2}\kappa_2 - \frac{1}{1+a_2^2}\kappa_2 + \frac{1}{a_2(1+a_2^2)^2}\frac{P_1}{\beta}$$

$$= \frac{3a_2^2}{1+a_2^2}\kappa_2 + \frac{1}{a_2\left(1+a_2^2\right)^2}\frac{P_1}{\beta}$$

Thus, the above is positive, which implies that $da_2 < 0$. Then the first FOA IC constraint implies that $da_1 > 0$.

To rule out the optimality of full disclosure, we can start from full revelation, i.e., $p_Q = v_Q$. In this case, $P_1 = \beta$ and $P_2 = 0$. Now, consider a rating that pools observations with quantiles above j. Then, the FOA ICs are given by

$$\begin{aligned}
\kappa_1 a_1\left(1+a_2^2\right) &= \int_0^j \Phi^{-1}(i)^2\, di + \frac{\phi\left(\Phi^{-1}(j)\right)^2}{1-j} \\
\kappa_2 a_2\left(1+a_2^2\right) &= \int_0^j \Phi^{-1}(i)^2\, di + \frac{\phi\left(\Phi^{-1}(j)\right)^2}{1-j} \\
&\quad + \frac{a_2}{\sqrt{1+a_2^2}}\left[-\phi\left(\Phi^{-1}(j)\right)\left(1+\Phi^{-1}(j)^2\right) + \Phi^{-1}(j)\frac{\phi\left(\Phi^{-1}(j)\right)^2}{1-j}\right]
\end{aligned}$$

Let us define $m(x) = \frac{\int_x^\infty (z-x)\phi(z)dz}{1-\Phi(x)} = \frac{\phi(x)}{1-\Phi(x)} - x$. Note that as $x \to \infty$, $m(x) \to 0$, and $m(x)\,x \to 1$ and $(1-m(x)\,x)\,x^2 \to 2$ both from below. Using $m = m\left(\Phi^{-1}(j)\right) = m(y)$, this allows us to rewrite the above FOA IC as

$$\begin{aligned}
\kappa_1 a_1\left(1+a_2^2\right) &= 1-(1-j)\left[1-m(m+y)\right] \\
\kappa_2 a_2\left(1+a_2^2\right) &= 1-(1-j)\left[1-m(m+y)\right] - \frac{a_2}{\sqrt{1+a_2^2}}(1-j)(m+y)(1-ym)
\end{aligned}$$

Obviously, at $j < 1$, a_2 is less than its value under $j = 1$. This is because the term multiplying $a_2/\sqrt{1+a_2^2}$ is positive. Moreover, as $j \to 1$,

$$(1-j)\left[1-m(m+y)\right] \approx \frac{1-j}{y^2}$$

In contrast, using the last equation for $j \to 1$ gives us

$$\kappa_2 a_2\left(1+a_2^2\right) \approx 1 - \frac{1-j}{y^2} - \frac{a_2}{\sqrt{1+a_2^2}}(1-j)(1/y+y)\frac{2}{y^2}$$

Since $y \to \infty$, the slowest converging term to 0 dominates, and thus we can write

$$\kappa_2 a_2\left(1+a_2^2\right) \approx 1 - D \times \frac{1-j}{y}$$

for some $D > 0$. For a_2 close enough to the full-disclosure benchmark, the above implies that $1+a_2^2$ also falls by something proportional to $\frac{1-j}{y}$. Now, by the IC for a_1, the RHS declines by $\frac{1-j}{y^2}$ while $1+a_2^2$ declines by $\frac{1-j}{y}$. Hence, for y large enough, a_1 increases and a_2 decreases. Since the return to a_1 is always less than $\beta < 1$, an increase in a_1 increases welfare, while the decrease in a_2 does the same. This establishes the desired property.

C Validity of the First-Order Approach

In this section, we describe conditions that make the first-order approach valid. To do so, we use an approach similar to that of Jewitt (1988) and, more recently, Chade and Swinkels (2020). More specifically, it is sufficient to show that, given our optimal ratings (lower- or upper-censorship), the payoff of the agent is quasi-concave in her effort. To show quasi-concavity of the payoff, it is sufficient to show that $U'(a) = 0$ implies $U''(a) < 0$ if U is twice continuously differentiable. This would imply that U cannot have more than one local maximum, i.e., it is single-peaked, and is thus quasi-concave.

To see this, suppose that there are two points $a_1 < a_2$ such that $U'(a_1) = U'(a_2) = 0$. Since $U''(a_1) < 0$, it must be that there is an interval of values above a_1 for which $U'(a) < 0$. Now, without loss of generality, let us assume that $a_2 = \inf_{a' > a_1, U'(a')=0} a'$. Since U is assumed to be twice continuously differentiable, U' is continuous and thus $U'(a_2) = 0$, and since $U'(a) < 0$ for an interval above a_1, $a_1 < a_2$. Moreover, we must have that for all $a \in (a_1, a_2)$, $U'(a) < 0$. This implies that $U'(a) < U'(a_2) = 0$ for values of a below a_2. Since U is assumed to be twice continuously differentiable, we must have that $U''(a_2) \geq 0$, which is a contradiction. This implies that U is single-peaked.

Throughout the section, we assume that the cost $c(a)$ is twice differentiable, increasing, and convex.

C.1 A Single Task

The ratings that arise in Sections 3 and 5 – full disclosure, upper- and lower-censorship, and mid-censorship – are all deterministic monotone partitions. The following lemma provides a condition for the validity of FOA that applies to all of them at once:

Lemma 4. *Let p be a non-decreasing interim price schedule and suppose that for all $a, \hat{a} \in A$,*

$$\int p(y; \hat{a}) \frac{\partial^2 g(y|a)}{\partial a^2} dy < \frac{c''(a)}{c'(a)} \int p(y; \hat{a}) \frac{\partial g(y|a)}{\partial a} dy$$

Then the payoff of the agent is quasi-concave in effort, and the FOA is valid under p.

Proof. The payoff of the agent is $U(a) = \int p(y; \hat{a}) g(y|a) dy - c(a)$. Suppose that $U'(a) = 0$ at some effort level a. Then $\int p\, g_a dy = c'(a) > 0$ and

$$U''(a) = \int p\, g_{aa} dy - c''(a) = \int p\, g_{aa} dy - \frac{c''(a)}{c'(a)} \int p\, g_a dy < 0$$

by the assumption of the lemma. Given our argument above, U is single-peaked and thus quasi-concave. □

For upper- and lower-censorship ratings, integration by parts turns the condition of Lemma 4 into a condition on the distribution function directly. This gives the following:

Lemma 5. *Suppose that the family of distributions $\{G(y|a)\}_{a\in A}$ is twice continuously differentiable and A is a convex subset of $\mathbb{R}$. Suppose further that G satisfies the following properties:*

$$0 > \int_{-\infty}^{\hat{y}} \left(\overline{v}(y;\hat{a}) - \mathbb{E}\left[\overline{v}|y \geq \hat{y}\right]\right) \frac{\partial}{\partial a}\frac{g_a(y|a)}{c'(a)} dy, \forall a, \hat{a} \in A, \hat{y} \in \mathbb{R},$$
$$0 > \int_{\hat{y}}^{\infty} \left(\overline{v}(y;\hat{a}) - \mathbb{E}\left[\overline{v}|y \leq \hat{y}\right]\right) \frac{\partial}{\partial a}\frac{g_a(y|a)}{c'(a)} dy, \forall a, \hat{a} \in A, \hat{y} \in \mathbb{R}.$$

Then the FOA is valid under lower- and upper-censorship policies.

