Optimal Rating Design under Moral Hazard*

Maryam Saeedi
Carnegie Mellon University
msaeedi@andrew.cmu.edu

Ali Shourideh
Carnegie Mellon University
ashourid@andrew.cmu.edu

July 24, 2023

Abstract

We examine the design of optimal rating systems in the presence of moral hazard. First, an intermediary commits to a rating scheme. Then, a decision-maker chooses an action that generates value for the buyer. The intermediary then observes a noisy signal of the decision-maker's choice and sends the buyer a signal consistent with the rating scheme. Here we fully characterize the set of allocations that can arise in equilibrium under any arbitrary rating system. We use this characterization to study various design aspects of optimal rating systems. Specifically, we study the properties of optimal ratings when the decision-maker's effort is productive and when the decision-maker can manipulate the intermediary's signal with a noise. With manipulation, rating uncertainty is a fairly robust feature of optimal rating systems.

1 Introduction

Several markets rely on information disclosure or ratings to facilitate trade and incentivize quality provision. For example, ESG ratings aim to incentivize companies to improve their environmental and social impact. Similarly, certification by government or private agencies is used in the labor market for college graduates (university course grades), credit

*This paper was previously circulated under the title "Optimal Rating Design". We thank Nageeb Ali, James Best, Aislinn Bohren, Odilon Camara, Maryam Farboodi, Hugo Hopenhayn, Emir Kamenica, Navin Kartik, Alexey Kushnir, Jacopo Perego, Ilya Segal, Vasiliki Skreta, Alex Wolitzky and Ariel Zetlin-Jones as well as numerous seminar and conference participants for their helpful comments.

markets (credit scores), etc. Although these tools are increasingly used in markets suffering from moral hazard, several questions are yet to be answered: What are the key trade-offs involved in the design of rating systems when market participants can react to them? What happens when the participants can *manipulate* the information observed by the certifier in their favor? What are the certifier's incentives in designing the rating systems? What are the trade-offs between transparent and opaque rating systems?

In this paper, we answer these questions by using a parsimonious model of rating under moral hazard, and develop the theoretical techniques required to study its optimal design. In our setting, an intermediary (e.g., an ESG rating firm) observes some information about the actions of a decision maker (DM) (e.g., a company seeking an ESG rating) and decides how to convey this information to a third party (e.g., market or a buyer—henceforth "market"). We assume that the DM's actions are costly and *some* of them are valued by the market. Finally, we assume that the market is willing to pay its expected value for the item or service based on the signal it received from the intermediary and its prior.

To answer the above questions in this setting, we must first describe the set of achievable outcomes. Perhaps surprisingly, relatively little is known about this question. The key difficulty is that the DM's choice of action is endogenous and is affected by the rating system. Additionally, when this system involves random signals ("*rating uncertainty*"), the DM must form beliefs about the market's expectation; these are the DM's second-order beliefs. Our first result provides a simple characterization of the set of achievable outcomes, which relies on the concept of interim prices, defined as the DM's expected prices given the state and the market's belief. They are thus the second-order expectation of the state and, generally, cannot be fully characterized in a simple fashion. However, we obtain a sharp result when interim prices and market valuation are comonotone (i.e., move in the same direction). In this case, the existence of a signal structure is equivalent to interim prices being a mean-preserving contraction of market valuations. We can thus cast the problem of rating design as a mechanism-design problem with transfers (i.e., interim prices), where transfers must satisfy a certain feasibility constraint (i.e., second-order stochastic dominance relative to market valuations).

We then proceed to study several applications to draw on the key implications of optimal rating design under moral hazard. Our key criterion for the evaluation of rating systems is the intermediary's revenue from charging a fee for information disclosure. In our setup, non-trivial rating leads to reshuffling of payoffs across states, which is desirable for two reasons: First, there are different types of DMs, and the intermediary wishes to encourage participation to maximize revenue. Second, the DM can ex post manipulate the signals. This creates a need for non-trivial ratings, because full revelation can generate too much manipulation ex post.

Our analysis identifies three broad lessons for optimal rating design stemming directly from moral hazard. First, absent any ability to manipulate the signals ex post, optimal signal structures are deterministic, monotone, and alternating partitions (i.e., they alternate between full revelation and pooling intervals). Second, when the DM can manipulate the signal observed by the intermediary, the optimal signal structure necessarily mixes randomization and full revelation so that the interim price function is continuous. Third, the extent of the mixture between full revelation and uncertainty is determined by the cost of manipulation.

When the DM cannot manipulate ratings, the optimal rating design is determined by the shape of the weights placed on different states. When these weights are increasing in market valuation, full revelation is optimal. For arbitrary weights, optimal ratings are monotone and alternate between intervals of full revelation and no information. The number of such intervals depends on the effort costs across different DM types and the distribution of market values.

Next, we consider an environment where the DM can manipulate the signal observed by the intermediary. Specifically, the DM can alter the realization of market valuation at a cost. A trade-off between effort and manipulation exists. In order to incentivize the DM to exert effort, interim prices should be higher for higher realizations of market values, but this can lead to manipulation. Hence, withholding some information reduces the incentives for manipulation. The optimal rating system balances the incentives for effort with the cost of manipulation; so, allowing for manipulation in equilibrium can provide incentives for (ex ante) effort.

Assuming that the cost of manipulation is a quadratic function of the difference between market valuation and the intermediary's observed signal, we show that interim prices should be continuous, so there should be no price jump as a function of observed market valuations. Thus, the rating system should have randomness as its main feature. This is in contrast with our result from the case where there was no possibility of manipulation because the optimal rating system was monotone partitions.

Finally, under some distributional assumptions, we can fully characterize optimal rating systems in the presence of costly manipulation. More specifically, we show that any optimal rating system requires an interval of intermediate values for which the state is perfectly revealed. For extreme values of the state, the optimal rating system is random and involves uncertainty. The extent of revelation depends on the cost of manipulation. Perhaps counterintuitively, as this cost declines, the full-revelation interval expands, mainly because of the value of manipulation for the provision of ex ante incentives. When this cost is low, a high degree of manipulation and a high dispersion of interim prices are required to implement a particular ex ante effort. Because interim prices need to be less dispersed than market valuations, full revelation for the intermediate values of the state is required.

Beyond its technical contributions, our paper has implications for the regulation and design of rating systems and certification. Rating uncertainty can be interpreted as the opaqueness of the rating system, where it is not fully disclosed by the intermediary. There are various examples of rating opaqueness: One is in the context of consumer credit scores—though the rough statistics that increase these scores can be determined, their cutoffs and formulas are unclear. Another is in the context of e-commerce platforms: In an experiment using eBay data, Nosko and Tadelis (2015) use a particular measure of seller quality in the search result ranking without the sellers' knowledge, showing it can improve eBay's reputation system. Our analysis sheds light on the optimal design of such experiments where full-revelation rating systems should be combined with random ones.

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 (2019), and Perez-Richet and Skreta (2022)). In contrast with Ball (2019) and Frankel and Kartik (2019), our mathematical result on second-order expectations allows us to study a larger class of problems without any restrictions on information structures. Our analysis, thus, identifies both the precise shape of the optimal information structure and when it is optimal to use uncertain rating systems. Furthermore, our application where the DM can manipulate the outcome observed by the intermediary is similar to the falsification model in Perez-Richet and Skreta (2022). However, given that we use different optimal information structures, our analysis emphasizes the need for information revelation. Whereas in Perez-Richet and Skreta (2022), the main goal of information revelation is a more accurate decision by an uninformed receiver, in our setup, it is to induce a more efficient effort choice by the DM.

A closely related paper to ours is Boleslavsky and Kim (2020). They study a model of Bayesian persuasion with moral hazard, similar to ours, in which an agent (corresponding to our DM) chooses an effort level that affects the distribution of the state, and a sender affects a receiver's action using an information structure. However, they study this problem using the concavification method of Kamenica and Gentzkow (2011), but because it relies on formulating the problem in terms of the beliefs, it is generally difficult to handle large state spaces. In contrast, our mathematical result in Section 3 allows us to use standard optimal control methods and work with arbitrary state spaces.[1]

Our paper is also related to the Bayesian persuasion literature (e.g., Kamenica and Gentzkow (2011), Dworczak and Martini (2019), and Kolotilin et al. (2023)). Theoretically, our solution method is closer to that of Dworczak and Martini (2019) and Kolotilin et al. (2017), because our characterization result in Section 3 allows us to pose the information design problem using majorization ordering and use optimal control techniques. In terms of information structure, our optimal rating design with rating uncertainty is different from most of the literature and it identifies the role of moral hazard.

From a technical perspective, our results are related to a result in mathematics and statistics shown by Strassen (1965), among others, on the existence of joint distributions with given marginals. However, our paper is different because we are concerned with second-order expectations, which are not straightforward marginals of the joint distribution of the signal and the state. Moreover, to the extent that our result holds only under a particular assumption about the correlation of interim prices and the state (i.e., comonotonicity), it is novel to this literature.

In our formulation, we use the majorization ranking for the functions representing interim prices and action profiles by the DM. Thus our rating design problem is equivalent to a mechanism design problem with transfers in which the transfer function majorizes the market valuations function. Similar to this problem, Kleiner et al. (2021) solve a class of problems where majorization appears as a constraint. Their solution method uses the characterization of extreme points of the set of functions that majorizes a certain function. Although we use their result in our first application, in general, the presence of incentive

[1]Unlike Boleslavsky and Kim (2020), in most of our analysis, we assume that the DM is paid the value of market's expectation. In the online Appendix, we show how to extend our analysis to allow for an action by the market.

constraints makes their approach not exactly applicable to our environment. In contrast, our solution of the mechanism design problem involves the calculus of variations because of the lack of linearity in our model.

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), and Zapechelnyuk (2020).[2] 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 (2021) also study variants of this problem, but they focus on *deterministic* rating systems. As we will show, rating systems with uncertainty are important when signals can be manipulated.[3]

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. (2020) and Nosko and Tadelis (2015)), health insurance markets (Vatter (2022)), 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 rest of the paper is organized as follows: We start with an example in Section 2; in Section 3 we set up the model; in Sections 4 and 5 we describe two applications of the model; in Section 6 we consider some extensions of our model; finally, in Section 7 we present our conclusions. All the proofs are relegated to the Appendix unless otherwise indicated.

2 Simple Example

Before diving into the model, we explore a simple example which illustrates the trade-offs involved in rating under moral hazard. A DM chooses to exert effort $a \in [0,1]$, which determines the distribution of quality, $y \in [0,1]$, given by the following distribution:

$$G(z|a) = \Pr(y \le z|a) = z^{\frac{1}{1-a}-1}.$$

[2] Our paper is also related to the extensive and growing literature that studies the problem of certification and information disclosure absent moral hazard (e.g., Lizzeri (1999) and Hopenhayn and Saeedi (2020) in static settings and Horner and Lambert (2020) and Kovbasyuk and Spagnolo (2023).)

[3] 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.

The DM can be of two possible types, $\theta \in \{1/2, 4/5\}$, which determines the cost of exerting effort, $c(a,\theta) = a^2/(2\theta)$. Suppose that $\Pr(\theta = .5) = 1/2$ and that θ is private information to the DM, not observed by the other market participants.

An intermediary, such as a certifier or a platform, observes the realization of quality y and commits to an information structure $(S, \pi(\cdot|y))$, where $\pi(\cdot|y) \in \Delta(S)$.[4] The intermediary charges the DM a tariff, t, for this information. The market's payoff is y, but it observes only the signal realization sent by the intermediary. It then uses its prior and the signal from the intermediary to update its beliefs, and pays its posterior mean, $\hat{p}(s) = \mathbb{E}_\pi[y|s]$ to the DM. The DM's payoff from choosing effort a is thus given by

$$\int_0^1 \int_S \hat{p}(s)\, \pi(ds|y)\, dG(y|a) - t - c(a,\theta).$$

Suppose that the DM's outside option has a payoff of 0. The intermediary wishes to maximize its own revenue by choosing the optimal signal structure and the tariff level.

First, consider the full-revelation rating system, i.e., $S = [0,1]$, $\pi(\{y\}|y) = 1$. In this case, $\hat{p}(y) = y$. Hence, conditional on participation, the marginal benefit of effort is $\int y g_a dy = 1$. Setting this equal to marginal cost, each DM's profit is maximized at $a_L^* = 1/2, a_H^* = 4/5$. Given these choices, the before-tariff payoffs of each DM type are $u_L = 1/4$ and $u_H = 2/5$. This implies that the intermediary can charge either u_L and both types participate or u_H and only $\theta = 4/5$ participates. Comparing the two cases shows the optimal tariff and the intermediary's expected revenue are given by $t = 1/4$ and $\text{Revenue} = 1/4$, respectively.

Next, we examine if the intermediary can increase its revenue by hiding some information. Consider a partial pooling rating system where the intermediary pools the realizations of y above 0.353 and fully discloses the lower realizations. Pooling high realizations of y allows a reshuffling of profits from the low-cost DM to the high-cost one while changing the DM types' incentives for exerting effort. This effect can potentially lead to a higher level of tariff that can keep both DM types engaged. The equilibrium efforts under this rating system are $a_L^* = 0.458, a_H^* = 0.59$, $u_L^* = 0.259 = \text{Revenue}$, which is higher than the revenue of the intermediary under full information.

Several questions arise from this exercise. Given that it is optimal for the intermediary to hide some information, what is the optimal information structure? Would randomized signals ever be optimal? What happens when the DM can manipulate the signal observed

[4] We use $\Delta(S)$ to denote the set of all Borel probability measures over the set S.

by the intermediary? In what follows, we develop techniques for a general solution of this problem. In Section 4, we show that the optimal information structure pools values of y for some intermediate interval and otherwise it is fully revealing.

3 General Model and Interim Prices

We now describe our general model of rating design and provide a sharp characterization result for the set of feasible payoffs. In general, we are interested in settings in which an intermediary observes some information about a DM's chosen actions and decides how to convey this information to a third party, henceforth "the market", who then pays its posterior mean as a price to the DM.

More specifically, consider a DM who chooses an action $a \in A \subset \mathbb{R}^N$, which creates a possibly random realization $y \in Y \subset \mathbb{R}^M$ with distribution $\sigma(\cdot|a) \in \Delta(Y)$. The action and the outcome generate a value of $v(a, y)$ for the market, who then pays its expected payoff $\mathbb{E}[v(a, y)|s]$ to the DM. This expectation is calculated using the information available, s, and the common belief about equilibrium play.[5] The intermediary first commits to a signal structure $(S, \pi(\cdot|y))$, where S is a set of signal realizations and $\pi(\cdot|y) \in \Delta(S)$ for all $y \in Y$.[6] Subsequently and with a full knowledge of the intermediary's choice, the DM chooses her action, which generates a signal for the intermediary and the market. Figure 1 depicts the structure of the model and actions.

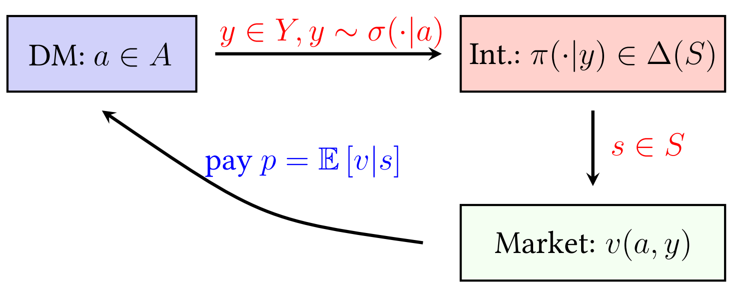

Figure 1: General structure of the model

[5]We maintain the assumption that the buyers are on the long side of the market, thus willing to pay their expected value. One can extend our analysis by allowing the market to have positive outside options or positive bargaining power.

[6]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 an 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 and a compact space for S. Hence, we drop the σ-algebra in our analysis. Additionally, when describing subsets, we refer to Borel subsets.

The DM has a type $\theta \in \Theta$, with the probability distribution given by $F \in \Delta(\Theta)$.[7] Type θ affects the cost of exerting effort, $c(a, \theta)$. Hence, the DM's payoff is given by

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

In a pure-strategy equilibrium, the DM chooses $a(\theta)$ to maximize (1).

