# Bayesian Persuasion and Moral Hazard*

Raphael Boleslavsky† Kyungmin Kim‡

March 2021

**Abstract**

We consider a three-player Bayesian persuasion game in which the sender designs a signal about an unknown state of the world, the agent exerts a private effort that affects the underlying state, and the receiver takes an action after observing the signal and its realization. The sender must not only persuade the receiver to select a desirable action, but also incentivize the agent's effort. We develop a general method of characterizing an optimal signal in this environment. We apply our method to derive concrete results in several natural examples and discuss their economic implications.

JEL Classification Numbers: C72, D82, D83, D86, M31.

Keywords: Bayesian persuasion; moral hazard; information design

*We thank Ricardo Alonso, Eduardo Faingold, George Georgiadis, Marina Halac, Johannes Hörner, Ilwoo Hwang, Ina Taneva, and Alex Wolitzky for helpful comments, along with seminar audiences at the Miami Economic Theory Conference, Econometric Society European Meeting, Transparency in Procurement Conference, Southern Economics Association, the NBER Organizations Workshop, SAET, NSF/NBER/CEME Mathematical Economics Conference, Duke, Florida State, and Wisconsin (Finance). Kyungmin Kim is supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5A2A03043516).

†University of Miami, r.boleslavsky@miami.edu.

‡Emory University, kyungmin.kim@emory.edu

# 1 Introduction

We study optimal information design in the presence of moral hazard, introducing an additional player—the agent—into the Bayesian persuasion framework. As in the standard setting, the sender designs a signal structure that transmits information about an unknown state to a receiver, who selects an action. Unlike the standard setting, the distribution of the underlying state is determined by the agent's unobservable effort. Thus, the signal structure influences the receiver's beliefs through two distinct channels: it determines the *prior belief* by incentivizing the agent's effort, and it affects the *posterior belief* by generating information about the realized state. Therefore, in our model, the sender is concerned with both information and incentive provision. In this paper, we study the tradeoff between these two objectives and explore its implications for optimal information design.

This tradeoff is a central feature in a variety of contexts. For example, universities assign grades to market their students while also providing them with incentives to work. Retailers' advertising strategies are designed both to ensure product quality and to increase sales. Information and incentive provision are also inherent in the functioning of the judiciary, which shapes citizens' behavior and reveals information about their actions to society. Similarly, an organization's method of forecasting a project's return influences its own future funding decisions and the project manager's initial effort.

To understand the underlying issues more clearly, consider the following example, which is borrowed from Kamenica and Gentzkow (2011) (KG, hereafter), but cast into a different context. A school (the sender) wishes to place a student (the agent) in a job. The student's ability to perform the job is uncertain: he may be skilled (type $s$) or unskilled (type $u$). The school and student both obtain payoff 1 if the student is hired and payoff 0 otherwise. However, the firm prefers to hire only skilled students. Thus, the firm (receiver) offers the job if and only if it believes that the student is skilled with probability at least $1/2$. Initially, the school and firm believe that the student is skilled with probability $3/10$. Thus, the firm will not hire the student based solely on the prior. The school commits to a grading policy, which assigns a student either grade $g$ or $b$, where $(\pi_s, \pi_u)$ represent the probabilities that the student is issued grade $g$ given his type.

Because the prior belief is exogenous, the school is not concerned with providing incentives, only information. Applying ideas from KG, it is easy to show that the optimal grading policy "inflates" the grades of unskilled students. Because it brings bad news about the student's skill, grade $b$ never generates a job offer. In contrast, $g$ generates a job offer if and only if it conveys sufficient good news. Thus, the school's goal is to assign grade $g$ as often as it can, while maintaining sufficient informativeness to generate an offer. Clearly, it is optimal to always assign $g$ to a skilled student, $\pi_s = 1$. By increasing $\pi_s$, the school increases both the frequency of $g$ and the good news it conveys. In contrast, by increasing $\pi_u$, the school increases the frequency of grade

$g$ but reduces its informativeness. In this example, maintaining the informativeness of grade $g$ constrains $\pi_u \leq 3/7$. Therefore, the optimal grading policy is $\pi_s = 1$ and $\pi_u = 3/7$. Under the optimal grading policy, the student gets the job with probability $3/5$, even though he is skilled with probability $3/10$. The probability that the student is skilled is $1/2$ conditional on grade $g$ and $0$ conditional on grade $b$. Therefore, either the firm is indifferent between offering the job and not (if the grade is $g$), or it strictly prefers not to offer the job (if $b$). Regardless, its payoff is the same as if it acts on its prior belief and rejects the student. In other words, observing the grade does not improve the firm's payoff.