Proof. Consider an upper-censorship policy that pools realizations of y above $\hat{y}$. If the market believes the agent chooses effort $\hat{a}$, then the payoff of the agent is given by

$$U(a) = \int_{-\infty}^{\hat{y}} \overline{v}(y;\hat{a})\, g(y|a)\, dy + \frac{\int_{\hat{y}}^{\infty} \overline{v}(y;\hat{a})\, g(y|\hat{a})\, dy}{1 - G(\hat{y}|\hat{a})} \left(1 - G(\hat{y}|a)\right) - c(a)$$

If we let $p_H = \mathbb{E}\left[\overline{v}(y;\hat{a})\,|y \geq \hat{y}\right] > \overline{v}(\hat{y};\hat{a})$, we can write

$$\begin{aligned} U'(a) &= \int_{-\infty}^{\hat{y}} \overline{v}(y;\hat{a})\, g_a(y|a)\, dy - p_H G_a(\hat{y}|a) - c'(a) \\ &= \int_{-\infty}^{\hat{y}} \left(\overline{v}(y;\hat{a}) - p_H\right) g_a(y|a)\, dy - c'(a) \end{aligned}$$

Now, suppose that $U'(a_1) = 0$ at some effort level a_1. Then,

$$\begin{aligned} U''(a_1) &= \int_{-\infty}^{\hat{y}} \left(\overline{v}(y;\hat{a}) - p_H\right) g_{aa}(y|a_1)\, dy - c''(a_1) \\ &= \int_{-\infty}^{\hat{y}} \left(\overline{v}(y;\hat{a}) - p_H\right) g_{aa}(y|a_1)\, dy \\ &\quad - \frac{c''(a_1)}{c'(a_1)} c'(a_1) \\ &= \int_{-\infty}^{\hat{y}} \left(\overline{v}(y;\hat{a}) - p_H\right) g_{aa}(y|a_1)\, dy \\ &\quad - \frac{c''(a_1)}{c'(a_1)} \int_{-\infty}^{\hat{y}} \left(\overline{v}(y;\hat{a}) - p_H\right) g_a(y|a_1)\, dy \\ &= \int_{-\infty}^{\hat{y}} \left(\overline{v}(y;\hat{a}) - p_H\right) \left[g_{aa}(y|a_1) - \frac{c''(a_1)}{c'(a_1)} g_a(y|a_1)\right] dy \\ &= \int_{-\infty}^{\hat{y}} \left(\overline{v}(y;\hat{a}) - p_H\right) c'(a_1) \frac{\partial}{\partial a}\frac{g_a(y|a_1)}{c'(a_1)} dy \end{aligned}$$

Since $c'(a_1) > 0$, the assumption on G in the statement of the lemma guarantees that $U''(a_1) < 0$. Given our argument above, this implies that U is quasi-concave, and thus FOA is valid. The argument for lower-censorship ratings is the mirror of this argument. □

The essence of the conditions in Lemma 5 is that they put a restriction on how convex the cost is, captured by $c''(a)/c'(a)$, relative to that of the expected interim price. Indeed, the conditions can be rewritten as

$$\frac{\int_{-\infty}^{\infty} p(y;\hat{y},\hat{a})\, g_{aa}(y|a)\, dy}{\int_{-\infty}^{\infty} p(y;\hat{y},\hat{a})\, g_a(y|a)\, dy} < \frac{c''(a)}{c'(a)}, \forall a \in A$$

where in the case of lower and upper censorship, respectively, interim prices are

$$p(y;\hat{y},\hat{a}) = \begin{cases} \overline{v}(y;\hat{a}) & y \geq \hat{y} \\ \mathbb{E}_{\hat{a}}[v|y \leq \hat{y}] & y \leq \hat{y} \end{cases}$$

$$p(y;\hat{y},\hat{a}) = \begin{cases} \mathbb{E}_{\hat{a}}[v|y \geq \hat{y}] & y \geq \hat{y} \\ \overline{v}(y;\hat{a}) & y \leq \hat{y} \end{cases}$$

In the special case where $\frac{\partial}{\partial y}\overline{v}(y;a) = \beta(a)$ and $y = a + \varepsilon$ with $\varepsilon \sim H(\varepsilon)$ and density $h(\varepsilon) = H'(\varepsilon)$, these become

$$\frac{\int_{-\infty}^{\hat{y}} (y - \mathbb{E}_{\hat{a}}[y|y \geq \hat{y}])\, g_{aa}(y|a)\, dy}{\int_{-\infty}^{\hat{y}} (y - \mathbb{E}_{\hat{a}}[y|y \geq \hat{y}])\, g_a(y|a)\, dy} < \frac{c''(a)}{c'(a)}$$

$$\frac{\int_{\hat{y}}^{\infty} (y - \mathbb{E}_{\hat{a}}[y|y \leq \hat{y}])\, g_{aa}(y|a)\, dy}{\int_{\hat{y}}^{\infty} (y - \mathbb{E}_{\hat{a}}[y|y \leq \hat{y}])\, g_a(y|a)\, dy} < \frac{c''(a)}{c'(a)}$$

and we have

$$\int_{-\infty}^{\hat{y}} (y - \mathbb{E}_{\hat{a}}[y|y \geq \hat{y}])\, g_a(y|a)\, dy =$$

$$\int_{-\infty}^{\hat{y}} (y - \hat{y} - \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}])\, dG_a(y|a) =$$

$$-\int_{-\infty}^{\hat{y}} G_a(y|a)\, dy - \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}]\, G_a(\hat{y}|a) =$$

$$-\int_{-\infty}^{\hat{y}} \frac{\partial}{\partial a} H(y - a)\, dy - \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}]\, \frac{\partial}{\partial a} H(\hat{y} - a) =$$

$$H(\hat{y} - a) + \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}]\, h(\hat{y} - a)$$

$$\int_{-\infty}^{\hat{y}} (y - \mathbb{E}_{\hat{a}}[y|y \geq \hat{y}])\, g_{aa}(y|a)\, dy =$$

$$-h(\hat{y} - a) - \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}]\, h'(\hat{y} - a)$$

Let us make the following assumption on H:

Assumption 4. *The cumulative distribution function $H(\varepsilon)$ satisfies:*

1. $\log(1 - H(\varepsilon))$ *is concave in* ε,
2. $\log H(\varepsilon)$ *is concave in* ε,

3. $\forall d > 0$, $\frac{h(x)}{1-H(x)} - \frac{h(x-d)}{1-H(x-d)} \leq \kappa_1(d)$, $\frac{h(x-d)}{H(x-d)} - \frac{h(x)}{H(x)} \leq \kappa_2(d)$

The first two conditions are standard log-concavity conditions, while the last condition implies that the variations in the derivative of $\log(1-H(x))$ are uniformly bounded above. A sufficient condition for the latter is that the expressions $h'(x)/(1-H(x)) + h(x)^2/(1-H(x))^2$ and $h'(x)/H(x) - h(x)^2/(H(x))^2$ are bounded above.

