

Optimal Disclosure When Agents Game the Score*

Gregory Dannay[†]

October 21, 2026

Abstract

Organizations that score individuals sometimes disclose how much a variable matters for the score, even though individuals can use this information to game it. I study a principal who scores an agent based on a report that the agent can distort at a cost. The principal privately knows her valuation, which determines how strongly the score responds to the report. The agent distorts more when he expects a higher valuation, precisely when scoring errors are more costly. As a result, opacity strictly dominates any nontrivial monotone disclosure policy. This conclusion changes when the principal can also remain silent. The optimal policy announces that the valuation lies in an intermediate interval and remains silent otherwise. By pooling low and high valuations, silence limits distortion when the valuation is high. This policy strictly improves upon opacity and, in numerical examples, captures most of the gains from unconstrained disclosure.

Keywords: Scoring, Manipulation, Information Design.

JEL Codes: D82, D83, C72.

*I thank Laurent Mathevet, Giacomo Calzolari, and Eduardo Perez-Richet for their guidance. I also thank Özlem Bedre-Defolie, Zeinab Aboutaleb, Matthew Mitchell and Zvika Neeman for insightful discussions and comments at various stages of the project as well as participants at various seminars and conferences.

[†]European University Institute

1 Introduction

Institutions increasingly rely on predictive models to guide decisions or to determine outcomes outright. Algorithms rank websites in search results, determine borrowers' credit scores, screen job applicants' resumes, and set flight prices. The inner workings of these models, and the data they draw on, are often kept secret, ostensibly to prevent manipulation of the inputs.¹ Once people learn which variables matter to a model, they can adjust their behavior, falsify documents, or simply lie to secure more favorable outcomes. This is the logic behind Goodhart's law: "When a measure becomes a target, it ceases to be a good measure." Yet organizations commonly release broad indications of how their algorithms work. The FICO credit score is one example: its website reports that the model weighs an individual's payment history, amounts owed, length of credit history, new credit, and credit mix.² This information can in turn be used to game the score. One website advises spreading balances across cards to keep the utilization of each card low, a strategy that costs little and does not change a person's credit risk.³ Motivated by such practices, I study how an organization can optimally disclose information about its predictive model to limit harm caused by gaming.

Revealing which variables matter tells individuals where to concentrate their gaming efforts. If an organization reveals that a variable matters little, individuals have little incentive to game it, but this brings little benefit, since distorting that variable does little harm. If, instead, it reveals that a variable matters a lot, individuals have a strong incentive to distort it, and the decision is then made on falsified data. This suggests a natural rationale for opacity. This paper shows, however, that partial disclosure strictly improves upon opacity: the organization discloses when a variable's importance is intermediate and stays silent when the variable matters very little or very much. Silence then leaves individuals uncertain about

¹How Much Do We Really Know About Google's Ranking Algorithm?, <https://medium.com/swlh/how-much-do-we-really-know-about-googles-ranking-algorithm-ef031586681b>

²What's in my FICO Scores?, <https://www.myfico.com/credit-education/whats-in-your-credit-score>

³How to Improve Your FICO Score: 15 Proven Strategies, <https://creditficoscores.com/fico-scores/how-to-improve>

which of the two extremes holds, which limits their incentive to game when the variable matters most.

I study the optimal disclosure policy in a principal-agent model in which the principal scores an agent based on a report about his privately known quality. The principal holds a private valuation that determines how much the score should respond to quality, and she commits in advance to a policy for disclosing this valuation. The valuation can be thought of, for instance, as the weight of an observable variable in a linear regression that predicts an underlying characteristic. The agent forms a belief about the valuation and can inflate his report about his quality at a cost, which represents, for example, the price of forging documents, the risk of fines if caught, or the psychological cost of lying. The principal observes the report and assigns a score. She suffers a quadratic loss from the distance between this score and the ideal score, her valuation times the agent’s quality, so that her scoring errors are most costly when her valuation is high. However, the agent distorts more the higher he expects the valuation to be, since his gain from inflating his report is then higher. Disclosing that the valuation is high, therefore, triggers gaming exactly where it hurts most.

Institutional or organizational constraints may limit the set of disclosure strategies available to the principal. An organization can hardly announce that a variable “matters either very little or a lot”, or randomize over what it says. Its statements must instead make sense on their own, such as that a variable matters “a lot” or “moderately”, which amounts to revealing an interval of weights. FICO, for instance, publishes approximate weights for the five broad categories of credit information that enter its score. Regulation codifies the same practice. Providers of online search engines and intermediation services must disclose the main parameters determining their rankings and the relative importance of these parameters.⁴ Online platforms face a similar requirement for their recommender systems.⁵

⁴Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services, Article 5.

⁵Regulation (EU) 2022/2065 (Digital Services Act), Article 27.

However, the former regulation exempts providers from disclosing algorithms or information that would enable consumers to be deceived through the gaming of search results.

Drawing on the literature on constrained information design, I therefore first restrict the principal to monotone signals, which pool valuations into intervals. Such signals cannot improve upon opacity. Pooling two ordered messages raises the posterior mean, and hence gaming, for the low valuations, but lowers it for the high valuations, whose scoring errors are more costly, and the gain always outweighs the loss (Proposition 2). Opacity therefore strictly dominates every monotone signal, including full transparency (Theorem 1).

In a second part, I give the principal one extra degree of freedom: the ability to remain silent. I introduce *monotone-or-mute* signals, in which all messages but one are ordered, and the remaining message, silence, is unrestricted. In practice, this represents an organization that describes the importance of a variable on an ordered scale, but may also say nothing. Additional ordered categories never help (Corollary 1), and Theorem 2 shows that the optimal monotone-or-mute signal announces that the valuation lies in an intermediate interval and remains silent otherwise, and that it strictly improves upon opacity. Silence pools the lowest valuations, which matter little, with the highest ones. An agent who hears nothing, therefore, expects a lower valuation than under opacity and distorts less, precisely when the principal’s errors are most costly.

As a benchmark, I characterize the optimal unconstrained disclosure policy using the results of Kolotilin et al. (2025) (henceforth KCW). The optimal signal pairs each high valuation with a low one: every message reveals that the valuation is one of two values, and wider pairs induce lower posterior means, so that the highest valuations face the least gaming. This signal is hard to implement, since it uses a continuum of messages, each pooling two distant valuations, and such disclosure is rarely observed in practice.

Finally, I quantify these gains numerically, measuring each policy by the change in the principal’s loss relative to opacity. In every example, the optimal monotone-or-mute signal achieves 78% to 79% of the reduction achieved by the unconstrained optimum. The gain

from silence is negligible when valuations vary by less than a factor of three. It is substantial when the highest valuation is a hundred times the lowest and most valuations are low: the loss then falls by 13%. It remains substantial even when the agent has greater distortion ability: the loss still falls by 8% when manipulation accounts for 42% of the report’s variance. Full transparency, by contrast, raises the loss by 31% to almost 7000%.

The rest of the paper is organized as follows. The remainder of this section reviews the related literature. Section 2 presents the model, characterizes the linear equilibrium, and reduces the principal’s payoff to a function of the first two moments of her posterior. Section 3 studies monotone signals and shows that opacity dominates this class of signals. Section 4 introduces monotone-or-mute signals and characterizes the optimal one. Section 5 characterizes the unconstrained optimum using the results of KCW. Section 6 quantifies the gains from silence and compares them with those of the unconstrained optimum. Section 7 concludes.

1.1 Literature review

This paper is related to the literature on strategic manipulation, in which an agent can exert a costly and socially wasteful effort to obtain a better outcome. The principal, in these setups, incurs a loss because she either makes a decision based on falsified data or must design a mechanism to account for this distortion. This situation has been studied through the lens of optimal contracting (Lacker and Weinberg, 1989), scoring (Frankel and Kartik, 2019, 2022; Ball, 2025), test design (Perez-Richet and Skreta, 2022, 2026), price discrimination (Tan, 2025), and, in computer science, classification algorithms (Hardt et al., 2016). These papers assume that the agent knows how his report will be used. An exception is Ederer et al. (2018), who show that a principal can deter gaming in a multitask incentive problem by keeping the agent uncertain about which task will be rewarded. In their setting, this uncertainty arises from the randomization of the incentive scheme. In this paper, the weight on the agent’s report is the principal’s private information, and she commits to a

policy for disclosing it rather than to the scoring rule itself.

Two papers study closely related questions. Wang et al. (2023) compares the two extreme cases of complete opacity and full transparency in a model where agents can invest in themselves, and shows that transparency can benefit the principal under certain conditions. In my model, agents can only wastefully distort their reports. Full transparency is then strictly worse than opacity, but partial disclosure strictly improves upon it. Sun (2025) also studies optimal disclosure with wasteful distortion. In that model, the agent’s type is binary, and the principal decides whether to lend, with the payoff linear in the principal’s private valuation. He shows that the optimal disclosure policy mixes continuous and discrete signals and collapses to opacity under monotonicity when the agent’s type is skewed toward low types. In my model, the principal instead assigns a continuous score to an agent with a continuum of types and has a quadratic loss. I show that the collapse to opacity under monotonicity then holds for every prior over valuations and every compactly supported elliptical distribution of the agent’s type with positive distortion ability. My main departure is to show that a single unrestricted message, silence, overturns this collapse: the optimal monotone-or-mute signal strictly improves upon opacity.

This paper also relates to the literature on voluntary disclosure of verifiable information. When the sender’s information is verifiable, costless to disclose, and known to exist, unraveling leads to full disclosure (Grossman, 1981; Milgrom, 1981). Nondisclosure arises in equilibrium when the receiver is unsure whether the sender is informed (Dye, 1985) or when disclosure is costly (Verrecchia, 1983). Silence then pools the senders with the least favorable information, and disclosure follows a threshold rule. In my setting, silence plays a different role. The principal commits to her disclosure policy, and silence pools the lowest and highest valuations, leaving the agent uncertain whether a variable matters little or matters a lot. Commitment is essential: since every type of principal prefers that the agent believes her valuation is low, ex post disclosure of verifiable information would unravel into full transparency, which is strictly worse than opacity.