In the above, the ex post market price, $\mathbb{E}[v|s]$, not only depends on the information structure, $\pi(\cdot|\cdot)$, but also on the market's prior about the distribution of (a, y), which depends on the DM's strategy profile (i.e., is an equilibrium object). More specifically, the market uses its prior about the distribution of θ and its beliefs about the equilibrium strategies of the DM types, $a(\theta)$, to form a prior $\mu \in \Delta(A \times Y)$ and uses Bayes rule to form the posterior expectation $\mathbb{E}[v|s]$ satisfying

$$\int_{A\times Y} \int_{S'} \mathbb{E}[v|s]\, d\pi(s|y)\, d\mu = \int_{A\times Y} v(a, y)\, \pi(S'|y)\, d\mu, \forall S' \subset S. \tag{2}$$

The above defines an equilibrium given the information structure. More specifically, given an information structure (S, π), an equilibrium is an action profile $a(\theta)$ by different DM types where given market beliefs $\mu \in \Delta(A \times Y)$ and $\mathbb{E}[v|s]$, $a(\theta)$ maximizes (1), and given $a(\theta)$, the market beliefs satisfy

$$\mu(A' \times Y') = \int_{A'} \int_{Y'} d\sigma(y|a(\theta))\, dF(\theta), \forall A' \subset A, Y' \subset Y$$

together with Bayesian updating as defined in (2).

Examples

To clarify the scope and applicability of our analysis, we describe a few examples of the above environment:

1. **Reputation mechanisms in online platforms:** Online platforms such as Airbnb and eBay face challenges in designing their reputation systems because of adverse selection and moral hazard. These platforms have access to performance data about

[7]We will often assume that $\Theta \subset \mathbb{R}$ has either a discrete distribution over a finite set of types or a continuous distribution with c.d.f. F. Using F to denote the probability measure governing θ is a slight abuse of notation, to avoid clutter.

providers (i.e., hosts on Airbnb and sellers on eBay) not available to the market.[8] The platform's certification policy, such as Airbnb's Superhost or eBay's Top Rated Seller, is based on performance measures. According to Hui et al. (2020), the details of this policy influence provider behavior, and they can be seen as the information structure in our model. Our model examines the resulting issues and trade-offs for both platform and providers. Reviews often serve as signals of past experiences, and to capture this, we can set $v(y, a) = a$ according to which past signals' performance measures are only a signal of the value created for the market.

2. **Ratings in asset markets:** Certification in financial markets relies on proprietary data and forecasting models used by rating agencies (e.g., Moody's, Fitch, and S&P). The issuer pays for the rating, which is then made public. One can thus view our analysis as the effect of credit rating models on issuer behavior. While certain efforts to increase value for bond holders may not be fully observable by rating agencies, their models still influence issuer behavior, as in our model. An important topic is the issue of regulation of the credit rating models, as discussed in Rivlin and Soroushian (2017). We demonstrate that a deterministic rating model is desirable when manipulation is not a concern (see Section 4). However, in the presence of manipulation, rating uncertainty becomes desirable, suggesting regulators should allow for some degree of uncertainty in certification policies (see Section 5). Our model provides insights into the nature of this opacity.

3. **Manipulation of ratings:** Rating-system manipulations are widespread, often involving misrepresentation of data by the certified party.[9] Online platforms are frequently plagued by data manipulation by providers.[10] For example, some third-party sellers on Amazon pay customers for positive reviews and higher ratings, He et al. (2022). In our model, the decision-maker can take costly actions to increase an observed indicator y without affecting market valuation. This creates a trade-off

[8] As documented by Saeedi (2019), Hui et al. (2016), and Nosko and Tadelis (2015), there are many performance indicators available to eBay that are not conveyed to the market directly, such as total quantities sold, and previous claims and their outcomes.

[9] In recent year, several lawsuits have involved rating manipulation in different industries, such as education (e.g., the case of Temple Unniversity, Temple Business School Dean Fraud, accessed August 16, 2022, and the case of Columbia University in NYT on Columbia's ranking manipulation and Michael Thadeuss on ranking manipulation, accessed August 16, 2022) and financial markets (e.g., the case of Greenwashing by Deutsche Bank, accessed August 16, 2022).

[10] Feedback manipulation has long been a debated issue on e-commerce platforms (e.g., Hui et al. (2017)).

in rating design, because information provision incentivizes productive actions but also raises incentives for data manipulation (details in Section 5).

3.1 Interim Prices: Definition and Characterization

In this section, we introduce a mathematical object, interim prices, to simplify the problem of rating design in the environment described above. Our first major result is a simple characterization of these interim prices that allows us to solve the problem of rating design in various applications.

The notion of interim prices is simple. They are the mathematical object that determines the DM's incentives in choice of effort and will be present in the incentive constraints for the DM. Specifically, we define *interim prices* as

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

Given that $\mathbb{E}[v|s]$ is an equilibrium object that depends on the market's beliefs about the DM's action profile, so is $p(y)$. Nevertheless, it is a sufficient statistics for the information structure from the DM's perspective. Specifically, the DM's payoff is given by

$$\int p(y) \, d\sigma(y|a) - c(a, \theta).$$

Interim prices are essentially the DM's beliefs about the beliefs of the market (or buyer)—second-order beliefs. More precisely, at the interim time of y realization and before signal realization, the DM faces a distribution over the realization of signals—when random signals are used—and thus over the market's beliefs. One can thus interpret them as second order beliefs of the DM.

Example 1. The following examples give a sense of interim prices and their relationship with an information structure. Suppose that $A = Y = [0,1], v(a,y) = y$; that is, the market values only the DM's signal realization, and the intermediary observes it. Then an example of an information structure with deterministic signals is one in which the values of $y \in [\underline{y}, \overline{y}]$ are revealed while those above and below this interval are pooled. In this case, interim price $p(y)$ coincides with y when $y \in [\underline{y}, \overline{y}]$, and is the conditional mean of y when $y \leq \underline{y}$ for lower values; higher values are similar.

Another example of an information structure is a *partially mixing* one in which the state is revealed with probability $\alpha(y)$, and otherwise a generic signal with probability $1-\alpha(y)$ is sent. Here, the DM faces uncertainty regarding its ratings, and the interim prices are given by

$$p(y)=\alpha(y)\,y+(1-\alpha(y))\,\frac{\int[1-\alpha(y)]\,yd\mu}{\int[1-\alpha(y)]\,d\mu}.$$

Given our definition of interim prices, instead of viewing an equilibrium as an action profile $a(\theta)$ and the distribution of market prices, $\mathbb{E}[v|s]$, it induces, we can view it as an action profile $a(\theta)$ and an interim price function $p(y)$. Evidently, given $p(y)$, $a(\theta)$ must be incentive compatible, i.e.,

$$a(\theta)\in\arg\max_{a\in A}\int p(y)\,d\sigma(y|a)-c(a,\theta).$$

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.

First, we examine interim prices. Suppose that the sets Y, A, and S are finite so we can easily write conditional expectations. Interim prices are given by

$$p(y)=\sum_{s\in S}\frac{\sum_{(a,\hat{y})\in A}v(a,\hat{y})\,\mu(a,\hat{y})\,\pi(s|\hat{y})}{\sum_{(a,\hat{y})\in A}\mu(a,\hat{y})\,\pi(s|\hat{y})}\pi(s|y).$$

Since $\sum_s\pi(s|y)=1$, $p(y)$ is a weighted average of the values of $v(a,\hat{y})$, where the weights depend on y. The above can be written as

$$p(y)=\sum_{\hat{y}\in Y}\overline{v}(\hat{y})\sum_{s\in S}\frac{\pi(s|\hat{y})\,\pi(s|y)\,\mu_y(\hat{y})}{\sum_{y'\in Y}\mu_y(y')\,\pi(s|y')},\tag{4}$$

where $\overline{v}(\hat{y})=\mathbb{E}[v(a,y)\,|\hat{y}]$ is the mean of $v(a,y)$ conditional on the realization of y, and $\mu_y(\hat{y})=\sum_{a\in A}\mu(a,\hat{y})$ is the marginal distribution of μ along the y-direction. We make the following assumption about $\overline{v}$:

Assumption 1. *The range of $\overline{v}(\cdot)$, i.e., $\overline{v}(Y)$, is a finite collection of closed subintervals of $\mathbb{R}$.*

Assumption 1 is a technical assumption that allows us to prove our main result on the characterization of interim prices, Theorem 1. It holds, for example, if Y is a finite collection of disjoint connected sets and $\overline{v}(\cdot)$ is continuous.

According to (4), $p(y)$ is a weighted average of $\overline{v}(y)$. Hence, $p(y)$ is a less dispersed version of $\overline{v}(y)$, i.e., a mean-preserving contraction. Indeed, we have the following lemma:

Lemma 1. *For any information structure* (S, π) *and* $p(y)$ *defined by (3),* $p(\cdot)$ *second-order stochastically dominates* $\overline{v}(\cdot)$*, i.e., for all concave and increasing functions* $u : \mathbb{R} \to \mathbb{R}$*,*

$$\sum_{y\in} \mu_y(y)\, u(\overline{v}(y)) \leq \sum_{y\in} \mu_y(y)\, u(p(y))$$
$$\sum_{y\in} \mu_y(y)\, \overline{v}(y) = \sum_{y\in} \mu_y(y)\, p(y)\,.$$

While the above result is a necessary requirement for interim prices, in general its reverse is not true, as the following example shows:

Example 2. Suppose that $A = Y = \{0, 1, 3\}$, $v(a, y) = \overline{v}(a) = a$, $\sigma(Y'|a) = \mathbf{1}[a \in Y']$, and $\mu(\{a\}) = 1/3$. In words, the consumer cares only about the action of the seller, and y coincides with it. Figure 2 depicts the values of $p(0)$ and $p(1)$; the sum of the three interim prices always equals 4, given Bayes rule. Area A shows the set of vectors $\mathbf{x} = (x_1, x_2, x_3)$ that second-order stochastically dominate $(0, 1, 3)$. Each random variable is represented by (x_1, x_2); the third element is the distance from the $x_1 + x_2 = 4$ line. The conditions are $0 \leq x_i \leq 3, 1 \leq x_i + x_j \leq 4$, for all i, j, and $x_1 + x_2 + x_3 = 4$. However, the set of interim prices does not coincide with the set A, and is depicted by set B.[11] Moreover, interim prices are not necessarily monotone. A signal that pulls $a = 0, 3$ and reveals $a = 1$ leads to an interim price of $3/2$ for $a = 0$ and 1 for $a = 1$, depicted by point c in Figure 2.

The above example illustrates the difficulties associated with identifying the set of all interim prices for all information structures. Nevertheless, we can show a somewhat general result when market valuations have the same ranking as interim prices. Our main mathematical result is that when market valuations $\overline{v}(y)$ and $p(y)$ are comonotone, then the existence of a signal structure is equivalent to second-order stochastic dominance, as stated by the following theorem:

[11] To find the set of all interim prices, one can characterize the extreme points of B. These points are associated with full revelation of some of the states and pooling of others.

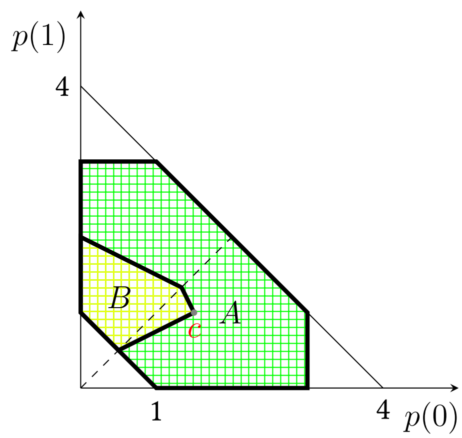

Figure 2: The set of interim prices and mean-preserving contractions of market valuations for Example 2. The green area, A, represents the three state random variables that are a mean-preserving contraction of a. The yellow area, B, is the set of interim prices arising from some information structure. The point c depicts a non-monotone interim price vector.

Theorem 1. *Consider an action profile $a(\theta)$ and its associated $\overline{v}(y)$ as defined in (4). Suppose that $p(y)$ is a function that maps Y into $\mathbb{R}$ such that*

1. *$p(\cdot)$ is comonotone with $\overline{v}(\cdot)$, i.e., $p(y) > p(y') \Rightarrow \overline{v}(y) > \overline{v}(y')$, and*
2. *$p(\cdot)$ second-order stochastically dominates $\overline{v}(\cdot)$.*

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

To see the intuition, suppose that $Y = \{y_1, \cdots, y_n\}$ is a finite set and $\overline{v}(y_{i+1}) > \overline{v}(y_i)$. We show the result by induction on the size of Y. The result is trivial for $|Y| = 2$, as we can generate any monotone interim price by sending a low signal when the state is y_1 and a high signal with some probability when the state is y_2. For any finite set Y, we consider the state y_j that has the lowest value of $\lambda = \frac{p(y_{j+1}) - p(y_j)}{\overline{v}(y_{j+1}) - \overline{v}(y_j)}$. We then consider the convex combination $p(y_i) = \lambda \tilde{p}(y_i) + (1-\lambda)\overline{v}(y_i)$. Given the value of λ, it can be shown that $\tilde{p}(y_j) = \tilde{p}(y_{j+1})$. We can thus consider a modified state space $\tilde{Y} = Y \setminus \{y_{j+1}\}$ and set the market value at y_j equal to average of $\overline{v}(y_j)$ and $\overline{v}(y_{j+1})$. For such market values and state space, it can be shown that $\tilde{p}(y)$ is comonotone with and second-order stochastically dominates $\overline{v}(y)$. Hence by the hypothesis of induction, $\tilde{p}$ can be generated by a rating system. The rating system that generates p is one that randomizes between full revelation (with probability $1-\lambda$) and a the signal for $\tilde{p}$ while pooling the states y_j

and y_{j+1}.Note that, this proof effectively identifies an algorithm for construction of one rating system by a sequence of randomizations between full revelation and signals that pool consecutive states.

For arbitrary compact Y, we approximate the distribution $\mu_y(\cdot)$ with a sequence of discrete distributions whose supports are ordered according to the subset order (i.e., they are a filtration). We can then apply the result from the finite case to construct an information structure associated with each of these discrete approximations. The main result then follows from the compactness of the space of measures over the posterior mean and y and the use of the martingale convergence theorem. We formalize this argument in the Appendix.

The above theorem implies that we can characterize the comonotone equilibria of the game for arbitrary information structures with an action profile $\{a(\theta)\}_{\theta\in\Theta}$ and interim prices $p(y)$ such that:

1. The action profile is incentive compatible,

$$a(\theta) \in \arg\max \int p(y)\, d\sigma(y|a) - c(a,\theta), \forall\theta \in \Theta. \tag{5}$$

2. Interim prices $p(y)$ dominate $\overline{v}(y) = \mathbb{E}[v(a,y)|y]$ according to the second-order stochastic order.

3. Interim prices and market valuations are comonotone.

Remark on Theorem 1 The result in Theorem 1 is reminiscent of the result of Blackwell (1953) and Rothschild and Stiglitz (1970), whose general version 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 ways. 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 $\mathbb{E}[\overline{v}|s]$ must be used to generate $\mathbb{E}[\mathbb{E}[\overline{v}|s]|y]$. Second, as illustrated by Example 2, the equivalent of Blackwell's result does not hold in general and can be shown only when $\overline{v}$ and p are comonotone.

Majorization In the rest of the paper, we will use Theorem 1 to characterize optimal rating systems in various applications. When $Y \subset \mathbb{R}$, the majorization formulation (see Hardy et al. (1934)) of second-order stochastic dominance helps us use a Lagrangian method to solve for the optimal rating systems. When $Y = \mathbb{R}$ and both $\overline{v}(y)$ and $p(y)$ are increasing and comonotone, we can write

$$p \succcurlyeq_{SOSD} \overline{v} \iff \int_{-\infty}^{y} p(\hat{y})\, d\mu_y(\hat{y}) \geq \int_{-\infty}^{y} \overline{v}(\hat{y})\, d\mu_y(\hat{y}), \forall y \in \mathbb{R}. \tag{6}$$

4 Application 1: Rating Design without Manipulation

Our first application of rating design is a general version of the model in Section 2, where effort generates a random market valuation that is observable by the intermediary and cannot be manipulated. We will show that revenue-maximizing rating systems are deterministic. We will then provide a specific characterization of optimal rating systems and how they depend on the outcome distribution.

