Optimal Rating Design under Moral Hazard*

Maryam Saeedi † Ali Shourideh ‡

August 10, 2026

Abstract

We study optimal 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. Whether optimal ratings reveal or censor information depends on how effort shifts the outcome distribution: 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 with window dressing, a monotone relative informativeness property again delivers upper- or lower-censorship ratings. In an application to redistributive test design, optimal tests can feature mid-censorship. These results follow from a general characterization of optimal ratings via concavification of a gain function, accommodating violations of monotone likelihood ratios and distributional concerns.

Keywords: Information Design, Moral Hazard, Window Dressing, Manipulation, 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 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 welfare 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 where a platform has concerns for fairness or redistribution.[1]

Our main results shed light on optimal provision of incentives for one- and multi-dimensional effort. First, we show how interactions between the distribution of the indicators 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, with the censored tail determined by the relative informativeness of the indicator about the two efforts.

The main technical challenge in formulating the optimal rating problem is the interplay between information structures and the agent incentive constraints. The main object of interest is the interim payoff of the agent who 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 market's posteriors—as the sufficient statistic that determines incentives from the agent's perspective. 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.[2] 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 information design problem. Our key theoretical result, Proposition 1, shows that when these interim prices are comonotone with market values, then 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, 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—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. When the intermediary's objective is to maximize effort (absent distributional motives), the design problem reduces to finding the rating that achieves the highest level of effort. 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 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

[2] 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 senders – 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.

downside risk, i.e., 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 and under CLRP, optimal ratings are upper censorship, pooling high realizations while revealing low ones, which punishes poor outcomes and encourages variance-reducing 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 extreme high outcomes are not especially "earned" by 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 affecting market values) which differentially impact the observed indicator and market value. An important determinant of optimal ratings is whether the technology is reducible, i.e., the distribution of the indicator is governed by a single function of effort.[3] When this is the case, the indicator is not differentially informative of the two actions. The multi–task model, then, is *reducible* to the single task model and results from Section 3 apply.

[3]Mathematically, this is equivalent to the set of equilibrium actions for arbitrary ratings having dimension one.

When the multi–task model is *irreducible*, different values of the indicator have different informational content about the two actions undertaken by the agent. Thus, our analysis identifies a condition on the information technology, what we refer to as the *monotone relative informativeness property* (MRIP). Under MRIP, the informativeness of the indicator about productive effort relative to window dressing is monotone across realizations; that is, high realizations of the indicator are either always more informative of window dressing or of productive activity. Our main result is that with MRIP, welfare-maximizing ratings are either upper or lower censorship. In a class of location–scale technologies, 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 indicators. Notably, censorship requires no deviation from MLRP: even when the indicator is informative about each effort in the standard way, the optimal rating censors, since the designer attaches weights of opposite signs to the two incentive constraints.

In the final section of our paper and to illustrate the range of problems that can be addressed by our formulation, we apply our framework to the problem of redistributive test design with heterogeneous students, showing that optimal tests may involve "mid-censorship" to balance incentive provision with redistributive motives towards disadvantaged groups.

Our analysis thus 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 fresh start which could be interpreted as lower censorship; others such as Airbnb have too many high ratings (see for example Zervas et al. (2021)) that can be interpreted as upper censorship. Our results suggest these differences may reflect optimal responses to underlying differences in how effort affects outcome distributions as we discussed above; Table 1 in Section 3 summarizes the resulting taxonomy. More broadly, the paper provides a toolkit for evaluating rating policies across domains—from ESG certification to educational testing to online marketplaces—by connecting observable 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 by the choice of the information structure (e.g., Frankel and Kartik (2019), Ball (2025), and Perez-Richet and Skreta (2022)). In contrast with 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. 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. Our analysis, thus, identifies both the precise shape of the optimal information structure and when it is optimal to use uncertain rating systems. In our model, the presence of window-dressing incentives is similar to the falsification model in Perez-Richet and Skreta (2022). The main difference with 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 concern model where the information structure only affects the agent's incentive. 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. The two papers are different in scope and techniques. 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 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

[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.

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 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 and 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 and 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. Proofs are relegated to the 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," who 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 $c(a)$. This action 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 $G(y|a)$.

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.[6] Throughout the paper we impose the following monotonicity assumption:

Assumption 1. *For all (deterministic) market beliefs $\hat{a}$, the market value $\overline{v}(y; \hat{a})$ is increasing in the indicator y.*

This assumption states that market values are ranked based on the values of the indicator. Without any other assumption on the distribution function $G(y|a)$, e.g., first-order stochastic dominance or MLRP, this assumption is innocuous as one can always relabel the values of the indicator according to the market values. While our main characterization results – Proposition 1 and Theorem 1 – hold without Assumption 1, we maintain this assumption for tractability and convenience.

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 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 $\hat{a}$.[7] This expectation is calculated using the information available,

[6] Throughout the paper, we focus on equilibria in which the agent plays a pure effort strategy. Our main characterization results, Proposition 2 and Theorem 1, hold when allowing for mixed effort strategy by the agent.

[7] 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.

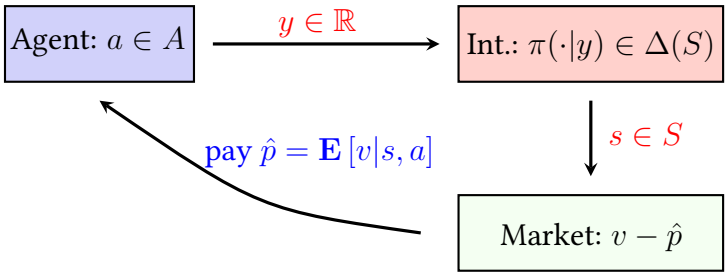


Figure 1: General structure of the model

s, and the common belief about equilibrium play.[8] 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 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).

We now turn to the intermediary's problem. 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.

The intermediary chooses an information structure to maximize an objective that may differ

[8] 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 as a compact subset of some Euclidean space, and $\mathscr{S}$ as the Borel σ-algebra associated with topology induced by the Euclidean norm S. Henceforth, we drop references to σ-algebra in our analysis. Additionally, when describing subsets, we refer to Borel subsets.

from total surplus. We consider a class of objectives in the form

$$W(a) + \int p(y)\,\alpha(y)\,dG(y|a)\,, \tag{3}$$

where $W(a)$ captures externalities or direct preferences over effort, and $\alpha(y) \geq 0$ represents distributional weights on agent payoffs. This class of objective functions fits various examples: the term $W(a)$ captures direct externalities—as in the ESG and career-concerns examples—so that $W(a) - V(a)$, with $V(a) = \mathbb{E}[v|a] - c(a)$ the total surplus, represents the external effects that are not captured by the market; the weights $\alpha(y)$ capture distributional concerns, such as a platform aiming to guarantee a minimum payoff level for its sellers or a college reweighting test outcomes for its admission policies.

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); we develop this application in Section 4. Finally, when market values vary with the agent's action, as in models of career concerns, equilibrium effort is inefficient and rating design shapes this inefficiency. We characterize welfare maximizing ratings in settings with a single or multiple tasks.

2.1 Interim Prices: Definition and Characterization

In this section, we introduce a mathematical object, *interim prices*, that allows us to simplify the problem of rating design in the environment described above. The notion of interim price is simple. This mathematical object determines the agent's incentives in choice of effort and will be present in the incentive constraints for the agent. Specifically, we define *interim prices* as

$$p(y) = \int \mathbb{E}[v|s]\,d\pi(s|y)\,. \tag{4}$$

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.[9] It can also be interpreted as the agent's "second-order belief": their belief about the belief of the market on values.

Critically, interim prices are a sufficient statistic for the information structure from the agent's

[9]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.

perspective. Specifically, for any choice of a, the agent's payoff is given by

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

Thus, 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 $p \succeq_{\text{cv}} \overline{v}$, i.e., p is a mean preserving contraction of $\overline{v}$.[10] Intuitively, a rating can 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 existence of ratings. Below we show that this is indeed true when interim prices and market values are comonotone.[11] 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. *$p \succeq_{\text{cv}} \overline{v}$.*

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

Proposition 1 is the main technical tool that allows us to significantly simplify the problem of rating design into the constrained concavification in Theorem 1. The main challenge is the

[10] The relation $p \succeq_{\text{cv}} \overline{v}$ represents the concave order which implies that for all concave functions $\phi : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[\phi(p)] \geq \mathbb{E}[\phi(\overline{v})]$. 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.

[11] In 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.

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 comonotonicity to construct a sequence of garblings each of which is a randomization between full revelation and pooling of consecutive states.[12]

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

$$a \in \arg\max_{\hat{a}\in A} \int p(y)\, dG(y|\hat{a}) - c(\hat{a}) \tag{5}$$

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.

Remark on Comonotonicity In the rest of the paper, we focus on ratings that satisfy this 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.[13]

2.2 The Main Characterization

Given this class of objectives and the comonotonicity restriction, the problem of optimal rating design can be stated as maximizing the objective in (3) subject to incentive compatibility (5), 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,

$$v_Q(i) = \overline{v}(G^{-1}(i|a); a). \tag{6}$$

For ease of notation, we have dropped a from the expression of quantile value. Similarly, let $p_Q(i)$

[12] In our proof, we first establish the result when there are finitely many market values and then approximate arbitrary distribution of market values by discretizing the values and use compactness arguments.

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

be the quantile representation of the interim price. Given the comonotonicity assumption 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 unconstrained objective 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 the 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 (5) requires that a maximizes $-\int_0^1 F(i|\hat{a};a)\,dp_Q(i) - c(\hat{a})$ over $\hat{a} \in A$.

The First Order Approach. In order to sidestep many of the complications that typically arise in moral hazard problems, we will use the first order approach (FOA) throughout the paper. That is, we replace the incentive constraint (5) 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 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. function $G(y|a)$ with respect to a to that of the cost function $c(a)$ and are typically satisfied provided that $c(a)$'s are sufficiently convex.[14]

The following provides a full characterization of optimal ratings under FOA:

Theorem 1. *If w^* is the highest value of the objective (3) and under the validity of FOA, there exists a real vector $\lambda \in \mathbb{R}^N$ such that*

$$w^* = \max_a W(a) + \int cav\Gamma(i;\lambda,a)\,dv_Q(i) \qquad \text{(D)}$$
$$\text{s.t. } i = G\left(\{y : \overline{v}(y;a) \leq v_Q(i)|a\}\right)$$

where

$$\Gamma(i;\lambda,a) = \int_{\{y:\overline{v}(y;a)>v_Q(i)\}} \alpha(y)\,dG - \sum_{n=1}^{N} \lambda_n \left[\frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\Big|_{\hat{a}=a} + \frac{\partial}{\partial a_n} c(a)\right]$$

[14] The literature on verification of 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).

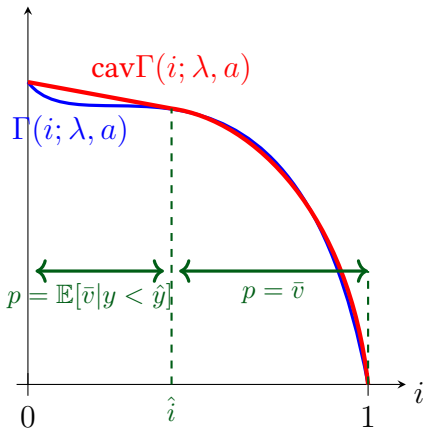


Figure 2: An illustration of the concavification of the gain function Γ and construction of optimal ratings

We refer to Γ as the gain function. It summarizes the weight that the intermediary puts on a particular realization of the indicator and the associated IC conditions. As it is demonstrated in Figure 2, the concave envelope of the gain function determines the optimal ratings. When it coincides with Γ, optimal rating should be fully revealing while over intervals where the concave envelope is linear, the optimal rating is pooling the realizations within that interval. Theorem 1 implies that the rating design problem can be solved by solving a one-dimensional concavification problem of the gain function and then finding optimal values of effort and the multipliers associated with the incentive compatibility constraints (5).