Under the above assumptions, $h(\varepsilon)/(1-H(\varepsilon))$ is increasing in ε, and therefore,

$$\begin{aligned}\mathbb{E}_{\hat{a}}[y-\hat{y}|y \geq \hat{y}] &= \frac{\int_{\hat{y}}^{\infty}(y-\hat{y})\,dH(y-\hat{a})}{1-H(\hat{y}-\hat{a})} \\ &= \frac{\int_{\hat{y}-\hat{a}}^{\infty}\frac{1-H(z)}{h(z)}dH(z)}{1-H(\hat{y}-\hat{a})} \leq \frac{1-H(\hat{y}-\hat{a})}{h(\hat{y}-\hat{a})}\end{aligned}$$

We then have that

$$\begin{aligned}&\frac{\int_{-\infty}^{\hat{y}}(y-\mathbb{E}_{\hat{a}}[y|y\geq\hat{y}])\,g_{aa}(y|a)\,dy}{\int_{-\infty}^{\hat{y}}(y-\mathbb{E}_{\hat{a}}[y|y\geq\hat{y}])\,g_a(y|a)\,dy} = \\ &-\frac{h(\hat{y}-a)+\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]\,h'(\hat{y}-a)}{H(\hat{y}-a)+\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]\,h(\hat{y}-a)} = \\ &-\frac{\frac{1}{\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]}+\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{1}{\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}}\frac{h(\hat{y}-a)}{H(\hat{y}-a)} = \left(-1+\frac{\frac{h(\hat{y}-a)}{H(\hat{y}-a)}-\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{1}{\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}}\right)\frac{h(\hat{y}-a)}{H(\hat{y}-a)}\end{aligned}$$

By $\log$-concavity of H, $h/H - h'/h \geq 0$, and hence, we have the following inequality

$$-1+\frac{\frac{h(\hat{y}-a)}{H(\hat{y}-a)}-\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{1}{\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}} \leq -1+\frac{\frac{h(\hat{y}-a)}{H(\hat{y}-a)}-\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{h(\hat{y}-\hat{a})}{1-H(\hat{y}-\hat{a})}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}} = -\frac{\frac{h(\hat{y}-\hat{a})}{1-H(\hat{y}-\hat{a})}+\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{h(\hat{y}-\hat{a})}{1-H(\hat{y}-\hat{a})}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}}$$

by the above property of $\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]$. If we define $\hat{y}-a=x$ and $d=\hat{a}-a$, then

$$\begin{aligned}\frac{\int_{-\infty}^{\hat{y}}(y-\mathbb{E}_{\hat{a}}[y|y\geq\hat{y}])\,g_{aa}(y|a)\,dy}{\int_{-\infty}^{\hat{y}}(y-\mathbb{E}_{\hat{a}}[y|y\geq\hat{y}])\,g_a(y|a)\,dy} &\leq -\frac{\frac{h(x-d)}{1-H(x-d)}+\frac{h'(x)}{h(x)}}{\frac{h(x-d)}{1-H(x-d)}+\frac{h(x)}{H(x)}}\frac{h(x)}{H(x)} \\ &= -\frac{\frac{h(x-d)}{1-H(x-d)}+\frac{h'(x)}{h(x)}}{\frac{h(x-d)}{1-H(x-d)}\frac{H(x)}{h(x)}+1}\end{aligned}$$

Since $1-H$ is log-concave, $h'/h \geq -h/(1-H)$, and the right-hand side of the above inequality satisfies

$$-\frac{\frac{h(x-d)}{1-H(x-d)}+\frac{h'(x)}{h(x)}}{\frac{h(x-d)}{1-H(x-d)}\frac{H(x)}{h(x)}+1} \leq \frac{\frac{h(x)}{1-H(x)}-\frac{h(x-d)}{1-H(x-d)}}{\frac{h(x-d)}{1-H(x-d)}\frac{H(x)}{h(x)}+1}$$

Note that in the above, if $d<0$, since $h/(1-H)$ is increasing (log-concavity of $1-H$), the RHS of the above inequality is negative, which is guaranteed to be less than $c''(a)/c'(a)$ since c is convex.

By Assumption 4, the RHS of the above is less than $\kappa_1(d)$. If $A=[0,\overline{a}]$, then the highest value

of d is $\overline{a}$, which means that it is sufficient for c to satisfy

$$\max_{d\leq\overline{a}}\kappa_1(d)\leq\frac{c''(a)}{c'(a)}$$

In a similar fashion, we can show that for lower-censorship ratings to satisfy the requirement of Lemma 5, we must also have

$$\max_{d\leq\overline{a}}\kappa_2(d)\leq\frac{c''(a)}{c'(a)}$$

Thus, the validity of FOA is guaranteed by c having a high enough curvature.

It remains to consider the ratings of Section 3. Full disclosure, which is optimal under MLRP (Proposition 3), satisfies the condition of Lemma 4 automatically in the location family: $\int\overline{v}\,g_{aa}dy=0$ while $\int\overline{v}\,g_a dy>0$. The censorship ratings of Proposition 4, by contrast, arise under ELRP and CLRP, and these properties cannot hold in a location family, where $\partial^2\log g/\partial a\partial y$ is single-signed; effort must also move the dispersion of the indicator, as in the examples of Section 3. For location–scale technologies with market values affine in the indicator, the validity of FOA under upper- and lower-censorship ratings follows from the computation of Section C.3 below, applied with a single task.

C.2 Multiple Tasks

The argument extends to the multi-task model of Section 4 once derivatives are replaced by directional derivatives. For $z\in\mathbb{R}^N$, write $D_z=\sum_n z_n\partial/\partial a_n$ and $D_z^2=\sum_{n,m}z_nz_m\partial^2/\partial a_n\partial a_m$. The observation that does the work is that single-peakedness is a property of restrictions to lines, so that the one-dimensional argument above can be applied direction by direction.