More specifically, suppose that $A = [\underline{a}, \overline{a}]$ for some $\overline{a} > 0$, $Y = [0,1]$, $v_a(y,a), v_y(y,a) \geq 0$, and $\Theta = \{\theta_1, \cdots, \theta_m\}$. We assume that the cost function $c(a,\theta)$ is decreasing in θ, increasing in a, and submodular. This implies that higher types (θ) are more efficient in exerting effort. Additionally, we assume that y conditional on a is distributed according to a continuously differentiable (c.d.f.) $G(y|a)$. We make the following assumption:

Assumption 2. *The distribution of y conditional on a satisfies the following conditions:*

1. *Its likelihood function is monotone increasing, i.e.,* $\frac{\partial^2}{\partial a \partial y} \log g(y,a) = \frac{\partial}{\partial y}\frac{g_a}{g} \geq 0.$
2. *For any finite sequence $(a_1, \cdots, a_K)$ and any nonempty interval $I \subset [0,1]$, the functions $\{g(y|a_i), g_a(y|a_i)\}_{1 \leq i \leq K}$ are linearly independent over I.*

This first part of Assumption 2 is the standard MLRP (monotone likelihood ratio assumption) common in the moral hazard literature, and it will help us show that optimal interim price functions should be increasing. Its second part is used to show that the optimal interim price function has to be an *extreme point* of the set of functions that are a mean-preserving contraction of $v(a,y)$ (see Kleiner et al. (2021)).

We aim to characterize revenue-maximizing information structures. Given any fee charged by the intermediary, t, and considering $c(a,\theta)$ is submodular and decreasing in θ, there must exist a marginal type $\hat{\theta}$ below which the DM does not participate, because the DM's profit becomes negative. The intermediary's revenue is $t\sum_{\theta\geq\hat{\theta}} f(\theta)$. Hence, given $\hat{\theta}$, t coincides with the utility of the marginal type, $\int pdG - c\left(a,\hat{\theta}\right)$. Thus, given $\hat{\theta}$, the problem of optimal rating is to choose $p(y)$ and $a(\theta)$ to maximize

$$\max_{p(y),a(\theta)} \int p(y)\, dG\left(y|a\left(\hat{\theta}\right)\right) - c\left(a\left(\hat{\theta}\right),\hat{\theta}\right)$$

subject to the incentive compatibility constraint (5), monotonicity of $p(y)$, and majorization constraint as stated in (6). In the above, we have assumed monotonicity of $p(\cdot)$; in the online Appendix, we establish that this is without loss of generality.[12]

Our first result illustrates that Pareto optimal rating rules are always deterministic:

Proposition 1. *Suppose that Assumption 2 holds and that the first-order approach is valid. Then the revenue-maximizing rating system is a monotone partition.*

The idea behind the proof of this theorem can be understood by realizing that under the first-order approach, Lagrange multipliers $\gamma(\theta), \theta \geq \hat{\theta}$, must exist such that the choice of optimal $p(\cdot)$ is equivalent to maximizing

$$\int \left[g\left(y|a\left(\hat{\theta}\right)\right) + \sum_{\theta\geq\hat{\theta}} \gamma(\theta)\, g_a(y|a(\theta))\right] p(y)\, dy \tag{7}$$

subject to monotonicity and majorization. In other words, $p(\cdot)$ must be an extreme point of the monotone orbit of y. Hence the result of Kleiner et al. (2021) applies and the result in Proposition 1 follows. The second part of Assumption 2 is used to show that the maximum is indeed achieved by an extreme point and not a support point.

We assume that the first-order approach is valid, because in our proof of this theorem, we use the existence of Lagrange multipliers on the incentive compatibility constraint. The first-order approach ensures that such Lagrange multipliers exist and our proof is valid. In the online Appendix, we provide sufficient conditions on the distribution function $G(y|a)$ so that the first-order approach is valid.

This proposition illustrates that not only optimal rating systems are *deterministic* but they are also *monotone partitions*. In other words, for any y, either y is revealed perfectly or there exists an interval around y where it is revealed that y belongs to this interval. In what

[12] If $p(\cdot)$ is decreasing over a subinterval followed by an increasing segment, we can increase prices for low values of y and decrease them for high values of y. This keeps the incentives intact and redistributes profits to the lowest type. If, on the other hand, $p(\cdot)$ is hump-shaped, a similar perturbation works.

follows, we further investigate the properties of optimal rating systems by considering certain classes of distributions.

Before proving our formal results, we provide a heuristic analysis of the trade-offs involved in optimal rating. As discussed above, finding the optimal interim price is equivalent to solving the following optimization:

$$\max_{p(\cdot)} \int \Gamma(y)\, p(y)\, h(y)\, dy$$

subject to the majorization and monotonicity constraints. In this formulation, $h(\cdot)$ is the measure of y among the types that participate, i.e., $h(y) = \sum_{\theta \geq \hat{\theta}} f(\theta)\, g(y|a(\theta))$. The function $\Gamma(y)$ is the *gain* function derived in 7 and is given by

$$\Gamma(y) = \frac{g\left(y|a\left(\hat{\theta}\right)\right)}{h(y)} \left(1 + \sum_{\theta \geq \hat{\theta}} \gamma(\theta) \frac{g_a(y|a(\theta))}{g\left(y|a\left(\hat{\theta}\right)\right)}\right). \tag{8}$$

When $g(\cdot)$ satisfies the monotone likelihood ratio property, then the ratio $g\left(y|a\left(\hat{\theta}\right)\right)/h(y)$ is a decreasing function of y. Hence, this term captures the *redistributive* forces present in the design of rating systems. The second term in the gain function represents the importance of the *incentive* provision for the participating types. When $g(\cdot)$ satisfies the monotone likelihood ratio property, this term is increasing, which then creates benefits for full revelation.

This implies that $\Gamma(\cdot)$ can be thought of as somewhat of a *sufficient statistic* that determines the optimal disclosure policy. The function $\Gamma(\cdot)$ is not an exact sufficient statistic because the effort levels chosen by DM types are endogenous to the rating system, which in turn affects $\Gamma(\cdot)$.

Despite the issue of endogeneity, $\Gamma(\cdot)$ can be a useful determinant of optimal rating systems for two reasons: First, given an allocation of efforts, it can determine a local perturbation that improves the objective. Second, it is possible to provide general properties for its shape for a certain class of distribution functions $g(\cdot|\cdot)$. In what follows, we provide a few characterization results in this direction, to help us understand better when the two forces present in Γ dominate each other.

In general, the key relevant property of Γ is the sign of its derivative. In Lemma (2) in the Appendix, we show that if $\Gamma'(y)$ changes sign k times, then the optimal interim price is comprised of at most k partitions alternating between full revelation and pooling.

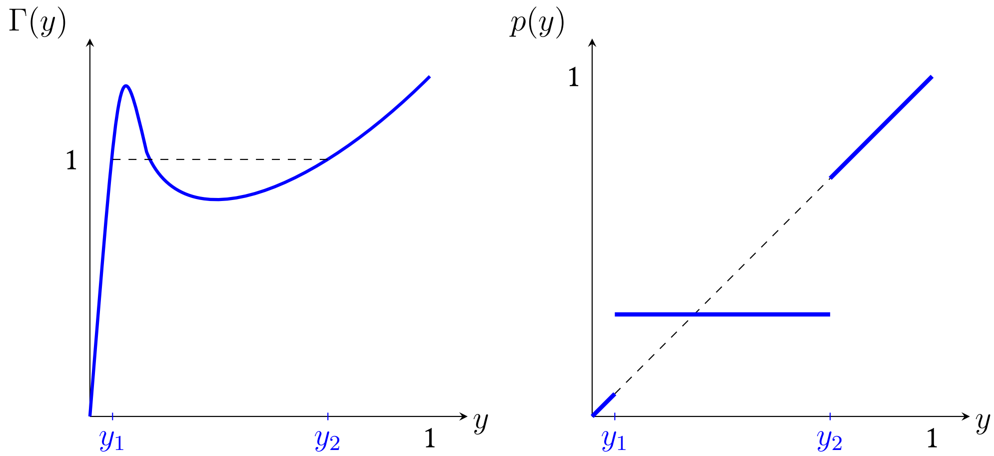

Figure 3: Right figure shows The gain function for the example in Section 2; the optimal efforts are $a_L^* = 0.401, a_H^* = 0.635$. Left figure shows the optimal interim price.

Moreover, full revelation happens only if $\Gamma' > 0$. In what follows, by assuming that there are two types of DM, we present some sufficient properties of the distribution function $g(\cdot)$ that limits the number of times Γ' has a change of sign. As a result, optimal ratings take a simple form.

Proposition 2. *Suppose that $G(\cdot)$ satisfies Assumption 2, and its log-density satisfies $\log g = r(y) m(a) + b(a)$, where r and a are increasing functions. If $m = 2$, then there exist two thresholds $y_1 < y_2$ such that the optimal rating is pooling for values of $y \in [y_1, y_2]$ and fully revealing for values of y below y_1 and above y_2.*

Under the assumption in Proposition 2, the incentive effect in the gain function dominates for high and low values of y, and the redistributive effect dominates for mid-values. A special case is that of the example in Section 2, where $\log g = (1/(1-a) - 2)\log y + \log(1/(1-a) - 1)$. In that example, Proposition 2 implies that we need to consider only mid-pooling information structures, and the optimal one has the bounds 0.0492 and 0.6636; the revenue of the intermediary is 0.269. The gain function and the optimal interim price are depicted in Figure 3, which illustrates the optimality of this information structure.

In sum, our first application highlights the importance of our characterization result in 3. Theorem 1 allows us to simply formulate the problem and identify the key determinants of optimal rating systems. The key determinant is the shape of the likelihood function

g_a/g, similar to the literature on moral hazard following Holmström (1979). Moreover, the likelihood function affects the optimal rating through its effect on the shape of the gain function defined in (7). [13]

5 Application 2: Rating Design with Manipulation

In this section, we consider another application of our result in Section (3) to a setting where the DM can manipulate the statistic observed by the intermediary. More specifically, consider a special case of the model in Section (4) where there is only one type, θ. Suppose that market valuation is $v(y,a) = y$ and that $Y = [0,1]$, and the intermediary does not observe the true realization of y, but instead observes x, which the DM can manipulate at a cost. In particular, the DM, after observing y, can pay a cost and reveal x to the intermediary. The cost of manipulation is given by $c_m(x-y) = k\frac{(x-y)^2}{2} + \tau|x-y|$, where $k, \tau \geq 0$ and $\tau < 1$. We assume that the intermediary wishes to maximize the surplus generated by the DM, which is given by $\int [y - c_m(\hat{x}(y) - y)]\, dG(y|a) - c(a)$, where $\hat{x}(y)$ is the manipulation strategy of the DM for each realization of $y \in [0,1]$.[14]
In this setup, a rating system is a signal structure $(S, \pi(s|x))$, i.e., an information structure that maps manipulated values x into signals for the market (or buyers). In equilibrium, there is common knowledge of strategies by the DM and thus the market's interpretation of the signals depends on the manipulation strategy of the DM. This updating takes the form of

$$\mathbb{E}[y|s] = \mathbb{E}\left[\hat{x}^{-1}(x)|s\right],$$

where $\hat{x}^{-1}$ is the inverse correspondence of the DM's equilibrium manipulation strategy.[15]
The DM's interim price from reporting x' to the intermediary is given by

$$\mathbb{E}[\mathbb{E}[y|s]|x'] = \mathbb{E}\left[\mathbb{E}\left[\hat{x}^{-1}(x)|s\right]|x'\right] = \hat{p}(x'). \tag{9}$$

The payoff function of the DM being supermodular between x and y implies that the

---

[13] In the online Appendix, we provide another example of a separable distribution function $g(y,a) = A_1(a) + A_2(a)r(y)$. In this case, the gain function can be shown to be always monotone.

[14] This is a special case of the setup in Section 3, where the valuation of the market, y, is not observed by the intermediary, who instead observes a signal x, controlled by the DM.

[15] When there are multiple y's that report x to the intermediary, set $\hat{Y} \subset [0,1]$, the conditional expectation $\mathbb{E}\left[\hat{x}^{-1}(x)|s\right]$ pools them together and treats them as one observation, with its value given by the conditional expectation of $\mathbb{E}\left[y|y \in \hat{Y}\right]$.

equilibrium manipulation function $\hat{x}(y)$ is increasing in y. Moreover, equilibrium interim price $\hat{p}(\hat{x}(y))$ is also increasing. Therefore, the comonotonicity assumption of Theorem 1 holds. Hence, the existence of a signal structure that satisfies (9) is equivalent to $p(y) = \hat{p}(\hat{x}(y))$ dominating y according to second-order stochastic dominance. Thus, we have the following corollary:

Corollary 1. *Consider any manipulation strategy $\hat{x}(y)$ together with an information structure (π, S). Then $\hat{x}(\cdot)$ is an equilibrium strategy if and only if there exists an increasing interim price function, $p(y)$, such that:*

1. *The function $p(y)$ second-order stochastically dominates y (given the distribution of y, $G(y|a)$).*
2. *The pair of functions $p(\cdot), \hat{x}(\cdot)$ satisfies the incentive compatibility*

$$p(y) - c_m(\hat{x}(y) - y) \geq p(y') - c_m(\hat{x}(y') - y), \forall y, y' \in [0,1]. \tag{10}$$

Corollary 1 implies that the problem of optimal rating design in this application is given by

$$\max_{p(y), \hat{x}(y), a} \int [p(y) - c_m(\hat{x}(y) - y)]\, g(y|a)\, dy - c(a) \tag{P1}$$

subject to the ex post incentive compatibility constraint (10), the optimality of effort a, and monotonicity and majorization of $p(y)$ by y. As there is only one type of DM, the revenue maximization problem of the intermediary is equivalent to welfare maximization for the DM because this way, the intermediary can charge the highest possible fee.

In this environment, there is no reason for the intermediary to rule out manipulation in equilibrium. When the cost of manipulation is high, e.g., τ is close to 1, it is optimal for the interim prices to have a slope of τ, as we will show. Thus the marginal cost of manipulation at $\hat{x}(y) = y$ is equal to the increase in the interim prices, ruling out manipulation in equilibrium. However, when τ is low, e.g., $\tau = 0$, manipulation always occurs in equilibrium.

We make the following assumption about the distribution $G(\cdot|\cdot)$ and the validity of the first-order approach:

Assumption 3. *The following conditions are satisfied:*

1. *The c.d.f. of y, $G(\cdot|\cdot)$, is a C^2 function of y and a.*

2. *An increase in a shifts the distribution of y upwards, i.e., $G_a(y|a) \leq 0$.*

3. *The first-order approach is valid so that optimality of effort a can be replaced by*

$$\int [p(y) - c_m(\hat{x}(y) - y)] g_a(y|a)\, dy = c'(a)\,. \quad (11)$$

The first and second parts of the above assumption help us prove our main result about the shape of optimal ratings in Theorem 2. The last part allows us to construct a local perturbation argument which satisfies the constraint and improves the objective in presence of a discontinuity in interim prices. While the validity of the first-order approach is not without loss of generality, in the online Appendix, we provide conditions such that the solution of the relaxed problem is indeed incentive compatible.

We proceed by stating our main result of this section:

Theorem 2. *Suppose that Assumption 3 holds. If $p(\cdot)$ is an interim price function that achieves the maximum in (P1), then $p(\cdot)$ is continuous.*

The intuition of the proof relies on the continuity of $G_a = \partial G(y|a)/\partial a$ together with the trade-offs involved in allowing for manipulation. The cost of manipulation is destroyed surplus, and its benefit is that it provides incentives for ex ante effort. Because the pricing function must be incentive compatible ex post, i.e., (10), a simple application of the envelope theorem and integration by part implies that an increase in $x(y)$ relaxes (11) by $-kG_a(y|a)$. Hence, the marginal benefit of manipulation is continuous in y, and its marginal cost is strictly convex. Therefore, optimal manipulation and, as a result, the interim price function should be continuous. This continuity implies that partitions should not be used at the optimum.