Optimality of Threshold Based Ratings

The first implication of Theorem 1 is that optimal rating systems are simple. In fact, since the function to be *concavified* is only a function of the quantile i, by Caratheodory theorem any convex combination of values of $\Gamma(i;\lambda,a)$ can be achieved by using at most two points. This logic establishes that the optimum 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.

While the optimum value in (D) is always achieved by an interim price associated with a deterministic monotone partitional signal, such a rating should also satisfy the incentive compatibility. To ensure that this is indeed possible, we make the following assumption on the distribution function $G(y|a)$:

Assumption 2. ***Independence.*** *For all $a \in A$:*

1. *$G(y|a)$ has full support over a convex subset of $\mathbb{R}$.*

2. *For any interval $I \subset Supp G(y|a)$, the function $\alpha(y)\, g(y|a)$ cannot be written as a non–zero*

linear combination of $g(y|a)$, $\left\{\frac{\partial g(y|a)}{\partial a_n}\right\}_{n=1}^{N}$ *for all values of* $y \in I$.

The independence assumption ensures that there is enough variation in y conditional on effort a so that only coarse moments of $p(y)$ matter for incentives.[15] An example of a family that satisfies the above is when $\alpha = 0$ and G is the scale–location family $G(y|a) = H\left(\frac{y-\mu(a)}{\sigma(a)}\right)$ with H having a support of the entire real line with density $h(\varepsilon)$ such that $h(\varepsilon)$, $h'(\varepsilon)$, and $\varepsilon h'(\varepsilon)$ are linearly independent. Given Assumption 2, Theorem 1, and Lemma 1 (in the Appendix), we have the following proposition:

Proposition 2. *Suppose that Assumption 2 holds. Then the optimal interim price in (D) is always associated with a deterministic monotone partitional rating. Moreover, whenever* $cav\Gamma(i;\lambda,a) = \Gamma(i;\lambda,a)$, *optimal rating reveals the value* $\overline{v} = v_Q(i)$ *to the market. When* $cav\Gamma(i;\lambda,a) > \Gamma(i;\lambda,a)$, *then there exists an interval* $i \in [i_1, i_2]$ *such that optimal rating reveals that* $\overline{v} \in [v_Q(i_1), v_Q(i_2)]$.

Theorem 1 and Proposition 2 together provide a full characterization of optimal ratings. In what follows, we provide general properties of optimal ratings with single and multiple efforts/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 function 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. Table 1 at the end of this section summarizes the resulting taxonomy. In this part, we focus on problems in which effort is one dimensional. In Section 4, we study multiple tasks.

Targeting An Action

When does an intermediary that wishes to maximize effort benefit from concealing information? To answer this question, we start our analysis by considering objectives that only target an action, i.e., $\alpha(y) = 0$ in (3). In this case, the problem of solving optimal rating design boils down to a characterization of the set of implementable efforts. As we have shown, an effort $a \in A$ is implementable when there exists an interim price function $p(y)$ such that $p(y)$ is a mean-preserving contraction of market values $\overline{v}(y;a)$ and a is incentive compatible given $p(y)$.

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

[15] An example that violates Assumption 2 is one in which $y = \overline{y}(a)$, an increasing function of a. In this case, any change in the interim price function affects the incentives of the agent. In a previous version of this paper, we have established that if Assumption 2 is violated, optimal ratings can involve randomization.

Assumption 3. *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 $I = \left[\underline{y}, \overline{y}\right]$ and $g(y|a)$ is twice differentiable.*
3. *Cost function $c(a)$ is non-negative, strictly convex, increasing and twice differentiable for all $a \in A$.*
4. *For any effort, $a \in A$, $\overline{v}(y; a)$ is increasing in y.*

The first three parts of Assumption 3 are fairly common in the moral hazard literature. The last assumption notably implies that y is an indicator that is positively correlated with market values. This assumption on its own is innocuous since the indicator y itself is not payoff relevant.

By Theorem 1, under FOA, the optimal rating is found by a concavification of the function $\Gamma(i; \lambda, a) = -\lambda \left[\frac{\partial F(i|\hat{a};a)}{\partial \hat{a}} \Big|_{\hat{a}=a} + c'(a) \right]$ where λ is the Lagrange multiplier on 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 $G(y|a)$:

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

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 implication for optimal rating.

Let us start from 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 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.[16]

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

[16] 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 1C.

Proposition 3 (Full Disclosure under MLRP)**.** *Suppose Assumptions 2 and 3 hold, FOA is valid, and $g(y|a)$ satisfies MLRP. Then the highest implementable effort is associated with interim price $p(y) = \overline{v}(y; a)$. 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 indicators.

We should note that the definition of highest implementable effort a_{FI} involves 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 straight 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, 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, 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, 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:

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

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

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 $y/a \leq e^{-1/\gamma}$.

A version of these examples, with mean parameter a in place of $\log a$, is depicted in Figure 3. As can be seen, in 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 single crossing of MLRP.

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

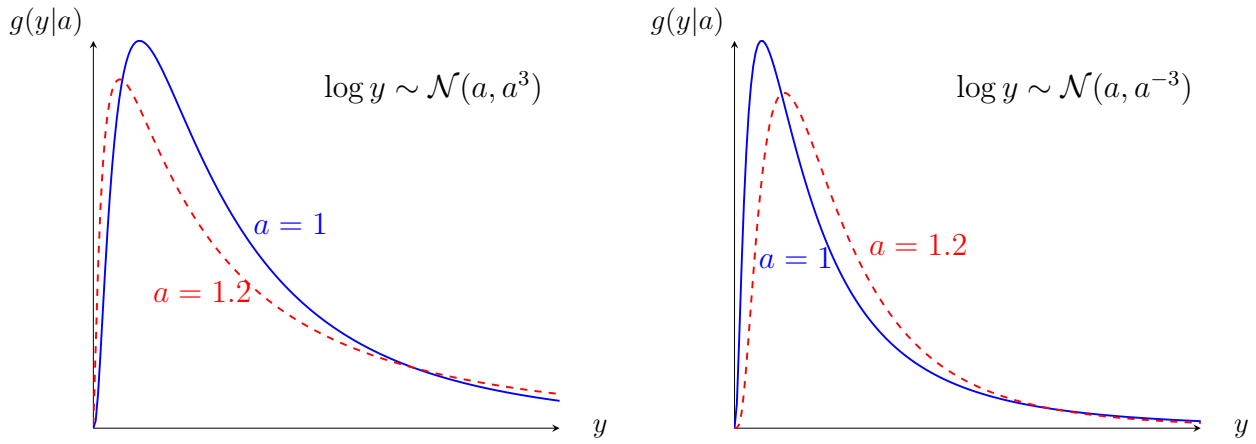
Proposition 4 (Censorship under ELRP and CLRP)**.** *Suppose Assumptions 2 and 3 hold and FOA is valid.*

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.*

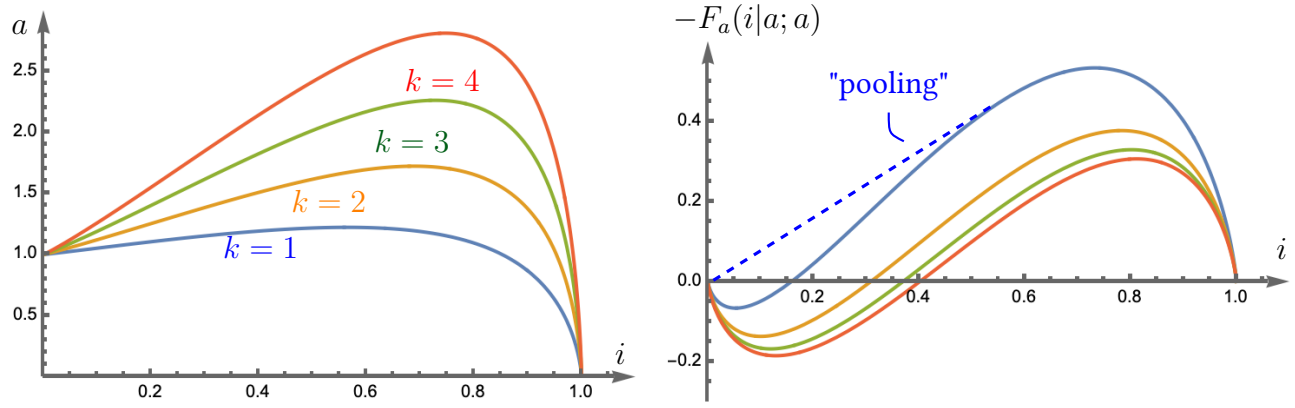
As the above proposition establishes, optimal ratings for ELRP and CLRP distributions are fairly simple. They involve either lower censorship (in case of ELRP) or upper censorship (in case of CLRP). In what follows we provide an example and discuss its implications for various tasks and technologies.

Suppose that $y \sim \mathcal{N}\left(a, (ka)^2\right)$, that y determines market values, i.e., $\overline{v}(y; a) = y$, and that cost is $c(a) = a^2/2$. Under a full information rating, $p(y) = y$, 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 $G(y|a)$ 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



(a) ELRP (b) CLRP

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



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

Figure 4: Optimal efforts for lower-censorship policies

of i lowest realizations of the indicator y. This is depicted in Figure 4 (left panel) for values of $k = 1, 2, 3, 4$. At the lowest value, $i = 0$, optimal effort is $a_{FI} = 1$. Optimal effort peaks at some threshold, 0.56, 0.69, 0.73, 0.75 respectively, and falls to zero as i tends to 1. 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, $-F_a(i|a;a)$, and its concavification (in the case of $k = 1$). The threshold for pooling on the right coincides with the peak of the left plot since optimal ratings take the form of lower censorship.

We should note that 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 $y =$

Technology	Economic interpretation	Optimal rating	Examples
MLRP	Effort shifts outcomes uniformly upward	Full disclosure (Proposition 3)	Routine tasks
ELRP	Effort increases tail risk; innovative activities, R&D	Lower censorship (Proposition 4)	Low-rating forgiveness (Shipt, Instacart)
CLRP	Effort reduces tail risk; maintenance activities	Upper censorship (Proposition 4)	Compressed top ratings (Airbnb)

Table 1: A taxonomy of optimal ratings

$f(\sigma(a)\varepsilon + m(a))$ 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 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.

The above findings point to a practical property of optimal ratings in targeting an action. Namely, how the variance of the indicator interacts with the desired action determines the best way to incentivize it. Specifically, for an activity where more effort leads to a more precise outcome (CLRP), such as maintenance activities, full revelation at low values and pooling at higher values encourages higher effort to avoid punishments for low outcomes. On the other hand, if higher effort leads to a riskier outcome (ELRP), such as innovative activities, then pooling of low realizations via lower-censorship ratings provides insurance against possible downsides and encourages risk taking (increasing effort). When effort does not affect variance (MLRP), optimal rating is full revelation and no pooling is needed.

We should end this section by emphasizing that while we have focused on the highest possible effort that is implementable, any lower value of effort can also be targeted by rating systems. The analysis in this section is specifically 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.

It is useful to conclude this section with a summary of the results. Table 1 collects the taxonomy that emerges from our analysis: the properties of the technology $g(y|a)$ and the objective of the intermediary determine both whether information should be censored and where in the distribution censorship occurs.

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), the multi-task principal-agent models have become the workhorse of analyzing incentives in settings where agents can use several actions to affect the observed outcomes – see also Baker (1992) and Dewatripont et al. (1999).[17] Despite this prevalence of multi–task moral hazard in applied settings and the importance of disclosure/rating policies in it, a unified theoretical analysis of rating policies is essentially non–existent. In this section, 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, a few general lessons will arise and in some special cases, lessons from Section 3 go through.

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 4 below:

Assumption 4 (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 strictly increasing in* a_1 *and strictly decreasing in* a_2*.*
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

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

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

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

[17] Since Dewatripont et al. (1999) analyze a variant of Holmström (1999)'s career concern 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 (2022) for incentives of teachers, and Mayzlin et al. (2014) and Hui et al. (2025) for incentives on online platforms.

where $\beta = \sigma_{vy}/\sigma_y^2 > 0$. Since a_2 must be a window dressing effort, we should have $b_{v,2} = 0, b_{y,2} > 0$.