This paper contributes to the literature on Bayesian persuasion initiated by Rayo and Segal (2010) and Kamenica and Gentzkow (2011). In many applications, the receiver’s behavior depends on the posterior only through its mean, so that the sender effectively chooses a distribution of posterior means (Gentzkow and Kamenica, 2016; Kolotilin, 2018; Dworzak and Martini, 2019). This approach has been extended to a privately informed receiver (Kolotilin et al., 2017; Guo and Shmaya, 2019), and Arieli et al. (2023) characterize optimal signals through bi-pooling policies when the sender’s payoff depends only on the posterior mean. In my model, the principal’s payoff depends nonlinearly on both the posterior mean and the valuation itself. My characterization of the unconstrained optimum relies on Kolotilin et al. (2025), who use optimal transport to solve nonlinear persuasion problems. This paper also relates to the literature on constrained information design, which studies monotone signals that pool states into intervals (Mensch, 2021; Onuchic and Ray, 2023; Kolotilin et al., 2024). I introduce monotone-or-mute signals, which add a single unrestricted message to a monotone signal, and show that this minimal departure is enough to make disclosure valuable.

2 Model

2.1 Setup

A principal (she) assigns a score to an agent (he). She wants the score to match the agent’s quality, weighted by her private valuation of that quality. Before scoring, she commits to a costless signal that discloses information about her valuation to the agent, who then sends a costly report about his quality. By choosing what the agent learns about her valuation, the principal can influence the agent’s incentives to distort his report.

Principal. The setup extends the model of Frankel and Kartik (2019). The principal with

valuation θ obtains payoff

$$u_P = -(s - \theta\eta)^2$$

from assigning score $s \in \mathbb{R}$ to an agent with quality η .⁶ The valuation $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$, with $0 < \underline{\theta} < \bar{\theta}$, is privately known to the principal and drawn from an atomless prior $\phi_0 \in \Delta(\Theta)$ with full support. Before θ is realized, the principal commits to a signal about θ . As is standard since Kamenica and Gentzkow (2011), choosing a signal is equivalent to choosing a distribution of posteriors $\tau \in \Delta(\Delta(\Theta))$ that is Bayes-plausible, i.e., such that $\int_{\Delta(\Theta)} \phi d\tau(\phi) = \phi_0$. For a posterior ϕ , I write $\mu_\phi = E_\phi[\theta]$ for the posterior mean, and $\mu_0 = E_{\phi_0}[\theta]$ for the prior mean.

Agent. The agent privately knows his type (η, δ) , where η is his *natural* quality and $\delta > 0$ is his *distortion* ability. He sends a report $x \in \mathbb{R}$ at cost $(x - \eta)^2/(2\delta)$, so that his payoff is

$$u_A = s - \frac{(x - \eta)^2}{2\delta}$$

Reporting his natural quality is costless, and the marginal cost of inflating the report decreases with his distortion ability. The type (η, δ) is drawn from a distribution F , independently of θ . The distribution F is elliptical, with means μ_η and $\mu_\delta > 0$, variances $\sigma_\eta^2, \sigma_\delta^2 \in (0, \infty)$, and zero covariance. Its support is compact and contained in $\mathbb{R} \times (0, \infty)$, ensuring that $\delta > 0$ for every type.⁷

Timing. The game unfolds as follows.

1. The principal commits to a signal τ .

⁶This setup is closely related to the one-variable model of Ball (2025).

⁷Elliptical distributions guarantee that, for every $b \in \mathbb{R}$, the conditional expectation of η given $\eta + b\delta$ is affine, as shown in Gómez et al. (2003).

2. The valuation θ and the type (η, δ) are drawn independently.
3. The agent observes the signal realization and his type, and sends a report x .
4. The principal observes θ , the signal realization, and the report x , and assigns a score s .

The principal commits to her disclosure policy, but not to her scoring rule. Since her payoff is quadratic and (η, δ) is independent of θ and of the signal, she optimally assigns $s = \theta E_F[\eta|x]$, where $E_F[\eta|x]$ denotes the expectation of the agent's quality given his report, under his equilibrium strategy following the realized signal.

2.2 Linear equilibrium

As in Frankel and Kartik (2022) and Ball (2025), I focus on Bayes–Nash linear equilibria, in which the scoring rule is affine in the agent's report. This restriction makes the equilibrium tractable and gives the slope a natural interpretation: it measures how sensitively the principal adjusts the score in response to the report.

Fix a signal realization, and let ϕ denote the posterior it induces. Since behavior after the realization depends on it only through ϕ , I index the equilibrium objects by ϕ .

Definition 2.1. Given posterior ϕ , a *linear equilibrium* consists of a scoring rule $s_\phi(x, \theta) = \beta_\phi(\theta)x + \beta_{0,\phi}(\theta)$ and a reporting strategy $x_\phi(\eta, \delta)$ such that

- (i) for every type (η, δ) , the report $x_\phi(\eta, \delta)$ maximizes $E_\phi[s_\phi(x, \theta)] - (x - \eta)^2/(2\delta)$ over $x \in \mathbb{R}$, and
- (ii) for every valuation $\theta \in \Theta$ and every type (η, δ) , $s_\phi(x_\phi(\eta, \delta), \theta) = \theta E_F[\eta | x_\phi(\eta, \delta)]$.

Given a linear scoring rule, the agent's objective is strictly concave in x , and his unique best response is

$$x_\phi(\eta, \delta) = \eta + b_\phi \delta, \quad \text{where } b_\phi = E_\phi[\beta_\phi(\theta)] \tag{1}$$

The agent inflates his report in proportion to his distortion ability δ and to his expected marginal return to inflating the report b_ϕ . Given this strategy, (η, x_ϕ) is elliptical by Gómez et al. (2003), and the conditional expectation of quality is affine in the report:

$$E_F[\eta|x] = \mu_\eta + k_\phi(x - \mu_\eta - b_\phi\mu_\delta)$$

where

$$k_\phi = \frac{\text{Cov}(\eta, x)}{\text{Var}(x)} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + b_\phi^2 \sigma_\delta^2}$$

Substituting this expression into condition (ii) yields, for every $\theta \in \Theta$ and every equilibrium report x ,

$$\beta_\phi(\theta) x + \beta_{0,\phi}(\theta) = \theta k_\phi x + \theta(\mu_\eta - k_\phi(\mu_\eta + b_\phi\mu_\delta))$$

Both sides are affine in x and must agree at every report sent in equilibrium, meaning that $\beta_\phi(\theta) = \theta k_\phi$. The principal's sensitivity to the report is therefore proportional to her valuation. The factor k_ϕ does not depend on θ because the agent does not observe the valuation. His report, and hence what it reveals about his quality, is the same whatever the true θ . The principal therefore forms the same estimate $E_F[\eta|x]$ of quality at every valuation, and her valuation only multiplies this estimate in the score $s = \theta E_F[\eta|x]$. Since $\text{Cov}(\eta, x) = \sigma_\eta^2$, the coefficient $k_\phi = \text{Cov}(\eta, x)^2 / (\text{Var}(\eta)\text{Var}(x))$ is also the population R^2 of the regression of quality on the report, that is, the share of the variance of quality explained by the report.

Consistency between the two players' strategies requires $b_\phi = E_\phi[\theta k_\phi] = \mu_\phi k_\phi$, that is,

$$b_\phi = \mu_\phi \frac{\sigma_\eta^2}{\sigma_\eta^2 + b_\phi^2 \sigma_\delta^2}$$

or, equivalently,

$$\sigma_\delta^2 b_\phi^3 + \sigma_\eta^2 b_\phi - \sigma_\eta^2 \mu_\phi = 0 \quad (2)$$

The intensity of distortion depends on the posterior only through its mean. For $y > 0$, let $h(y)$ denote the unique real root b of $\sigma_\delta^2 b^3 + \sigma_\eta^2 b - \sigma_\eta^2 y = 0$, that is, of (2) with $\mu_\phi = y$.

Proposition 1. *For every posterior ϕ , there exists a unique linear equilibrium. In this equilibrium, $b_\phi = h(\mu_\phi)$ and $k_\phi = h(\mu_\phi)/\mu_\phi \in (0, 1)$. The intensity of distortion b_ϕ is strictly increasing in μ_ϕ , and the principal's sensitivity to the report, θk_ϕ , is strictly decreasing in μ_ϕ for every $\theta \in \Theta$.*

Consistent with Frankel and Kartik (2019) and Ball (2025), the agent distorts more when he believes the valuation is high, since his expected marginal return from inflating the report is then higher. The report becomes a noisier signal of quality, and the principal, anticipating this, discounts it more heavily.

2.3 The principal's problem

The equilibrium characterization reduces the principal's payoff to a function of the first two moments of the posterior.

Lemma 1. *In the linear equilibrium following posterior ϕ , the principal's expected payoff is*

$$W(\phi) = -E_\phi[\theta^2] G(\mu_\phi)$$

where

$$G(y) = \frac{\sigma_\eta^2 \sigma_\delta^2 h(y)^2}{\sigma_\eta^2 + \sigma_\delta^2 h(y)^2} = \sigma_\eta^2 \left(1 - \frac{h(y)}{y} \right) \quad (3)$$

The function $G : (0, \infty) \rightarrow (0, \sigma_\eta^2)$ is continuously differentiable with $G' > 0$, and satisfies $\lim_{y \rightarrow 0} G(y) = 0$ and $\lim_{y \rightarrow \infty} G(y) = \sigma_\eta^2$.

The payoff is the product of two terms. The first, $E_\phi[\theta^2]$, captures the *stakes*: a principal with a higher valuation suffers more from a given scoring error. The second, $G(\mu_\phi) = E_F[\text{Var}(\eta|x)]$, is the *residual uncertainty* about the agent's quality after observing his report, which equals the prior variance σ_η^2 times the share $1 - k_\phi$ left unexplained by the report. A higher posterior mean induces greater distortion, making the report less informative and increasing residual uncertainty. The principal's payoff, therefore, decreases in both moments of the posterior. The central tension is the following: disclosing that the valuation is high raises the posterior mean, and hence distortion, precisely when the stakes are highest.

The principal chooses a Bayes-plausible distribution of posteriors to maximize her expected payoff,

$$\max_{\tau \in \Delta(\Delta(\Theta))} \int_{\Delta(\Theta)} W(\phi) d\tau(\phi) \quad \text{s.t.} \quad \int_{\Delta(\Theta)} \phi d\tau(\phi) = \phi_0 \quad (\text{P})$$

Problem (P) is an infinite-dimensional persuasion problem in which the principal's payoff depends on both her valuation and the agent's posterior mean. Section 3 introduces monotone signals and shows that none of them improves upon opacity. Section 4 shows that this changes when the principal can also remain silent: she can then strictly improve upon opacity.