A key insight of Theorem 22 is that we can divide the domain $[0, 1]$ into a collection of subintervals where the optimal interim price function alternates between the identity function—for which the majorization constraint is binding—and one in which majorization is slack and thus involves rating uncertainty. In other words, partitions often used by various disclosure mechanisms where various states are pooled are not optimal. Additionally, the theorem illustrates an algorithm to find optimal rating via optimization over alternating intervals.

The result in Theorem 2 is in sharp contrast with that in Proposition 1, in which we established that optimal information structures are monotone partitions. This means that the DM does not face any uncertainty when determining its rating. In contrast, when $p(\cdot)$

is continuous and not all information is revealed, then the rating system must involve randomization or rating uncertainty.

To provide two further characterizations of optimal rating under manipulation, we make the following assumption:

Assumption 4. *In addition to Assumption 3, the c.d.f of y, $G(\cdot|\cdot)$, satisfies the following properties:*

1. *Effort is mean y, i.e., $\int_0^1 y dG(y|a) = a$.*
2. *The function $\frac{G_a(y|a)}{g(y|a)}$ is bounded below.*
3. *$G(\cdot|\cdot)$ satisfies the monotone likelihood ratio property, i.e., g_a/g is strictly increasing.*

The first part of Assumption 4 normalizes the choice of effort as choosing the mean of the distribution y. Because this distribution as a function of a is ranked according to first-order stochastic dominance, this is without loss of generality as long as cost is convex in the mean of the distribution. The second part ensures that the change in the c.d.f. from an increase in a is not too large relative to the density $g(y|a)$. The last part allows us to prove existence of Lagrange multipliers in Theorem 3.

Our first main result is on optimal rating when manipulation is costly (i.e., its marginal cost at 0 is high).

Proposition 3. *Suppose that Assumption 4 holds. Then, there exists $\overline{\tau} \in (0,1)$ such that for all $\tau \geq \overline{\tau}$, there is no manipulation under optimal rating, i.e., $\hat{x}(y) = y$, and the optimal rating satisfies*

$$S = \{N\} \cup [0,1], \pi(\{s\}|y) = \begin{cases} \tau & s = y \\ 1-\tau & s = N \end{cases}. \tag{12}$$

The idea behind Proposition 3 is quite simple. When τ is large enough, the marginal cost of manipulation is too high and thus its benefit in incentivizing ex ante effort is too small. Hence, we can think about the steepest interim price function that implements no manipulation. This is a pricing function that satisfies $p'(y) = \tau$ for all $y \in [0,1]$. Similar to Example 1, one way to achieve this is using a partial mixing information structure, as described in (12).

By imposing more restrictions on the distribution function $g(\cdot)$, we can provide a full characterization of optimal ratings. We make the following assumption about the distribution function $G(y|a)$:

Assumption 5. *The log-density of the distribution function* $l(y,a) = \log g(y|a)$ *satisfies the following conditions:*

1. $l_{ya} \geq 0, l_{yy} \geq 0, l_y \geq 0$
2. $l_{yya} \geq l_y l_{ya}$, $l_{yyy} \geq l_y l_{yy}$, *and* $2l_{yy} \geq (l_y)^2$,

where $l_y, l_{yy}, \cdots$ *are partial derivatives of* l.

The following theorem provides a characterization of optimal rating systems under the above assumption about the distribution function:

Theorem 3. *Suppose that Assumptions 4 and 5 hold. Then for any given effort* a*, optimal rating systems satisfy the following properties:*

1. *There exists an interval* $[y_1, y_2]$ *(possibly empty) where the rating system is fully revealing for values of* $x \in [y_1, y_2]$ *and partially mixing for values of* x *in* $[0, y_1]$ *and* $[y_2, 1]$.
2. *Optimality can be achieved by using a signal space of* $[0,1] \cup \{L, H\}$.

Theorem 3 provides the conditions so that at the optimum, the majorization constraint binds for at most one interval. The key implication of Assumption 5 is that absent the majorization constraint, the optimal interim price is the steepest for mid-values of y. Thus, when the majorization constraint is violated, it is violated for an interval of values of y. Figure 4 depicts the optimal manipulation and interim prices for various k values. When the value of k is high, the optimal level of manipulation required to provide ex-ante incentives is low, and as a result interim prices are not very dispersed. For low values of k, higher manipulation and higher dispersion of prices are needed. This implies that majorization is binding for mid-values of y. For low values of y, the optimal manipulation is 0 because the marginal cost of manipulation is positive and equal to τ while the benefit of manipulation converges to 0 as y converges to 0.

Our distributional assumptions in Theorem 2 are sufficient but not necessary. In general, it is possible to weaken the assumptions, but we omit it here for brevity. In the class of distribution functions that satisfy $l(y,a) = r(y)\,m(a) + b(a)$, Assumption 5 holds if $r(y) = -\log(1-y)$ and $m(a)$ satisfies $1 \geq m \geq 0, m' \geq 0$. More generally, r has to be convex enough for Assumption 5 to hold.

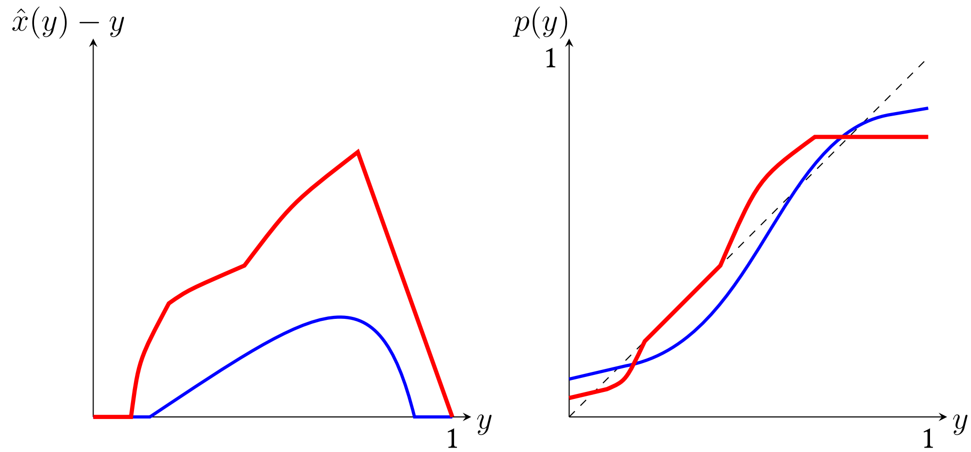

Figure 4: Optimal manipulation and interim prices for low (red) and high (blue) values of $k = \frac{\partial^2}{\partial x^2} c_m (x; y)$.

Finally, regarding the second part of Theorem 3, when the majorization is slack, Assumption 5 implies that the optimal interim price function is convex for low values of y and concave for high values of y. As a result, there are at most three values at which $p(y) = y$. In the proof of Theorem 3, we show that such an interim price function can always be implemented by a rating system similar to a partial mixing one, as in Example 1. The rating system differs for low and high values of y: when y is low, it reveals the state with some probability and otherwise generates the signal L; when y is high, it reveals the state and signals L and H probabilistically.

6 Implications and Extensions

In this section, we discuss the key implications of our model on the design of rating systems under moral hazard.

Changes in the Cost of Manipulation and Design of Ratings Arguably, with the increased ease in communication on the internet, rating manipulation and fake reviews are easier to achieve. In Section 3, we discussed various studies that illustrate the prevalence of manipulation. But how should certifiers and platforms react to this decline in cost? While efforts to detect fake reviews can address this problem, there are limits to this practice. Our model shows how the rating system itself needs to change in response to the decline in the cost of manipulation. Interestingly, it identifies a force which implies that a lower cost of manipulation should be accompanied by more informative rating systems because of the interplay of ex ante incentives and limitations implied by stochastic dominance.

Whereas rating uncertainty is a natural response to rating manipulation, perhaps counterintuitively, a decline in the cost of manipulation leads to more informative ratings (see Figure 4). A decline in k leads to a steeper interim price schedule and eventually full revelation for intermediate values of y. A decline in τ works similarly; it leads to an increase in manipulation and thus steeper interim prices.

Noisiness of Optimal Ratings at the Extremes A robust feature of optimal ratings with manipulation is that they are noisy at extreme realizations of y. In general it is difficult to verify this property in the data, but there is some suggestive evidence that noisiness at the extremes could be present. At the low end, we note that several platforms often provide a form of rating forgiveness. For example, Instacart, a delivery platform, allows shoppers to apply for removal of their previous low ratings by customers.[16] This in turn creates more uncertainty for shoppers with low ratings. On the high end, there is some evidence of upward bias in rating systems used by several platforms—a feature that creates more perceived noisiness for high ratings. Raval (2023) uses data on businesses' ratings across several platforms such as Google, Facebook, Yelp, and BBB to estimate quality, and suggests an upward bias is present in all platforms.

6.1 Extensions

Allowing for Market Action

In our focus on the effect of rating systems on moral hazard, we have abstracted from several realistic features of these systems. Most importantly, for ease of analysis, we have abstracted from situations in which ratings have value for the market, e.g., helping market participants not make a purchase when the quality is low. In online Appendix B.5.1, we show that an extension of our main result on the majorization of interim prices holds in an environment with a market action. In such an environment, we need to allow for the obedience constraint by the market. Nevertheless, we show that if one focuses on the DM's second-order expectation conditional on each action by the market, a similar result holds and comonotinicity implies majorization.

[16] Changing How Ratings Work (accessed July 7, 2023)

Different Priors

Market participants may have different perceptions about the informativeness of rating systems. In online Appendix B.5.2, we consider a version of our model where the market has a dogmatic prior about the joint distribution of (a, y) which is not necessarily consistent with the intermediary's rating policy. Nevertheless, we show that the existence of rating systems for interim prices that are comonotone with market values (appropriately calculated) is equivalent to second-order stochastic dominance. We also characterize optimal ratings in some special cases of this dogmatic bias.

7 Conclusion

We have explored the design of optimal rating systems in the presence of moral hazard. Methodologically, we showed that interim prices—an informed decision-maker's expectation of market expectation—can be used to simplify the design problem. It thus becomes a mechanism design problem with transfers, where transfers have to be a mean-preserving contraction of market values.

Using two applications, we have established that the interim price methodology helps us derive substantive results on the design of rating systems. Specifically, we characterized optimal ratings when effort is valued by the market, either directly or through the realization of a random valuation. Additionally, we considered a model where in addition to making an effort valued by the market, the decision-maker can make an ex post costly effort to manipulate the signal observed by the rating agency. This model has important implications for the design of rating systems in the presence of easier manipulation.

References

Albano, G. L. and A. Lizzeri (2001): "Strategic certification and provision of quality," *International economic review*, 42, 267–283. 6

Ali, S. N., N. Haghpanah, X. Lin, and R. Siegel (2022): "How to sell hard information," *The Quarterly Journal of Economics*, 137, 619–678. 6

Aliprantis, C. D. and K. Border (2013): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer-Verlag Berlin and Heidelberg GmbH & Company KG. 33, 36

Ball, I. (2019): "Scoring strategic agents," *arXiv preprint arXiv:1909.01888.* 4

Barahona, N., C. Otero, and S. Otero (2023): "Equilibrium Effects of Food Labeling Policies," *Econometrica (forthcoming).* 6

Berg, F., J. F. Koelbel, and R. Rigobon (2022): "Aggregate confusion: The divergence of ESG ratings," *Review of Finance*, 26, 1315–1344. 6

Blackwell, D. (1953): "Equivalent comparisons of experiments," *The annals of mathematical statistics*, 265–272. 15

Boleslavsky, R. and K. Kim (2020): "Bayesian persuasion and moral hazard," *Working Paper, Emory University.* 5

Doob, J. L. (1994): *Measure theory*, Springer Science & Business Media. 34

Dworczak, P. and G. Martini (2019): "The simple economics of optimal persuasion," *Journal of Political Economy*, 127, 1993–2048. 5

Frankel, A. and N. Kartik (2019): "Muddled information," *Journal of Political Economy*, 127, 1739–1776. 4

Hardy, G., J. Littlewood, and G. Polya (1934): *Inequalities*, Cambridge Universtiy Press, Cambridge, UK. 16

He, S., B. Hollenbeck, and D. Proserpio (2022): "The market for fake reviews," *Marketing Science.* 10

Holmström, B. (1979): "Moral hazard and observability," *The Bell journal of economics*, 74–91. 20

Hopenhayn, H. and M. Saeedi (2020): "Optimal Quality Ratings and Market Outcomes," *National Bureau of Economic Research Working Paper.* 6

Horner, J. and N. S. Lambert (2020): "Motivational ratings," *Review of Economic Studies. forthcoming.* 6

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 (2020): "Raising the Bar: Certification Thresholds and Market Outcomes","," *Working Paper, Carnegie Mellon University*. 6, 10

Hui, X., M. Saeedi, and N. Sundaresan (2017): "Adverse Selection or Moral Hazard: An Empirical Study," *Working paper*. 10

Kamenica, E. and M. Gentzkow (2011): "Bayesian persuasion," *American Economic Review*, 101, 2590–2615. 5

Kleiner, A. and A. Manelli (2019): "Strong Duality in Monopoly Pricing," *Econometrica*, 87, 1391–1396. 46

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89, 1557–1593. 5, 16, 17, 36

Kolotilin, A., R. Corrao, and A. Wolitzky (2023): "Persuasion with Non-Linear Preferences," . 5

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017): "Persuasion of a privately informed receiver," *Econometrica*, 85, 1949–1964. 5

Kovbasyuk, S. and G. Spagnolo (2023): "Memory and Markets," *The Review of Economic Studies*, rdad067. 6

Lizzeri, A. (1999): "Information revelation and certification intermediaries," *The RAND Journal of Economics*, 214–231. 6

Luenberger, D. G. (1997): *Optimization by Vector Space Methods*, John Wiley & Sons. 35

Mitter, S. K. (2008): "Convex Optimization in Infinite Dimensional Spaces," *Recent advances in learning and control*, 161–179. 46

Myerson, R. B. (1981): "Optimal auction design," *Mathematics of operations research*, 6, 58–73. 39, 43

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. 4, 6, 10

Onuchic, P. and D. Ray (2021): "Conveying value via categories," *arXiv preprint arXiv:2103.12804*. 6

Perez-Richet, E. and V. Skreta (2022): "Test design under falsification," *Econometrica*, 90, 1109–1142. 4

Raval, D. (2023): "Do Gatekeepers Develop Worse Products? Evidence from Online Review Platforms," *FTC Working Paper*. 26

Rivlin, A. M. and J. B. Soroushian (2017): "Credit rating agency reform is incomplete," *Brookings Institution, https://www. brookings. edu/research/credit-rating-agency-reform-is-incomplete*. 10

Rothschild, M. and J. E. Stiglitz (1970): "Increasing risk: I. A definition," *Journal of Economic theory*, 2, 225–243. 15

Saeedi, M. (2019): "Reputation and adverse selection: theory and evidence from eBay," *The RAND Journal of Economics*, 50, 822–853. 10

Strassen, V. (1965): "The existence of probability measures with given marginals," *The Annals of Mathematical Statistics*, 36, 423–439. 5, 15

Vatter, B. (2022): "Quality disclosure and regulation: Scoring design in medicare advantage," . 6

Zapechelnyuk, A. (2020): "Optimal quality certification," *American Economic Review: Insights*, 2, 161–76. 6

Zubrickas, R. (2015): "Optimal grading," *International Economic Review*, 56, 751–776. 6

A Proofs

A.1 Proof of Theorem 1

We will first prove the theorem when Y is finite. We will then show that the theorem 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 $\overline{v}(y_1) \le \overline{v}(y_2) \le \cdots \le \overline{v}(y_n)$. When $p(y)$ is co-monotone with $\overline{v}(y)$, we must have that $p(y_1) \le p(y_2) \le \cdots \le p(y_n)$. 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

$$\begin{aligned} &f_1\overline{v}_1 + f_2\overline{v}_2 = f_1 p_1 + f_2 p_2, \\ &0 \le p_2 - p_1 \le \overline{v}_2 - \overline{v}_1. \end{aligned}$$

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

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

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

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

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 y_1, y_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 $p_1 \le \cdots \le p_i \le p_{i+2} \le \cdots \le p_n$ distributed according to $f_1, \cdots, f_{i-1}, f_i + f_{i+1}, f_{i+2}, \cdots, f_n$ and market values of $\overline{v}_1 \le \cdots \le \overline{v}_{i-1} \le \frac{f_i\overline{v}_i + f_{i+1}\overline{v}_{i+1}}{f_i + f_{i+1}} \le \overline{v}_{i+2} \le \cdots \le \overline{v}_n$. 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 $\lambda = \min_{i \le n-1} \frac{p_{i+1} - p_i}{\overline{v}_{i+1} - \overline{v}_i}$. 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

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

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

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

This concludes the proof.