Throughout the section, our focus will be on welfare maximizing rating design. 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. In Section 4.1, we discuss cases in which it is impossible to differentially affect incentives for productive and window dressing efforts; *reducible* multi–tasking. Section 4.2 then discusses *irreducible* multi–tasking technologies and establishes new general insights in this case.

4.1 Reducible Multi–Tasking

We say that a multi-task technology is *reducible* if the effort vector a affects the distribution of the indicator solely through a one–dimensional index $m(a)$, i.e., $G(y|a) = \hat{G}(y|m(a))$ for some family of distributions $\hat{G}$, such as in the above example. In this case, the multi-task problem collapses to a single–action problem: given any level of the index, the agent chooses the cost-minimizing vector of efforts, and the rating design problem is reducible to a choice of an interim price function $p(\cdot)$ and the index m. As a result, the analysis of Section 3 applies directly and the MLRP principles derived there drive optimal ratings.

Formally, with a reducible technology, one can simply contract on m. This is because the cost of effort is separable from its expected benefit and we can define the following indirect cost function:

$$C(\hat{m}) = \min_{a \in [0,\overline{a}]^2, m(a) = \hat{m}} c(a)$$

and a selection of its solution be $\tilde{a}(m)$. Given this, the problem becomes essentially the same as that in Section 3 in which the intermediary is choosing the index m and interim prices $p(y)$. As a result, we can simply apply the same ideas around MLRP (ELRP and CLRP), in this case applied to $\hat{G}$.

4.2 Irreducible Multi–Tasking

The key benefit of the setup above was that it was reducible to the single effort setup characterized in Section 3. Here we discuss the case in which the technology is not reducible to a single effort and its implication on optimal rating design. The setup is the same as before, i.e., Assumption 4 holds, except for the fact that $G(y|a)$ is no longer reducible. Specifically, let us denote the *score* of task n by

$$\ell_n(y|a) = \frac{\partial g(y|a)}{\partial a_n} / g(y|a).$$

Then the multi–tasking technology is *irreducible* when $\ell_1(y|a)/\ell_2(y|a)$ is *not* independent of y. In this case, different parts of the indicator distribution carry different information about the two tasks, and censorship can be used to discriminate between them.

Before proceeding, it is worth pointing out an important difference between the censorship results below and those of Section 3. In the single effort model, censorship was optimal only when the technology deviated from MLRP, that is, when effort expanded (ELRP) or compressed (CLRP) the distribution of the indicator. In the multi-task model, no such deviation is needed: even when the indicator is informative about each effort in the standard way, the welfare-maximizing rating censors, since the designer attaches weights of opposite signs to the two incentive constraints.

To make this precise, let us apply Theorem 1. Under welfare maximization, the objective of rating design places no distributional weights on the realization of the indicator, i.e., $\alpha(y) = 0$ in (3), and is thus independent of $p(y)$. Moreover, the shape of the optimal rating is determined by a weighted value of the marginal change in the quantiles:

$$\Gamma(i|a) = -\lambda_1 \left.\frac{\partial G\left(G^{-1}(i|\hat{a})\,|a\right)}{\partial a_1}\right|_{\hat{a}=a} - \lambda_2 \left.\frac{\partial G\left(G^{-1}(i|\hat{a})\,|a\right)}{\partial a_2}\right|_{\hat{a}=a}$$

and its concavification. A simple calculation implies that the derivative of Γ with respect to i evaluated at $i = G(y|a)$ is:

$$\Gamma'(i|a)|_{i=G(y|a)} = -\lambda_1 \ell_1(y|a) - \lambda_2 \ell_2(y|a) \tag{7}$$

and thus the local concavity of Γ is determined by whether the above is increasing (convex) or decreasing (concave) in y.

Suppose that $\lambda_1 > 0 > \lambda_2$ which means that it is optimal to discourage a_2 and encourage a_1, which is the natural case to consider. Note that under MLRP, which means that the $\ell_n(y)$'s are increasing in y, the above is the difference of two increasing functions and thus the logic from Section 3 does not hold. Indeed, under MLRP where both $\ell_n(y|a)$'s are increasing, the local concavity of Γ changes exactly where the ratio

$$r(y|a) = \frac{\partial \ell_1(y|a)/\partial y}{\partial \ell_2(y|a)/\partial y}$$

crosses $|\lambda_2|/\lambda_1$. Whether this happens once, and on which side it begins, is what we take up next. To understand this better, consider the following example.

Example 1. Assume:

$$v = a_1(\varepsilon_1 + 1), y = ba_1(\varepsilon_1 + 1) + a_2 + \varepsilon_2$$

with $\varepsilon_1, \varepsilon_2$ being two standard normal and independent random variables. In this case, two increasing functions $\mu(a)$, $\sigma(a)$ exist such that $G(y|a) = \Phi\left(\frac{y-\mu(a)}{\sigma(a)}\right)$ with Φ the c.d.f. of standard

normal.[18] Simple calculations show that in this case, the expression in (7) is the sum of two quadratic functions in y which means that its derivative is linear and thus changes sign only once. In other words, $\Gamma(i|a)$ is either convex–concave or concave–convex. This implies that optimal rating is either upper or lower censorship. This can be seen in terms of

$$r(y|a) = b + 2\frac{y - \mu(a)}{\sigma(a)^2} b^2 a_1,$$

which is linear and monotone in y. This together with ℓ_2 being increasing in y, implies that $\Gamma(i|a)$ is either convex–concave or concave–convex.

The technology in this example satisfies MLRP with respect to both efforts, and involves no expansion or compression of the tails in the sense of Section 3. In the single effort model such a technology calls for full disclosure.

This example is a special case of a more general class of distributions that we define below. The ratio $r(y|a)$ introduced above can be interpreted as the relative informativeness of an increase in y about a_1 relative to a_2. In fact, we can define $I_n(i|a) = -\left.\frac{\partial G\left(G^{-1}(i|\hat{a})|a\right)}{\partial a_n}\right|_{\hat{a}=a}$ and this is the local Lehmann informativeness, Lehmann (1988) and Persico (2000), of an increase in a_n. Simple application of chain rule says that

$$I_n'(i|a) = -\ell_n\left(G^{-1}(i|a)|a\right)$$

which implies that r coincides with I_1''/I_2''.

For $r(y|a)$ to be an appropriate notion of relative informativeness of y for a_1 and a_2, MLRP has to hold. For general irreducible technologies, the following defines the key property used in this section:

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

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

always has the same sign, i.e., it is non–negative or non–positive for all such y_1, y_2 and a.

Using the Lehmann informativeness analogy, one can interpret MRIP as stating that a higher value of y provides uniformly more, or uniformly less information about a_1 relative to a_2. When $G(y|a)$ exhibits MLRP with respect to a_2, i.e., $\ell_2(y|a)$ is increasing in y, G exhibits MRIP when $r(y|a)$ is monotone, which is the case we verified in the example above. Formally, $D(y_1, y_2|a)$ having the same sign is equivalent to $\Gamma''(i|a)$ changing sign at most once for any λ_1, λ_2. Hence, we have the following theorem:

[18] For this example, $\mu(a) = ba_1 + a_2, \sigma(a) = \sqrt{(ba_1)^2 + 1}$.

Theorem 2. *Suppose that $G(y|a)$ is an irreducible technology that exhibits MRIP and FOA is valid. Then welfare maximizing ratings are either upper or lower censorship.*

The significance of Theorem 2 is that it ties a natural notion of informativeness, relative Lehmann informativeness between the productive and the window-dressing efforts, to general properties of welfare–maximizing ratings. We should also note that MRIP is not merely a convenience. When $r(y|a)$ crosses $|\lambda_2|/\lambda_1$ more than once, the local concavity of Γ alternates and its concavification pools more than one interval, which leads to the mid-censorship ratings we study in Section 5.

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 have the following proposition:[19]

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. Then the welfare–maximizing rating is upper censorship if and only if*

$$\frac{\frac{\partial \mu}{\partial a_2}\frac{\partial c}{\partial a_1} - \frac{\partial \mu}{\partial a_1}\frac{\partial c}{\partial a_2}}{\frac{\partial \mu}{\partial a_1}\frac{\partial \sigma}{\partial a_2} - \frac{\partial \sigma}{\partial a_1}\frac{\partial \mu}{\partial a_2}} \geq 0 \tag{8}$$

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 (8) is positive – then when window dressing has a larger impact on the dispersion of the indicator relative to its mean, 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 (8) is negative and the direction of censorship is reversed. We provide the derivations behind Proposition 5 in Appendix B.

[19] MRIP fails for H being Cauchy or Frechet but neither of these have log–concave densities. While MRIP and log–concavity of h are not equivalent and it is possible to come up with examples that are log–concave but not MRIP, among the textbook classes of distributions this is the case.

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 is 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 (8) is reversed, the same logic delivers lower censorship.

The above insight 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 (8), 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, and it is optimal to do so exactly where the indicator mainly reflects manipulation.

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.[20] These observations have prompted a movement to relax the requirement that students take such tests. This includes several universities' policies to make the SAT op-

[20] In 2023, the California Board of Education passed 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.

tional (see Dessein et al. (2025)) and attempts at making mathematics education more accessible and easier. Inspired by this debate, in this section, we provide an alternative answer to this debate in the form of optimal test design. Namely, we will show that our main characterization results in Section 2 can be used to shed light on this question.

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 which 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 where $0 < k_R < k_P$. Finally, let $v = y$ so that the market values are simply the value of the indicators and let $p(y)$ be the interim price function that is increasing.

Consider a rating designer who wishes to maximize the following objective

$$\alpha_P f_P \left[\int_I p(y)\, dG(y|a_P) - k_P c(a_P)\right] + \alpha_R f_R \left[\int_I p(y)\, dG(y|a_R) - k_R c(a_R)\right] \tag{9}$$

where $\alpha_P f_P + \alpha_R f_R = 1$ and $\alpha_P > 1 > \alpha_R$. Let us assume that an increase in a leads to an increase in $g(y|a)$ in the sense of first order stochastic dominance.

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 formulation with weights $\alpha(y)$ on realizations of the indicator in Section 3 can thus be viewed as the reduced form of the problem studied here when the designer cannot condition on types.

The problem of optimal rating design is then to find $p(y)$ and a_R, a_P to maximize the objective in (9) subject to incentive compatibility for both types and 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 constraint. As in the previous sections, we maintain Assumption 2 and the validity of FOA throughout this section.

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

Proposition 6. *Suppose that* $\log g(y|a) = f(y) + r(y)\, m(a) - b(a)$ *where* $r(y), m(a), b(a)$ *are increasing functions and* $m(a) > 0$. *Then the optimal rating that maximizes (9) always has at most four alternating intervals of pooling and revelation.*

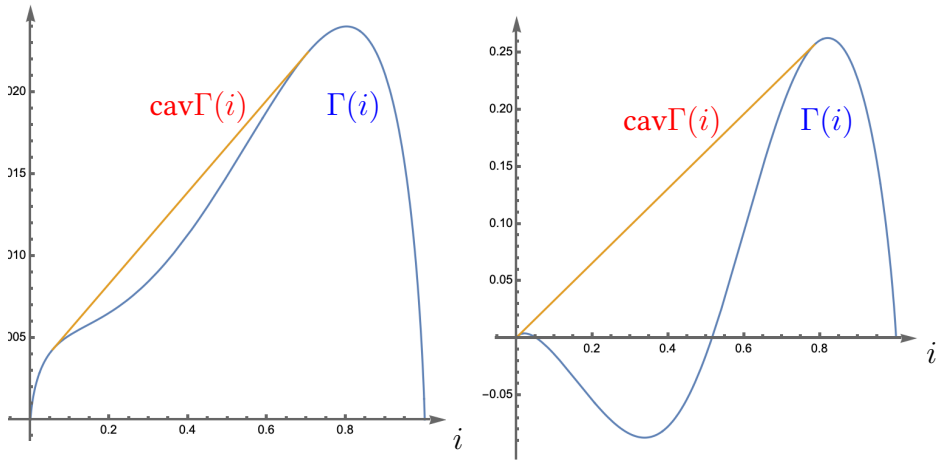
The class of distributions considered in Proposition 6 includes some of the fairly common ones that are used in applied work including 1. a normally distributed y where either the mean or the variance is controlled by the action a, 2. a log–normally distributed y such that a controls the mean of $\log y$, 3. when y is distributed according to an extreme-value distribution of types 1 and 2 (Gumbel and Fréchet) where the scale parameter is controlled by a, among others. Moreover, the assumption implies that $\log g$ is supermodular in (y, a), i.e., it satisfies MLRP.