Lemma 6. *Let U be twice continuously differentiable on a convex set $A\subset\mathbb{R}^N$ and suppose that for every $a\in A$ and every $z\neq0$, $D_zU(a)=0$ implies $D_z^2U(a)<0$. Then U is single-peaked along every line in A, and any $a^*\in A$ with $\nabla U(a^*)=0$ maximizes U over A.*

Proof. Fix $a\in A$ and $z\neq0$, and let $\phi(t)=U(a+tz)$, defined on an interval since A is convex. Then $\phi'(t)=D_zU(a+tz)$ and $\phi''(t)=D_z^2U(a+tz)$, so $\phi'=0$ implies $\phi''<0$, and ϕ is single-peaked by the argument at the beginning of this section. Now let $\nabla U(a^*)=0$ and take any $a\in A$. With $z=a-a^*$, $\phi'(0)=z\cdot\nabla U(a^*)=0$, so $t=0$ is the peak of ϕ and $U(a)=\phi(1)\leq\phi(0)=U(a^*)$. □

Since $D_{-z}U=-D_zU$ while $D_{-z}^2U=D_z^2U$, the requirement of Lemma 6 need only be verified for directions in a half-space; we use this to restrict attention to z with $D_zc(a)\geq0$. The analogue of Lemma 4 is then:

Lemma 7. *Let p be a non-decreasing interim price schedule and suppose that for all $a, \hat{a} \in A$ and all $z \neq 0$ with $D_z c(a) > 0$,*

$$\int p(y;\hat{a})\, D_z^2 g(y|a)\, dy < \frac{D_z^2 c(a)}{D_z c(a)} \int p(y;\hat{a})\, D_z g(y|a)\, dy$$

while $\int p\, D_z^2 g\, dy < D_z^2 c(a)$ for $z \neq 0$ with $D_z c(a) = 0$. Then the FOA is valid under p.

Proof. For any direction z, $D_z U(a) = \int p\, D_z g\, dy - D_z c(a)$ and $D_z^2 U(a) = \int p\, D_z^2 g\, dy - D_z^2 c(a)$. Suppose that $D_z U(a) = 0$ with $D_z c(a) > 0$. Then $\int p\, D_z g\, dy = D_z c(a)$, and substituting as in the proof of Lemma 4 delivers $D_z^2 U(a) < 0$. When $D_z c(a) = 0$, $D_z U(a) = 0$ requires $\int p\, D_z g\, dy = 0$, and the second assumption applies directly. Lemma 6 then implies that any effort satisfying the agent's first-order conditions maximizes her payoff. □

The condition of Lemma 7 has a simple reading: restricted to the line $t \mapsto a + tz$, the technology is the one-parameter family $\tilde{G}(y|t) = G(y|a+tz)$ with cost $\tilde{c}(t) = c(a+tz)$, and the displayed inequality is precisely the requirement of Lemma 4 applied to $\left(\tilde{G}, \tilde{c}\right)$. The FOA is thus valid in the multi-task model whenever it is valid in every single-task model obtained by restricting effort to a line: multi-tasking creates no new difficulty beyond requiring the single-task condition to hold uniformly across directions.

C.3 The Ratings of Section 4

Proposition 5 invokes FOA for upper- and lower-censorship ratings when G is a location–scale technology and $\overline{v}$ is affine in y. In this case, we can establish that as long as the cost function has a lowest curvature level above a threshold, FOA is valid and the value of the agent is quasi-concave in effort. We first make the following assumption:

Assumption 5. *In the location–scale technology with log-concave density, $h(\varepsilon)$, there exists $\overline{C}$ such that $\left(1 + |\varepsilon|^3\right) |h'(\varepsilon)| \leq \overline{C}$ for all values of $\varepsilon \in \mathbb{R}$.*

This assumption is satisfied by the normal distribution as well as by Logistic, Gumbel, Generalized Normal (and Gamma and Weibull for shape parameter no less than 2).

Under an upper-censorship rating with threshold $\hat{y}$ and market beliefs $\hat{a}$, the gross payoff of the agent is, up to a constant,

$$\Pi(a) = -\beta(\hat{a})\,\sigma(a)\,\mathcal{H}(\hat{\varepsilon}) - \Delta H(\hat{\varepsilon}), \qquad \mathcal{H}(x) = \int_{-\infty}^{x} H(t)\, dt$$

where $\hat{\varepsilon} = (\hat{y} - \mu(a))/\sigma(a)$ and $\Delta = \beta(\hat{a})\,\sigma(\hat{a})\, m(\hat{\varepsilon}_{\hat{a}})$ is the atom of the price schedule at the threshold, with $m(x) = \mathbb{E}[\varepsilon - x | \varepsilon \geq x]$. Writing $\mu_z = D_z\mu$, $\sigma_z = D_z\sigma$ and similarly for second derivatives, and using $D_z\hat{\varepsilon} = -(\mu_z + \hat{\varepsilon}\sigma_z)/\sigma$, we obtain

$$D_z^2\Pi(a) = -\beta\left[\sigma_{zz}\mathcal{H}(\hat{\varepsilon}) + 2\sigma_z H(\hat{\varepsilon})\, D_z\hat{\varepsilon} + \sigma\left(h(\hat{\varepsilon})\,(D_z\hat{\varepsilon})^2 + H(\hat{\varepsilon})\, D_z^2\hat{\varepsilon}\right)\right]$$

$$- \Delta \left[h' \left(\hat{\varepsilon} \right) \left(D_z \hat{\varepsilon} \right)^2 + h \left(\hat{\varepsilon} \right) D_z^2 \hat{\varepsilon} \right]$$

where $D_z^2 \hat{\varepsilon} = - \left(\mu_{zz} + \hat{\varepsilon} \sigma_{zz} \right) / \sigma - 2 D_z \hat{\varepsilon} \sigma_z / \sigma$, and a similar expression holds for lower censorship and the same argument below applies.

Substituting for $D_z^2 \hat{\varepsilon}$ in the first line implies that the terms involving $\sigma_z H \left(\hat{\varepsilon} \right) D_z \hat{\varepsilon}$ cancel, and we have

$$D_z^2 \Pi \left(a \right) = -\beta \left[\sigma_{zz} \left(\mathcal{H} \left(\hat{\varepsilon} \right) - \hat{\varepsilon} H \left(\hat{\varepsilon} \right) \right) + \sigma h \left(\hat{\varepsilon} \right) \left(D_z \hat{\varepsilon} \right)^2 - \mu_{zz} H \left(\hat{\varepsilon} \right) \right] - \Delta \left[h' \left(\hat{\varepsilon} \right) \left(D_z \hat{\varepsilon} \right)^2 + h \left(\hat{\varepsilon} \right) D_z^2 \hat{\varepsilon} \right]$$