2. When Y is an arbitrary compact subset of a Euclidean space.

Let $V = \overline{v}(Y)$ be the range of $\overline{v}$ and a subset of $\mathbb{R}$. Furthermore, let us define

$$\forall v \in V, \hat{p}(v) = p(y), \overline{v}(y) = v.$$

This function is well-defined since p is co-monotone with $\overline{v}$. That is, if for two values y_1 and y_2, $\overline{v}(y_1) = \overline{v}(y_2)$, then we must have that $p(y_1) = p(y_2)$. We also let $\mu_v \in \Delta(V)$ be the probability measure induced on V using μ_y and $\overline{v}(\cdot)$. Clearly, we also must have that $\hat{p}(\cdot)$ is a monotone function of v.

By Assumption 1, V is a finite collection of subintervals. For ease of exposition, we prove the claim when there is only one subinterval $[\underline{v}, \overline{v}]$. The proof with a finite number is almost identical but is more cumbersome. Consider a sequence of partitions $V^n = \{v_0^n = \underline{v} < v_1^n < \cdots < v_n^n = \overline{v}\}$ for $n = 1, 2, \cdots$ with $\min_{0 \le i \le n-1} v_{i+1}^n - v_i^n \to 0$ and $V^{n+1} \subset V^n$. We define

$$f_i^n = \mu_v\left(\left[v_{i-1}^n, v_i^n\right)\right), 1 \le i \le n-1; f_n^n = \mu_v\left(\left[v_{n-1}^n, \overline{v}\right]\right)$$

$$v_i^n = \begin{cases} \frac{\int v \mathbf{1}\left[v \in \left[v_{i-1}^n, v_i^n\right)\right] d\mu_v}{f_i^n} & f_i^n > 0, i \le n-1 \\ \frac{\int v \mathbf{1}\left[v \in \left[v_{n-1}^n, \overline{v}\right)\right] d\mu_v}{f_n^n} & f_n^n > 0, i = n \\ \frac{v_{i-1}^n + v_i^n}{2} & f_i^n = 0, i \ge 1 \end{cases} \qquad \hat{p}^n\left(v_i^n\right) = \begin{cases} \frac{\int \hat{p}(v) \mathbf{1}\left[v \in \left[v_{i-1}^n, v_i^n\right)\right] d\mu_v}{f_i^n} & f_i^n > 0, i \le n-1 \\ \frac{\int \hat{p}(v) \mathbf{1}\left[v \in \left[v_{n-1}^n, \overline{v}\right)\right] d\mu_v}{f_n^n} & f_n^n > 0, i = n \\ \frac{\hat{p}\left(v_{i-1}^n\right) + \hat{p}\left(v_i^n\right)}{2} & f_i^n = 0, i \ge 1. \end{cases}$$

In words, the above constructs a discretization of the buyer values v and the DM's interim prices $\hat{p}(v)$. Since $\hat{p}^n(v)$ is an increasing function of v and, by construction, $\hat{p}^n \succsim_{\text{SOSD}} v^n$ and V^n is finite, we can apply the result from the first part. That is, an information structure (S^n, π^n) exists where $\pi^n : V^n \to \Delta(S^n)$ such that $\hat{p}^n(v_i^n) = \sum_{s \in S^n} \pi^n(\{s\} | v_i^n) \mathbb{E}[\overline{v}|s]$.

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

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

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

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

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

where the last equality follows because $\mathbb{E}\left[\hat{p}\left(v\right)|\mathcal{F}^n\right] = \hat{p}^n, \mathbb{E}\left[v|\mathcal{F}^n\right] = v^n$ given the definition of $\hat{p}^n$ and 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, $\mathbb{E}_\zeta\left[\int v d\mu|v\right] = \hat{p}\left(v\right)$. This concludes the proof.

A.2 Proof of Proposition 1

Proof. We first show that at the optimum, for all y, either the monotonicity constraint or the majorization constraint, Equation (6), is binding. Suppose to the contrary that this does not hold. Note that a change in $p\left(y\right)$ for a measure zero set of y's does not affect the objective and the majorization constraint. This implies that in order to achieve a contradiction, we need to rule out an interval in which neither the majorization nor the monotonicity constraint is binding. Suppose that there exists an interval $I = [y_1, y_2]$ for which majorization and monotonicity are slack. Assuming the first-order approach is valid, given any effort profile $a\left(\theta\right)$, the optimal rating system must be a solution to the following planning problem:

$$\max_{p(\cdot)} \sum_{\theta\in\Theta} f\left(\theta\right)\lambda\left(\theta\right)\left[\int_0^1 p\left(y\right) g\left(y|q\left(\theta\right)\right) dy - c\left(a\left(\theta\right), \theta\right)\right] \tag{P1}$$

subject to

$$\int_0^1 p(y)\, g_q(y|a(\theta))\, dy = c_a(a(\theta), \theta),$$

$$\sum_{\theta\in\Theta} f(\theta) \int_0^y [p(y') - y']\, g(y'|a(\theta))\, dy' \geq 0, \forall y \in [0,1],$$

$$\sum_{\theta\in\Theta} f(\theta) \int_0^1 [p(y) - y]\, g(y|a(\theta))\, dy = 0,$$

$$p(y) - p(y') \geq 0, \forall y \geq y'.$$

By combining the Theorems 1 in Sections 9.3 and 9.4 of Luenberger (1997) together with the fact that we have finitely many types and thus finitely many linear equality constraints, there must exist Lagrange multipliers $\gamma(\theta)$ – for the incentive compatibility constraint – so that $p(y)$ satisfies

$$p \in \arg\max_{\hat{p}} \int_0^1 \hat{p}(y) \sum_{\theta} f(\theta) [\lambda(\theta)\, g(y|a(\theta)) + \gamma(\theta)\, g_a(y|a(\theta))]\, dy. \tag{13}$$

subject to $\hat{p} \succcurlyeq_{\text{SOSD}} y$, and $\hat{p}(\cdot)$ is monotone. Let us define $h(y) = \sum_{\theta} f(\theta)\, g(y|a(\theta))$ and

$$\alpha(y) = \frac{\sum_{\theta} f(\theta) [\lambda(\theta)\, g(y|a(\theta)) + \gamma(\theta)\, g_a(y|a(\theta))]}{h(y)},$$

The Lagrangian associated with (13) is given by

$$\begin{aligned}
\mathcal{L} = & \int_0^1 \hat{p}(y)\, \alpha(y)\, h(y)\, dy + \int_0^1 \int_0^y (\hat{p}(y') - y')\, h(y')\, dy' dM(y) \\
& - m \int_0^1 (\hat{p}(y') - y')\, h(y')\, dy' \\
= & \int_0^1 \hat{p}(y)\, \alpha(y)\, h(y)\, dy + \int_0^1 (\hat{p}(y) - y) [M(1) - M(y)]\, h(y)\, dy \\
& - m \int_0^1 (\hat{p}(y') - y')\, h(y')\, dy',
\end{aligned}$$

where $M(y)$ is an increasing function y. Moreover,

$$\int_0^1 \int_0^y (\hat{p}(y') - y') \sum_{\theta} f(\theta)\, g(y'|a(\theta))\, dy' dM(y) = 0.$$

From the result in Kleiner et al. (2021), we know that there exists $p_e(y)$, an extreme point of the set $\{p : p \succcurlyeq_{\text{SOSD}} y, p : \text{monotone}\}$ that maximizes the objective in (13), and a collection of disjoint intervals $\left[\underline{y}_i, \overline{y}_i\right)$ exists such that

$$p_e(y) = \begin{cases} y & y \notin \bigcup_i \left[\underline{y}_i, \overline{y}_i\right) \\ \frac{\int_{\underline{y}_i}^{\overline{y}_i} y \sum_\theta f(\theta) g(y|a(\theta)) dy}{\int_{\underline{y}_i}^{\overline{y}_i} \sum_\theta f(\theta) g(y|a(\theta)) dy} & y \in \left[\underline{y}_i, \overline{y}_i\right). \end{cases}$$

Note that optimality conditions implied by the Lagrangian are that if $y \in \left(\underline{y}_i, \overline{y}_i\right)$, then

$$\alpha(y) - m + M(1) - M(y) = 0.$$

In other words, $\alpha(y)$ must be weakly increasing. Moreover, if $(z_1, z_2) \subset [0,1] \setminus \bigcup_i \left[\underline{y}_i, \overline{y}_i\right)$, then $M(y)$ has to be constant.

If the solution of the optimization problem (P1) is not an extreme point of $\{p : p \succcurlyeq_{\text{SOSD}} y, p : \text{monotone}\}$, then by Krein-Milman (see Aliprantis and Border (2013)), it must be a convex combination of the extreme points of the set $\{p : p \succcurlyeq_{\text{SOSD}} y, p : \text{monotone}\}$. Hence, there must exist another extreme point $\tilde{p}_e(y)$ that also achieves the optimum in (13). If $p_e \neq \tilde{p}_e$, there must exist $y \in (0,1)$ so that $p_e(y) = y$ for an interval around y and $\tilde{p}_e(y)$ is constant for an interval around y. By optimality, it must be that

$$\alpha(y') - m + M(1) - M(y') = 0, M(y') = M(y)$$

for $y' \in I$, an interval around y. This means that there must exist a constant, $c = m - M(1) + M(y)$, so that for all $y' \in I$

$$\sum_\theta f(\theta)\left[\lambda(\theta) g(y'|a(\theta)) + \gamma(\theta) g_a(y'|a(\theta))\right] = c \sum_\theta f(\theta) g(y'|a(\theta))$$

or

$$\sum_\theta f(\theta)\left[(\lambda(\theta) - c) g(y'|a(\theta)) + \gamma(\theta) g_a(y'|a(\theta))\right] = 0,$$

which then implies that $\{g(y|a(\theta)), g_a(y|a(\theta))\}_{\theta \in \Theta}$ are linearly dependent over I', which is in contradiction with our assumption. This concludes the proof. $\square$

A.3 Proof of Proposition 2

We first prove the following lemma:

Lemma 2. *Consider the optimization problem*

$$\max_{p:p\succsim_{SOSD}y,\,monotone}\int_0^1 p\left(y\right)\Gamma\left(y\right)h\left(y\right)dy.$$

Suppose that $\Gamma\left(x\right)$ is continuously differentiable and that its derivative changes sign $k<\infty$ times, i.e., we can partition $[0,1]$ into k intervals, where in each interval $\Gamma'\left(x\right)$ has the same sign but not in two consecutive intervals. Then, an optimal information structure is an alternating partition (between full revelation and pooling) with at most k intervals. Moreover, if it is separating over an interval, then $\Gamma'\geq 0$ over this interval.

Proof. As in the proof of Proposition 1, we know that $\int_0^1 p\Gamma dH$ is maximized at extreme point p_e of the set $\{p : p \succsim_{\text{SOSD}} y, p : \text{monotone}\}$ and hence p_e is a collection of disjoint intervals over which we either pool or fully separate the values of y. Thus, to prove the optimality, it is sufficient to show that we cannot have two consecutive pooling intervals and if $\Gamma'\left(y\right)<0$, we cannot have separation at y.

First, suppose that $\Gamma'\left(\hat{y}\right)<0$ and $p_e\left(y\right)=y$ for an interval $\left(y_1,y_2\right)$ around $\hat{y}$. By continuous differentiability of Γ, we can assume that $\Gamma'<0$ over the entire interval. Since Γ is decreasing, Γ and y are negatively correlated over this interval and thus

$$\text{cov}\left(y,\Gamma|y_1<y<y_2\right)<0\rightarrow\frac{\int_{y_1}^{y_2}y\Gamma\left(y\right)dH}{H\left(y_2\right)-H\left(y_1\right)}<\frac{\int_{y_1}^{y_2}ydH\int_{y_1}^{y_2}\Gamma\left(y\right)dH}{\left(H\left(y_2\right)-H\left(y_1\right)\right)^2},$$

which means that pooling over this interval improves the objective.

Now suppose that $\forall y\in I_1, p\left(y\right)=\mathbb{E}\left[y|I_1\right]$ and $\forall y\in I_2, p\left(y\right)=\mathbb{E}\left[y|I_2\right]$ for two neighboring intervals $I_1\leq I_2$. If $\mathbb{E}\left[\Gamma\left(y\right)|I_1\right]\geq\mathbb{E}\left[\Gamma\left(y\right)|I_2\right]$. In this case, the same argument as above implies that pooling I_1 and I_2 improves the objective. Suppose, on the other hand, $\mathbb{E}\left[\Gamma|I_1\right]<\mathbb{E}\left[\Gamma|I_2\right]$. Let $\hat{y}=\min I_2=\max I_1$. If $\Gamma\left(\hat{y}\right)>\mathbb{E}\left[\Gamma|I_1\right]$, then separating a small subinterval of I_1 around $\hat{y}$ improves the objective. Therefore, we must have that $\Gamma\left(\hat{y}\right)\leq\mathbb{E}\left[\Gamma|I_1\right]<\mathbb{E}\left[\Gamma|I_2\right]$, which then implies that separating a subinterval of I_2 around $\hat{y}$ improves the objective. This proves the claim. □

Now, we can use Lemma 2 to prove Proposition 2.