Now suppose that the probability that the student is skilled depends on his unobservable effort. In particular, after the grading policy is set by the school, the student privately chooses whether to shirk or work. In the former case, the student is unskilled with probability 1, while in the latter case, the student is skilled with probability $3/10$. The student's disutility of work is $c = 1/5$.

Because the student's private effort determines the distribution of his type, the school must be concerned with both incentive and information provision when it designs its grading policy. Indeed, if the grading policy fails to provide incentives for the student to work, then the firm infers that the student is unskilled and never offers the job, resulting in the worst outcome for both school and student. Furthermore, even if the student works, he will never get an offer if grade $g$ does not convey sufficient good news. Therefore, the school must design its grading policy to ensure both that the student prefers to work,

$$\underbrace{\frac{3}{10}\pi_s + \frac{7}{10}\pi_u - c}_{\text{Work}} \geq \underbrace{\pi_u}_{\text{Shirk}},$$

and that grade $g$ is sufficiently informative to generate an offer when the firm anticipates that the student works ($\pi_u \leq 3/7$).

Both incentive and information provision shape the optimal grading policy. On one hand, increases in $\pi_s$ are beneficial to both information and incentive provision, resulting in a more informative grade $g$ and a higher payoff difference between work and shirk. Therefore, $\pi_s = 1$ is also optimal with moral hazard. On the other hand, increases in $\pi_u$ are detrimental to both information and incentive provision, reducing the informativeness of $g$ and the payoff difference between work and shirk. Provided $g$ leads to an offer, inducing student effort requires $\pi_u \leq 1/3$, which is more restrictive than the condition on informativeness. Thus, the optimal grading policy is $\pi_s = 1$ and $\pi_u = 1/3$. In order to eliminate shirking, the school inflates grades *less* than in the preceding case ($1/3 < 3/7$). Under the optimal policy, the probability of a job placement is equal to $8/15$. This placement probability falls short of the optimal outcome in the absence of moral hazard ($3/5$), demonstrating the cost of moral hazard for the student and school. In contrast, moral hazard benefits the firm: its belief conditional on grade $g$ is $9/16 (> 1/2)$ and, therefore, it now has a strict preference to hire a student with grade $g$.

In what follows, we explore this interaction of incentive and information provision in the Bayesian persuasion framework. Our model has the same basic structure as the preceding example, but it is considerably more general, allowing for any finite number of underlying states, arbitrary preferences, and a continuous effort choice for the agent. Our analysis proceeds in two steps. We first develop a general characterization of the sender's optimal signal. We then apply it to derive additional results in two tractable environments.

Our characterization of the optimal signal extends the elegant concavification method in Aumann and Maschler (1995) and KG. In KG, this method works because the sender's problem can be reformulated as a constrained optimization problem in which both the objective function (the sender's expected utility) and the constraint (Bayes-Plausibility) can be written as expectations taken with respect to the distribution of posteriors. In our model, the sender faces an additional constraint that, as in standard moral hazard models, ensures that the agent has an incentive to choose the effort level intended by the sender. We show that this incentive compatibility constraint can also be expressed as an expectation with respect to the posterior belief distribution. Exploiting this feature, we describe how to concavify the sender's objective function and the incentive constraint simultaneously, characterizing the optimal signal geometrically and analytically.[1]

Two general results highlight the role of moral hazard, which distinguishes our analysis from the most relevant literature (Kamenica and Gentzkow, 2011; Alonso and Câmara, 2016). First, absent the need to provide incentives, if the sender's utility is concave in the receiver's posterior belief, then it is optimal for the sender to reveal no information. In our model, such a signal leads to zero effort by the agent and, therefore, is not optimal in most economically relevant environments.[2] Second, absent the need to provide incentives, if there are $N$ possible states, then an optimal signal utilizes at most $N$ signal realizations. In our model, the number increases by $1$; that is, an optimal signal may require $N + 1$ realizations. This difference arises from the incentive constraint, which necessitates an extra degree of freedom.