Interestingly, one might expect redistributive motives to always favor pooling low realizations, since pooling at the bottom shields the disadvantaged type. However, the difference here is that there are two incentive constraints. Under certain conditions on the likelihood function g_a/g for low realizations of y – specifically as it becomes arbitrarily large as $y \to \underline{y}$, 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 2. 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 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 three times 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 5 depicts the optimal rating and the concavification of the gain function described above for different values of costs for type P.[21] In the left panel, the cost of type P is closer to that of type R, $k_P = 3/4$. In this case, optimal rating pools observations approximately between the 6th and 71st percentiles of the y distribution. When the cost for type–P increases to 5/4, the optimal rating pools observations between the 1st and the 78th 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.

Viewed through the lens of the current debate on standardized testing, our analysis suggests an alternative to making tests optional: rather than discarding the information altogether which is done by making the tests optional, a designer with redistributive motives can pool intermediate scores while preserving the incentive effects of the extremes.

[21] In order for the difference between the concavification and the function to be more visible, we subtract $(\alpha_R f_R + \alpha_P f_P)(1 - i)$ 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, $k_P = 3/4$ (b) High cost for type P, $k_P = 5/4$

Figure 5: Determining the optimal rating for Example 2

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 value. This converts rating design into the 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 by 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.

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

Aliprantis, C. D. and K. Border (2013): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer-Verlag Berlin and Heidelberg GmbH & Company KG. 35

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

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. 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

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24, 265–272. 56

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. 13, 47

Conlon, J. R. (2009): "Two new conditions supporting the first-order approach to multisignal principal–agent problems," *Econometrica*, 77, 249–278. 13

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. 27

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

Doob, J. L. (1994): *Measure theory*, Springer Science & Business Media. 36

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

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. 6

Gentzkow, M. and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review*, 106, 597–601. 11, 37

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

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

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. 21

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. 12

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56, 1177–1190. 13, 47

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. 13

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89, 1557–1593. 6, 40

Kolotilin, A. (2018): "Optimal Information Disclosure: A Linear Programming Approach," *Theoretical Economics*, 13, 607–635. 37

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

Lehmann, E. L. (1988): "Comparing location experiments," *The Annals of Statistics*, 16, 521–533. 24

Luenberger, D. G. (1997): *Optimization by Vector Space Methods*, John Wiley & Sons. 40

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, 21

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. 13

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. 6

Persico, N. (2000): "Information acquisition in auctions," *Econometrica*, 68, 135–148. 24

Rayo, L. (2013): "Monopolistic Signal Provision," *The B.E. Journal of Theoretical Economics*, 13, 27–58. 7

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton University Press. 40

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. 13

Rothschild, M. and J. E. Stiglitz (1970): "Increasing Risk: I. A Definition," *Journal of Economic theory*, 2, 225–243. 11, 56

Shaked, M. and J. G. Shanthikumar (2007): *Stochastic Orders*, Springer Science & Business Media. 11, 16

Strassen, V. (1965): "The Existence of Probability Measures with Given Marginals," *The Annals of Mathematical Statistics*, 36, 423–439. 56

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

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

A Proofs

A.1 Proof of Proposition 1

Proof. The if side of the statement, that is if p is constructed from some information structure (π, S) then $p \succcurlyeq_{\text{cv}} \overline{v}$ is immediate from the text. For the reverse, we will first prove the proposition

when the support of the indicator Y is finite. We will then show that the proposition can be extended to the case when Y is a compact Euclidean space.

1. When Y is finite. Let $Y = \{y_1, y_2, \cdots, y_n\}$ such that $\overline{v}(y_1) \leq \overline{v}(y_2) \leq \cdots \leq \overline{v}(y_n)$. When $p(y)$ is co-monotone with $\overline{v}(y)$, we must have that $p(y_1) \leq p(y_2) \leq \cdots \leq p(y_n)$. In this case, $\mathcal{S} \subset \mathbb{R}^n$. For simplicity, we also let $f_i = \mu_y(\{y_i\})$ and $\overline{v}_i = \overline{v}(y_i)$ and $p_i = p(y_i)$.

We prove the claim by induction on n.

First step: The claim holds for $n = 2$.

If $n = 2$, then majorization implies that

$$f_1\overline{v}_1 + f_2\overline{v}_2 = f_1p_1 + f_2p_2,$$
$$0 \leq p_2 - p_1 \leq \overline{v}_2 - \overline{v}_1.$$

Consider a signal structure that only sends a low signal when the state is y_1. When the state is y_2, it sends the low signal with probability α and otherwise a high signal. Let

$$0 \leq \alpha = \frac{p_1 - \overline{v}_1}{\overline{v}_2 - p_1}\frac{f_1}{f_2} \leq 1.$$

Then, the ex post price upon observing the low signal is

$$\frac{\overline{v}_1f_1 + \alpha\overline{v}_2f_2}{f_1 + \alpha f_2} = \frac{\overline{v}_1f_1 + \frac{p_1-\overline{v}_1}{\overline{v}_2-p_1}f_1\overline{v}_2}{f_1 + \frac{p_1-\overline{v}_1}{\overline{v}_2-p_1}f_1} = \frac{\overline{v}_1f_1(\overline{v}_2 - p_1) + (p_1 - \overline{v}_1)f_1\overline{v}_2}{f_1(\overline{v}_2 - p_1) + (p_1 - \overline{v}_1)f_1} = p_1.$$

Thus the interim price when the state y_1 is p_1. The mean of p_1 and p_2 is the same as that of $\overline{v}_1, \overline{v}_2$ which means that p_2 is the second order expectation of $\overline{v}$ when the state is y_2. This proves the claim.

Second Step: If the claim is true for $n-1$, then it is true for n.

Consider an interim price function $\{p_i\}$. If for some i, $p_i = p_{i+1}$, we can reduce the number of states by considering the interim price function $p_1 \leq \cdots \leq p_i \leq p_{i+2} \leq \cdots \leq p_n$ distributed according to $f_1, \cdots, f_{i-1}, f_i + f_{i+1}, f_{i+2}, \cdots, f_n$ and market values of $\overline{v}_1 \leq \cdots \leq \overline{v}_{i-1} \leq \frac{f_i\overline{v}_i + f_{i+1}\overline{v}_{i+1}}{f_i+f_{i+1}} \leq \overline{v}_{i+2} \leq \cdots \leq \overline{v}_n$. Since by the induction assumption, an information structure π exists that generates this interim price function, simply pooling y_i and y_{i+1} and using the information structure π generates the original interim price function.

Suppose, on the other hand, that $p_i < p_{i+1}$ for all i. Let $\lambda = \min_{i \leq n-1} \frac{p_{i+1}-p_i}{\overline{v}_{i+1}-\overline{v}_i}$. Let $\hat{p}_j$ be defined by

$$\hat{p}_j = \frac{p_j - \lambda\overline{v}_j}{1-\lambda}$$

Then $(1-\lambda)(\hat{p}_{j+1} - \hat{p}_j) = p_{j+1} - p_j - \lambda(\overline{v}_{j+1} - \overline{v}_j) \geq 0$. Moreover,

$$\sum_{j=1}^{k} f_j(\hat{p}_j - \overline{v}_j) = \sum_{j=1}^{k} f_j\frac{p_j - \overline{v}_j}{1-\lambda} \geq 0,$$

and, finally, if $\lambda = (p_{i+1} - p_i)/(\overline{v}_{i+1} - \overline{v}_i)$, then $\hat{p}_i = \hat{p}_{i+1}$. This implies that an argument

similar to the above shows that an information structure $\left(\hat{\pi}, \hat{S}\right)$ exists that generates $\hat{p}$ as its interim price. Now consider an information structure that reveals the state with probability λ and otherwise it is the same as $\hat{\pi}$. That is

$$\pi\left(s|y_j\right) = \begin{cases} \lambda & s = y_j \\ \left(1-\lambda\right)\hat{\pi}\left(s|y_j\right) & s \in \hat{S}. \end{cases}$$

Then, since the set of signals that reveal the state does not overlap with $\hat{S}$, we must have that

$$\sum_s \pi\left(s|y_j\right)\frac{\sum_i \pi\left(s|y_i\right) f_i \overline{v}_i}{\sum_i \pi\left(s|y_i\right) f_i} =$$

$$\sum_{s\in\hat{S}} \left(1-\lambda\right)\hat{\pi}\left(s|y_j\right)\frac{\sum_i \hat{\pi}\left(s|y_i\right) f_i \overline{v}_i}{\sum_i \hat{\pi}\left(s|y_i\right) f_i} + \lambda\overline{v}_j =$$

$$\left(1-\lambda\right)\hat{p}_j + \lambda\overline{v}_j = p_j.$$

This concludes the proof.

2. When Y is a subset of $\mathbb{R}$. Let Y be the support of the indicator y and a subset of $\mathbb{R}$. Consider a sequence of partitions $Y^n = \{\min Y = y_0^n < y_1^n < \cdots < y_n^n = \max Y\}$ for $n = 1, 2, \cdots$ with $\max_{0\leq i\leq n-1} y_{i+1}^n - y_i^n \to 0$ and $Y^n \subset Y^{n+1}$. We define

$$g_i^n = \Pr\left(y \in \left[y_{i-1}^n, y_i^n\right)\right), 1 \leq i \leq n$$

$$\overline{v}_i^n = \begin{cases} \frac{\int \overline{v}(y)\mathbf{1}\left[y\in\left[y_{i-1}^n, y_i^n\right)\right]dG}{g_i^n} & g_i^n > 0, i \leq n \\ \frac{\overline{v}\left(y_{i-1}^n\right)+\overline{v}\left(y_i^n\right)}{2} & g_i^n = 0, i \geq 1 \end{cases} \quad p_i^n = \begin{cases} \frac{\int p(y)\mathbf{1}\left[y\in\left[y_{i-1}^n, y_i^n\right)\right]dG}{g_i^n} & g_i^n > 0, i \leq n \\ \frac{p\left(y_{i-1}^n\right)+p\left(y_i^n\right)}{2} & g_i^n = 0, i \geq 1. \end{cases}$$

In words, the above constructs a discretization of the buyer values v and the intermediary's interim prices p. Since $p\left(y\right)$ is comonotone with $\overline{v}\left(y\right)$, so are $\overline{v}^n$ and p^n. Moreover, $p^n \succcurlyeq_{\text{SOSD}} \overline{v}^n$ and Y^n is finite, we can apply the result from the first part. That is, an information structure $\left(S^n, \pi^n\right)$ exists where $\pi^n : Y^n \to \Delta\left(S^n\right)$ such that $p^n\left(y_i^n\right) = \sum_{s\in S^n} \pi^n\left(\{s\}\,|y_i^n\right)\mathbb{E}\left[\overline{v}|s\right]$.

Note that each $\left(S^n, \pi^n\right)$ induces a distribution over posterior beliefs of the buyers given by $\tau^n \in \Delta\left(\Delta\left(Y^n\right)\right)$, since any probability measure in $\Delta\left(Y^n\right)$ can be embedded in $\Delta\left(Y\right)$. This is because for any $\mu \in \Delta\left(Y^n\right)$, we can construct $\hat{\mu} \in \Delta\left(Y\right)$ defined by $\hat{\mu}\left(A\right) = \sum_{i=1}^n \mu_i^n \mathbf{1}\left[y_i^n \in A\right]$, where A is an arbitrary Borel subset of Y. Similarly, we can find $\hat{\tau}^n \in \Delta\left(\Delta\left(Y\right)\right)$, which is equivalent to τ^n.