Here, we argue that the above is bounded above for $\| z \| = 1$. Since μ and σ are twice continuously differentiable on the compact action space, μ_{zz} and β are bounded and so is $\beta \mu_{zz} H \left(\hat{\varepsilon} \right)$ – the last term in the first bracket. The second term in the first bracket, $-\beta \sigma h \left(\hat{\varepsilon} \right) \left(D_z \hat{\varepsilon} \right)^2$ is bounded above since h is log-concave. The first term in the first bracket is proportional to

$$\mathcal{H} \left(\hat{\varepsilon} \right) - \hat{\varepsilon} H \left(\hat{\varepsilon} \right) = \int_{-\infty}^{\hat{\varepsilon}} H \left(t \right) dt - \hat{\varepsilon} H \left(\hat{\varepsilon} \right) = - \int_{-\infty}^{\hat{\varepsilon}} t dH \left(t \right)$$

which is bounded since ε has a finite first moment. Finally, the second bracket is bounded for the following reasons: 1. in the very last term Δ and $D_z^2 \hat{\varepsilon}$ grow at most proportional to $|\hat{\varepsilon}|$; this implies that the last term grows proportional to $|\hat{\varepsilon}|^2 h \left(\hat{\varepsilon} \right)$ which is bounded due to log-concavity of h. 2. for the term, $-\Delta h' \left(D_z \hat{\varepsilon} \right)^2$, we note that both Δ and $D_z \hat{\varepsilon}$ grow proportional to $|\hat{\varepsilon}|$ and thus, this term stays bounded by Assumption 5. Therefore, we have directly established that

$$\overline{\overline{\Xi}} \equiv \sup \left\{ \frac{D_z^2 \Pi \left(a \right)}{\| z \|^2} \middle| a, \hat{a} \in A, \ \hat{y} \in \mathbb{R}, \ z \neq 0 \right\} < \infty$$

and if $D^2 c \left(a \right) - \overline{\overline{\Xi}} I$ is positive definite for every $a \in A$, then $D_z^2 U \left(a \right) \leq \overline{\overline{\Xi}} \| z \|^2 - D_z^2 c \left(a \right) < 0$ for all a and all $z \neq 0$: the payoff of the agent is strictly concave, and the FOA is valid a fortiori. As in the single-task case, the validity of FOA amounts to the cost having enough curvature, now in every direction.

For the technology of Example 1, where $\mu \left(a \right) = a_1 + a_2$ and $\sigma \left(a \right) = \sqrt{a_2^2 + 1}$, the curvature bound can be computed explicitly: $\mu_{zz} = 0$, $\sigma_z = a_2 z_2 / \sigma$, $\sigma_{zz} = z_2^2 / \sigma^3$, and the slope of market values is $\beta \left(\hat{a} \right) = 1 / \left(\hat{a}_2^2 + 1 \right)$. Evaluating $\overline{\overline{\Xi}}$ numerically over all censorship thresholds, directions, actions, and beliefs in $A = [0, \overline{a}]^2$ gives $\overline{\overline{\Xi}} \approx 0.20$ for upper-censorship ratings and $\overline{\overline{\Xi}} \approx 2.21$ for lower-censorship ratings when $\overline{a} = 1$, and 0.40 and 2.77 when $\overline{a} = 2$. Hence, with the quadratic cost of the example, $c(a) = \frac{1}{2}(\kappa_1 a_1^2 + \kappa_2 a_2^2)$, for which $D^2 c = \text{diag}(\kappa_1, \kappa_2)$, the FOA is valid under upper censorship, the optimal rating in this example, whenever $\min\{\kappa_1, \kappa_2\} > 0.20$ (resp. 0.40), and under both censorship forms whenever $\min\{\kappa_1, \kappa_2\} > 2.21$ (resp. 2.77). The required curvature grows with the size of the action space. Finally for the computations, we can verify the validity of the FOA directly by checking the payoff of the agent over $[0, \overline{a}]^2$ and they are concave.

C.4 The Ratings of Section 5

In the redistributive test design problem of Section 5, each type of student solves a single-effort problem given the rating, and the mid-censorship ratings of Proposition 6 are deterministic monotone partitions, so Lemma 4 applies type by type; the cost normalization k_θ cancels from the condition. Given the functional form $\log g(y|a) = \zeta(y) + r(y)\, m(a) - b(a)$, the score is $\psi(y|a) = r(y)\, m'(a) - b'(a)$ and $g_{aa} = (\psi^2 + r\, m'' - b'')\, g$, so the condition of Lemma 4 reads

$$\frac{\int p(y) \left[\psi(y|a)^2 + r(y)\, m''(a) - b''(a)\right] g(y|a)\, dy}{\int p(y)\, \psi(y|a)\, g(y|a)\, dy} < \frac{c''(a)}{c'(a)}.$$

In this setting, we can provide two conditions. First, if r is bounded above, the above holds when

$$\frac{c''}{c'} \geq m' \sup_y r(y) - b' + \frac{m''}{m'}$$

When r is unbounded, we make the following assumptions:

1. $m'(a), m''(a)$ are both bounded on A,
2. $\mathbb{E}\left[(1+|y|)\left(1+r(y)^2\right)|a\right]$ is bounded on A,
3. $\mathbb{E}[y - s | y \geq s, a]$ and $\mathbb{E}[s - y | y \leq s, a]$ are at most $\overline{C}(1+|s|)$ for some $\overline{C}$.

In this case, Ξ exists such that if $k_\theta c''(a) \geq \Xi$, then FOA is valid. These assumptions are satisfied for the class of distributions mentioned in Section 5. To see this, note that majorization together with monotonicity implies that for any value of t,

$$p(t)\left(1 - \overline{G}(t|a_P, a_R)\right) \leq \int_t^\infty y d\overline{G}(y|a_P, a_R), \int_{-\infty}^t y d\overline{G}(y|a_P, a_R) \leq p(t)\, \overline{G}(t|a_P, a_R)$$

and hence,

$$\begin{aligned} &\mathbb{E}[y|y \leq t, a_P, a_R] \leq p(t) \leq \mathbb{E}[y|y \geq t, a_P, a_R] \\ \Rightarrow\ &\mathbb{E}[y - t|y \leq t, a_P, a_R] + t \leq p(t) \leq \mathbb{E}[y - t|y \geq t, a_P, a_R] + t \end{aligned}$$

This implies that

$$|p(t)| \leq \left(1 + \overline{C}\right)(1 + |t|)$$

We also have

$$\begin{aligned} \left|\frac{g_{aa}}{g}\right| &= \left|\psi^2 + r\, m'' - b''\right| \\ &= \left|(m')^2 \left[(r - \mathbb{E}[r|a])^2 - \text{Var}[r|a]\right] + m''(r - \mathbb{E}[r|a])\right| \end{aligned}$$

where in the above we have used $b' = m'\mathbb{E}[r|a]$ and $b'' = m''\mathbb{E}[r|a] + (m')^2 \operatorname{Var}[r|a]$. Given the assumptions above, there must exist $C_1 > 0$ so that

$$\left|\frac{g_{aa}}{g}\right| \leq C_1 \left(1 + r(y)^2\right).$$

Hence,

$$\left|\int p g_{aa} dy\right| \leq \int |p| \left|\frac{g_{aa}}{g}\right| dG \leq \left(1 + \overline{C}\right) C_1 \int (1 + |y|) \left(1 + r(y)^2\right) dG \leq \Xi$$

for some $\Xi > 0$. Hence, if $k_\theta c''(a) \geq \Xi$ for all $a \in A$, the payoff of each type of agent is concave and FOA is valid.