Proof. Recall that the gain function is given by

$$\Gamma(y)=\frac{1+\sum_{j\geq i}\lambda_j\frac{g_a(y|a_j)}{g(y|a_i)}}{f_i+\sum_{j>i}f_j\frac{g(y|a_j)}{g(y|a_i)}}=\frac{\Gamma_1(y)}{\Gamma_2(y)}.$$

Note that since the cost function $c(a,\theta)$ is submodular, any incentive compatible effort choice should be weakly monotone. Moreover, it must be the case that $\lambda_i \geq 0$ for all i because we can assume that all the local incentive compatibility constraints hold with equality such that the marginal benefit of effort is weakly higher than its marginal cost. Now suppose that both types are participating under the optimal rating system. Then, the gain function is given by

$$\Gamma_1(y)=1+\lambda_1\left[r(y)\,m(a_1)+b(a_1)\right]+\lambda_2 e^{r(y)(m(a_2)-m(a_1))+b(a_2)-b(a_1)}\left(r(y)\,m(a_2)+b(a_2)\right)$$
$$\Gamma_2(y)=f_1+f_2e^{r(y)(m(a_2)-m(a_1))+b(a_2)-b(a_1)}.$$

If we define $z=e^{r(y)(m(a_2)-m(a_1))}$, we have

$$\Gamma(y)=\frac{\lambda+\alpha\log z(y)+\beta z(y)\log z(y)}{f_1+\gamma z(y)}=\hat{\Gamma}(z(y))$$

for some values of $\alpha,\beta,\gamma>0$ (because $m'\geq 0$ and $a_2\geq a_1$) and since $z'(y)>0$, $\hat{\Gamma}'(z(y))$ and $\Gamma'(y)$ have the same sign. Then

$$\frac{d}{dz}\left(\left(f_1+\gamma z\right)^2\hat{\Gamma}'(z)\right)=\frac{(-\alpha+\beta z)(f_1+\gamma z)}{z^2}.$$

Since $f_1+\gamma z>0$, there are three possibilities: 1. $\beta z(\overline{y})<\alpha$. In this case, the above expression is always negative, which then implies that $\hat{\Gamma}'(z)=0$ has at most one solution. Below this point, $\hat{\Gamma}'(z)>0$ and above it, $\hat{\Gamma}'(z)<0$. By Lemma 2, the optimal rating system is upper-censorship, 2. $\beta z(\underline{y})>\alpha$. In this case, the above expression is always positive and the same logic as before implies that the optimal rating system is lower-censorship, 3. $\beta z(\overline{y})>\alpha>\beta z(\underline{y})$. In this case, $\hat{\Gamma}'(z)$ is zero at most at two points. If it is zero at one point, then we have again lower- or upper-censorship. If, on the other hand, it is zero at two points, z_1 and z_2, then $\hat{\Gamma}'$ is positive for values of z below z_1, negative in between, and positive at the top. This implies that the optimal rating system structure is mid-censorship, which concludes the proof. □

A.4 Proof of Theorem 2

Before proceeding with the proof, we describe how to simplify the constraint set of our optimization problem. More specifically, the ex post incentive compatibility is equivalent to (see Myerson (1981))

$$u(y) = p(y) - c_m(\hat{x}(y) - y) = \underline{u} + \int_0^y c_m'(\hat{x}(z) - z)\, dz \tag{14}$$

$$\hat{x}(y)\text{: monotone.}$$

Replacing this into the ex ante incentive compatibility and using integration by parts implies that

$$\int_0^1 u(y)\, g_a(y|a)\, dy = -\int_0^1 c_m'(\hat{x}(y) - y)\, G_a(y|a)\, dy,$$

where the equality uses $G_a(1|a) = 0$ because $G(1|a) = 0$ for all values of a. We can also replace this into the majorization constraint. Since, the mean of $p(z)$ and z are the same, we can write the majorization constraint as

$$\int_y^1 [z - p(z)]\, dG(z|a) \geq 0,$$

Replacing for p from above and using integration by parts yields the following:

$$\int_y^1 \left[z - c_m(z) - c_m'(z)\frac{1 - G(z|a)}{g(z|a)}\right] dG(z|a) \tag{15}$$
$$- \left(\underline{u} + \int_0^y c_m'(z)\, dz\right)[1 - G(y|a)] \geq 0.$$

Proof. Suppose that, contrary to the claim, $\hat{x}(y)$ and $p(y)$ exhibit an upward jump at $\tilde{y}$, i.e., $\hat{x}(\tilde{y}+) > \hat{x}(\tilde{y})$. Note that without loss of generality, we can assume that $\hat{x}(\cdot), p(\cdot)$ are left continuous. In what follows, we construct a new allocation that improves the objective. Recall from above that we can replace the incentive compatibility for ex-ante effort with

$$-\int c_m'(x(y) - y)\, G_a(y|a)\, dy \geq c'(a). \tag{16}$$

Note that the inequality can be imposed because if it is slack, an increase in a increases the objective.

Consider now a perturbation of $\hat{x}$ given by

$$\delta\hat{x}\left(y\right)=\begin{cases}\varepsilon_1 & y\in\left[\tilde{y}-\delta,\tilde{y}\right)\\ -\varepsilon_2 & y\in\left[\tilde{y},\tilde{y}+\delta\right]\\ 0 & y\in\left[0,1\right]\setminus\left[\tilde{y}-\delta,\tilde{y}+\delta\right],\end{cases}$$

where $\varepsilon_1,\varepsilon_2,\delta>0$ are small enough. Note that since there is a jump at $\tilde{y}$, $c_m'\left(\hat{x}\left(\tilde{y}+\right)-\tilde{y}\right)-c_m'\left(\hat{x}\left(\tilde{y}\right)-\tilde{y}\right)=\Delta>0$ and hence $\delta>0$ can be chosen so that

$$c_m'\left(\hat{x}\left(y_2\right)-y_2\right)-c_m'\left(\hat{x}\left(y_1\right)-y_1\right)>\Delta/2,\forall y_1\in\left[\tilde{y}-\delta,\tilde{y}\right),y_2\in\left[\tilde{y},\tilde{y}+\delta\right],$$

Moreover, we assume that ε_1 and ε_2 satisfy

$$\varepsilon=\varepsilon_1\int_{\tilde{y}-\delta}^{\tilde{y}}\left[c_m'\left(y\right)g\left(y\right)+k\left(1-G\left(y\right)\right)\right]dy=\varepsilon_2\int_{\tilde{y}}^{\tilde{y}+\delta}\left[c_m'\left(y\right)g\left(y\right)+k\left(1-G\left(y\right)\right)\right]dy. \tag{17}$$

The above implies that this perturbation is budget balanced. Moreover, since $\hat{x}\left(y\right)$ is increasing, and it has a jump at $\tilde{y}$, the resulting perturbed function is also increasing for small values of $\varepsilon_1,\varepsilon_2>0$. Finally, the approximate change in prices is given by

$$\delta p\left(y\right)=\begin{cases}0 & y<\tilde{y}-\delta\\ \varepsilon_1 c_m'\left(y\right)+\varepsilon_1 k\left(y-\tilde{y}+\delta\right) & y\in\left[\tilde{y}-\delta,\tilde{y}\right)\\ -\varepsilon_2 c_m'\left(y\right)+\varepsilon_1 k\delta-\varepsilon_2 k\left(y-\tilde{y}\right) & y\in\left[\tilde{y},\tilde{y}+\delta\right]\\ -\varepsilon_2 k\delta & y>\tilde{y}+\delta.\end{cases}$$

Given (17), we must have that $\int_0^1\delta p\left(y\right)dG=0$. Additionally,

$$\begin{aligned}&\forall y\in\left[\tilde{y}-\delta,\tilde{y}\right),\delta p\left(y\right)=\varepsilon_1\left[k\left(\hat{x}\left(y\right)-\tilde{y}\right)+\tau\right]+\varepsilon_1 k\delta,\\ &\forall y\in\left[\tilde{y},\tilde{y}+\delta\right],\delta p\left(y\right)=-\varepsilon_2\left[k\hat{x}\left(y\right)+\tau\right]+\varepsilon_1 k\delta+\varepsilon_2 k\tilde{y}.\end{aligned}$$

Since $\hat{x}$ is an increasing function of y, the above implies that $\delta p\left(y\right)$ is positive for values of y below a certain threshold and negative for values of y above it. Hence, $\int_0^y\delta p\left(z\right)dG$ is a single-peaked function of y, which then implies that $\int_0^y\delta p\left(z\right)dG\geq 0$ for all values of y and $\int_0^y\delta p\left(z\right)dG>0$ for all values of $y\in\left[\tilde{y}-\delta,1\right)$. Thus the perturbed allocation is feasible.

The change in the LHS of the modified incentive constraint above is given by

$$\Delta_1 = \varepsilon_2 \int_{\tilde{y}}^{\tilde{y}+\delta} \lambda k G_a \left(y|a\right) dy - \varepsilon_1 \int_{\tilde{y}-\delta}^{\tilde{y}} \lambda k G_a \left(y|a\right) dy,$$

while the change in the objective is given by

$$\begin{aligned}
\Delta_2 = & -\varepsilon_1 \int_{\tilde{y}-\delta}^{\tilde{y}} c_m' \left(y\right) g\left(y|a\right) dy + \varepsilon_2 \int_{\tilde{y}}^{\tilde{y}+\delta} c_m' \left(y\right) g\left(y|a\right) dy \\
= & -\varepsilon_2 \frac{\int_{\tilde{y}-\delta}^{\tilde{y}} c_m' \left(y\right) g\left(y|a\right) dy \times \int_{\tilde{y}}^{\tilde{y}+\delta} \left[c_m' \left(y\right) g\left(y\right) + k\left(1 - G\left(y\right)\right)\right] dy}{\int_{\tilde{y}-\delta}^{\tilde{y}} \left[c_m' \left(y\right) g\left(y\right) + k\left(1 - G\left(y\right)\right)\right] dy} \\
& + \varepsilon_2 \int_{\tilde{y}}^{\tilde{y}+\delta} c_m' \left(y\right) g\left(y|a\right) dy.
\end{aligned}$$

Since the set of discontinuity points of a monotone function are countable, we can approximate all the above integrals by their integrand evaluated at $\tilde{y}+$ (right limit at $\tilde{y}$) and $\tilde{y}-$ (left limits at $\tilde{y}$) multiplied by δ – the length of the interval – as δ approaches 0. So for small values of δ, we have

$$\begin{aligned}
\Delta_1 \int_{\tilde{y}-\delta}^{\tilde{y}} \left[c_m' \left(y\right) g\left(y\right) + k\left(1 - G\left(y\right)\right)\right] dy \approx \\
-\varepsilon_2 \delta^2 \lambda k G_a \left(\tilde{y}\right) \times \left[c_m' \left(\tilde{y}+\right) g\left(\tilde{y}\right) + k\left(1 - G\left(\tilde{y}\right)\right)\right] \\
+\varepsilon_2 \delta^2 \left[c_m' \left(\tilde{y}-\right) g\left(\tilde{y}\right) + k\left(1 - G\left(\tilde{y}\right)\right)\right] \times \lambda k G_a \left(\tilde{y}\right) = \\
-\varepsilon_2 \delta^2 G_a \left(\tilde{y}\right) g\left(\tilde{y}\right) \left(c_m' \left(\tilde{y}+\right) - c_m' \left(\tilde{y}+\right)\right) \geq 0,
\end{aligned}$$

$$\begin{aligned}
\Delta_2 \int_{\tilde{y}-\delta}^{\tilde{y}} \left[c_m' \left(y\right) g\left(y\right) + k\left(1 - G\left(y\right)\right)\right] dy \approx \\
-\varepsilon_2 \delta^2 c_m' \left(\tilde{y}-\right) g\left(\tilde{y}\right) \times \left[c_m' \left(\tilde{y}+\right) g\left(\tilde{y}\right) + k\left(1 - G\left(\tilde{y}\right)\right)\right] \\
+\varepsilon_2 \delta^2 \left[c_m' \left(\tilde{y}-\right) g\left(\tilde{y}\right) + k\left(1 - G\left(\tilde{y}\right)\right)\right] c_m' \left(\tilde{y}+\right) g\left(\tilde{y}\right) = \\
\varepsilon_2 \delta^2 k g\left(\tilde{y}\right) \left[c_m' \left(\tilde{y}+\right) - c_m' \left(\tilde{y}-\right)\right] \left(1 - G\left(\tilde{y}\right)\right) > 0,
\end{aligned}$$

where the above are positive because c_m' has a jump at $\tilde{y}$ and $G_a \leq 0$. This implies that for small enough values of δ, $\Delta_1 \geq 0, \Delta_2 > 0$. In other words, this perturbation increases the value of the objective while satisfying the constraints. This gives us the desired contradiction. □

A.5 Proof of Proposition 3

Proof. If we ignore the majorization constraint and solve out for p from the ex-post incentive compatibility, our constraint set is linear in $\hat{x}$. Thus, we conjecture the solution and verify that it satisfies the first-order conditions. Specifically, we set $\hat{x}(y) = y$, $c'(\hat{a}) = \tau$, $p(y) = \tau y + (1-\tau)\int y dG(y|\hat{a})$, and $\lambda = \frac{1-\tau}{c''(\hat{a})}$. Note that for this to be optimal, we need

$$0 \geq -\lambda G_a(y|\hat{a})\, c''_m(\hat{x}(y) - y) - g(y|\hat{a})\, c'_m(\hat{x}(y) - y) = \\ -\lambda G_a(y|\hat{a})\, k - g(y|\hat{a})\, \tau,$$

In words, the social value of manipulation should be negative for all realizations of y. As a result,

$$\frac{\tau}{1-\tau} c''\left(\left(c'\right)^{-1}(\tau)\right) \geq -\frac{kG_a(y|\hat{a})}{g(y|\hat{a})}, \forall y \in [0,1].$$

By Assumption 4, the right-hand side is bounded above, while as τ gets closer to 1, the left-hand side converges to ∞. Moreover, given this choice of λ, the choice of $\hat{a}$ is optimal because the effect of an increase in $\hat{a}$ on the marginal cost a is $c''(\hat{a})$,, while for $p(y)$ defined above, the marginal benefit is zero. Hence, the marginal total benefit of effort $1 - c'(\hat{a})$ should be equated with its marginal cost $\lambda c''(\hat{a})$. This concludes the proof. □

A.6 Proof of Theorem 3

Proof. Without loss of generality, we can write the optimization problem in terms of the extent of manipulation, $m(y) = \hat{x}(y) - y$. Since our optimization problem is in general non-convex, we first solve a more relaxed version given by

$$\max_{m \in \hat{X}} \int_0^1 \left[y - c_m(m(y))\right] dG(y|a) \tag{P2}$$

subject to (16) and (15). In the above, $\hat{X}$ is the set of functions m for which $m(y) + y$ is increasing and $1 - y \geq m(y) \geq 0$. Note that the constraints in (15) exclude the constraint that states that the mean of $p(\cdot)$ and y are the same (including this constraint makes the constraint set non-convex) and instead takes the utility of the $y = 0$, $\underline{u}$, as given. Later, we will show that by considering another optimization over $\underline{u}$, we achieve the optimum in the main optimization. We prove the claim of the theorem for the solution of the optimization (P2).

Now, consider the objective in (P2). Absent the majorization and the local incentive constraint, this objective is maximized at a function $\overline{m}$, which is monotone (this maximization possibly needs ironing a lá Myerson (1981) since a pointwise optimization of the objective can lead to a non-monotone $y + m(y)$.) If the incentive constraint and the majorization constraints are satisfied at $\overline{m}$, then this coincides with the solution of the problem and proves our claim.

The strong duality result in the Online Appendix, Lemma (3) in Section (B.1), establishes that solving the optimization problem above is equivalent to the following:

$$\max_{m \in \hat{X}} - \int_0^1 c_m g dy + \lambda \left[- \int c_m' G_a dy - c'(a) \right] + \int_0^1 s(y; m) \, d\Lambda(y) \tag{18}$$

where $s(y; m)$ is the LHS of (15), $\lambda > 0$ is a real number, and Λ is a positive Borel measure over $[0, 1]$. Moreover, at the optimum, $\int s(y; m) \, d\Lambda = 0$. Using a similar argument as in the proof of 2, we can show that the solution to the above optimization is a continuous function. Thus, in order to prove the claim, it is sufficient to show that at the optimum, there cannot exist an interval $I = [y_1, y_2]$ for which (15) is slack in its interior and binding at y_1 and y_2. Suppose to the contrary that this is the case.

Let m^* be the optimal manipulation strategy and p^* its associated interim price given by (14). Continuity of m^* and p^* together with majorization being binding at y_1 and y_2 implies that $p^*(y_1) = y_1 < p^*(y_2) = y_2$. Moreover, for values of $y > y_1$ and close to y_1, we must have that $p^*(y)$ is increasing and $p^*(y) > y$. Similarly, for values of $y < y_2$, $p^*(y) < y$ and p^* is increasing. If this is not the case, majorization is violated. Note further that by complementarity slackness, $\Lambda(y)$ is constant for values of $y \in (y_1, y_2)$.