Now consider the probability measure ζ^n representing the joint distribution of y^n and posterior mean $\mathbb{E}_\mu\left[v\right] = \int v d\mu$ for any $\mu \in \text{Supp}\left(\hat{\tau}^n\right)$ induced by $\hat{\tau}^n$. Note that $\zeta^n \in \Delta\left(Y \times Y\right)$. By an application of Riesz representation theorem (see Theorem 14.12 in Aliprantis and Border (2013)), $\Delta\left(Y \times Y\right)$ is compact according to the weak-* topology. This implies that the sequence $\{\zeta^n\}$ must have a convergent subsequence whose limit is given by $\zeta \in \Delta\left(Y \times Y\right)$. Let $\mathcal{G}^n$ be the σ-field generated by the sets $\left\{\left[y_i^n, y_{i+1}^n\right)\right\}_{i\leq n}$ and let $\mathcal{F}^n = \mathcal{G}^n \times \{\emptyset, \Delta\left(Y\right)\}$. In words, $\mathcal{F}^n$ conveys

the information that $y \in \left[y_i^n, y_{i+1}^n\right)$. Note that $\mathcal{F}^n \subset \mathcal{F}^{n+1}$ because $Y^n \subset Y^{n+1}$. Moreover,

$$\mathbb{E}\left[\zeta^n | \mathcal{F}^n\right] = \left(\overline{v}^n, p^n\right),$$

where $\left(\overline{v}^n, p^n\right)$ is the random variable with values $\left(\overline{v}_i^n, p_i^n\right)$ with probability g_i^n. Note that the above holds by the construction of τ^n and ζ^n. As a result

$$\mathbb{E}\left[\zeta^{n+1} | \mathcal{F}^n\right] = \mathbb{E}\left[\mathbb{E}\left[\zeta^{n+1} | \mathcal{F}^{n+1}\right] | \mathcal{F}^n\right] = \mathbb{E}\left[\left(\overline{v}^{n+1}, p^{n+1}\right) | \mathcal{F}^n\right] = \left(\overline{v}^n, p^n\right),$$

where the last equality follows because $\mathbb{E}\left[p\left(y\right) | \mathcal{F}^n\right] = p^n, \mathbb{E}\left[\overline{v}\left(y\right) | \mathcal{F}^n\right] = \overline{v}^n$ given the definition of p^n and $\overline{v}^n$ above. All of this implies that $\mathcal{F}^n$ is a filtration and $\left(\zeta^n, \mathcal{F}^n\right)$ forms a bounded martingale (for a definition see Doob (1994)). Hence by Doob's martingale convergence theorem (see Theorem XI.14 in Doob (1994)), we must have that

$$\lim_{n \to \infty} \mathbb{E}\left[\zeta^n | \mathcal{F}^n\right] = \mathbb{E}\left[\zeta | \mathcal{F}\right].$$

Therefore, $\mathbb{E}_\zeta\left[\int \overline{v}\left(y'\right) d\mu | y\right] = p\left(y\right)$. This concludes the proof. □

A.2 Proof of Theorem 1

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

Lemma 1. *Let $h\left(y\right)$ be an integrable function. Suppose that $H\left(i\right) = \int_{\left\{y : \overline{v}(y) > v_Q(i)\right\}} h\left(y\right) dG$ and let $cavH$ be the concave envelope of H, i.e., the lowest concave function that is higher than $H\left(i\right)$. Then*

$$\max_{\substack{p \,:\, p \succcurlyeq_{cv} \overline{v}, \\ p, \overline{v} \,:\, comonotone}} \int h\left(y\right) p\left(y\right) dG = \int_0^1 cavH\left(i\right) dv_Q\left(i\right) \tag{10}$$

Moreover, an optimal p that achieves the above value satisfies $p\left(y\right) = \overline{v}\left(y; a\right)$ when $H\left(G\left(y|a\right)\right) = cavH\left(G\left(y|a\right)\right)$. Moreover, if for some maximal interval $I \subset [0,1]$, $cavH\left(i\right) < H\left(i\right), i = G\left(y|a\right) \in intI$, $p\left(y\right) = \mathbb{E}\left[\overline{v} | G\left(y|a\right) \in I\right]$.

Proof. In this proof, we assume that $-\infty < v_Q\left(0\right) < v_Q\left(1\right) < \infty$. The cases with $v_Q\left(1\right) = \infty$ or $v_Q\left(0\right) = -\infty$ can be proved using a limiting argument. Before proceeding, we prove the following lemma:

Lemma 2. *Let $p, \overline{v}$ be comonotone and $p_Q\left(i\right), v_Q\left(i\right)$ be their associated quantile representation as defined in (6). If $F_p\left(v\right), F_v\left(v\right)$ are the cumulative distribution functions of $p, \overline{v}$ respectively, then $p \succcurlyeq_{cv} \overline{v}$ if and only if $F_v\left(v\right) \succcurlyeq_{cv} F_p\left(v\right)$ where v is uniformly distributed over $V = \left[v_Q\left(0\right), v_Q\left(1\right)\right]$. In other words,*

$$\int_0^1 \phi\left(p_Q\left(i\right)\right) di \geq \int_0^1 \phi\left(v_Q\left(i\right)\right) di, \forall \phi : V \to \mathbb{R} : concave \Leftrightarrow$$

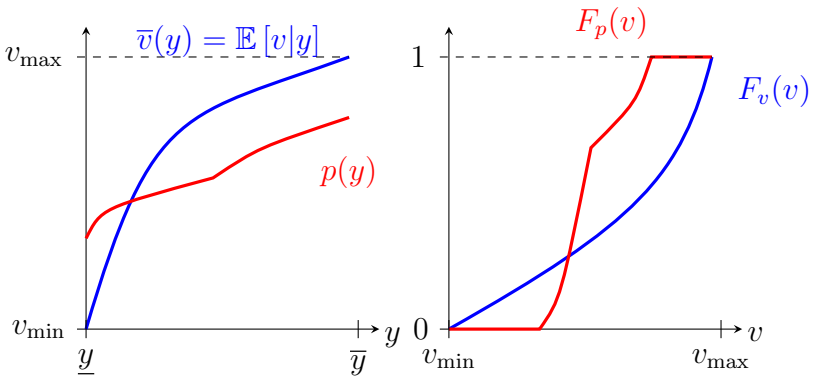


Figure 6: Example of $p \succcurlyeq_{\text{cv}} \overline{v}$ (left) and their associated CDF's (right) that satisfy $F_v \succcurlyeq_{\text{cv}} F_p$

$$\int_V \psi\left(F_v\left(v\right)\right) dv \geq \int_V \psi\left(F_p\left(v\right)\right) dv, \forall \psi : [0,1] \to \mathbb{R} : \textit{concave}$$

An example that illustrates Lemma 2 is depicted in Figure 6. On the left, we have an example of $p, \overline{v}$ (not their quantile version) where p is a mean–preserving contraction of $\overline{v}$ while this is reversed for their c.d.f.'s. The idea behind Lemma 2 is simple. Since c.d.f.'s are inverses of the quantile functions mean preserving contraction for one implies mean preserving spread for the other. We provide its proof in the Online Appendix, Section E. In the above, we can view $i = F_v\left(v\right)$ as having a distribution according to $\frac{v_Q(i)-v_Q(0)}{v_Q(1)-v_Q(0)}$ and $j = F_p\left(v\right)$ as having a distribution $\frac{p_Q(j)-v_Q(0)}{v_Q(1)-v_Q(0)}$ with probability $\frac{v_Q(1)-p_Q(1)}{v_Q(1)-v_Q(0)}$ on $j = 1$.

Now, if $p \succcurlyeq_{\text{cv}} \overline{v}$, Lemma 2 implies that $F_v \succcurlyeq_{\text{cv}} F_p$. By Blackwell's theorem (see also Kolotilin (2018) and Gentzkow and Kamenica (2016)), there must exist a signal structure (or a garbling) where $F_v\left(v\right) = i$ is the conditional mean of $F_p = j$ upon realization of the signal with i having distribution $dv_Q\left(i\right) / \left(v_Q\left(i\right) - v_Q\left(0\right)\right)$ and similarly for j. Let $\mu\left(\cdot|i\right) \in \Delta\left[0,1\right]$ be the posterior associated with $F_v\left(v\right) = i$. Since the distribution of i is given by $dv_Q\left(i\right) / \left(v_Q\left(i\right) - v_Q\left(0\right)\right)$ we can write Bayes plausibility as

$$\int \mu\left(A|i\right) \frac{dv_Q\left(i\right)}{|V|} = \frac{\int_A dp_Q\left(j\right)}{|V|}, A \subset [0,1] \tag{11}$$

such that

$$i = \int j d\mu\left(j|i\right), \forall i \in [0,1] \tag{12}$$

where in the above $|A|$ is the Lebesgue measure of A and $|V| = v_Q\left(1\right) - v_Q\left(0\right)$. Equation (12) simply states that $i = F_v\left(v\right)$ is the posterior mean of $j = F_p\left(v\right)$ according to the distribution $\mu\left(\cdot|i\right)$.

We can then write

$$\int_Y p\left(y\right) h\left(y\right) dG = -\int_0^1 p_Q\left(j\right) dH\left(j\right)$$

$$
\begin{aligned}
&= -\int_0^1 \mu\left([0,j]\,|i\right) dv_Q\left(i\right) dH\left(j\right) \text{ by setting } A\text{=}[0,j] \text{ in (11)} \\
&= -\int_0^1 \int_0^1 \mu\left([0,j]\,|i\right) dH\left(j\right) dv_Q\left(i\right) \text{ Fubini's Theorem} \\
&= -\int_0^1 \left[H\left(1\right) - \int_0^1 H\left(j\right) d\mu\left(j|i\right)\right] dv_Q\left(i\right) \text{ Int. by parts for inside integral} \\
&= \int_0^1 \int_0^1 H\left(j\right) d\mu\left(j|i\right) dv_Q\left(i\right) \text{ since by construction } H\left(1\right) = 0
\end{aligned}
$$

Now, consider the expression $\int_0^1 H\left(j\right) d\mu\left(j|i\right)$. In this expression $\mu\left(\cdot|i\right)$ is a probability distribution over j whose mean value is i. In other words, this is a convex combination of values of $H\left(j\right)$ with average of i. Given the definition of the concave envelope, we have that $\int_0^1 H\left(j\right) d\mu\left(j|i\right) \leq \text{cav}H\left(i\right)$. Moreover, since the space of measures over $[0,1]$ is compact according to weak−* topology (Banach-Alaoglu theorem) for each i there must exist $\mu\left(\cdot|i\right) \in \Delta\left[0,1\right]$ such that $\int_0^1 H\left(j\right) d\mu\left(j|i\right) = \text{cav}H\left(i\right)$. We can then use the above procedure to construct an interim price function that delivers $\int \text{cav}H\left(i\right) dv_Q\left(i\right)$. This proves the equality (10).