For the example of Section 5 (Example 3), where $y = a + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0,1)$ and $c(a) = a^2/2$, we can verify FOA numerically: for the rating and efforts obtained from the concavification, we check that $U'_\theta(a)$ changes sign only once on A, from positive to negative at $a = a_\theta$, for both types. For the parameter values of the example, this holds for both types and for both values of k_P. In fact $\int p g_{aa} dy \leq 0.4 < k_R$ for both values of k_P, so the payoff of each type is strictly concave.

D Importance of Comonotonicity in Proposition 1

The following counterexample demonstrates that there exist price schedules satisfying the mean-preserving contraction property that cannot be generated by any information structure.

Example 4. Suppose that $A = \{0, 1/3, 1\} = \{a_1, a_2, a_3\}$, $v = a_i$. The indicator is deterministic: $G(\{a_i\}|a_i) = 1$, and the prior is uniform $\mu(\{a_i\}) = 1/3$. In words, the market cares only about the action of the agent, and y coincides with it. Figure 9 depicts the feasible interim prices in the space of $(p(a_1), p(a_2))$ (the third coordinate is pinned down by Bayes' rule since $\mathbb{E}[p] = \mathbb{E}[v] = \frac{4}{9}$). Area A shows the set of random variables that are mean-preserving contractions of $(0, 1/3, 1)$. They are depicted by their first two variables, while the third is again pinned down by Bayes' rule.[20] The set of interim prices is denoted by area B in Figure 9. We find this set by solving the optimization problem associated with the highest and lowest values of $p(a_2)$ as a function of $p(a_1)$.[21] Evidently, the set B does not coincide with A. This is mainly due to the restrictions imposed by the second-order expectations. For example, it can be easily shown that the coefficient of $\overline{v}(a)$ in $p(a)$ is at least $1/3$, which means that $p(a_3)$ cannot become lower than $4/9$.

[20] The conditions are $0 \leq x_i \leq 1, 1/3 \leq x_i + x_j \leq 4/3$, for all i, j, and $x_1 + x_2 + x_3 = 4/3$.

[21] The upper and lower bounds of $p(a_2)$ can be found via a standard concavification method. The lower bound is given by $2(4 - p(a_1)) / \left(10 - 3p(a_1) + \sqrt{(8 - 3p(a_1))^2 - 12}\right)$ and the upper bound is $3 - (6 - 2p(a_1)) / (6 - 3p(a_1))$.

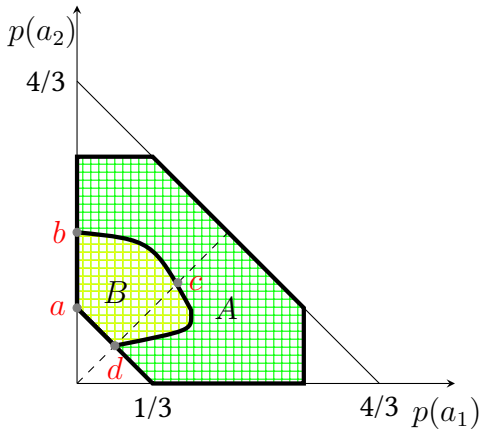


Figure 9: The set of interim prices and mean-preserving contractions of market valuations for Example 4. The green area, A, represents the three-state random variables that are mean-preserving contractions of market values. The yellow area, B, is the set of feasible interim prices under some information structure.

Finally, the points a, b, c, d are associated with deterministic ratings that either separate or pool the states and for which $p(a_1) \leq p(a_2) \leq p(a_3)$. Interestingly, if we consider the set of random variables whose realizations are less dispersed than $\overline{v}(a)$ and satisfy monotonicity, this coincides with the convex hull of the points a, b, c, and d. In Proposition 1, we show that this insight holds generally and allows us to significantly simplify the problem of rating design under a comonotonicity condition. ■

The result in Proposition 1 is reminiscent of the result of Blackwell (1953) and Rothschild and Stiglitz (1970), the general version of which can be found in Strassen (1965). That result states that for any two random variables x and y, there exists a random variable s such that $\mathbb{E}[x|s]$ has the same distribution as y if and only if y second-order stochastically dominates x.

While similar, our result is different in two respects. First, it is stated for the second-order *conditional* expectation, and thus Blackwell's result cannot be applied. The key intricacy is that the same signal structure that generates the random variable $\mathbb{E}[v|s]$ must be used to generate $\mathbb{E}[\mathbb{E}[v|s]|y]$. Second, as illustrated by Example 4, the equivalent of Blackwell's result does not hold in general and can be shown only when v and p are comonotone.

We also note that, since the indicator y matters for the market and payoffs only through its effect on $\overline{v}(y; a)$, we can relabel the indicator to be $\overline{v}(y; a)$. Under this reformulation, comonotonicity implies that interim prices $p(y)$ are a non-decreasing function of $\overline{v}(y; a)$.

E Interior Incentive-Compatible Efforts Are Open

In this section, we show the following lemma to be used in the proof of Theorem 1:

Lemma 8. *The set $U = \{a : 0 \in int T_a \mathcal{P}_a\}$ is open.*

Proof. To show this, consider an effort level $\tilde{a} \in U$ such that $0 \in \text{int} T_{\tilde{a}} \mathcal{P}_{\tilde{a}}$. This means that there exist $\varepsilon > 0$ and $p_Q^{j+}, p_Q^{j-} \in \mathcal{P}_{\tilde{a}}$ such that

$$T_{\tilde{a}} p_Q^{j+} = \varepsilon e_j, T_{\tilde{a}} p_Q^{j-} = -\varepsilon e_j, j = 1, \cdots, N$$

and given this construction, 0 is in the convex hull of $\left\{T_{\tilde{a}} p_Q^{j+}, T_{\tilde{a}} p_Q^{j-}\right\}_{j=1}^N$.