We provide additional concrete results in two tractable environments: one with two states (and many actions for the receiver) and another with two actions for the receiver (and many states). In both environments, we characterize the set of implementable effort levels under some natural economic assumptions. In the binary-state environment, a fully informative signal maximizes the agent's effort, but in the binary-action environment, this is not the case. Furthermore, we explicitly characterize the optimal signal. In the binary-state environment, we derive conditions under

[1] See Doval and Skreta (2018) and Le Treust and Tomala (2019) for related contributions.

[2] Providing no information is optimal, for example, if the agent has fully opposing preferences from those of the sender (i.e., the sender wishes to minimize the agent's utility).

which the optimal signal garbles information about only one state and discuss its economic implications. In the binary-action environment, we show that the optimal signal is a binary partition, and illustrate how it is affected by moral hazard.

One particularly interesting question is the effects of transparency which allows the receiver to observe the agent's effort. Intuitively, it is reasonable to expect that transparency "crowds out" incentive provision, reducing the informativeness of the equilibrium signal. At the same time, because the agent's effort is observed, the agent can directly affect the receiver's inference by increasing his effort, which gives an additional incentive for the agent to work. Thus, it is natural to expect that transparency introduces a tradeoff for the receiver: more effort, but less information. This tradeoff appears in some of the environments we consider, but not in all of them. In particular, we provide an example in which transparency reduces both the informativeness of the equilibrium signal and the agent's effort. We also demonstrate that transparency can be harmful to the receiver. These results offer a perspective on the adverse consequences of transparency that does not stem from pandering incentives (Prat, 2005; Levy, 2007).

Since a pioneering contribution by Kamenica and Gentzkow (2011), the literature on Bayesian persuasion has been growing rapidly. The basic framework has been extended to accommodate, for example, multiple senders (Boleslavsky and Cotton, 2015; Li and Norman, 2018; Au and Kawai, 2020; Gentzkow and Kamenica, 2017; Boleslavsky and Cotton, 2018), multiple receivers (Alonso and Câmara, 2016; Chan et al., 2019), a privately informed receiver (Kolotilin et al., 2017; Kolotilin, 2018; Guo and Shmaya, 2019), dynamic environments (Ely, 2017; Renault et al., 2017), and the possibility of falsification (Perez-Richet and Skreta, 2017). More broadly, optimal information design has been incorporated into various economic contexts, including price discrimination (Bergemann et al., 2015), monopoly pricing (Roesler and Szentes, 2017), and auctions (Bergemann et al., 2017).

Two contemporary papers, Rodina (2017) and Rodina and Farragut (2017), study a similar three-player game to ours. The main difference lies in the sender's objective.[3] In our model, the sender has general preferences over the receiver's actions. The sender is indirectly concerned with the agent's effort, because the receiver's posterior beliefs (and the actions they induce) depend on the receiver's conjecture about the agent's effort. In contrast, in both Rodina (2017) and Rodina and Farragut (2017), the sender is concerned only with maximizing the agent's effort.[4] On one hand, this objective can be accommodated in our analysis by specifying that the sender's payoff is linear

[3]See also Bloedel and Segal (2018), Habibi (2020), and Zapechelnyuk (2020). These papers also study the tension between incentive and information provision in the Bayesian persuasion framework. Habibi (2020) and Zapechelnyuk (2020) consider specific applications, self-motivation and certification, respectively. In Bloedel and Segal (2018), the tension between incentive and information provision arises from the receiver's limited attention.

[4]In this sense, these papers are related to Hörner and Lambert (2020), who characterize the rating system that maximizes the agent's effort in a dynamic career concerns model with various information sources.

in the receiver's posterior belief. On the other hand, these authors provide a more thorough analysis of this case than we do, analyzing multiple settings with different assumptions about information asymmetry and the production technology.[5]

Also related is Boleslavsky and Cotton (2015), who analyze a Bayesian persuasion model of school competition. In their baseline model, students are passive, but they also consider an extension in which each student must exert effort in order to acquire skill. Their main focus is on a tradeoff between school investment and loose academic standards. Furthermore, their analysis relies heavily on the assumptions of binary actions for the receiver (evaluator in their model) and binary effort for the students.