3 Monotone disclosure and the case for opacity

In many settings, organizations communicate the importance of a variable through an ordered scale such as “low,” “moderate,” and “high”. Such communication is relevant when an organization evaluates an agent's report about a characteristic whose importance depends on the organization's own objectives. I capture the restriction to ordered categories by requiring

that the messages correspond to non-overlapping intervals of the valuation. This restriction rules out pooling distant valuations without the valuations in between, and therefore allows me to study whether conventional ordered disclosure can improve upon opacity.

Definition 3.1. Two posteriors ϕ and ϕ' are *ordered* if $\max \text{supp } \phi \leq \min \text{supp } \phi'$ or $\max \text{supp } \phi' \leq \min \text{supp } \phi$. A signal τ is *monotone* if any two distinct posteriors in $\text{supp } \tau$ are ordered.

Monotone messages are statements of the form "the weight of this variable lies between a and b ". Regulation asks for such statements when it requires providers to disclose the relative importance of their main parameters.

To state the next result, identify each message with the posterior it induces. *Pooling* a set $\Phi \subseteq \text{supp } \tau$ of messages means that, whenever the principal would have sent one of them, she sends a single common message instead, and leaves all other messages unchanged. The pooled message then induces the average posterior $\psi = \frac{1}{\tau(\Phi)} \int_{\Phi} \phi d\tau(\phi)$.

Proposition 2. *Let τ be a signal, and let τ' be the signal obtained from τ by pooling a set $\Phi \subseteq \text{supp } \tau$ of pairwise ordered posteriors with $\tau(\Phi) > 0$. If $\tau' \neq \tau$, then τ' yields a strictly higher payoff than τ .*

To understand the logic, consider two ordered messages. Pooling them raises the posterior mean, and hence distortion, when the valuation lies in the lower message, and lowers it when the valuation lies in the upper one. Two forces make the gain outweigh the loss. First, I show that the function $A(\mu_\phi) = \mu_\phi G(\mu_\phi)$ is convex, which shows that pooling would be profitable if the stakes were proportional to the valuation. Second, the loss of each message is $A(\mu_\phi)$ scaled by $E_\phi[\theta^2]/E_\phi[\theta]$, a weighted average of the valuations in the message. This factor is larger for the upper message, which is precisely the message in which pooling reduces distortion. The second force, therefore, reinforces the first. The proof extends the reasoning to an arbitrary collection of messages.

Pooling all the messages of a monotone signal yields opacity, so Proposition 2 leads to the following result.

Theorem 1. *Opacity strictly dominates every monotone signal other than opacity itself. In particular, opacity strictly dominates full transparency.*

Theorem 1 formalizes the conventional case for opacity. Under monotone disclosure, information and stakes are aligned in the worst possible way: the agent learns that the valuation is high precisely when scoring errors are most costly, and distorts accordingly. Full transparency is the extreme case of this alignment, but any monotone signal, however coarse, suffers from it. This result parallels Sun (2025), in whose model the optimal signal collapses to opacity under monotonicity when the agent’s type is skewed toward low types. Here, the collapse holds for every prior and every compactly supported elliptical distribution of the agent’s type with positive distortion ability.

4 The value of silence

Theorem 1 relies on a single feature of monotone signals: their messages are ordered, so that a message inducing a low posterior mean contains only low valuations, and hence low stakes. To improve upon opacity, the principal must be able to pool high valuations with low ones within the same message, so that the agent’s uncertainty dampens distortion precisely when the valuation is high. This section shows that the possibility of remaining silent achieves exactly this objective and makes disclosure valuable for the principal. With this option, the optimal monotone-or-mute signal indicates that the valuation lies in an intermediate interval $[a, b]$ and remains silent otherwise.

4.1 Monotone-or-mute signals

I first introduce a new class of signals that allows the principal to remain silent.

Definition 4.1. A signal τ is *monotone-or-mute* if there exists a posterior $\phi_S \in \text{supp } \tau$, called the *silent message*, such that any two distinct posteriors in $\text{supp } \tau \setminus \{\phi_S\}$ are ordered.

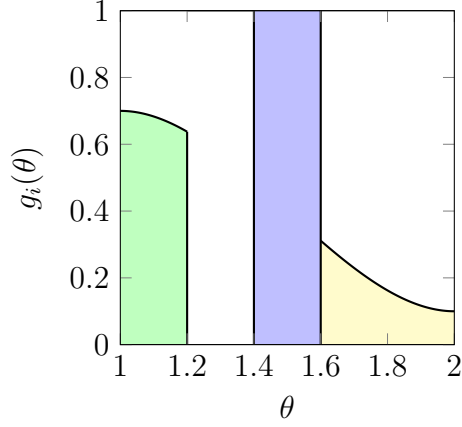
Like a monotone signal, such a signal communicates through a scale of categories that do not overlap, but it adds a single message whose support is unrestricted: silence. In models of verifiable disclosure, the sender can always withhold information (Grossman, 1981; Dye, 1985). Silence differs from the other messages in a fundamental way: it has no literal content, so its meaning is entirely determined by the disclosure policy, and the restriction to intervals does not apply to it. This is precisely what makes silence valuable. It is the natural way to pool distant valuations without making a disjunctive statement: silence may mean that the variable barely matters, or that it matters a lot. Moreover, there is only one way to say nothing, hence a single silent message. Monotone-or-mute signals are a minimal and tractable departure from monotone communication under which the disclosure problem is not trivial.

Figure 1 depicts two examples of signals. Each color is a message, $g_i(\theta)$ is the probability that message i is sent at valuation θ , and the white area is silence. In panel (a), the colored messages are pairwise ordered, so that the signal is monotone-or-mute. In panel (b), the green message is sent both below and above the blue one, and the white one is as well. Whichever message plays the role of silence, two of the remaining messages are not ordered, so that the signal is not monotone-or-mute.

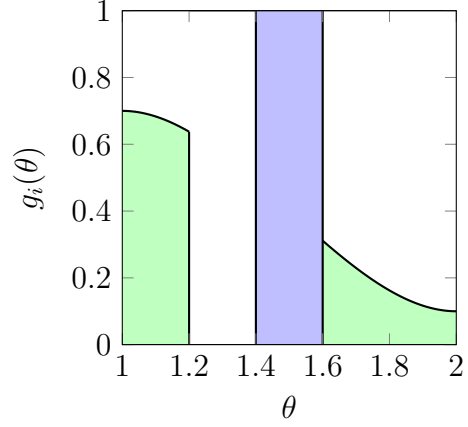
The messages of a monotone-or-mute signal other than silence are pairwise ordered. Proposition 2 therefore applies to them directly.

Corollary 1. *Every monotone-or-mute signal is weakly dominated by the signal that pools all its messages other than the silent message into a single message. This signal has at most two messages.*

A finer scale of categories, therefore, never helps: the principal gains nothing from distinguishing several categories, and only needs to decide at which valuations to speak and



(a) Monotone-or-mute signal.



(b) Not a monotone-or-mute signal.

Figure 1: Two signals for valuations on $[1, 2]$.

at which to remain silent. Figure 2 shows the signal obtained by pooling the non-silent messages of Figure 1(a), which yields a strictly higher payoff to the principal.

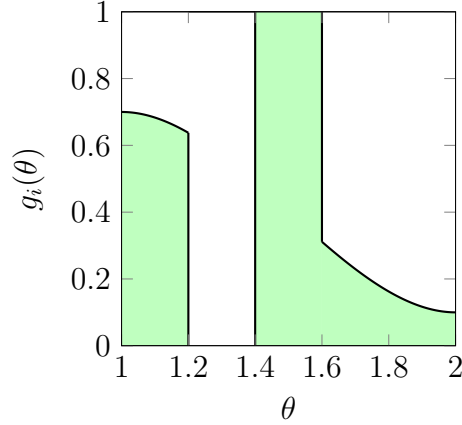


Figure 2: Signal obtained by pooling the non-silent messages of Figure 1(a).

4.2 Optimal interval disclosure

The next result characterizes the optimal monotone-or-mute signal.

Theorem 2. *Every optimal monotone-or-mute signal announces that $\theta \in [a^*, b^*]$, for some $\underline{\theta} < a^* < b^* < \bar{\theta}$, and remains silent otherwise. It strictly improves upon opacity, and the silent message induces a lower expected valuation than the announcement.*

Intuition. Corollary 1 reduces the problem to signals with two messages. Consider any such signal, and call L the message with the lower posterior mean and H the other one. Since the two messages together carry all the stakes, the principal's loss can be written as

$$E_{\phi_0}[\theta^2] G(\mu_H) - \Pr(L) E_L[\theta^2] (G(\mu_H) - G(\mu_L))$$

Holding fixed the probabilities and the posterior means of the two messages, the principal therefore wants to load as much of the stakes θ^2 as possible onto message L , in which the agent distorts least. For a given mean, the second moment is largest when the valuations are as spread out as possible, so that message L should consist of the two tails of the distribution, and message H of the intermediate valuations. Message L therefore contains the highest valuations despite having the lower mean, exactly the reverse of what monotone signals do.

The principal can always improve upon opacity in this way. Split the valuations into two tails and an interval such that both messages have the prior mean. This signal does not change the agent's behavior and is payoff-equivalent to opacity, but it loads more stakes onto the tails. Now tilt the tails slightly toward the lower valuations. The mean after the message sent on the tails falls, and the mean after the other message rises, by amounts that cancel on average. The agent distorts less after the first message and more after the second. Since the first carries greater stakes, the principal gains.

The optimal signal is deterministic and takes a simple form: the principal reveals that the variable matters moderately and says nothing when it matters little or much. Silence is therefore ambiguous about which of the two extremes holds, and the agent who hears nothing expects a lower valuation than the agent who hears that the valuation is intermediate. Figure 3 depicts the optimal signal for a uniform prior on $[1, 2]$ and $\sigma_\eta^2 = \sigma_\delta^2 = 1$. The principal announces that $\theta \in [1.19, 1.83]$ (message A), which happens with probability

0.64, and remains silent otherwise (message S). The expected valuation is 1.511 after the announcement and 1.479 after silence.⁸

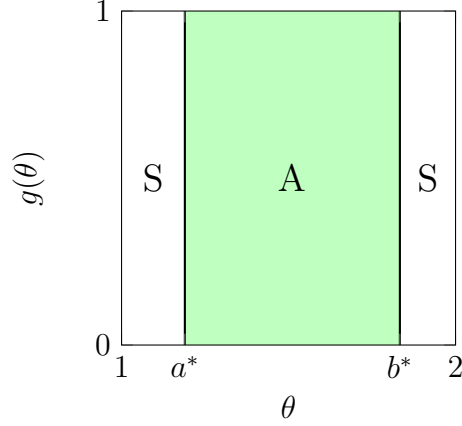


Figure 3: Optimal monotone-or-mute signal for a uniform prior on $[1, 2]$ and $\sigma_\eta^2 = \sigma_\delta^2 = 1$.