To prove the second part, by Caratheodory theorem, for any i either $\text{cav}H\left(i\right) = H\left(i\right)$ or that there exists $i_1 < i < i_2$ such that

$$
\text{cav}H\left(i\right) = \frac{i_2 - i}{i_2 - i_1} H\left(i_1\right) + \frac{i - i_1}{i_2 - i_1} H\left(i_2\right)
$$

In the first case, $\mu\left(\{i\}\,|i\right) = 1$ and in the second case $\mu\left(\{i_1\}\,|i\right) = \frac{i_2 - i}{i_2 - i_1} = 1 - \mu\left(\{i_2\}\,|i\right)$. In other words, concavification of H partitions the range of $[0,1]$ into subintervals $(i_{1,\alpha}, i_{2,\alpha})$, $[i_{1,\beta}, i_{2,\beta}]$ where for any α there exists β such that $i_{2,\alpha} = i_{1,\beta}$ and another β for which $i_{1,\alpha} = i_{2,\beta}$ where for all $i \in [i_{1,\beta}, i_{2,\beta}]$, $\mu\left(\{i\}\,|i\right) = 1$ and for all $i \in (i_{1,\alpha}, i_{2,\alpha})$, $\mu\left(\{i_{1,\alpha}\}\,|i\right) = \frac{i_{2,\alpha} - i}{i_{2,\alpha} - i_{1,\alpha}} = 1 - \mu\left(\{i_{2,\alpha}\}\,|i\right)$. From above we know that

$$
p_Q\left(j\right) = \int_0^1 \mu\left([0,j]\,|i\right) dv_Q\left(i\right)
$$

Given the construction of μ, we have

$$
\mu\left([0,j]\,|i\right) = \begin{cases} \mathbf{1}\left[j \geq i\right] & \text{cav}H\left(i\right) = H\left(i\right) \\ \frac{i_{2,\alpha} - i}{i_{2,\alpha} - i_{1,\alpha}} \mathbf{1}\left[i_{2,\alpha} > j \geq i_{1,\alpha}\right] + \mathbf{1}\left[j \geq i_{2,\alpha}\right] & \text{otherwise} \end{cases}
$$

Now, suppose that $j \in [i_{1,\beta}, i_{2,\beta}]$ for some β. This means that if $i > j$, $\mu\left([0,j]\,|i\right) = 0$ since all higher quantiles either put weights only on i or on values of the form $i_{1,\alpha}, i_{2,\alpha}$ which are higher than j. Then we can write

$$
\begin{aligned}
p_Q\left(j\right) &= \int_0^1 \mu\left([0,j]\,|i\right) dv_Q\left(i\right) \\
&= \int_0^j dv_Q\left(i\right) = v_Q\left(j\right)
\end{aligned}
$$

Moreover, if $j \in (i_{1,\alpha}, i_{2,\alpha})$ for some α, then it has to be that if $i > i_{2,\alpha}$, then $\mu([0,j]\,|i) = 0$. So we can write

$$\begin{aligned}
p_Q(j) &= \int_0^{i_{2,\alpha}} \mu([0,j]\,|i)\, dv_Q(i) \\
&= \int_0^{i_{1,\alpha}} dv_Q(i) + \int_{i_{1,\alpha}}^{i_{2,\alpha}} \mu([0,j]\,|i)\, dv_Q(i) \\
&= v_Q(i_{1,\alpha}) + \int_{i_{1,\alpha}}^{i_{2,\alpha}} \frac{i_{2,\alpha} - i}{i_{2,\alpha} - i_{1,\alpha}} dv_Q(i) \\
&= v_Q(i_{1,\alpha}) - v_Q(i_{1,\alpha}) + \int_{i_{1,\alpha}}^{i_{2,\alpha}} \frac{v_Q(i)}{i_{2,\alpha} - i_{1,\alpha}} di \\
&= \mathbb{E}\left[v_Q(i)\,|i \in (i_{1,\alpha}, i_{2,\alpha})\right]
\end{aligned}$$

This establishes the claim. □

A.2.1 Proof of the Theorem

Proof. Consider the problem of finding the best interim price function in quantile form for a given action a:

$$\max_{p_Q} W(a) - \int p_Q(i)\, dH(i;a) = \max_{p_Q} W(a) + T_H p_Q$$

subject to $p_Q \succcurlyeq_{\text{cv}} v_Q$, monotonicity of p_Q and the first order IC constraints. We will show that for any $a \in A$, Lagrange multipliers associated with the first order IC constraints exist so that the constrained optimization gives the same value as the unconstrained optimization of the Lagrangian over the space of p_Q's that are a mean preserving contraction of v_Q and are monotone. This combined with Lemma 1 implies the desired result.

Let us view p_Q as a member of the space $L_\infty([0,1])$. Let us refer to the first order IC constraints with respect to a_n as $T_n p_Q = 0$ where T_n is an affine transformation that maps $L_\infty([0,1])$ to $\mathbb{R}$ (same is true for T_H) and we can define $Tp_Q = (T_1 p_Q, \cdots, T_N p_Q)$ which is an affine transformation from $L_\infty([0,1])$ to $\mathbb{R}^N$. Finally, let us refer to the set of p_Q's that satisfy $p_Q \succcurlyeq_{\text{cv}} v_Q$ and monotonicity of p_Q as $\mathcal{P}$ and the subset of $\mathcal{P}$ that satisfies $Tp_Q = 0$ as $\mathcal{Q}$.

For any subset $S \subset \{1, \cdots, N\}$, let $S^c = \{1, \cdots, N\} \setminus S$ and let us consider the following sets

$$\mathcal{P}(S) = \{p_Q \in L_\infty([0,1])\,|p_Q \succcurlyeq_{\text{cv}} v_Q, p_Q \text{ increasing}, T_n p_Q \geq 0, n \in S, T_{n'} p_Q \leq 0, n' \in S^c\}$$

Lemma 3. *There exists S such that $\max_{p_Q \in \mathcal{P}(S)} T_H p_Q = \max_{p_Q \in \mathcal{Q}} T_H p_Q$.*

Proof. Since all members of $\mathcal{Q}$ satisfy $T_n p_Q = 0$, it must be that $\mathcal{Q} \subset \mathcal{P}(S)$ for all S. This implies that $\max_{p_Q \in \mathcal{P}(S)} T_H p_Q \geq \max_{p_Q \in \mathcal{Q}} T_H p_Q$. Now, suppose to the contrary that for all S, the left hand side is strictly higher than the right hand side. This implies that for any $S \subset \{2, \cdots, N\}$, there exists $p \in \mathcal{P}(S), p' \in \mathcal{P}(S \cup \{1\})$ such that $T_H p, T_H p' > \max_{p_Q \in \mathcal{Q}} T_H p_Q$. Since $T_1 p \leq$

$0 \leq T_1 p'$ there must exist λ such that $T_1\left(\lambda p + (1-\lambda) p'\right) = 0$. Let $p^{(1),S} = \lambda p + (1-\lambda) p'$ and recall that $p^{(1),S} \in \mathcal{P}(S)$. We must also have that $T_H p^{(1),S} > \max_{p_Q \in \mathcal{Q}} T_H p_Q$. Now, we know that $T_2 p^{(1),S} \leq 0 \leq T_2 p^{(1),S\cup\{2\}}$ for any $S \subset \{3, \cdots, N\}$. By using the same argument, we can find $p^{(1,2),S}$ such that $T_1 p^{(1,2),S} = T_2 p^{(1,2),S} = 0$, $p^{(1,2),S} \in \mathcal{P}(S)$ and $T_H p^{(1,2),S} > \max_{p_Q \in \mathcal{Q}} T_H p_Q$. By continuing this construction, we can find $p^{(1,2,\cdots,N)}$ such that $T_1 p^{(1,\cdots,N)} = \cdots = T_N p^{(1,\cdots,N)} = 0$ and that $T_H p^{(1,\cdots,N)} > \max_{p_Q \in \mathcal{Q}} T_H p_Q$ which is a contradiction since $p^{(1,\cdots,N)} \in \mathcal{Q}$. □

Now, suppose that $\hat{S} \subset \{1, \cdots, N\}$ satisfies the condition in Lemma 3. Let us define $\mathcal{P} = \{p_Q | p_Q \succcurlyeq_{\text{cv}} v_Q, p_Q \text{ increasing}\}$. Then, T maps members of $\mathcal{P}$ into $\mathbb{R}^N$. Moreover, T maps members of $\mathcal{P}\left(\hat{S}\right)$ into a convex cone. Since the image of $\mathcal{P}\left(\hat{S}\right)$ under T is convex in $\mathbb{R}^N$, it must have a non–empty relative interior.[22] This implies that we can apply standard results for existence of Lagrange multipliers (strong duality) – see for example, Theorem 8.3.1. in Luenberger (1997). Hence, it must be that $\lambda \neq 0 \in \mathbb{R}^N$ exists such that $\lambda_n \geq 0$ for all $n \in \hat{S}$ and $\lambda_n \leq 0$ for all $n \in \hat{S}^c$ such that

$$\max_{p_Q \in \mathcal{P}} W(a) - \int p_Q(i)\, dH(i;a) =$$

$$\max_{p_Q \in \mathcal{P}} W(a) + \int \left[H(i;a) - \sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\Big|_{\hat{a}=a} \right] dp_Q - \sum_{n=1}^{N} \lambda_n \frac{\partial c(a)}{\partial a_n}$$

The rest of the claim follows from Lemma 1. □

A.3 Proof of Proposition 2

Proof. Given the statement of Theorem 1, we know that the unconstrained objective in (D) can be achieved by a monotone partition. Note that by Kleiner et al. (2021), the extreme points of the convex set $\mathcal{P}$ – the set of p_Q's that are mean preserving contractions of v_Q and are monotone – are associated with the monotone partitions. In what follows, we show that under the Assumption 2, no two extreme points of $\mathcal{P}$ can deliver the same value of the Lagrangian

$$L(p_Q, \lambda; a) = W(a) - \int p_Q(i)\, d\left(H(i;a) - \sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\Big|_{\hat{a}=a} \right) - \sum_{n=1}^{N} \lambda_n \frac{\partial c(a)}{\partial a_n}$$

This would imply that $L(p_Q, \lambda; a)$ has a unique maximizer for any λ which establishes the claim.

Suppose that there are two interim price functions $p_1, p_2 \in \mathcal{P}$ that are associated with monotone partitions. If $p_1 \neq p_2$, then there must exist an interval $I \subset [0,1]$ so that all of its members satisfy $p_{1,Q}(i) = v_Q(i)$ and $p_{2,Q}(i)$ is constant for all $i \in I$. By Lemma 1, if $p_{1,Q}(i) = v_Q(i)$ is optimal for an interval I, then $\Gamma(i;a,\lambda) = H(i;a) - \sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\Big|_{\hat{a}=a}$ should coincide with its concave envelope. Moreover, suppose that the maximal interval containing I for which

[22] See for example Theorem 6.2 in Rockafellar (1970).

$p_{2,Q}$ is constant is $\tilde{I} = (i_1, i_2)$. Let the value of $p_{2,Q}$ be $\tilde{p}$ over $\tilde{I}$.

We show that this implies that $\Gamma(i; a, \lambda)$ is linear over the interval I. Suppose to the contrary that for some sub–interval $I' \subset I$, $\Gamma(i; a, \lambda)$ is strictly concave. Then,

$$-\int_{\tilde{I}} p_{2,Q}(i)\, d\Gamma(i; a, \lambda) = -\tilde{p}\left[\Gamma(i_2; a, \lambda) - \Gamma(i_1; a, \lambda)\right]$$

We also have that

$$v_Q(i_1) < \tilde{p} = \frac{\int_{\tilde{I}} v_Q(i)\, di}{i_2 - i_1} < v_Q(i_2)$$

Let j be an interior point of I'. Let us define

$$\hat{p}_Q(i) = \begin{cases} p_{2,Q}(i) & i \in [0,1] \setminus \tilde{I} \\ \underline{p} & i \in (i_1, j) \\ \overline{p} & i \in (j, i_2) \end{cases}$$

where

$$\underline{p} = \frac{\int_{i_1}^{j} v_Q(i)\, di}{j - i_1}, \overline{p} = \frac{\int_{j}^{i_2} v_Q(i)\, di}{i_2 - j}$$

We have

$$-\int \hat{p}_Q(i)\, d\Gamma(i; \lambda, a) + \int p_{2,Q}(i)\, d\Gamma(i; \lambda, a) =$$
$$-\frac{\int_{i_1}^{j} v_Q(i)\, di}{j - i_1}\left[\Gamma(j; a, \lambda) - \Gamma(i_1; a, \lambda)\right] - \frac{\int_{j}^{i_2} v_Q(i)\, di}{i_2 - j}\left[\Gamma(i_2; a, \lambda) - \Gamma(j; a, \lambda)\right]$$
$$+\frac{\int_{i_1}^{i_2} v_Q(i)\, di}{i_2 - i_1}\left[\Gamma(i_2; a, \lambda) - \Gamma(i_1; a, \lambda)\right]$$

In the above, since $\Gamma(i; a, \lambda)$ is strictly concave over parts of I, we must have that

$$\frac{\Gamma(j; a, \lambda) - \Gamma(i_1; a, \lambda)}{j - i_1} > \frac{\Gamma(i_2; a, \lambda) - \Gamma(j; a, \lambda)}{i_2 - j}$$