Next, we approximate each of these interim prices with interim prices that take finitely many values and use continuity to show that $\tilde{a}$ is interior to U. For any $m \geq 1$, let $\{I_k\}_{1 \leq k \leq m}$ be a partition of $[0,1]$ into m intervals of length $1/m$, and for any $p_Q \in \mathcal{P}_{\tilde{a}}$, we can define $p_{Q,m} \in \mathcal{P}_{\tilde{a}}$ by

$$p_{Q,m}(i) = \mathbb{E}\left[p_Q | I_k\right], i \in I_k.$$

By construction, we have $p_{Q,m} \succeq_{\text{c.v.}} p_Q$, and $p_{Q,m}$ is increasing and thus $p_{Q,m} \in \mathcal{P}_{\tilde{a}}$. Moreover, $\|p_{Q,m} - p_Q\|_2 \to 0$ as $m \to \infty$. Finally, by the Cauchy–Schwarz bound in Step 1 of the proof of Theorem 1, $T_{\tilde{a}}$ is continuous in $\|\cdot\|_2$, and thus $\|T_{\tilde{a}} p_{Q,m} - T_{\tilde{a}} p_Q\|_{\mathbb{R}^N} \to 0$. Given this, let $\tilde{p}_Q^{j+}, \tilde{p}_Q^{j-} \in \mathcal{P}_{\tilde{a}}$ be such interim prices for some large m that approximate p_Q^{j+}, p_Q^{j-}, respectively, such that

$$\left\|T_{\tilde{a}} \tilde{p}_Q^{j+} - \varepsilon e_j\right\|_{\mathbb{R}^N}, \left\|T_{\tilde{a}} \tilde{p}_Q^{j-} + \varepsilon e_j\right\|_{\mathbb{R}^N} \leq \frac{\varepsilon}{4\sqrt{N}}.$$

Now, consider the convex hull $\mathcal{D}(\tilde{a})$ of $\left\{T_{\tilde{a}} \tilde{p}_Q^{j+}, T_{\tilde{a}} \tilde{p}_Q^{j-}\right\}_{j=1}^N$. We can show that 0 must belong to the interior of $\mathcal{D}(\tilde{a})$. To see that, take any unit vector $u \in \mathbb{R}^N$ and suppose that $|u_j| \geq 1/\sqrt{N}$ for some j. If $u_j > 0$, then

$$u \cdot T_{\tilde{a}} \tilde{p}_Q^{j+} = u \cdot \varepsilon e_j + u \cdot \left(T_{\tilde{a}} \tilde{p}_Q^{j+} - \varepsilon e_j\right) = \varepsilon u_j + u \cdot \left(T_{\tilde{a}} \tilde{p}_Q^{j+} - \varepsilon e_j\right)$$

By the Cauchy–Schwarz inequality, we have

$$\left|u \cdot \left(T_{\tilde{a}} \tilde{p}_Q^{j+} - \varepsilon e_j\right)\right| \leq \|u\|_{\mathbb{R}^N} \left\|T_{\tilde{a}} \tilde{p}_Q^{j+} - \varepsilon e_j\right\|_{\mathbb{R}^N} \leq \frac{\varepsilon}{4\sqrt{N}}$$

and hence

$$u \cdot T_{\tilde{a}} \tilde{p}_Q^{j+} = \varepsilon u_j + u \cdot \left(T_{\tilde{a}} \tilde{p}_Q^{j+} - \varepsilon e_j\right) \geq \frac{\varepsilon}{\sqrt{N}} - \frac{\varepsilon}{4\sqrt{N}} = \frac{3\varepsilon}{4\sqrt{N}}$$

Similarly, we have

$$u \cdot T_{\tilde{a}} \tilde{p}_Q^{j-} = -\varepsilon u_j + u \cdot \left(T_{\tilde{a}} \tilde{p}_Q^{j-} + \varepsilon e_j\right) \leq -\frac{\varepsilon}{\sqrt{N}} + \frac{\varepsilon}{4\sqrt{N}} = -\frac{3\varepsilon}{4\sqrt{N}}$$

When $u_j < 0$, all of the above is reversed. This means that for any unit vector, there are two members $x, y \in \mathcal{D}(\tilde{a})$ such that $u \cdot x < 0 < u \cdot y$. Hence, 0 cannot be on the boundary or outside of $\mathcal{D}(\tilde{a})$ – by the separating hyperplane theorem. Thus, there exists $r > 0$ for which $B_r(0) \subset \mathcal{D}(\tilde{a})$.

Next, we use Assumption 1 to establish that result. To see this, we first describe how to find

another interim price close to $\tilde{p}_Q^{j\pm}$ in $\mathcal{P}_a$ for a close to $\tilde{a}$. The first complication is that a change in effort changes the set of feasible interim prices. To find a nearby interim price, consider the following perturbed interim price – explicitly being a function of a

$$\hat{p}_Q^{j\pm}(i;a) = \eta \times \left(\tilde{p}_Q^{j\pm}(i) - \int_0^1 v_Q(i;\tilde{a})\, di \right) + \int_0^1 v_Q(i;a)\, di$$

where $\eta \in (0,1)$ is to be chosen. Since $\tilde{p}_Q^{j+}, \tilde{p}_Q^{j-} \in \mathcal{P}_{\tilde{a}}$, the above implies that $\hat{p}_Q^{j\pm}(\cdot;a)$ has the same mean as $v_Q(\cdot;a)$, and $\hat{p}_Q^{j\pm}(i;a)$ is monotone in i. Moreover, for $\tilde{a}$, $\hat{p}_Q^{j\pm}(i;\tilde{a}) \in \mathcal{P}_{\tilde{a}}$, and the majorization constraint never binds at interior quantiles. For any a, we have

$$\begin{aligned}
&\int_0^{\frac{k}{m}} \left[\hat{p}_Q^{j\pm}(i;a) - v_Q(i;a) \right] di = \\
&\eta \int_0^{k/m} \left[\tilde{p}_Q^{j\pm}(i) - \int_0^1 v_Q(\cdot;\tilde{a}) \right] di \\
&+\frac{k}{m} \int_0^1 v_Q(\cdot;a) - \int_0^{k/m} v_Q(i;a)\, di
\end{aligned}$$

The first term is independent of a and non-positive, since $\tilde{p}_Q^{j\pm}$ is increasing, while the second term is non-negative. Since the expression is non-negative at $\eta = 1$ and $a = \tilde{a}$, for $\eta < 1$, it is at least $(1-\eta)$ times the second term at $\tilde{a}$ minus the change in the second term between $\tilde{a}$ and a. Since $0 \in \text{int} T_{\tilde{a}} \mathcal{P}_{\tilde{a}}$, $v_Q(\cdot;\tilde{a})$ is not constant, so the second term at $\tilde{a}$ is positive for $0 < k < m$. By Assumption 1, parts 2 and 4, the second term is continuous in a, and there are finitely many k. Hence, for any $\eta < 1$, there exists a neighborhood $B(\tilde{a})$ of $\tilde{a}$ on which the above is non-negative for all k. Note that since $\hat{p}_Q^{j\pm}(i;a)$ is constant on each interval $\left[\frac{k}{m}, \frac{k+1}{m}\right]$, $\int_0^t \left[\hat{p}_Q^{j\pm}(i;a) - v_Q(i;a) \right] di$ is, as a function of t, the difference between a linear and a convex function over $\left(\frac{k}{m}, \frac{k+1}{m}\right)$. Thus, it is concave, and since it is non-negative at the endpoints of the interval, it is non-negative on the whole interval. Therefore, $\hat{p}_Q^{j\pm}(i;a) \in \mathcal{P}_a$.