We shall note that the objective in (18) is of the form $\int F_1(m(y), y) \, dy$, where

$$\begin{aligned} F(m, y) = & - c_m(m) g(y) - \lambda c_m'(m) G_a(y) \\ & + \left[g(y)(y - c_m(m)) - c_m'(m)(1 - G(y|a)) \right] \Lambda(y) \\ & - c_m'(m) \int_y^1 (1 - G(z|a)) \, d\Lambda. \end{aligned}$$

Let $\tilde{m}(y) \in [0, 1-y]$ be the pointwise maximizer of the above for all y. The solution of (P2), $m^*(y)$, deviates from $\tilde{m}(y)$ only when $m^*(y) + y$ is constant for an interval above or below y. As we argued above, in intervals above y_1 and below y_2, p^* and as a result, $m^* + y$ are increasing, which means that $\tilde{m} = m^*$ for such intervals. Moreover, since $p^*(y_1 + \varepsilon) > p^*(y_1) + \varepsilon$ for some $\varepsilon > 0$, we must have that $m^*(y_1) > 0$. A similar argument shows that $m^*(y_2) < 1$. Therefore, for values of y close to y_1 and y_2, we must have that $m^* = \tilde{m}$ and $F_m(\tilde{m}, y) = 0$. Using integration by parts, we can write this as

$$\tilde{m}(y) = -\frac{\tau}{k} + \frac{-\lambda G_a(y|a) - \int_y^1 \Lambda(z)\, dG}{(1 + \Lambda(y))\, g(y)}.$$

For all $y \in (y_1, y_2)$, due to slackness of the majorization constraint, $\Lambda(y)$ is constant. Using this property and in what follows, we show that under Assumption P2, the interim price function associated with m^* cannot imply $p^*(y_2) = y_2$ and $p^*(y) < y$ for $y < y_2$, yielding a contradiction. The manipulation function $\tilde{m}(y)$ can be written as

$$\tilde{m}(y) + \frac{\tau}{k} = \frac{\int_y^1 [\lambda \ell_a - \Lambda(z)]\, g(z)\, dz}{g(y)(1 + \Lambda(y))}.$$

Since $\tilde{m} > 0$ and $\frac{d}{dy}\tilde{m} > 0$ for values of y below y_2, we must have that

$$\frac{d}{dy}\tilde{m}(y) = \frac{\Lambda(y) - \lambda \ell_a(y)}{1 + \Lambda(y)} - (\tilde{m} + \tau/k)\, \ell_y > 0,$$

where we have used the fact that $\Lambda(y)$ is constant over (y_1, y_2). By Assumption 5, we know that $\ell_y \geq 0$. Hence the above inequality implies that

$$\Lambda(y) - \lambda \ell_a(y) > 0 \tag{19}$$

for values of y close to y_2. Since by Assumption 5, $\ell_{ay} \geq 0$, the above implies that (19) should hold for all values of $y \geq y_1$. This in turn means that the derivative of the function $\int_y^1 [\lambda \ell_a(y) - \Lambda]\, dG$ is positive over (y_1, y_2) and, as a result, $\tilde{m}(y) + \tau/k > 0$ for all values of $y \in (y_1, y_2)$. Moreover,

$$\frac{d^2}{dy^2}\tilde{m}(y) = -\frac{\lambda \ell_{ay}(y)}{1 + \Lambda(y)} - (\tilde{m} + \tau/k)\, \ell_{yy} - \ell_y \frac{d}{dy}\tilde{m}.$$

By Assumption 5, we must have that $\ell_{yy}, \ell_{ay}, \ell_y \geq 0$, which implies that the above is negative and thus $\tilde{m}$ is concave. Since $\tilde{m}$ is increasing at y_2, concavity implies that it should be increasing for all values of $y \in [y_1, y_2]$. This implies that $\tilde{m} = m^*$ and the monotonicity constraint is slack for the entire interval. Finally,

$$\begin{aligned}\frac{d^3}{dy^3}\tilde{m} =& -\frac{\lambda \ell_{ayy}(y)}{1+\Lambda(y)} - (\tilde{m} + \tau/k)\,\ell_{yyy} - 2\ell_{yy}\frac{d}{dy}\tilde{m} - \ell_y \frac{d^2}{dy^2}\tilde{m} \\ =& -\lambda \frac{\ell_{ayy} - \ell_y \ell_{ay}}{1+\Lambda(y)} - (\tilde{m} + \tau/k)\left[\ell_{yyy} - \ell_y \ell_{yy}\right] - \left(2\ell_{yy} - (\ell_y)^2\right)\frac{d}{dy}\tilde{m}.\end{aligned}$$

By Assumption 5, all of the elements of the above are negative and thus $\frac{d^3}{dy^3}\tilde{m} \leq 0$. Hence,

$$\begin{aligned}\frac{d^3 p^*}{dy^3} =& \frac{d^2}{dy^2}(km^* + \tau)\left(1 + \frac{d}{dy}m^*\right) \\ =& k\frac{d^2 m^*}{dy^2}\left(1 + 3\frac{dm^*}{dy}\right) + (km^* + \tau)\frac{d^3 m^*}{dy^3} \leq 0.\end{aligned}$$

This implies that if at y, p^* is concave, it should be concave for all higher values. Now recall that using majorization, we had argued that $\frac{d}{dy}p^* \geq 1$ for values of y close to and above y_1. Since majorization is binding at y_1 and y_2 , we must have that $\int_{y_1}^{y_2}(p^*(y) - y)\,dG = 0$. Thus, for some intermediate value of $y \in (y_1, y_2)$, we must have that $\frac{d}{dy}p^* < 1$. This in turn means that $\frac{d}{dy}p^*(y_2) < 1$, which cannot hold if we are to have continuity of the interim price function and $p^*(y_2) = y_2$. This concludes the proof.

Thus, it remains to show that the solution of (P1) is the same as that of (P2). Consider the $\underline{u}^*$ associated with the solution of (P1) and suppose that the solutions of (P1) and (P2) are not the same. In this case, a similar argument to the perturbations considered in Lemma 3 establishes that we can improve upon the objective in (P1). □

B Online Appendix

B.1 Strong Duality for Optimal Ratings with Manipulation

In this section, we show that strong duality holds for the optimization problem in (P2). This would imply that Lagrange multipliers exist so that the constrained optimization is equivalent to unconstrainted optimization of the Lagrangian.

We use a result from Mitter (2008) (see also Kleiner and Manelli (2019)) to show existence of Lagrange multipliers for the optimization (P2). Mitter's result establishes that strong duality holds for an optimization problem of the form $\min\{f(u)|u \in C, g(u) \leq 0\} = V_0$ where C is a convex set, $\leq$ is associated with a convex cone and $g(\cdot)$ is convex if $\exists \varepsilon > 0, M > 0$ s.t. $f(u) \geq V_0 - M|b|$ for all $u \in C, g(u) \leq b, |b| \leq \varepsilon$. The objective $f(\cdot)$ is a convex function that is possibly infinite-valued. We state this in the following lemma:

Lemma 3. *The optimization in (P2) satisfies strong duality.*

Proof. To apply Mitter (2008)'s result, let us define the linear vector space

$$X = \{m(y) = \hat{x}(y) - y | m : [0,1] \to \mathbb{R}, m(1) = 0\}$$

equipped with the sup-norm. Let $C \subset X$ satisy (15), $m(y) \geq 0$, and $1 - y \geq m(y)$ and $m(y) + y$ is increasing.The set C is obviously convex since c_m is convex. Let $F(b) = \left\{m \in C| - \int_0^1 c_m'(m(y)) G_a dy - c'(a) \geq b\right\}$. Then our optimization is $V(b) = \min_{m \in F(b)} f(m)$ where $f(m)$ is the negative of the objective in (P2) and $b \in \mathbb{R}$. Let m^* be a feasible manipulation strategy for which $f(m^*) = V(0)$. Let $m \in F(b)$. If $b > 0$, then since $F(b) \subset F(0)$, we have that $f(m) \geq f(m^*) = V(0)$.

Now, suppose that b is negative. To prove the claim, we proceed as follows:

Step 1. We show that there exists $\hat{m} \in C$ such that $-\int_0^1 c_m'(\hat{m}) G_a dy = c'(a) + d$ for some $d > 0$. To see this, consider the point of optimality m^*. Since $m^* \in [0, 1-y]$ for values of y close to 1, the implied interim price must have a low slope and thus the majorization constraint must be slack. This in turn implies that there exists a highest value of $\tilde{y} < 1$ so that majorization is binding at $\tilde{y}$ and slack for values of $y > \tilde{y}$. Since we can use the argument in Theorem 2 to show that optimal manipulation and interim prices are continuous functions. This implies that for values of y close to $\tilde{y}$, $p(y)$ is increasing and $p(y) > y$. Consequently, $m^*(y) > 0$ and $m^*(y) + y$ is strictly increasing for such values. Let $I = [y_1, y_2]$ be an interval of y's for which these properties hold. Suppose contrary to the claim that $\forall m \in C$,

$$-\int c'_m(y)\, G_a dy \leq c'(a)$$

Consider the following set

$$A = \left\{ m : [y_1, y_2] \to \mathbb{R} \;\middle|\; \begin{array}{c} m(y_1) = m^*(y_1), m(y_2) = m^*(y_2) \\ m + y\text{increasing} \\ \int_I c'_m dy = \int_I c'_m(m^*)\, dy \\ , \int_{y_1}^{y_2} \left[c_m g(y) + c'_m(1 - G(y))\right] dy \leq \\ \int_{y_1}^{y_2} \left[c_m(m^*)\, g + c'_m(m^*)(1 - G)\right] dy \end{array} \right\}$$

This is a convex set. Moreover, let

$$B = \left\{ m : [y_1, y_2] \to \mathbb{R} | -\int c'_m(y)\, G_a dy \geq -c'(a) \right\}$$

B is a convex set with non-empty interior. Then our contrary assumption implies that $A \cap B = \{m^*\}$. Since $m^* + y$ is strictly increasing over I, it must be that every perturbation $\varepsilon(y)$ of m in every direction that satisfies $\int_I \left[c'_m g + c''_m(1 - G)\right] \varepsilon dy = 0$ and $\varepsilon(y_1) = \varepsilon(y_2) = 0 = \int_I \varepsilon dy$. As a result and by using separating hyperplane theorem, there must exist λ_1 and λ_2 such that

$$\forall y \in I, c'_m(m^*(y))\, g(y|a) + k(1 - G(y|a)) = \lambda_1 k - \lambda_2 G_a(y|a)\, k \tag{20}$$

Now consider the set

$$B' = \left\{ m | m \in B, \int_I \left[c'_m(m) - c'_m(m^*)\right](1 - G)\, dy \geq 0 \right\}$$

We can again show that $A \cap B' = \{m^*\}$. To see this, suppose there exists $m' \neq m, m' \in A \cap B$. Then strict convexity of c_m implies that if we set $\overline{m} = (m + m')/2$, then

$$\int_{y_1}^{y_2} \left[c_m(\overline{m}(y))\, g(y) + c'_m(\overline{m}(y))(1 - G(y))\right] dy <$$
$$\int_{y_1}^{y_2} \left[c_m(m^*(y))\, g(y) + c'_m(m^*(y))(1 - G(y))\right] dy$$

Since $\overline{m} \in B'$, we must have that

$$\int_I c'_m\left(\overline{m}\left(y\right)\right)\left(1-G\left(y|a\right)\right)dy \geq \int_I c'_m\left(m^*\right)\left(1-G\left(y|a\right)\right)dy$$

Adding the two above inequalities implies that

$$\int_I c_m\left(\overline{m}\right)gdy \leq \int_I c_m\left(m^*\right)gdy$$

Hence, the function $\tilde{m}\left(y\right)=m^*\left(y\right), y \notin I$ and $\tilde{m}\left(y\right)=\overline{m}\left(y\right), y \in I$ satisfies all the constraints in $F\left(0\right)$ and improves the objective in (P2) which is in contradiction with m^* being optimal. Thus $A \cap B' = \{m^*\}$. A similar argument as before shows that

$$\forall y \in I, c'_m\left(m^*\left(y\right)\right)g\left(y|a\right)+k\left(1-G\left(y|a\right)\right)=\mu_1 k-\mu_2 G_a\left(y|a\right)k+\mu_3\left(1-G\left(y|a\right)\right)k$$

Combining the above with (20) implies that

$$\mu_1 k-\mu_2 G_a\left(y|a\right)k+\mu_3\left(1-G\left(y|a\right)\right)k=\lambda_1 k-\lambda_2 G_a\left(y|a\right)k, \forall y \in I$$

If we take a derivative of this equation, we must have

$$\left(\mu_2-\lambda_2\right)g_a\left(y|a\right)+\mu_3 g\left(y|a\right)=0, \forall y \in I$$

Since MLRP is a strictly increasing function, this leads to a contradiction. Therefore, we must have that there exists $\hat{m} \in C$ such that $-\int c'_m\left(m\right)dG = c'\left(a\right)+d$ for some d positive.

Step 2. Given $\hat{m}$, for any $m \in F\left(b\right)$ for $b<0$, consider $\tilde{m}=\frac{d}{d-b}m+\frac{-b}{d-b}\hat{m}$. Since C is convex, we must have that $\tilde{m} \in C$. Moreover, since $c'_m\left(\cdot\right)$ is linear, we must have that

$$-\int_0^1 c'_m\left(\tilde{m}\left(z\right)\right)G_a\left(z|a\right)dz-c'\left(a\right)= \\ \frac{d}{d-b}b-\frac{b}{d-b}d=0$$

This implies that $\tilde{m} \in F(0)$. Moreover,

$$\begin{aligned}\|\tilde{m} - m\| &= \frac{-b}{d-b}\|\hat{m} - m\| \\ &= \frac{-b}{d-b}\sup_y |\hat{m}(y) - m(y)| \\ &\leq \frac{-b}{d-b} \leq \frac{|b|}{d}\end{aligned}$$

It is fairly straightforward to show that $f(\cdot)$ has a Frechet derivative with a bounded norm. Therefore, $N > 0$ exists such that for all m_1, m_2, $|f(m_1) - f(m_2)| \leq N\|m_1 - m_2\|$. Thus, we have

$$f(m) \geq f(\tilde{m}) - N\|m - \tilde{m}\| \geq f(\tilde{m}) - \frac{|b|}{d}N \geq f(m^*) - \frac{|b|}{d}N$$

which concludes the claim. □

B.2 Optimal Ratings without Manipulation with Separable Distributions

In this section, we characterize the optimal rating systems in section 4 for a special case of separable distributions.

Consider the problem in this section and assume that $g(y|a)$, the density of y given a satisfies the following separability

$$g(y|a) = 1 + \beta(a)m(y) \tag{21}$$

where $m(y)$ is an increasing function that satisfies $\int_0^1 m(y)\,dy = 0$ and $\beta(a)$ is increasing and concave. Note that under this specification, the marginal benefit of effort is given by

$$\int_0^1 p(y)\,g_a(y|a)\,dy = \beta'(a)\int_0^1 p(y)\,m(y)\,dy$$

Hence, if effort profile $a(\theta)$ is optimal, then

$$\beta'(a(\theta))\int_0^1 p(y)\,m(y)\,dy = c_a(a(\theta), \theta)$$

Hence, we must have

$$\frac{c_a\left(a\left(\theta\right),\theta\right)}{\beta'\left(a\left(\theta\right)\right)}=\frac{c_a\left(a\left(\theta'\right),\theta'\right)}{\beta'\left(a\left(\theta'\right)\right)}$$

Thus, if we choose the effort level of the lowest type $\underline{a}$, the above determines the effort level for all the other types. Let us refer to the solution of the above as $\hat{a}\left(\underline{a},\theta\right)$. Hence, the problem of optimal rating design is given by

$$\max_{\underline{a},p(\cdot)}\int_0^1 p\left(y\right)g\left(y|\underline{a}\right)dy-c\left(\underline{a},\underline{\theta}\right)$$

subject to

$$\int_0^1 p\left(y\right)g_a\left(y\right)dy=c_a\left(\underline{a},\underline{\theta}\right) \tag{22}$$

$$\int_0^y p\left(\hat{y}\right)\int g\left(\hat{y}|\hat{a}\left(\underline{a},\theta\right)\right)dFd\hat{y}\geq\int_0^y\hat{y}\int g\left(\hat{y}|\hat{a}\left(\underline{a},\theta\right)\right)dFd\hat{y},\forall y\in(0,1)$$

$$\int_0^1 p\left(\hat{y}\right)\int g\left(\hat{y}|\hat{a}\left(\underline{a},\theta\right)\right)dFd\hat{y}=\int_0^1\hat{y}\int g\left(\hat{y}|\hat{a}\left(\underline{a},\theta\right)\right)dFd\hat{y}$$

$$p\left(y\right)\geq p\left(y'\right),\forall y\geq y'$$

Similar to the analysis in subsection 4, given $\underline{a}$, the problem of solving for optimal interim prices is to maximize $\int p\left(y\right)\frac{g(y|\underline{a})+\gamma g_a(y|\underline{a})}{h(y)}h\left(y\right)dy$ subject to majorization and monotonicity where $h\left(y\right)=\int g\left(y|\hat{a}\left(\underline{a},\theta\right)\right)dF\left(\theta\right)$. Note that in this formulation γ is the Lagrange multiplier associated with the incentive constraint

Given the separability assumption on $g\left(\cdot|\cdot\right)$, we can show that the function $\frac{g(y|\underline{a})+\gamma g_a(y|\underline{a})}{h(y)}$ is either decreasing in y – when γ is low enough – or increasing – for high γ. Thus the solution of the above problem is either full pooling or full information. Since full pooling leads to marginal benefit of effort being 0, we have the following proposition:

Proposition 4. *Suppose that $g\left(\cdot|a\right)$ satisfies (21).Then optimal monopoly rating system is full disclosure.*

This result can be understood by considering the two effects identified before: redistributive and incentive. Given our specification of the distribution, the forces cannot be balanced. Since the redistributive force cannot dominate as it leads to no effort being taken by the DM, full disclosure should be optimal.