Silence, therefore, allows the principal to improve upon opacity even when her other messages must be ordered.

5 The unconstrained benchmark

Monotone-or-mute signals are simple to describe and to implement, but they restrict the principal. This section describes the optimal signal without any restriction, and Section 6 measures how much of its value the optimal monotone-or-mute signal captures.

Problem (P) belongs to the class of persuasion problems studied by KCW, and its solution follows from their results. To see the connection, write the principal's payoff after posterior ϕ as

$$W(\phi) = \int_{\Theta} V(\mu_\phi, \theta) d\phi(\theta), \quad \text{with} \quad V(y, \theta) = -\theta^2 G(y)$$

The posterior mean μ_ϕ plays the role of the receiver's action in KCW, and $V(y, \theta)$ is the sender's payoff when the action is y and the state is θ . In their terminology, this is a

⁸In this example, the gain over opacity is small, about 0.01% of the loss. Section 6 shows that it can exceed 13% when valuations are dispersed and skewed toward low values.

problem with a *linear receiver*, whose action is the posterior mean, and a sender whose payoff is nonlinear in the action and depends on the state.

Definition 5.1 (KCW). A signal τ is

1. *pairwise* if every posterior $\phi \in \text{supp } \tau$ has at most two valuations in its support,
2. *negative assortative* if it is pairwise and, for any two posteriors in $\text{supp } \tau$, the intervals spanned by their supports are nested, one inside the other, and
3. *single-peaked* if, for any posteriors $\phi, \phi' \in \text{supp } \tau$ and valuations $\theta_1 < \theta_2 < \theta_3$ with $\theta_1, \theta_3 \in \text{supp } \phi$ and $\theta_2 \in \text{supp } \phi'$, we have $\mu_\phi < \mu_{\phi'}$.

In a negative assortative signal, each message reveals that the valuation takes one of two values, a low one and a high one, and these pairs are nested: the lowest valuation is pooled with the highest, and the pairs shrink toward the middle. A single-peaked signal assigns lower posterior means to wider pairs.

Proposition 3 (KCW). *An optimal signal exists. Every optimal signal is pairwise, negative assortative, and single-peaked. If ϕ_0 has a density, the optimal signal is unique.*

Intuition. Each of the three properties has a simple interpretation in this model.

Pairwise. Any posterior can be split into posteriors with at most two valuations in their supports and the same mean. Such a split leaves the agent's behavior and the principal's payoff unchanged, since $W(\phi)$ is linear in ϕ for a given mean. KCW show that, under a nondegeneracy condition that holds here, posteriors with three or more valuations are never optimal.

Single-peaked. The cross-partial derivative $V_{y\theta}(y, \theta) = -2\theta G'(y)$ is negative and decreasing in θ : lowering the posterior mean reduces the agent's distortion, and the benefit is larger the higher the valuation. The highest valuations can only face a low posterior mean if they are pooled with low ones. The outer valuations are therefore pooled together and induce the

lowest posterior means, while the inner valuations induce higher ones. This is the logic of Theorem 2, which loads the stakes onto the message with the lower mean.

Negative assortative. Two pairs that do not overlap, such as $\{\theta_1, \theta_2\}$ and $\{\theta_3, \theta_4\}$ with $\theta_1 < \theta_2 < \theta_3 < \theta_4$, are ordered posteriors, so that pooling them strictly increases the principal's payoff by Proposition 2. Two pairs that cross, with $\theta_1 < \theta_3 < \theta_2 < \theta_4$, violate single-peakedness, since each pair contains a valuation that lies inside the other. The pairs must therefore be nested.

Under the optimal signal, every valuation except one, θ_m , is pooled with a partner on the other side of θ_m , and the partner of θ decreases with θ : the lowest valuation $\underline{\theta}$ is pooled with the highest valuation $\bar{\theta}$, and the pairs shrink towards θ_m . The wider the pair, the lower its posterior mean. The extreme valuations, low and high, therefore share the lowest posterior mean and face the least distortion.

The optimal monotone-or-mute signal is a two-message version of this structure. Silence pools the two tails, as the widest pairs do, and induces the lower posterior mean, while the announcement pools the intermediate valuations and induces the higher one. Figure 4 plots, for the uniform prior on $[1, 2]$, the posterior mean induced at each valuation under both signals. Under the unconstrained optimum, the posterior mean is single-peaked in the valuation, and the optimal monotone-or-mute signal approximates this curve with two steps.

6 How much does the simple policy lose?

Theorem 2 shows that a monotone-or-mute signal strictly improves upon opacity, and Proposition 3 describes the unconstrained optimum. This section studies how large these gains are, what determines their size, and how much of the unconstrained gain the simple signal captures. Gains are measured as a percentage of the loss under opacity. Section 6.1 studies the gains with respect to the distribution of valuations, and Section 6.2 with respect to the

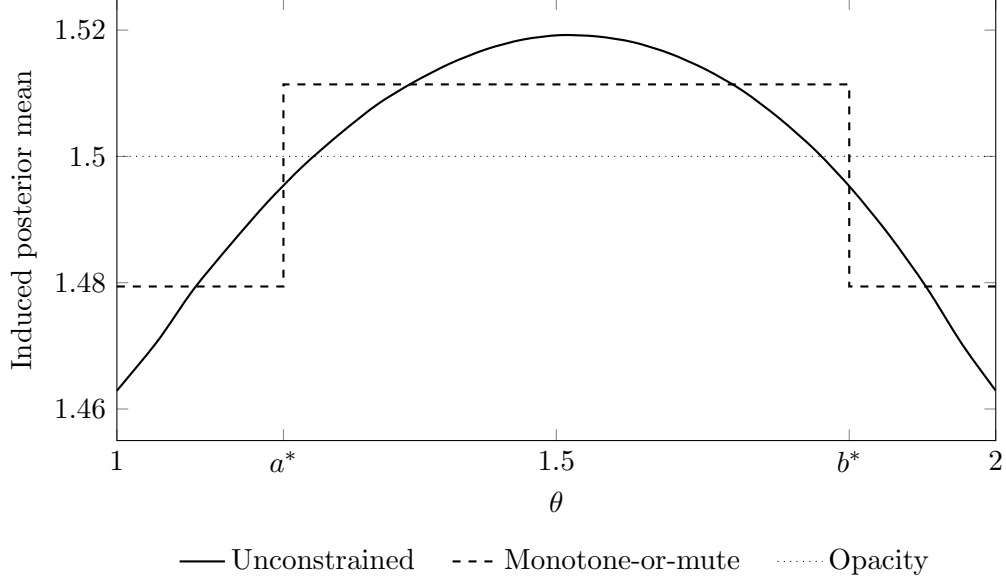


Figure 4: Posterior mean induced at each valuation under the unconstrained optimum, the optimal monotone-or-mute signal, and opacity, for θ uniform on $[1, 2]$ and $\sigma_\eta^2 = \sigma_\delta^2 = 1$.

distribution of the agent's type.

6.1 The distribution of valuations

For this subsection, I set $\bar{\theta} = 1$, and $\sigma_\eta^2 = \sigma_\delta^2 = 1$, and analyze the gain with different $\underline{\theta}$ and different distributions. Two features of the valuation distribution matter most for the gain from silence.

The first is *dispersion*. The stakes θ^2 differ across messages only if the valuations do. When $\bar{\theta}/\underline{\theta}$ increases, low valuations become smaller relative to the maximum valuation, increasing dispersion in the stakes θ^2 . The second is *skewness toward low valuations*. Shifting probability toward low valuations pulls the posterior mean down under the silent message without adding much to the loss. However, it also makes high valuations rarer. As long as the high valuations carry most of the stakes, additional low valuations further lower the posterior mean after silence, increasing the gain. Once high valuations become too rare, there is little left to protect from distortion. A moderate skewness, therefore, increases the gain, but too much skewness reduces it.

Figure 5 illustrates both effects. For three priors with increasing skewness, it plots the gain of the optimal monotone-or-mute signal against the dispersion $\bar{\theta}/\underline{\theta}$: the uniform prior, the log-uniform prior, and the inverse-uniform prior.⁹ In this range of skewness, the gain increases with skewness at every dispersion. The gain is negligible when valuations vary by less than a factor of three. It rises steeply with dispersion, and more so the more skewed the prior.

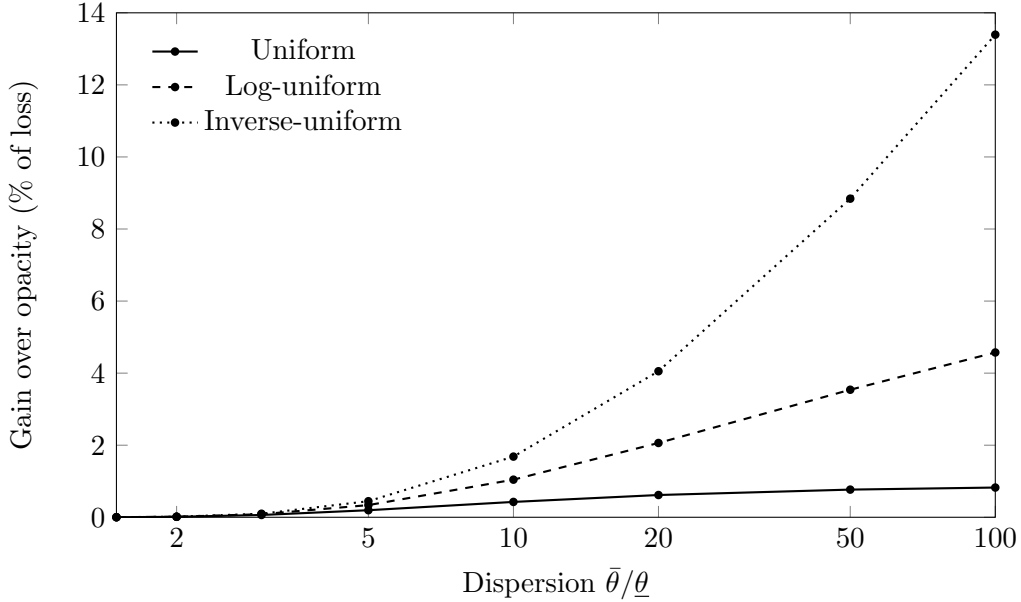


Figure 5: Gain of the optimal monotone-or-mute signal over opacity, as a percentage of the loss under opacity, for priors on $[\underline{\theta}, 1]$ and $\sigma_{\eta}^2 = \sigma_{\delta}^2 = 1$.