For $a \in B(\tilde{a})$, let $\mathcal{C}(a)$ be the convex hull of $\left\{ T_a \hat{p}_Q^{j+}(\cdot;a), T_a \hat{p}_Q^{j-}(\cdot;a) \right\}_{j=1}^N$. Given the construction of $\hat{p}_Q^{j\pm}(i;a)$, we have that

$$\begin{aligned}
T_{\tilde{a},n} \hat{p}_Q^{j\pm}(\cdot;\tilde{a}) &= -\eta \int \frac{\partial F}{\partial \hat{a}_n}(i|\tilde{a};\tilde{a})\, d\tilde{p}_Q^{j\pm}(i) - \frac{\partial c(\tilde{a})}{\partial a_n} \\
&= \eta T_{\tilde{a},n} \tilde{p}_Q^{j\pm} - (1-\eta) \frac{\partial c(\tilde{a})}{\partial a_n}
\end{aligned}$$

As a result,

$$\mathcal{C}(\tilde{a}) = \eta \mathcal{D}(\tilde{a}) - (1-\eta) \nabla_a c(\tilde{a})$$

Since $B_r(0) \subset \mathcal{D}(\tilde{a})$, $\mathcal{C}(\tilde{a})$ contains the ball of radius $r' = \eta r - (1-\eta) \|\nabla_a c(\tilde{a})\|$ around 0, and $r' > 0$ for η close to 1. We fix such an η and the corresponding $B(\tilde{a})$.

By Assumption 1, $T_a \hat{p}_Q^{j\pm}(\cdot;a)$ is continuous in a. Indeed, $T_a \hat{p}_Q^{j\pm}(\cdot;a) = \eta T_a \tilde{p}_Q^{j\pm} - (1-\eta) \nabla_a c(a)$,

where $T_a\tilde{p}_Q^{j\pm}$ is given by the same formula, and since $\tilde{p}_Q^{j\pm}$ jumps only at the points k/m,

$$\int \frac{\partial F}{\partial \hat{a}_n}(i|a;a)\, d\tilde{p}_Q^{j\pm}(i) = \sum_{k=1}^{m-1}\left[\tilde{p}_Q^{j\pm}\left(\tfrac{k}{m}+\right) - \tilde{p}_Q^{j\pm}\left(\tfrac{k}{m}-\right)\right]\frac{\partial F}{\partial \hat{a}_n}\left(\tfrac{k}{m}|a;a\right),$$

which is continuous in a by continuity of $\nabla_a g, \nabla_a G$ and $\nabla_a c$. Shrinking $B(\tilde{a})$ so that each $T_a\hat{p}_Q^{j\pm}(\cdot;a)$ is within $r'/2$ of $T_{\tilde{a}}\hat{p}_Q^{j\pm}(\cdot;\tilde{a})$, for every unit vector u some vertex x of $\mathcal{C}(a)$ satisfies $u\cdot x \geq r'/2$. Hence, $0 \in \text{int}\mathcal{C}(a)$. Finally, $\mathcal{C}(a) \subset T_a\mathcal{P}_a$ since $\mathcal{P}_a$ is convex and T_a is affine, so $a \in U$. This concludes the proof. □

F Extension of Kleiner et al. (2021) to Integrable Functions

In this section, we provide a lemma that allows us to extend the results of Kleiner et al. (2021) to integrable functions.

Lemma 9. *Let $v_Q(i) \in L_1([0,1])$ be monotone increasing and let P be the set of non-decreasing functions $p_Q : [0,1] \to \mathbb{R}$ with $p_Q \succeq_{\text{c.v.}} v_Q$. Then P is a convex and compact subset of $L_1([0,1])$.*

Proof. Convexity is evident. We prove the compactness in four steps:

Step 1. Uniform boundedness in $[\varepsilon, 1-\varepsilon]$. Consider $p_Q \in P$. Since p_Q is increasing and $p_Q \succeq_{\text{c.v.}} v_Q$, majorization implies that

$$xp_Q(x) \geq \int_0^x p_Q(i)\, di \geq \int_0^x v_Q(i)\, di,$$
$$\int_x^1 v_Q(i)\, di \geq \int_x^1 p_Q(i)\, di \geq (1-x)\, p_Q(x),$$

for all $x \in [\varepsilon, 1-\varepsilon]$. In other words,

$$\frac{\int_0^x v_Q(i)\, di}{x} \leq p_Q(x) \leq \frac{\int_x^1 v_Q(i)\, di}{1-x}$$

Since v_Q is integrable, the above implies that $p_Q(x)$ is uniformly bounded on $[\varepsilon, 1-\varepsilon]$.

Step 2. Uniform integrability. Let $K > 0$ and consider the function that maps x to $-\max\{|x| - K, 0\}$. This is a concave function, and since $p_Q \in P$ satisfies $p_Q \succeq_{\text{c.v.}} v_Q$, we must have that

$$\int_0^1 \max\{|p_Q(i)| - K, 0\}\, di \leq \int_0^1 \max\{|v_Q(i)| - K, 0\}\, di$$

Note that as $K \to \infty$, the RHS of the above converges to 0, and at $K = 0$, we have

$$\int_0^1 |p_Q(i)|\, di \leq \int_0^1 |v_Q(i)|\, di < \infty$$

which implies a uniform bound on $\int |p_Q|\, di$ for all $p_Q \in P$.

Step 3. Pointwise convergence. Consider a sequence $\{p_{Q,n}\} \subset P$. From Step 1, we know that, restricted to any $[1/k, 1-1/k]$ with $k \geq 3$, the sequence of functions is uniformly bounded, and thus by Helly's selection theorem, see for example Theorem 25.9 in Billingsley (1995), it must have a subsequence that is pointwise convergent for almost all $i \in [1/k, 1-1/k]$. For k, given the considered sequence so far, we extract a further subsequence that converges almost everywhere on $[1/k, 1-1/k]$. If we let n_k be the k-th index of the k-th subsequence, $\{p_{Q,n_k}\}$ converges almost everywhere on $(0,1)$. This means that there exists $\hat{p}$ such that $p_{Q,n_k}(i) \to \hat{p}(i)$ for almost all $i \in (0,1)$ with $\hat{p}$ monotone increasing.

Step 4. Convergence in $L_1([0,1])$ and P. The subsequence above converges almost everywhere in $[0,1]$ and is uniformly integrable. This means that, by Vitali's convergence theorem, see Section 4.6 in Royden and Fitzpatrick (2010), the convergence holds in $L_1([0,1])$ and $\hat{p} \in L_1([0,1])$. Since each $p_{Q,n_k} \succeq_{\text{c.v.}} v_Q$ and we have L_1 convergence, we must have that $\int_0^x p_{Q,n_k}(i)\, di \to \int_0^x \hat{p}(i)\, di$ for all $x \in [0,1]$, which implies that $\hat{p} \in P$. This implies that P is sequentially compact, and this concludes the proof.

□

Lemma 9 implies that Proposition 1 in Kleiner et al. (2021) holds for integrable functions. Since the proof of Theorem 1 in Kleiner et al. (2021) does not use the boundedness assumption, it will hold verbatim for integrable functions as well.