B.3 Monotonicity

In this section, we show that the monotonicity assumption on interim prices in section 4 is without loss of generality. We do this for continuously differentiable functions. More specifically, we show that if $\hat{p}$ is a continuously differentiable function that is decreasing in some subinterval of $[0,1]$, then there exists another $\tilde{p}$ which is "less decreasing" and delivers a higher value to the lowest participating types and does not decrease the marginal benefit of effort for all DM types. This implies that in the set of continuously differentiable functions, each interim price function can be replaced with a montone version of itself. Since by Stone-Weierstrass theorem, the set of smooth functions is dense in $L_\infty([0,1])$, this implies focusing on montone interim price functions is without loss of generality. Note also that we do this in the set of random variables – implied by $\hat{p}$ – that are a mean preserving contraction of y. Obviously any interim price belongs to this set. Thus, if we show that the optimal interim price in such a set should be increasing, we prove our claim.

Suppose that $\hat{p}$ is strictly decreasing on an interval $[y_1, y_2]$. Since $\hat{p}$ cannot be decreasing for the entire interval $[0,1]$, we can assume that $\hat{p}$ is increasing over $[y_0, y_1]$ and $\hat{p}(y_0) = \hat{p}(y_2)$ – an alternative would be that $\hat{p}$ is increasing over $[y_2, y_0]$ for some $y_0 > y_2$ but this is somewhat symmetric to our case and can be dealt with the same way. Moreover, let us choose the lowest such interval. This would imply that the majorization constraint is not binding for values of p in such an interval – this is because $\hat{p}$ is increaing below and thus the quantile representation of $\hat{p}$ is above that of y. This would imply that a spreading of values in a neighborhood of this point does not violate the mean preserving contraction property.[17]

Let y_1' be the lowest value of $y \in (y_0, y_1)$ such that $\hat{p}(y_1') = \hat{p}(y_1)$. Without loss of generality, let us assume that y_1 is the highest such value. Moreover, we can assume that $\hat{p}$ is strictly increasing over (y_0, y_1') and strictly decreasing over (y_1, y_2). Now, consider

[17]For any function $\hat{q}(y)$, the random variable $\hat{q}$ is a mean preserving spread of y if and only if

$$\int_0^i Q_q(j)\,dj \geq \int_0^i Q_y(j)\,dj, \forall i < 1$$
$$\int_0^1 Q_q(j)\,dj = \int_0^1 Q_y(j)\,dj$$

where Q_q and Q_y are quantile representation of $\hat{q}$ and y.

the perturbation

$$\tilde{p}(y) = \begin{cases} \hat{p}(y) + \varepsilon(y - y_0) & y \le \tilde{y} \\ \hat{p}(y) - \delta(y) & y \ge \tilde{y} \end{cases}$$

where in the above, $\hat{p}(\tilde{y}) = \hat{p}(y_1) - \varepsilon(\tilde{y} - y_0), \tilde{y} \le y_1$. Moreover, the function $\delta(y)$ is constructed such that $\delta(y)$ is constant over (y_1', y_1) and equal to $\varepsilon_2(y_2 - y)$ over (y_1, y_2). Furthermore, $\varepsilon \int_{y_0}^{\tilde{y}} (y - y_0)\, dH = \int_{\tilde{y}}^{y_2} \delta(y)\, dH$ where H is the distribution of y and $\delta(\tilde{y}) = -\varepsilon(y - y_0)$. Note that for ε small, the length of the interval $(\tilde{y}, y_1')$ is proportional to ε. Additionally, ε_2 is proportional to ε for small values of $\varepsilon, \varepsilon_2 > 0$The change in incentives for any type is given by

$$\begin{aligned} \int_0^1 (\tilde{p} - p)\, g_a dy &= -\int_0^1 G_a d(\tilde{p} - p) \\ &= -\int_{y_0}^{\tilde{y}} G_a \varepsilon dy - \int_{y_1}^{y_2} \varepsilon_2 G_a dy + o\left(\varepsilon^2\right) \\ &> 0 \end{aligned}$$

where the inequality follows from the fact that $G_a < 0$. This implies that marginal benefit of effort increases for all DM types. Finally, we need to check that the payoff of the lowest DM type increases. This is because the distribution of y, $H(y)$, puts more weight on higher realizations of y relative to that of the lowest participation type. This proves the claim.

B.4 First Order Approach

Consider the payoff of the DM given by

$$\int p(y)\, g(y|a)\, dy - c(a)$$

For the validity of the first order approach, it is sufficient to have this function be concave for all increasing functions $p(y)$. This holds if

$$\int p(y)\, g_{aa}(y|a)\, dy \le 0$$

or

$$\int G_{aa}(y|a)\, dp(y) \ge 0$$

Since $p(\cdot)$ is montone, the above implies that $G_{aa} \geq 0$. Thus if G is convex in a, then first order approach is valid. For the class of distribution functions of the form $\log g = r(y)\, m(a) + b(a)$ with m and r increasing, we have

$$\begin{aligned} G_{aa}(y|a) &= \int_0^y g(z|a)\left[(b'(a) + m'(a)\, r(z))^2 + m''(a)\, r(z) + b''(a)\right] dz \\ &= \int_0^y g(z|a)\left[(b'(a) + m'(a)\, r(z))^2 + m''(a)\, r(z) + b''(a)\right] dz \end{aligned}$$

If the expression in the bracket is decreasing in z, then $G_{aa} \geq 0$. This is the case when

$$\begin{aligned} 2(m')^2\, r(y) + 2b'(a)\, m'(a) + m''(a) \leq 0 \\ 2(m')^2 (r(1) - \mathbb{E}r) + m''(a) \leq 0 \end{aligned}$$

An example for this is when $G(y|a) = y^{m(a)}$ with $m(\cdot)$ increasing, concave and $2(m')^2 + mm'' \leq 0$.

Note that while $G_{aa} \geq 0$ is sufficient, it is not necessary. More specifically, it is possible that $G_{aa} \leq 0$ for some values of y but if at such values $p(y)$ is flat, then $\int G_{aa}(y|a)\, dp \geq 0$ for all values of a and hence the first order approach is valid. For example, for $G(y|a) = 1 - (1-y)^{1/a-1}$ with $m(a) \leq 0$ and increasing, the above inequality is violated. However, if p is pooling for y above $1 - e^{2a}$ then the objective is concave and thus the first order approach is valid.

B.5 Extensions

In this section, we expand on the extensions that were briefly discussed in section (6).

B.5.1 Allowing for Market Action

So far, our analysis of rating systems was limited to environments where the market simply pays its expectation to the DM. In several settings, the information provided is also valuable for the market since it allows it to make better decision.Here, we show that it is possible to extend our characterization result to such settings.

Specifically, suppose that market participants have a payoff of $e \cdot \left(y - \underline{y}\right)$ with $e \in \{0, 1\}$ being chosen by the market. Suppose as before that the price paid to the DM is $\left(\mathbb{E}[y|s] - \underline{y}\right) \mathbf{1}\left(\mathbb{E}[y|s] \geq \underline{y}\right)$. For any arbitrary rating system (S, π), we can define the following objects:

$$q(y) = \Pr\left(\left\{s : \mathbb{E}[y|s] \geq \underline{y}\right\}|y\right)$$
$$\hat{p}(y) = \frac{1}{q(y)}\mathbb{E}\left[\left(\mathbb{E}[y|s] - \underline{y}\right)\mathbf{1}\left(\mathbb{E}[y|s] \geq \underline{y}\right)|y\right], \text{if } q(y) > 0 \tag{23}$$

In words, $q(y)$ is the probability that market action is equal to 1 conditional on the state being equal to y. Furthermore, $p(y) = q(y)\hat{p}(y)$ is the interim price vector faced by the DM. Thus, $\hat{p}(y)$ is the interim price conditional on $e = 1$.

Note that given this change of variable, we can write

$$\hat{p}(y) = \sum_{s\in\hat{S}} \hat{\pi}(s|y) \frac{\sum_{y'\in Y} \hat{\pi}(s|y') \left(y' - \underline{y}\right) q(y') \mu_y(y')}{\sum_{y'\in Y} \hat{\pi}(s|y') q(y') \mu_y(y')}$$

where $\hat{S} = \left\{\mathbb{E}[y|s] \geq \underline{y}\right\}$ and $\hat{\pi}(s|y) = \frac{\pi(s|y)}{q(y)}$. This implies that $\hat{\pi}(\cdot|y) \in \Delta\left(\hat{S}\right)$. Thus, $\hat{p}(y)$ becomes the interim price function associated with the signal structure $(S, \hat{\pi}(\cdot|y))$ with the prior distribution of y given by $q(y)\mu_y(y)$.

The following proposition summarizes this logic while allowing for arbitrary probability measures and a direct application of Theorem 1 to $\hat{p}$ and $\hat{\pi}$:

Proposition 5. *Suppose that market payoff is given by $v(a,y)\cdot e$ where $e \in \{0,1\}$ is the action optimally taken by the market and that marginal distribution of y is given by a probability measure μ_y. Then if a <u>positive</u> measure ρ and a pricing function $\hat{p}(y)$ exists such that*

1. *$\rho\left(\hat{Y}\right) \leq \mu_y\left(\hat{Y}\right)$ for all Borel subsets $\hat{Y} \subset Y$,*

2. *interim price function satisfies $\hat{p}(y)$*

$$\int \hat{p}(y)\,d\rho = \int \overline{v}(y)\,d\rho$$
$$\int u(\hat{p}(y))\,d\rho \geq \int u(\overline{v}(y))\,d\rho, \forall u\text{: concave}$$

3. *$\hat{p}(y)$ is co-monotone with $\overline{v}(y)$ and $\min_{y\in Y}\hat{p}(y) \geq 0$*

4. *$\int \overline{v}(y)\,d\mu_y \leq \int \hat{p}(y)\,d\rho$*

Then, there exists (S, π) *such that*

$$\int_A \hat{p}(y)\, d\rho = \int_A \mathbb{E}\left[\mathbb{E}\left[\max_{\alpha \in \{0,1\}} v \cdot \alpha \middle| s\right] \middle| y\right] dy \tag{24}$$

for all Borel subsets A *of* Y.

Note that (24) is the measure theoretic version of (23). When ρ is a sum of continuous and discrete distributions, then it reduces to

$$\rho(y)\, \hat{p}(y) = \mathbb{E}\left[\mathbb{E}\left[\max_{e \in \{0,1\}} v \cdot e \middle| s\right] \middle| y\right]$$

where $\rho(y)$ is the density or probability of ρ at y.

It is worth mentioning that, first, the positive measure ρ can be thought of as the probability of state being y and posterior mean of $v(a, y)$ being positive. Since posterior mean of v can be sometimes negative – which leads to $e = 0$, ρ is not necessarily a probability but a positive measure. Moreover, by the first condition, ρ is absolutely continuous with respect to μ_y and thus by Radon-Nykodim theorem there must exist $0 \leq q \leq 1$ such that $\rho = \int q(y)\, d\mu_y$. The function q is then the probability of positive posterior mean conditional on $y \in \hat{Y}$. Second, an incentive compatibility for the market needs to be included which is given by part 4 in Proposition 5.

To summarize, Proposition (5) establishes that even in the presence of market action, a modified version of our characterization result can be applied. In this case and under the co-monotonicity restriction, the information structure can be summarized by two objects: 1. the probability that action $e = 1$ is chosen – represented by the measure ρ in the Proposition; 2. the interim price function $\hat{p}(\cdot)$ which is conditional on inducing $e = 1$ as the action. To the extent that this is a more concise summary of a rating system, our characterization significantly simplifies the information design problem in presence of moral hazard. Finally, we should note that while we have decided to focus on a two action problem, it is readily observed that this can be extended to many finite actions. With many but finite actions and under certain monotonicity constraints, each action is optimal for a certain interval of ex-post values for the market. Knowing this, Proposition 5 can be extended by considering multiple interim price functions and multiple probability measures for each action by the market.

B.5.2 Different Priors

Consider a variation of the model in section 3, wherein the market has dogmatic prior beliefs about the distribution of (a, θ); this is in contrast with the market having rational expectations which coincide with the equilibrium behavior of DM. More specifically, let $\phi \in \Delta (A \times \Theta)$ be the prior of the market and suppose that the market uses this prior and the true signal distribution to do Bayesian updating. When A, Θ, and Y are finite and we can write $\phi (a, \theta)$ as the probability of (a, θ) under market prior, interim prices are given by

$$p(y) = \sum_{s \in S} \frac{\sum_{(a,\hat{y}) \in A} v(a, \hat{y}) \phi(a, \hat{y}) \pi(s|\hat{y})}{\sum_{(a,\hat{y}) \in A} \phi(a, \hat{y}) \pi(s|\hat{y})} \pi(s|y).$$

The above is identical to interim prices in section 3 except for the fact that ϕ is used instead of the true distribution of (a, y). One can then conclude that the following holds:

Lemma 4. *If market prior is given by ϕ and $\overline{v}_\phi (y) = \mathbb{E}_{\phi,\sigma} [v (a, y) |y]$, if $p (y)$ is co-monotone with $\overline{v}_\phi (y)$ and $p (y) \succcurlyeq_{S.O.S.D} \overline{v}_\phi (y)$, both distributed according to ϕ_y, then there exists an information structure that induces $p (y)$.*

To see the benefit of this result, consider a simple setting in which there is only one type of DM who has a cost $c (a)$ and market has a biased belief that the density of y is given by $h (y)$. Moreover, suppose that $\mathbb{E}_\phi [v (a, y) |y] = \overline{v}_\phi (y) = \alpha y + (1 - \alpha) \overline{y}$. Then the problem of optimal rating design is to find $p (y)$ and a to solve the following:

$$\max_{p(\cdot),a} \int_0^1 p(y) g(y|a) dy - c(a)$$

subject to

$$\int_0^y p(y') h(y') dy' \geq \int_0^y p(y') h(y') dy'$$
$$\int_0^1 p(y) h(y) dy = \int_0^1 \overline{v}_\phi(y) h(y) dy$$

We can then use simple arguments from calculus of variations, similar to those in section 4, to show the following:

Proposition 6. *Optimal ratings are:*

1. *Upper-censorship if for all a, $h (y) / g (y|a)$ is hump-shaped, i.e., increasing-then-decreasing.*

Moreover, if a^* achieves the maximum of $\mathbb{E}[y|a] - c(a)$ and if $h(y)$ is strictly dominated by $g(y|a^*)$ according to first order stochastic dominance, then optimal rating is not full information.

2. *Lower-censorship if for all a, $h(y)/g(y|a)$ is U-shaped, i.e., decreasing-then-increasing. Moreover, if a^* achieves the maximum of $\mathbb{E}[y|a] - c(a)$ and if $h(y)$ strictly dominates $g(y|a^*)$ according to first order stochastic dominance, then optimal rating is not full information.*

Proposition 6 again illustrates the power of our result on characterization of interim prices. It describe how the shape of the bias determines the structure of optimal ratings. More specifically, in the class of the distributions considered, more pessimism leads to upper-censorship while more optimism leads to lower-censorship.