Table 1 reports, for several priors, the optimal monotone-or-mute signal and the change in the principal's loss relative to opacity under full transparency, under the optimal monotone-or-mute signal, and under the unconstrained optimum of Proposition 3.¹⁰ The last column is the share of the unconstrained gain captured by the optimal monotone-or-mute signal.

Three conclusions emerge. First, the value of silence can be substantial. For each family of priors, the gain increases with the dispersion of valuations. It is largest for the most

⁹On $[\underline{\theta}, \bar{\theta}]$, the log-uniform distribution, under which $\log \theta$ is uniform, has density $f(\theta) = 1/(\theta \log(\bar{\theta}/\underline{\theta}))$. The inverse-uniform distribution, under which $1/\theta$ is uniform, has density $f(\theta) = \underline{\theta} \bar{\theta} / ((\bar{\theta} - \underline{\theta}) \theta^2)$.

¹⁰The unconstrained optimum is computed by solving a discretized version of (P) as a linear program over pairwise posteriors, which is without loss by Proposition 3. The grid has 120 valuations and 150 posterior means per pair, and 300 posterior means per pair for the inverse-uniform prior on $[0.01, 1]$.

Prior	Support	$[a^*, b^*]$	Change in loss (%)			Share
			Full transp.	Monotone-or-mute	Unconstrained	
Uniform	[0.1, 1]	[0.28, 0.86]	+39	−0.43	−0.55	78%
Uniform	[0.01, 1]	[0.21, 0.85]	+55	−0.82	−1.06	78%
Log-uniform	[0.1, 1]	[0.20, 0.75]	+95	−1.04	−1.33	78%
Log-uniform	[0.01, 1]	[0.066, 0.61]	+413	−4.57	−5.77	79%
Inverse-uniform	[0.05, 1]	[0.078, 0.50]	+584	−4.05	−5.16	78%
Inverse-uniform	[0.01, 1]	[0.018, 0.30]	+6973	−13.4	−16.9	79%

Table 1: Change in the principal’s loss relative to opacity under full transparency, the optimal monotone-or-mute signal, and the unconstrained optimum, for $\sigma_\eta^2 = \sigma_\delta^2 = 1$. The interval $[a^*, b^*]$ is announced by the optimal monotone-or-mute signal, and Share is the share of the unconstrained gain that this signal captures.

dispersed and most skewed prior, the inverse-uniform prior on $[0.01, 1]$, for which the optimal monotone-or-mute signal reduces the principal’s loss by 13.4% relative to opacity. Second, the simple signal captures a stable share of the unconstrained gain, 78% or 79% in every example, although the gains themselves vary by a factor of thirty. Third, full transparency is far more costly than silence is valuable: it raises the loss by between 39% and almost 7000%. The largest figure reflects the very small loss under opacity when the prior mean is low, since the agent then barely distorts. The first-order recommendation is therefore to avoid monotone disclosure, as Theorem 1 shows.

6.2 The distribution of the agent’s type

I keep $\bar{\theta} = 1$ and $\sigma_\eta^2 = 1$ and vary σ_δ^2 . This covers every change in the variances. Dividing the cubic equation (2) by σ_η^2 shows that the agent’s distortion depends on the variances only through $\sigma_\delta^2/\sigma_\eta^2$. The loss of every signal is then proportional to σ_η^2 , which cancels in percentage gains. To separate the role of the agent’s type from that of the valuations, I fix the prior to the inverse-uniform distribution on $[0.01, 1]$, which yields the largest gain in Section 6.1. Table 2 reports the results for σ_δ^2 between 1 and 10 000. The second column gives the share of the variance of the report due to manipulation under opacity, $1 - h(\mu_0)/\mu_0$. This share is small when $\sigma_\delta^2 = 1$: since the prior mean is low, $\mu_0 = 0.047$, the agent has little

incentive to distort. The last column is the share of the unconstrained gain captured by the optimal monotone-or-mute signal.

σ_δ^2	Manip. share	$[a^*, b^*]$	Change in loss (%)			Share
			Full transp.	Monotone- or-mute	Uncon- strained	
1	0.2%	[0.018, 0.30]	+6973	−13.4	−16.9	79%
10	2%	[0.018, 0.31]	+1816	−13.3	−16.9	79%
100	14%	[0.019, 0.33]	+354	−12.4	−15.7	79%
1000	42%	[0.020, 0.38]	+90	−8.42	−10.7	79%
10 000	68%	[0.020, 0.42]	+31	−4.28	−5.49	78%

Table 2: Change in the principal’s loss relative to opacity under full transparency, the optimal monotone-or-mute signal, and the unconstrained optimum, for the inverse-uniform prior on $[0.01, 1]$ and $\sigma_\eta^2 = 1$. Manip. share is the share of the variance of the report due to manipulation under opacity, the interval $[a^*, b^*]$ is announced by the optimal monotone-or-mute signal, and Share is the share of the unconstrained gain that this signal captures.

The gain declines slowly as manipulation increases. When manipulation accounts for 14% of the variance of the report, the optimal monotone-or-mute signal reduces the loss by 12.4%, against 13.4% when $\sigma_\delta^2 = 1$. The reduction falls to 8.4% when manipulation accounts for 42% of the variance, and to 4.3% when it accounts for 68%. Even then, it exceeds the gain under the uniform priors of Table 1, where manipulation accounts for at most 17% of the variance.

Manipulation has little effect on the shape of the optimal signal. The optimal interval moves only from $[0.018, 0.30]$ to $[0.020, 0.42]$. The optimal monotone-or-mute signal captures 78% or 79% of the unconstrained gain in every row. Full transparency always raises the loss by at least 31%.

7 Concluding remarks

Organizations that rely on predictive models are often reluctant to reveal how these models work, for fear that individuals will game the inputs. This paper studies how much an organization should disclose about the weight of a variable that individuals can distort. The

answer depends on the form of disclosure. Disclosure on an ordered scale, which describes the weight as low, moderate, or high, always does worse than opacity: it tells individuals that the variable matters precisely when it matters most, and they distort accordingly. The option to remain silent overturns this conclusion. The optimal policy discloses that the weight is moderate and says nothing when it is low or high. Silence leaves individuals uncertain whether the variable matters little or a lot, and this uncertainty restrains their gaming precisely when the variable is important. In numerical examples, this simple policy captures close to 80% of the gain from unconstrained disclosure, and its gain over opacity is large when valuations are widely dispersed and skewed toward low values.

These results bear on Article 5 of the EU Platform-to-Business Regulation, which requires providers to disclose the main parameters determining ranking and their relative importance. Monotone signals are a stylized model of such ordered disclosure, not a literal description of the Regulation. They nonetheless capture a limitation of ordered disclosure: when agents can manipulate their reports, revealing that a variable matters a lot encourages manipulation precisely when scoring errors are most costly.

The Regulation also leaves room for silence. Providers need not disclose parameters that matter little, since only the main parameters must be disclosed, and they are exempt from disclosing information that would enable the manipulation of search results. In the model, remaining silent when a variable matters little or a lot, and announcing that it matters moderately otherwise, strictly improves upon opacity for well-chosen thresholds. Remaining silent only when a variable matters a lot does not, since the policy is then monotone. Exemptions from disclosure requirements may therefore help limit manipulation, but only if silence also covers the least important variables. Whether the Regulation induces such a policy in practice depends on its implementation and interpretation, which is beyond the scope of the model.

This analysis is a first step toward understanding optimal disclosure under strategic manipulation. Predictive models rarely rely on a single variable, and a natural extension

is to study disclosure about several correlated variables, possibly in nonlinear models. The paper also assumes that the organization commits to its disclosure policy but not to its scoring rule. An organization that can also commit to how it uses reports, as in Frankel and Kartik (2022) and Ball (2025), could combine both instruments, and how they interact remains an open question. Finally, distortion is purely wasteful here. When agents can also invest to improve their quality, as in Wang et al. (2023), disclosure may also encourage productive effort, and the optimal policy would trade off this benefit against the cost of gaming.

References

- ARIELI, I., Y. BABICHENKO, R. SMORODINSKY, AND T. YAMASHITA (2023): “Optimal Persuasion via Bi-pooling,” *Theoretical Economics*, 18, 15–36.
- BALL, I. (2025): “Scoring Strategic Agents,” *American Economic Journal: Microeconomics*, 17, 97–129.
- DWORCZAK, P. AND G. MARTINI (2019): “The Simple Economics of Optimal Persuasion,” *Journal of Political Economy*, 127, 1993–2048.
- DYE, R. A. (1985): “Disclosure of Nonproprietary Information,” *Journal of Accounting Research*, 23, 123–145.
- EDERER, F., R. HOLDEN, AND M. MEYER (2018): “Gaming and Strategic Opacity in Incentive Provision,” *The RAND Journal of Economics*, 49, 819–854.
- FRANKEL, A. AND N. KARTIK (2019): “Muddled Information,” *Journal of Political Economy*, 127, 1739–1776.
- (2022): “Improving Information from Manipulable Data,” *Journal of the European Economic Association*, 20, 79–115.

- GENTZKOW, M. AND E. KAMENICA (2016): “A Rothschild-Stiglitz Approach to Bayesian Persuasion,” *American Economic Review*, 106, 597–601.
- GÓMEZ, E., M. A. GÓMEZ-VILLEGAS, AND J. M. MARÍN (2003): “A Survey on Continuous Elliptical Vector Distributions,” *Revista matemática complutense*, 16, 345–361.
- GROSSMAN, S. J. (1981): “The Informational Role of Warranties and Private Disclosure about Product Quality,” *The Journal of Law and Economics*, 24, 461–483.
- GUO, Y. AND E. SHMAYA (2019): “The Interval Structure of Optimal Disclosure,” *Econometrica*, 87, 653–675.
- HARDT, M., N. MEGIDDO, C. PAPADIMITRIOU, AND M. WOOTTERS (2016): “Strategic Classification,” in *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, 111–122.
- KAMENICA, E. AND M. GENTZKOW (2011): “Bayesian Persuasion,” *American Economic Review*, 101, 2590–2615.
- KOLOTILIN, A. (2018): “Optimal Information Disclosure: A Linear Programming Approach,” *Theoretical Economics*, 13, 607–635.
- KOLOTILIN, A., R. CORRAO, AND A. WOLITZKY (2025): “Persuasion and Matching: Optimal Productive Transport,” *Journal of Political Economy*, 133, 1334–1381.
- KOLOTILIN, A., H. LI, AND A. ZAPECHELNYUK (2024): “On Monotone Persuasion,” *arXiv preprint arXiv:2412.14400*.
- KOLOTILIN, A., T. MYLOVANOV, A. ZAPECHELNYUK, AND M. LI (2017): “Persuasion of a Privately Informed Receiver,” *Econometrica*, 85, 1949–1964.
- LACKER, J. M. AND J. A. WEINBERG (1989): “Optimal Contracts under Costly State Falsification,” *Journal of Political Economy*, 97, 1345–1363.