Let us also define $\pi_1 = \frac{j - i_1}{i_2 - i_1} = 1 - \pi_2$. Since $\underline{p} < \overline{p}$,we can write

$$\pi_1 \underline{p} \frac{\Gamma(j; a, \lambda) - \Gamma(i_1; a, \lambda)}{j - i_1} + \pi_2 \overline{p} \frac{\Gamma(i_2; a, \lambda) - \Gamma(j; a, \lambda)}{i_2 - j} <$$
$$\left(\pi_1 \underline{p} + \pi_2 \overline{p}\right)\left(\pi_1 \frac{\Gamma(j; a, \lambda) - \Gamma(i_1; a, \lambda)}{j - i_1} + \pi_2 \frac{\Gamma(i_2; a, \lambda) - \Gamma(j; a, \lambda)}{i_2 - j}\right) =$$
$$\frac{\int_{i_1}^{i_2} v_Q(i)\, di}{i_2 - i_1} \frac{\Gamma(i_2; a, \lambda) - \Gamma(i_1; a, \lambda)}{i_2 - i_1}$$

which implies that

$$-\int \hat{p}_Q(i)\, d\Gamma(i; \lambda, a) + \int p_{2,Q}(i)\, d\Gamma(i; \lambda, a) > 0$$

and thus $p_{2,Q}$ cannot be optimal. Therefore,

$$\Gamma'(i;a,\lambda) = c, \forall i \in I$$

for some c. Using the definition of H and F, we have

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

This is indeed in contradiction with the independence assumption which establishes the claim. □

A.4 Proof of Proposition 3

Proof. In this case, the function $\Gamma(i;a,\lambda)$ satisfies

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

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 convex Γ leads to a fully uninformative rating. □

A.5 Proof of Proposition 4

Proof. Recall that $\Gamma(i;\lambda,a) = -\lambda \left.\frac{\partial F(i|\hat{a};a)}{\partial \hat{a}}\right|_{\hat{a}=a}$, where we drop the constant $-\lambda c'(a)$ term since it shifts Γ uniformly and does not affect its concavification. As we have shown in Section 3,

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

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

Suppose to the contrary that λ is negative. In this case, since $\Gamma(0;a,\lambda) = \Gamma(1;a,\lambda) = 0$, there are two possibilities: 1. $\Gamma'(0;a,\lambda) < 0$ in which case $\Gamma(i;a,\lambda)$ is non–positive for all values of i and its concave envelope is the zero function associated with no information; 2. $\Gamma'(0;a,\lambda) > 0$ 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 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, $\lambda \geq 0$ and thus 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 is given by

$$\frac{\Gamma''(i|a)|_{i=G(y|a)}}{g(y|a)} = -\lambda_1 \frac{\partial \ell_1(y|a)}{\partial y} - \lambda_2 \frac{\partial \ell_2(y|a)}{\partial y}.$$

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, optimal rating is either lower or upper censorship. To see this, consider three values $y_1 < y_2 < y_3$. Note that we can assume without loss of generality that $D \geq 0$ as when $D \leq 0$ the reverse argument would work. Let us define the vectors $\mathbf{z}_i = \left(\frac{\partial \ell_1(y_i|a)}{\partial y}, \frac{\partial \ell_2(y_i|a)}{\partial y}\right)$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ and for an arbitrary vector x let $x^+ = (x_2, -x_1)$ be the perpendicular vector of x; we know that $x^+ \cdot x = 0$. Then MRIP implies that

$$\mathbf{z}_2, \mathbf{z}_3 \in \left\{\mathbf{x}|\mathbf{z}_1^+ \cdot \mathbf{x} \geq 0\right\}, \mathbf{z}_3 \in \left\{\mathbf{x}|\mathbf{z}_2^+ \cdot \mathbf{x} \geq 0\right\}$$

This means that $\mathbf{z}_2$ belongs to the half space above the line through $\mathbf{z}_1$ and the same holds for $\mathbf{z}_3$ and $\mathbf{z}_2$. In other words, $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$ should relate to each other as in Figure 7 where $\mathbf{z}_2$ has a higher

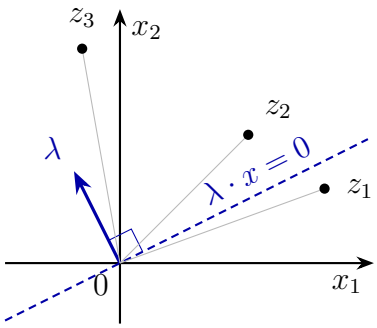


Figure 7: The vectors of likelihood changes under MRIP.

angle with the x–axis relative to $\mathbf{z}_1$ and $\mathbf{z}_3$ has a higher angle than $\mathbf{z}_2$. This means that $\mathbb{R}^2$ is partitioned by the following sets

$$\left\{\mathbf{x}|\mathbf{z}_1^+ \cdot \mathbf{x} \geq 0 > \mathbf{z}_2^+ \cdot \mathbf{x}\right\}, \left\{\mathbf{x}|\mathbf{z}_2^+ \cdot \mathbf{x} \geq 0 > \mathbf{z}_3^+ \cdot \mathbf{x}\right\}, \left\{\mathbf{x}|\mathbf{z}_3^+ \cdot \mathbf{x} \geq 0 > \mathbf{z}_1^+ \cdot \mathbf{x}\right\}, \left\{\mathbf{x}|0 > \mathbf{z}_1^+ \cdot \mathbf{x}\right\}.$$

This implies that $\boldsymbol{\lambda}^+$ has to belong in one of the above sets. Suppose that $\boldsymbol{\lambda}^+$ belongs to the first set (the situation depicted in Figure 7), then

$$\mathbf{z}_1^+ \cdot \boldsymbol{\lambda}^+ \geq 0 > \mathbf{z}_2^+ \cdot \boldsymbol{\lambda}^+ \Rightarrow \begin{cases} \frac{\partial \ell_2(y_1|a)}{\partial y}\lambda_2 + \frac{\partial \ell_1(y_1|a)}{\partial y}\lambda_1 \geq 0 \\ \frac{\partial \ell_2(y_2|a)}{\partial y}\lambda_2 + \frac{\partial \ell_1(y_2|a)}{\partial y}\lambda_1 < 0 \end{cases}$$

This means that if $\boldsymbol{\lambda}^+ = \alpha_1 \mathbf{z}_1 + \alpha_2 \mathbf{z}_2$, then

$$\mathbf{z}_1^+ \cdot \boldsymbol{\lambda}^+ = \alpha_2 \mathbf{z}_1^+ \cdot \mathbf{z}_2 \Rightarrow \alpha_2 \geq 0$$
$$\mathbf{z}_2^+ \cdot \boldsymbol{\lambda}^+ = \alpha_1 \mathbf{z}_2^+ \cdot \mathbf{z}_1 \Rightarrow \alpha_1 \geq 0$$

This in turn means that

$$\begin{aligned} \mathbf{z}_3^+ \cdot \boldsymbol{\lambda}^+ = \mathbf{z}_3 \cdot \boldsymbol{\lambda} &= \mathbf{z}_3^+ \cdot (\alpha_1 \mathbf{z}_1 + \alpha_2 \mathbf{z}_2) \\ &= \alpha_1 \mathbf{z}_3^+ \cdot \mathbf{z}_1 + \alpha_2 \mathbf{z}_3^+ \cdot \mathbf{z}_2 \leq 0 \end{aligned}$$

where the above holds because $\mathbf{z}_1^+ \cdot \mathbf{z}_3, \mathbf{z}_2^+ \cdot \mathbf{z}_3 \geq 0$. This means that $\mathbf{z}_1 \cdot \boldsymbol{\lambda} \geq 0 \geq \mathbf{z}_2 \cdot \boldsymbol{\lambda}, \mathbf{z}_3 \cdot \boldsymbol{\lambda}$ which is the desired result. The other cases can be shown the same way. □

A.7 Proof of Proposition 6

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

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

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

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

Replacing in the formula for Γ' implies

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

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 can argue that $\lambda_R > 0$. This is because if we consider the problem by replacing the IC for the R type with its inequality version imposing that the marginal return to a_R be higher than its cost. In this problem, if this constraint remains slack, one can simply increase a_R and shifts all $p(y)$'s upwards by the same amount and improve the payoffs. Hence, $\lambda_R \geq 0$. Since $m'_R \geq 0$ by assumption, we must have that $A_4 > 0$.

We then have

$$\begin{aligned} (B_1 + B_2 x)^2 \frac{\Gamma''\left(\overline{G}(y)\right)\overline{g}(y)}{\frac{dx}{dy}} &= (B_2 A_3 - B_1 A_4)\log x - B_1 A_3/x - B_2 A_4 x \\ &\quad + B_2(A_1 - A_3) - B_1(A_1 + A_4) \\ &= B_2 A_4(\alpha_1 \log x + \alpha_2/x - x + \alpha_3) \end{aligned}$$

Note that in the above $A_4 B_2 > 0$. The derivative of the above with respect to x is given by $\frac{\alpha_1}{x} - \alpha_2/x^2 - 1 = \frac{\alpha_1 x - \alpha_2 - x^2}{x}$. Suppose that $\alpha_2 > 0$. Since the numerator is a quadratic function, it has at most two roots and this means that Γ'' switches sign at most three times. This establishes the claim. □

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,

$$D(y_1, y_2|a) = \eta''(\varepsilon_1)\eta''(\varepsilon_2)\Psi(a)\left[\varepsilon_2 + \frac{\eta'(\varepsilon_2)}{\eta''(\varepsilon_2)} - \varepsilon_1 - \frac{\eta'(\varepsilon_1)}{\eta''(\varepsilon_1)}\right] \tag{13}$$

where $\varepsilon_n = \frac{y_n-\mu(a)}{\sigma(a)}$ and $\Psi(a) = \frac{\partial\mu}{\partial a_1}\frac{\partial\sigma}{\partial a_2} - \frac{\partial\sigma}{\partial a_1}\frac{\partial\mu}{\partial a_2}$. 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 4. *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 5. *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 is equivalent to $m(x) \leq (1 - H(x))/h(x)$; 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 interiority of the implemented action. Hence, by Theorem 2, the welfare-maximizing rating is an upper- or a lower-censorship rating. By Lemma 4, $P_2 = (\mu_1 c_2 - \mu_2 c_1)/\Psi$ at the implemented action. When (8) holds, $P_2 \leq 0$ and thus by Lemma 5 the optimal rating cannot be lower censorship; it is therefore upper censorship, degenerating to full disclosure when (8) holds with equality. When (8) fails, $P_2 > 0$ and the optimal rating cannot be upper censorship; it is therefore lower censorship. This establishes the claim. □

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 around 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.

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 6. *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 6 into a condition on the distribution function directly. This gives the following:

Lemma 7. *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*

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

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

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

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

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

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

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

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

Since $c'\left(a_1\right) > 0$, the assumption on G in the statement of the lemma guarantees that $U''\left(a_1\right) < 0$. 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 7 is that they put a restriction on how convex the cost is, captured by $c''\left(a\right)/c'\left(a\right)$ relative to that of expected interim price. Indeed, the conditions can be rewritten as

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

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

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

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

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

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

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

and we have

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

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

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

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

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

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

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

Let us make the following assumption on H:

Assumption 5. *The cumulative distribution function $H\left(\varepsilon\right)$ satisfies:*

1. $\log\left(1-H\left(\varepsilon\right)\right)$ *is concave in* ε,
2. $\log H\left(\varepsilon\right)$ *is concave in* ε,
3. $\forall d>0$, $\frac{h(x)}{1-H(x)}-\frac{h(x-d)}{1-H(x-d)}\leq\kappa_1\left(d\right)$, $\frac{h(x-d)}{H(x-d)}-\frac{h(x)}{H(x)}\leq\kappa_2\left(d\right)$

The first two conditions are standard log–concavity conditions while the last condition implies that the variations in the derivative of $\log\left(1-H\left(x\right)\right)$ 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 concave, then $h'/h\geq -h/(1-H)$ 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 5, the RHS of the above is less than $\kappa_1(d)$. If $A=[0,\overline{a}]$, then the highest value of d is $2\overline{a}$ which means that it is sufficient for c to satisfy

$$\max_{d\leq 2\overline{a}}\kappa_1(d)\leq\frac{c''(a)}{c'(a)}$$

In a similar fashion, we can show that for lower-censorship ratings to satisfy the requirement of Lemma 7, we must also have

$$\max_{d \leq 2\overline{a}} \kappa_2(d) \leq \frac{c''(a)}{c'(a)}$$

Thus validity of FOA is equivalent to 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 6 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 $D_z^2 = \sum_{n,m} z_n z_m \partial^2/\partial a_n \partial a_m$. 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 8. *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 $z \neq 0$, $D_z U(a) = 0$ implies $D_z^2 U(a) < 0$. 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 $z \neq 0$ and let $\phi(t) = U(a + tz)$, defined on an interval since A is convex. Then $\phi'(t) = D_z U(a + tz)$ and $\phi''(t) = D_z^2 U(a + tz)$, 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 $D_{-z} U = -D_z U$ while $D_{-z}^2 U = D_z^2 U$, the requirement of Lemma 8 need only be verified for directions in a half space; we use this to restrict attention to z with $D_z c(a) \geq 0$. The analogue of Lemma 6 is then:

Lemma 9. *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 6 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 8 then implies that any effort satisfying the agent's first order conditions maximizes her payoff. □

The condition of Lemma 9 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 6 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. In the reducible case of Section 4.1, where $G(y|a) = \hat{G}(y|m(a))$, the agent's payoff depends on effort only through the index and the requirement collapses to the condition of Lemma 6 applied to $\hat{G}$ with the indirect cost $C(m)$.