- MENSCH, J. (2021): “Monotone Persuasion,” *Games and Economic Behavior*, 130, 521–542.
- MILGROM, P. R. (1981): “Good News and Bad News: Representation Theorems and Applications,” *The Bell Journal of Economics*, 380–391.
- ONUCHIC, P. AND D. RAY (2023): “Conveying Value via Categories,” *Theoretical Economics*, 18, 1407–1439.
- PEREZ-RICHET, E. AND V. SKRETA (2022): “Test Design under Falsification,” *Econometrica*, 90, 1109–1142.
- (2026): “Falsification-proof Non-market Allocation Mechanisms,” *arXiv preprint arXiv:2610.18625*.
- RAYO, L. AND I. SEGAL (2010): “Optimal Information Disclosure,” *Journal of political Economy*, 118, 949–987.
- SUN, J. (2025): “Model Transparency with Manipulable Input,” .
- TAN, T. Y. (2025): “Price Discrimination with Manipulable Observables,” *The RAND Journal of Economics*.
- VERRECCHIA, R. E. (1983): “Discretionary Disclosure,” *Journal of Accounting and Economics*, 5, 179–194.
- WANG, Q., Y. HUANG, S. JASIN, AND P. V. SINGH (2023): “Algorithmic Transparency with Strategic Users,” *Management Science*, 69, 2297–2317.

A Appendix

A.1 Proof of Proposition 1

Fix a posterior ϕ , and recall that $\mu_\phi \geq \underline{\theta} > 0$. The derivation in Section 2.2 shows that, in any linear equilibrium, the agent reports $x_\phi(\eta, \delta) = \eta + b_\phi \delta$, the scoring rule satisfies

$$\beta_\phi(\theta) = \theta k_\phi \quad \text{and} \quad \beta_{0,\phi}(\theta) = \theta(\mu_\eta - k_\phi(\mu_\eta + b_\phi \mu_\delta)) \quad \text{for every } \theta \in \Theta \quad (4)$$

and b_ϕ solves (2). Conversely, suppose that b_ϕ solves (2) and that the scoring rule is given by (4). Then $E_\phi[\beta_\phi(\theta)] = \mu_\phi k_\phi = b_\phi$, so the agent's unique best response is $x_\phi(\eta, \delta) = \eta + b_\phi \delta$, and condition (ii) of Definition 2.1 holds. These strategies, therefore, form a linear equilibrium. Existence and uniqueness of the linear equilibrium are thus equivalent to existence and uniqueness of a real root of (2).

Existence and uniqueness. The polynomial $Q(b) = \sigma_\delta^2 b^3 + \sigma_\eta^2 b - \sigma_\eta^2 \mu_\phi$ satisfies $Q'(b) = 3\sigma_\delta^2 b^2 + \sigma_\eta^2 > 0$, $Q(0) = -\sigma_\eta^2 \mu_\phi < 0$, and $Q(\mu_\phi) = \sigma_\delta^2 \mu_\phi^3 > 0$. It therefore has a unique real root $b_\phi = h(\mu_\phi)$, which lies in $(0, \mu_\phi)$. Hence, the linear equilibrium exists and is unique, and $k_\phi = b_\phi / \mu_\phi \in (0, 1)$.

Comparative statics. By the implicit function theorem, h is continuously differentiable, with

$$h'(y) = \frac{\sigma_\eta^2}{3\sigma_\delta^2 h(y)^2 + \sigma_\eta^2} > 0$$

Hence $b_\phi = h(\mu_\phi)$ is strictly increasing in μ_ϕ . Since h is positive and strictly increasing, $k_\phi = \sigma_\eta^2 / (\sigma_\eta^2 + h(\mu_\phi)^2 \sigma_\delta^2)$ is strictly decreasing in μ_ϕ , and so is θk_ϕ for every $\theta > 0$.

A.2 Proof of Lemma 1

Conditional on the signal realization, θ is distributed according to ϕ , (η, δ) is distributed according to F , and the two are independent. The report $x = \eta + b_\phi \delta$ depends only on the agent's type. With $s = \theta E_F[\eta|x]$, the principal's expected payoff is therefore

$$-E \left[\theta^2 \left(\eta - E_F[\eta|x] \right)^2 \right] = -E_\phi[\theta^2] E_F \left[\left(\eta - E_F[\eta|x] \right)^2 \right]$$

Using the expression for $E_F[\eta|x]$ from Section 2.2, with $V = \text{Var}(x) = \sigma_\eta^2 + b_\phi^2 \sigma_\delta^2$ and $k_\phi = \sigma_\eta^2/V$,

$$\begin{aligned} E_F \left[\left(\eta - E_F[\eta|x] \right)^2 \right] &= \sigma_\eta^2 - 2k_\phi \sigma_\eta^2 + k_\phi^2 V \\ &= \sigma_\eta^2 - \frac{\sigma_\eta^4}{V} \\ &= \frac{\sigma_\eta^2 \sigma_\delta^2 b_\phi^2}{\sigma_\eta^2 + \sigma_\delta^2 b_\phi^2} \\ &= \sigma_\eta^2 (1 - k_\phi) \end{aligned}$$

By Proposition 1, $b_\phi = h(\mu_\phi)$ and $k_\phi = h(\mu_\phi)/\mu_\phi$, so this expression equals $G(\mu_\phi)$, in both forms given in (3).

For the properties of G : by the proof of Proposition 1, h is continuously differentiable on $(0, \infty)$, with $h'(y) = \sigma_\eta^2/(3\sigma_\delta^2 h(y)^2 + \sigma_\eta^2) > 0$. Since $0 < h(y) < y$, $h(y) \rightarrow 0$ as $y \rightarrow 0$. Since $\sigma_\eta^2 y = \sigma_\delta^2 h(y)^3 + \sigma_\eta^2 h(y)$, h is unbounded, as the left-hand side tends to infinity with y . Since h is increasing, $h(y) \rightarrow \infty$ as $y \rightarrow \infty$. Next, $G = \gamma \circ h$ with

$$\gamma(c) = \frac{\sigma_\eta^2 \sigma_\delta^2 c^2}{\sigma_\eta^2 + \sigma_\delta^2 c^2} = \frac{\sigma_\eta^2}{1 + \sigma_\eta^2/(\sigma_\delta^2 c^2)}$$

The map γ takes values in $(0, \sigma_\eta^2)$, tends to 0 as $c \rightarrow 0$ and to σ_η^2 as $c \rightarrow \infty$, and is continuously differentiable with $\gamma'(c) = 2\sigma_\eta^4 \sigma_\delta^2 c/(\sigma_\eta^2 + \sigma_\delta^2 c^2)^2 > 0$. Hence G takes values in $(0, \sigma_\eta^2)$, tends to 0 as $y \rightarrow 0$ and to σ_η^2 as $y \rightarrow \infty$, and is continuously differentiable with

$$G'(y) = \gamma'(h(y)) h'(y) > 0.$$

A.3 Proof of Proposition 2

The proof relies on the following Lemma

Lemma 2. *The function $A(y) = yG(y) = \sigma_\eta^2(y - h(y))$ is strictly increasing and strictly convex on $(0, \infty)$.*

Proof. By Lemma 1, $G(y) = \sigma_\eta^2(1 - h(y)/y)$, so that

$$A(y) = \sigma_\eta^2(y - h(y))$$

Monotonicity. By Lemma 1, G is positive and strictly increasing on $(0, \infty)$, so A , the product of y and $G(y)$, is strictly increasing.

Convexity. The left-hand side of $\sigma_\delta^2 h^3 + \sigma_\eta^2 h - \sigma_\eta^2 y = 0$ is a polynomial in (h, y) whose derivative with respect to h , $3\sigma_\delta^2 h^2 + \sigma_\eta^2$, is strictly positive. By the implicit function theorem, h is infinitely differentiable. Differentiating $\sigma_\delta^2 h(y)^3 + \sigma_\eta^2 h(y) = \sigma_\eta^2 y$ once yields

$$(3\sigma_\delta^2 h(y)^2 + \sigma_\eta^2) h'(y) = \sigma_\eta^2$$

so that $h'(y) > 0$, and differentiating a second time yields

$$(3\sigma_\delta^2 h(y)^2 + \sigma_\eta^2) h''(y) + 6\sigma_\delta^2 h(y) h'(y)^2 = 0$$

that is,

$$h''(y) = -\frac{6\sigma_\delta^2 h(y) h'(y)^2}{3\sigma_\delta^2 h(y)^2 + \sigma_\eta^2} < 0$$

since $h(y) > 0$. Hence $A''(y) = -\sigma_\eta^2 h''(y) > 0$, and A is strictly convex. □

Let $\Phi \subseteq \text{supp } \tau$ denote the set of pooled posteriors, which are pairwise ordered, and τ' the signal that pools these messages into one message and leaves the remaining messages unchanged. Since $\tau' = \tau$ when $\tau(\Phi) = 0$, I assume that $p = \tau(\Phi) > 0$, so that the pooled message induces the posterior $\psi = \frac{1}{p} \int_{\Phi} \phi d\tau(\phi)$. Pooling leaves the signal unchanged if and only if a single posterior $\phi \in \Phi$ satisfies $\tau(\{\phi\}) = p$, so I assume that this is not the case.

For every posterior ϕ , write $\mu_{\phi} = E_{\phi}[\theta]$ and $m_{\phi} = E_{\phi}[\theta^2]$. By linearity,

$$p \mu_{\psi} = \int_{\Phi} \mu_{\phi} d\tau(\phi) \quad \text{and} \quad p m_{\psi} = \int_{\Phi} m_{\phi} d\tau(\phi)$$

Since the messages outside Φ are unaffected, and since $W(\phi) = -m_{\phi} G(\mu_{\phi})$, the difference in payoff between τ' and τ equals

$$\int_{\Phi} m_{\phi} G(\mu_{\phi}) d\tau(\phi) - p m_{\psi} G(\mu_{\psi}) = \int_{\Phi} m_{\phi} (G(\mu_{\phi}) - G(\mu_{\psi})) d\tau(\phi)$$

I therefore need to show that

$$\int_{\Phi} m_{\phi} (G(\mu_{\phi}) - G(\mu_{\psi})) d\tau(\phi) > 0 \tag{5}$$

Step 1: Ordered posteriors have ordered means. Let ϕ and ϕ' be two distinct posteriors with $\max \text{supp } \phi \leq \min \text{supp } \phi'$. Then $\mu_{\phi} \leq \max \text{supp } \phi \leq \min \text{supp } \phi' \leq \mu_{\phi'}$. Equality throughout would require ϕ and ϕ' to be concentrated on the same point, contradicting $\phi \neq \phi'$. Hence $\mu_{\phi} < \mu_{\phi'}$. Since the posteriors in Φ are pairwise ordered, two distinct posteriors in Φ therefore never have the same mean. Hence μ_{ϕ} is not τ -almost everywhere constant on Φ : if it were equal to some c , all the probability of Φ would be carried by the only posterior in Φ with mean c , a case ruled out above.

Step 2: First moments. By Lemma 2, A is strictly convex. Since μ_{ϕ} is not τ -almost everywhere constant on Φ , Jensen's inequality, applied with the weights $d\tau/p$ on Φ , holds

strictly:

$$\frac{1}{p} \int_{\Phi} A(\mu_{\phi}) d\tau(\phi) > A\left(\frac{1}{p} \int_{\Phi} \mu_{\phi} d\tau(\phi)\right) = A(\mu_{\psi})$$

Multiplying by p , and using $A(\mu_{\phi}) = \mu_{\phi} G(\mu_{\phi})$ and

$$p A(\mu_{\psi}) = p \mu_{\psi} G(\mu_{\psi}) = \int_{\Phi} \mu_{\phi} d\tau(\phi) G(\mu_{\psi})$$

this inequality reads

$$X = \int_{\Phi} \mu_{\phi} (G(\mu_{\phi}) - G(\mu_{\psi})) d\tau(\phi) > 0 \quad (6)$$

Inequality (6) coincides with (5), except that each posterior is weighted by its first rather than its second moment.

Step 3: From first to second moments. Let $\Phi_- = \{\phi \in \Phi : \mu_{\phi} < \mu_{\psi}\}$ and $\Phi_+ = \{\phi \in \Phi : \mu_{\phi} > \mu_{\psi}\}$. The set Φ_- is nonempty: otherwise, $\mu_{\phi} \geq \mu_{\psi}$ for τ -almost every $\phi \in \Phi$, and since $\int_{\Phi} (\mu_{\phi} - \mu_{\psi}) d\tau(\phi) = 0$, the mean μ_{ϕ} would equal μ_{ψ} for τ -almost every $\phi \in \Phi$, contradicting Step 1. Let $\phi \in \Phi_-$ and $\phi' \in \Phi_+$. The two posteriors are ordered, and by Step 1, ϕ' cannot lie below ϕ , since this would imply $\mu_{\phi'} < \mu_{\phi}$. Hence $\max \text{supp } \phi \leq \min \text{supp } \phi'$, and the threshold

$$t = \sup_{\phi \in \Phi_-} \max \text{supp } \phi \geq \underline{\theta} > 0$$

satisfies $\max \text{supp } \phi \leq t$ for every $\phi \in \Phi_-$ and $\min \text{supp } \phi' \geq t$ for every $\phi' \in \Phi_+$.

Let $r_{\phi} = m_{\phi}/\mu_{\phi}$, which is well defined since $\underline{\theta} > 0$. Since $\theta^2 \leq \theta \max \text{supp } \phi$ and $\theta^2 \geq \theta \min \text{supp } \phi$ on the support of ϕ , integrating with respect to ϕ yields $\min \text{supp } \phi \leq r_{\phi} \leq \max \text{supp } \phi$. Hence $r_{\phi} \leq t$ on Φ_- and $r_{\phi} \geq t$ on Φ_+ . Since G is strictly increasing,

$G(\mu_\phi) - G(\mu_\psi)$ is negative on Φ_- , positive on Φ_+ , and zero elsewhere. In all three cases,

$$m_\phi(G(\mu_\phi) - G(\mu_\psi)) = r_\phi \mu_\phi(G(\mu_\phi) - G(\mu_\psi)) \geq t \mu_\phi(G(\mu_\phi) - G(\mu_\psi))$$

Integrating over Φ with respect to τ and using (6) yields

$$\int_{\Phi} m_\phi(G(\mu_\phi) - G(\mu_\psi)) d\tau(\phi) \geq t X > 0$$

which is (5).

A.4 Proof of Theorem 1

Let τ be a monotone signal other than opacity, and pool all of its messages, that is, apply Proposition 2 with $\Phi = \text{supp } \tau$, which satisfies $\tau(\Phi) = 1$. By Bayes plausibility, the pooled message induces the posterior $\int \phi d\tau(\phi) = \phi_0$, so that τ' is opacity. Since τ is not opacity, $\tau' \neq \tau$, and opacity yields a strictly higher payoff than τ .

A.5 Proof of Corollary 1

Let τ be a monotone-or-mute signal with silence ϕ_S . The posteriors in $\Phi = \text{supp } \tau \setminus \{\phi_S\}$ are pairwise ordered. By Proposition 2, pooling the messages in Φ yields a signal τ' such that either $\tau' = \tau$ or τ' yields a strictly higher payoff than τ . By construction, τ' consists of the silent message and at most one other message.

A.6 Proof of Theorem 2

Throughout the proof, let $F(\theta) = \phi_0([\underline{\theta}, \theta])$ denote the prior distribution function. Since ϕ_0 is atomless and has full support, F is continuous and strictly increasing, and its inverse $q = F^{-1} : [0, 1] \rightarrow \Theta$ is continuous and strictly increasing. If U is uniform on $[0, 1]$, then

$q(U)$ has distribution ϕ_0 , so that for $0 \leq \alpha \leq \beta \leq 1$ and $n \in \{0, 1, 2\}$,

$$\int_{q(\alpha)}^{q(\beta)} \theta^n d\phi_0(\theta) = \int_{\alpha}^{\beta} q(u)^n du \quad (7)$$

Since q is continuous, this expression is continuously differentiable in α and β .

For a measure $\nu \leq \phi_0$, let $M_n(\nu) = \int_{\Theta} \theta^n d\nu(\theta)$ for $n \in \{0, 1, 2\}$, and define the loss $\mathcal{L}(\nu) = M_2(\nu) G(M_1(\nu)/M_0(\nu))$ if $\nu \neq 0$, and $\mathcal{L}(0) = 0$. A message sent with probability p and inducing the posterior ϕ contributes $-\mathcal{L}(p\phi)$ to the principal's payoff. For $\underline{\theta} \leq a \leq b \leq \bar{\theta}$, the *interval signal* $[a, b]$ is the two-message signal that sends one message if $\theta \in [a, b]$, and the other message otherwise.

Lemma 3. *Consider a signal τ with two messages whose posterior means are $\mu_L \leq \mu_H$. There exists an interval $[a, b] \subseteq \Theta$ such that the two-message signal τ' in which the message with mean μ_H is sent if and only if $\theta \in [a, b]$ yields a weakly higher payoff than τ . If $\mu_L < \mu_H$ and $\tau' \neq \tau$, the payoff is strictly higher.*

Proof. Let $g(\theta)$ denote the probability that τ sends the message with mean μ_L at valuation θ , and let $\nu_L(B) = \int_B g(\theta) d\phi_0$ and $\nu_H(B) = \int_B (1 - g(\theta)) d\phi_0$ for every Borel set B . Let $m = M_0(\nu_L)$. If $m \in \{0, 1\}$, τ sends a single message with probability one and is payoff-equivalent to opacity, and the claim holds with $[a, b] = \Theta$. Suppose that $m \in (0, 1)$.

Step 1: Tails with the same probability and mean. For $\alpha \in [0, m]$, let $T_{\alpha} = [\underline{\theta}, q(\alpha)) \cup (q(1 - m + \alpha), \bar{\theta}]$, a lower tail of probability α and an upper tail of probability $m - \alpha$, and let

$$M_n^T(\alpha) = \int_{T_{\alpha}} \theta^n d\phi_0(\theta) = \int_0^{\alpha} q(u)^n du + \int_{1-m+\alpha}^1 q(u)^n du$$

by (7). The upper tail $T_0 = (c, \bar{\theta}]$, with $c = q(1 - m)$, has the largest first moment among all messages of probability m . Indeed, $\phi_0(T_0) = m$. If $\theta > c$, then $\mathbb{1}_{T_0}(\theta) - g(\theta) = 1 - g(\theta) \geq 0$, and if $\theta \leq c$, then $\mathbb{1}_{T_0}(\theta) - g(\theta) = -g(\theta) \leq 0$. Hence $(\theta - c)(\mathbb{1}_{T_0}(\theta) - g(\theta)) \geq 0$ for every θ ,

and integrating against ϕ_0 gives

$$0 \leq \int_{\Theta} \theta \mathbb{1}_{T_0}(\theta) d\phi_0 - \int_{\Theta} \theta g(\theta) d\phi_0 - c \underbrace{\int_{\Theta} (\mathbb{1}_{T_0} - g) d\phi_0}_{=0} = M_1^T(0) - M_1(\nu_L)$$

Symmetrically, $M_1(\nu_L) \geq M_1^T(m)$. Since M_1^T is continuous, there exists $\hat{\alpha} \in [0, m]$ such that $M_1^T(\hat{\alpha}) = M_1(\nu_L)$. Let $a = q(\hat{\alpha})$ and $b = q(1 - m + \hat{\alpha})$, and let τ' send the first message if and only if $\theta \in T_{\hat{\alpha}} = \Theta \setminus [a, b]$, that is, with probability $g^*(\theta) = \mathbb{1}_{T_{\hat{\alpha}}}(\theta)$, and let $\nu_L^*(B) = \int_B g^*(\theta) d\phi_0$. Since ν_L^* and ν_L have the same probability and the same first moment, τ' has the same posterior means μ_L and μ_H as τ .

Step 2: The tails carry more stakes. The function $(\theta - a)(\theta - b)$ is positive on $T_{\hat{\alpha}}$, where $g^*(\theta) - g(\theta) = 1 - g(\theta) \geq 0$, and nonpositive on $[a, b]$, where $g^*(\theta) - g(\theta) = -g(\theta) \leq 0$. Since ν_L^* and ν_L have the same probability and first moment,

$$M_2(\nu_L^*) - M_2(\nu_L) = \int_{\Theta} (\theta - a)(\theta - b)(g^*(\theta) - g(\theta)) d\phi_0 \geq 0$$

with a strict inequality unless $g^*(\theta) = g(\theta)$ outside a set of probability zero.