C.3 The Ratings of Section 4

Theorem 2 and Proposition 5 invoke FOA for upper- and lower-censorship ratings when G is a location–scale technology and $\overline{v}$ is affine in y. In this case the condition of Lemma 9 can be verified from a closed form. 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 $D_z\hat{\varepsilon} = -(\mu_z + \hat{\varepsilon}\sigma_z)/\sigma$,

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

where $D_z^2\hat{\varepsilon} = -(\mu_{zz} + \hat{\varepsilon}\sigma_{zz})/\sigma + 2\sigma_z(\mu_z + \hat{\varepsilon}\sigma_z)/\sigma^2$, and the mirror expression holds for lower censorship. Since μ and σ are twice continuously differentiable on the compact action space, the only terms that require care are those involving Δ and the growth of $\mathcal{H}$: the former because $m(x)$ grows without bound as the threshold falls, the latter because $\mathcal{H}(x)$ grows linearly as it rises. Log–concavity of h controls both: it implies $m(x)\,h(x) \leq 1 - H(x) \leq 1$, and since a and $\hat{a}$ lie in a compact set, $\hat{\varepsilon}$ and $\hat{\varepsilon}_{\hat{a}}$ differ by a bounded amount, so that the Δ terms are uniformly bounded; and the linear growth of $\mathcal{H}$ enters only multiplied by second derivatives of μ and σ,

remaining bounded as the payoff approaches its full-disclosure limit. Therefore

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

and if $D^2 c(a) - \overline{\overline{\Xi}} I$ is positive definite for every $a \in A$, then $D_z^2 U(a) \leq \overline{\overline{\Xi}} \|z\|^2 - D_z^2 c(a) < 0$ 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, validity of FOA amounts to the cost having enough curvature, now in every direction.

For the technology of Example 1, where $\mu(a) = ba_1 + a_2$ and $\sigma(a) = \sqrt{b^2 a_1^2 + 1}$, the curvature bound can be computed explicitly: $\mu_{zz} = 0$, $\sigma_z = b^2 a_1 z_1 / \sigma$, $\sigma_{zz} = b^2 z_1^2 / \sigma^3$, and the slope of market values is $\beta(\hat{a}) = b\hat{a}_1^2 / (b^2 \hat{a}_1^2 + 1)$. Evaluating $\overline{\overline{\Xi}}$ numerically over all censorship thresholds, directions, actions and beliefs in $A = [0, \overline{a}]^2$ gives

$$\overline{\overline{\Xi}} \approx 0.26 \ \left(b = \tfrac{1}{2}, \overline{a} = 1\right), \quad 0.85 \ \left(b = \tfrac{1}{2}, \overline{a} = 2\right), \quad 0.72 \ (b = 1, \overline{a} = 1), \quad 1.50 \ (b = 1, \overline{a} = 2)$$

Hence with separable quadratic costs $c(a) = k(a_1^2 + a_2^2)/2$, for which $D^2 c = kI$, the FOA is valid whenever $k > \overline{\overline{\Xi}}$: symmetric quadratic costs ($k = 1$) validate the FOA for $b = \frac{1}{2}$ on both action spaces and for $b = 1$ when $\overline{a} = 1$. The required curvature grows with the passthrough b from productive effort to the indicator – which is what makes the technology irreducible – and with the size of the action space.

C.4 The Ratings of Section 5

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

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

When m is concave and b is convex the middle terms are non-positive, and if in addition $\psi(y|a) \geq 0$ on the support – that is, $b'(a) \leq m'(a) \inf_y r(y)$ – then $\psi^2 \leq \left(\sup_y \psi\right) \psi$ pointwise, so that

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

is sufficient for the validity of FOA for both types. Once again, the requirement is that the cost be convex enough, here relative to the range of the score.

For the example of Section 5 (Example 2), where $y = a + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, 1)$ and $c(a) = a^2/2$, we can quantify the required curvature. In this case $r(y) = y$, $m(a) = a$, $b(a) = a^2/2$ and $\psi =$

$y-a$, and the Gaussian identities $\mathbb{E}\left[(y-a)\,\mathbf{1}\left\{y>s\right\}\right]=h\left(s-a\right)$ and $\mathbb{E}\left[\left((y-a)^2-1\right)\mathbf{1}\left\{y>s\right\}\right]=(s-a)\,h\left(s-a\right)$ reduce the condition of Lemma 6 to

$$\int\left(s-a\right)h\left(s-a\right)dp\left(s\right)<\frac{c''\left(a\right)}{c'\left(a\right)}\int h\left(s-a\right)dp\left(s\right)$$

where dp has unit density on the revealed regions and atoms at the pooling boundaries. For a rating that pools $[y_1,y_2]$ and reveals elsewhere – the mid-censorship shape of Proposition 6 – writing $s_i=y_i-a$ and q for the pooled price net of a, the condition is $\Lambda\left(s_1,s_2,q\right)<c''\left(a\right)/c'\left(a\right)$ where

$$\Lambda\left(s_1,s_2,q\right)=\frac{h\left(s_2\right)-h\left(s_1\right)+s_1\left(q-s_1\right)h\left(s_1\right)+s_2\left(s_2-q\right)h\left(s_2\right)}{H\left(s_1\right)+1-H\left(s_2\right)+\left(q-s_1\right)h\left(s_1\right)+\left(s_2-q\right)h\left(s_2\right)}$$

Maximizing Λ numerically over all pooling intervals contained in $[-T,T]$ and all admissible pooled prices gives $\Lambda^*\approx 0.54$ for $T=1$, $\Lambda^*\approx 1.01$ for $T=1.5$ and $\Lambda^*\approx 1.49$ for $T=2$; restricting the pooled price to be Bayes-consistent with a belief error $|\hat{a}-a|\leq 1$ tightens the bound for $T=2$ to $\Lambda^*\approx 0.83$. The requirement is thus $c''\left(a\right)/c'\left(a\right)\geq\Lambda^*$: with the quadratic cost of the example, $c''/c'=1/a$, FOA is valid at every effort $a\leq 1/\Lambda^*$ – approximately 1 for pooling within ± 1.5 standard deviations, and above the equilibrium efforts of both types once the Bayes-consistent bound applies. For a power cost $c\left(a\right)\propto a^\gamma$ the condition reads $\left(\gamma-1\right)/a\geq\Lambda^*$, i.e., curvature $\gamma\geq 1+a\Lambda^*$. Since Λ^* grows with the width of the pooling region, the curvature needed for the validity of FOA scales with how much of the distribution the optimal rating pools.

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 3. Suppose that $A=\{0,1/3,1\}=\{a_1,a_2,a_3\}$, $v=a_i$. The indicator is deterministic: $G\left(\{a_i\}\,|a_i\right)=1$, and prior is uniform $\mu\left(\{a_i\}\right)=1/3$. In words, the market cares only about the action of the seller, and y coincides with it. Figure 8 depicts the feasible interim prices in the space of $\left(p\left(a_1\right),p\left(a_2\right)\right)$; (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. [23] The set of interim prices is denoted by area B in Figure 8. We find this set by solving the optimization problem associated with the highest and lowest value of $p\left(a_2\right)$ as a function of $p\left(a_1\right)$.[24] Evidently, the set B does not coincide with A. This is mainly due to the restrictions put by the second order expectations. For example, it can be easily shown that the coefficient of $\overline{v}\left(a\right)$

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

[24] The upper and lower bound of $p\left(a_2\right)$ can be found via standard concavification method. The lower

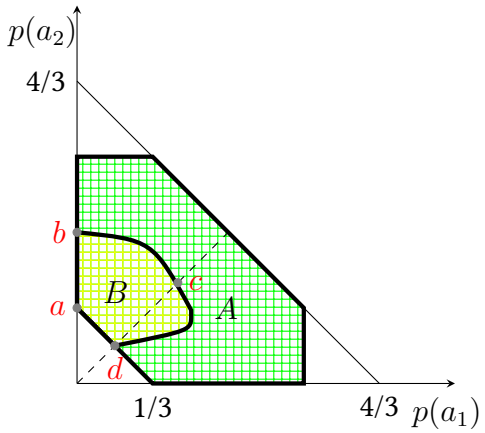


Figure 8: The set of interim prices and mean-preserving contractions of market valuations for Example 3. The green area, A, represents the three state random variables that are a mean-preserving contraction of market values. The yellow area, B, is the set of feasible interim prices under some information structure.

in $p(a)$ is at least $1/3$ which means that $p(a_3)$ cannot become lower than $4/9$.

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 aspects. 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 3, 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 the comonotonicity is effectively a form of uni-dimensionality for the indicator. 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

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 $3 - (6 - 2p(a_1)) / (6 - 3p(a_1))$.

interim prices $p(y)$ are a well-defined function of $\overline{v}(y;a)$. In other words, comonotonicity means that the indicator can always be reduced to a one-dimensional signal.

E Proof of Lemma 2

Proof. Suppose that $p_Q(i)$ is majorized by $v_Q(i)$ – both of these are non–decreasing functions. We will show that the same is true for F_v and F_p respectively, i.e., F_v is majorized by F_p. Then, for any $z \in [\underline{v}, \overline{v}]$ – the range of values of $v_Q(i)$, since $h(x) = \max\{x, z\}$ is a convex function, its expected value under v_Q must be higher than its expected value under p_Q. That is,

$$\int_0^1 \max\{z, v_Q(i)\} di \geq \int_0^1 \max\{z, p_Q(i)\} di$$

We can rewrite the above as

$$\int_{\underline{v}}^{\overline{v}} \max\{z, v\} dF_v(v) \geq \int_{\underline{v}}^{\overline{v}} \max\{z, v\} dF_p(v)$$

or

$$F_v(z) z + \int_z^{\overline{v}} v dF_v(v) \geq F_p(z) z + \int_z^{\overline{v}} v dF_p(v)$$

Using integration by parts, we can write

$$\int_z^{\overline{v}} v dF_v(v) = \overline{v} - z F_v(z) - \int_z^{\overline{v}} F_v(v) dv$$
$$\int_z^{\overline{v}} v dF_p(v) = \overline{v} - z F_p(z) - \int_z^{\overline{v}} F_p(v) dv$$

Replacing in the above inequality implies that

$$-\int_z^{\overline{v}} F_v(v) dv \geq -\int_z^{\overline{v}} F_p(v) dv \Rightarrow \int_z^{\overline{v}} F_p(v) dv \geq \int_z^{\overline{v}} F_v(v) dv$$

Since v_Q and p_Q are equal in expectation, we must have that the above holds with equality at $z = \underline{v}$ and therefore,

$$\int_{\underline{v}}^{z} F_v(v) dv \geq \int_{\underline{v}}^{z} F_p(v) dv$$

which is the same as saying F_v is majorized by F_p. The proof of the reverse follows a similar logic and thus we establish the claim. □