Step 3: Comparison of losses. Since $M_2(\nu_H) = M_2(\phi_0) - M_2(\nu_L)$, the loss of τ is

$$\mathcal{L}(\nu_L) + \mathcal{L}(\nu_H) = M_2(\phi_0) G(\mu_H) - M_2(\nu_L)(G(\mu_H) - G(\mu_L))$$

and the loss of τ' is the same expression with $M_2(\nu_L^*)$ in place of $M_2(\nu_L)$. Since G is increasing, τ' does weakly better than τ , and strictly better if $\mu_L < \mu_H$ and $\tau' \neq \tau$. $\square$

Existence. For $0 \leq \alpha \leq \beta \leq 1$, let $\Lambda(\alpha, \beta)$ denote the loss of the interval signal $[q(\alpha), q(\beta)]$. By (7), the moments of its two messages are continuous in (α, β) . Since G is continuous and bounded by σ_{η}^2 by Lemma 1, $0 \leq \mathcal{L}(\nu) \leq \sigma_{\eta}^2 \bar{\theta}^2 M_0(\nu)$, so Λ is continuous on the compact set

$\{0 \leq \alpha \leq \beta \leq 1\}$ and attains its minimum at some (α^*, β^*) . Let $a^* = q(\alpha^*)$ and $b^* = q(\beta^*)$. By Corollary 1, every monotone-or-mute signal is weakly dominated by a signal with at most two messages, and by Lemma 3, every such signal is weakly dominated by an interval signal. Hence, the interval signal $[a^*, b^*]$ is optimal among monotone-or-mute signals and among signals with at most two messages.

Strict improvement upon opacity. Fix $m \in (0, 1)$, and let T_α and M_n^T be defined as in Step 1 of the proof of Lemma 3. The upper tail T_0 has a mean above μ_0 and the lower tail T_m has a mean below μ_0 . Since M_1^T is continuous, there exists $\alpha_0 \in (0, m)$ such that $M_1^T(\alpha_0) = m\mu_0$. Let $a_0 = q(\alpha_0)$ and $b_0 = q(1 - m + \alpha_0)$, so that $\underline{\theta} < a_0 < b_0 < \bar{\theta}$. Both messages of the interval signal $[a_0, b_0]$ have mean μ_0 , so that its loss equals the loss under opacity. Let E_T and E_I denote the expectations after the messages sent on T_{α_0} and on $[a_0, b_0]$. For $i \in \{T, I\}$, since both messages have mean μ_0 ,

$$E_i[(\theta - a_0)(\theta - b_0)] = E_i[\theta^2] - (a_0 + b_0) E_i[\theta] + a_0 b_0 = E_i[\theta^2] - (a_0 + b_0) \mu_0 + a_0 b_0.$$

The last two terms do not depend on i . Since $(\theta - a_0)(\theta - b_0)$ is positive on T_{α_0} and nonpositive on $[a_0, b_0]$, it follows that

$$E_T[\theta^2] - E_I[\theta^2] = E_T[(\theta - a_0)(\theta - b_0)] - E_I[(\theta - a_0)(\theta - b_0)] > 0$$

I now tilt the tails. For α close to α_0 , the loss of the interval signal $[q(\alpha), q(1 - m + \alpha)]$ is

$$\ell(\alpha) = M_2^T(\alpha) G\left(\frac{M_1^T(\alpha)}{m}\right) + (M_2(\phi_0) - M_2^T(\alpha)) G\left(\frac{M_1(\phi_0) - M_1^T(\alpha)}{1 - m}\right)$$

By (7), M_n^T is differentiable, with derivative $q(\alpha)^n - q(1 - m + \alpha)^n$, equal to $a_0^n - b_0^n$ at α_0 . At α_0 , both means equal μ_0 , so that the two terms involving the derivative of M_2^T cancel,

and

$$\begin{aligned}\ell'(\alpha_0) &= G'(\mu_0) (a_0 - b_0) \left(\frac{M_2^T(\alpha_0)}{m} - \frac{M_2(\phi_0) - M_2^T(\alpha_0)}{1 - m} \right) \\ &= -G'(\mu_0) (b_0 - a_0) (E_T[\theta^2] - E_I[\theta^2]) < 0\end{aligned}$$

Hence, for α slightly above α_0 , the interval signal $[q(\alpha), q(1 - m + \alpha)]$ has a strictly lower loss than opacity, and so does $[a^*, b^*]$.

Interior endpoints. If $a^* = \underline{\theta}$, $b^* = \bar{\theta}$ or $a^* = b^*$, the signal $[a^*, b^*]$ is monotone, and by Theorem 1, it does weakly worse than opacity. Hence $\underline{\theta} < a^* < b^* < \bar{\theta}$. For the same reason, the posterior means μ_I after the announcement and μ_S after silence differ, since otherwise both equal μ_0 and the signal does as well as opacity.

Silence has the lower mean. Suppose that $\mu_I < \mu_S$. The message with the higher mean, silence, is then sent on $\Theta \setminus [a^*, b^*]$. Since $\underline{\theta} < a^* < b^* < \bar{\theta}$ and ϕ_0 has full support, this set differs from every interval on an event of positive probability. The signal given by Lemma 3, in which the message with the higher mean is sent on an interval, therefore differs from $[a^*, b^*]$ and yields a strictly higher payoff, a contradiction. Hence $\mu_S < \mu_I$.

Every optimal signal is of this form. Let τ be an optimal monotone-or-mute signal. By Proposition 2, τ has at most two messages, since otherwise pooling its messages other than silence would strictly improve it. Since $[a^*, b^*]$ strictly improves upon opacity, τ is not payoff-equivalent to opacity, so that it has two messages with distinct means. By Lemma 3, τ coincides with the interval signal that the lemma associates with it, since otherwise this signal would strictly improve upon τ . Hence τ is an interval signal, and the two arguments above, which only use its optimality, show that its endpoints are interior and that silence has the lower mean.

A.7 Proof of Proposition 3

KCW normalize states and actions to $[0, 1]$. Let $c = \bar{\theta} - \underline{\theta}$ and, for $x \in [0, 1]$, let $\theta(x) = \underline{\theta} + cx$. In normalized units, the receiver's marginal utility is $u(y, x) = x - y$, so that the action at posterior ϕ is its normalized mean, and the sender's payoff is $\hat{V}(y, x) = -\theta(x)^2 G(\theta(y))$. KCW's Assumptions 1 to 4 hold. First, $\hat{V}$ and u are three times differentiable, since the derivative $3\sigma_\delta^2 b^2 + \sigma_\eta^2$ of the cubic equation (2) in b is positive, so that h , and hence G , is smooth by the implicit function theorem. Second, $u_y = -1 < 0$. Third, $\min_x u(0, x) = \max_x u(1, x) = 0$. Fourth, $u_x = 1 > 0$, and the requirement $\hat{V}_y > 0$ is not needed with a linear receiver (KCW, footnote 7). An optimal signal exists by KCW's Lemma 1.

Pairwise. By KCW's Theorem 2, every optimal signal is pairwise if the twist condition holds: for every action y and states $x_1 < x_2 < x_3$, the matrix whose rows are $(\hat{V}_y(y, x_i))_i$, $(u(y, x_i))_i$, and $(u_y(y, x_i))_i$ is nonsingular. Here $\hat{V}_y(y, x) = -c\theta(x)^2 G'(\theta(y))$, $u(y, x) = x - y$, and $u_y = -1$. Subtracting y times the third row from the second turns the second row into $(x_i)_i$. Factoring $-cG'(\theta(y))$ out of the first row, and using $\theta(x_i)^2 = \underline{\theta}^2 + 2\underline{\theta}cx_i + c^2x_i^2$, adding $\underline{\theta}^2$ times the third row and subtracting $2\underline{\theta}c$ times the second row turns the first row into $(c^2x_i^2)_i$. Factoring c^2 out of the first row and -1 out of the third, and swapping these two rows, yields the Vandermonde matrix with rows $(1)_i$, $(x_i)_i$, and $(x_i^2)_i$, whose determinant is $(x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$. Hence the determinant equals

$$-c^3 G'(\theta(y)) (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$$

which is nonzero since $c > 0$ and $G' > 0$.

Single-peaked. KCW's Theorem 4 gives a sufficient condition: if u_{yx}/u_x is constant and $\hat{V}_{yx}(y, x)/u_x(y, x)$ is strictly decreasing in x for every y , then their contact set, which contains the support of every optimal signal, is strictly single-peaked, and so is every optimal

signal.¹¹ Here $u_x = 1$ and $u_{yx} = 0$, and $\hat{V}_{yx}(y, x) = -2c^2 \theta(x) G'(\theta(y))$ is strictly decreasing in x since $G' > 0$.

Negative assortative. By KCW's Theorem 6, every optimal signal is negative assortative if, first, the contact set is strictly single-peaked, and second, for any two states $x_1 < x_2$, pooling them with some probability $r \in (0, 1)$ yields a strictly higher payoff than revealing them. Moreover, if the prior has a density, the optimal signal is unique. The first condition holds by the previous step. For the second, KCW's Corollary 3 shows that, when the condition of their Theorem 4 holds, as shown in the previous step, it is equivalent to a local condition which, with a linear receiver, reads $\hat{V}_{yy}(y, y) + 2\hat{V}_{yx}(y, y) \leq 0$ for every $y \in [0, 1]$. Here, with $t = \theta(y)$,

$$\hat{V}_{yy}(y, y) + 2\hat{V}_{yx}(y, y) = -c^2 t(tG''(t) + 4G'(t)) = -c^2 t(A''(t) + 2G'(t)) < 0$$

since $A''(t) = tG''(t) + 2G'(t) > 0$ by Lemma 2 and $G' > 0$ by Lemma 1.

¹¹KCW state Theorem 4 for the mirror-image property, single-dippedness, under the condition that the ratio is strictly increasing. Reversing the order of the states and of the actions, $x \mapsto 1 - x$ and $y \mapsto 1 - y$, keeps the receiver linear, turns a decreasing ratio into an increasing one, and turns single-dipped signals into single-peaked ones. The same remark applies to Theorem 6 and Corollary 3 below, since the reversal preserves nested pairs and leaves $\hat{V}_{yy}$ and $\hat{V}_{yx}$ unchanged.