Rosar (2017) studies the design of an optimal test when a privately informed agent chooses whether or not to participate. The participation decision signals some of the agent's private information, which leads to an endogenous prior belief. Focusing on an environment with binary states, the author derives conditions under which the participation constraint can be summarized by a single indifference condition for the threshold type and characterizes an optimal test via concavification of the Lagrangian, similar to what we do in this paper. He finds that the optimal test is a "no false-positive" test. In contrast, in our binary-state environment, both "no false-positive" and "no false-negative" signals can be optimal.

The remainder of this paper is organized as follows. Section 2 introduces our general model. Section 3 explains how to reformulate the sender's problem as a constrained optimization problem and characterize the solution to the problem. Section 4 considers the case where there are two states, and Section 5 analyzes the case where there are two actions. Section 6 addresses two economic questions particularly relevant to our model/analysis, and Section 7 concludes.

# 2 The Model

**The game.** There are three players, sender ($S$), agent ($A$), and receiver ($R$), and an unobservable state $\omega \in \Omega \equiv \{1, ..., N\}$ whose distribution is determined by the agent's effort. The game unfolds in three stages. In the first stage, the sender designs and publicly reveals a signal structure $\pi$, which consists of a message space $\Sigma$ and a set of conditional probability distributions $\{\pi_\omega(\cdot)\}_{\omega\in\Omega}$ over $\Sigma$, where $\pi_\omega(s)$ represents the probability that message $s \in \Sigma$ is realized conditional on state $\omega \in \Omega$. As in KG, we impose no structural restriction on the sender's choice of $\pi$; that is, the sender can choose any finite set $\Sigma$ and any conditional probabilities $\pi_\omega(\cdot)$ over this set. In the second stage,

[5] Two other papers combine elements of information design and moral hazard in novel ways. Barron et al. (2020) consider a canonical principal-agent model in which the agent can add mean-preserving noise after observing an intermediate output, and show that if the agent is risk neutral, then a linear contract is optimal. Georgiadis and Szentes (2020) introduce endogenous monitoring into a dynamic principal-agent model. By applying ideas from information design, they show that the optimal contract has a simple binary structure.

the agent observes the chosen signal structure $\pi$ and chooses an effort $e \in [0,1]$. Given $e$, the state is drawn according to the probability distribution $\eta_e \in \Delta(\Omega)$.[6] We use $\eta_e(\omega)$ to denote the probability that state $\omega$ is realized conditional on effort $e$. In the third stage, a message $s \in \Sigma$ is realized according to the given signal structure $\pi$. The receiver observes $\pi$ and $s$, and chooses an action $x$ from a compact set of feasible actions $X$. Note that $e$ is the agent's private effort, so the receiver does not observe it.

The sender's utility, $u_S(x,\omega)$, and the receiver's utility, $u_R(x,\omega)$, depend on the receiver's action and the underlying state. The agent's utility depends on the receiver's action and his own effort $e$.[7] For convenience, we assume that the agent's utility function is additively separable and given by $u_A(x) - c(e)$. We impose standard restrictions to ensure that the agent's problem is well-behaved: $u_A(\cdot)$ is non-negative and bounded. $c(\cdot)$ is strictly increasing, strictly convex, and twice continuously differentiable, with $c(0) = c'(0) = 0$ and $c'(1)$ sufficiently large.[8] All agents are risk neutral and maximize their expected utility.

**Reformulation.** Let $\mu \in \Delta(\Omega)$ denote the receiver's belief about the state, where $\mu(\omega)$ denotes the probability that the state is $\omega$. For any $\mu$, let $x^*(\mu)$ denote the set of the receiver's optimal mixed actions; that is, $x^*(\mu) \equiv \Delta(x(\mu))$ where $x(\mu) \equiv argmax_{x \in X} \mathbb{E}_\mu[u_R(x,\omega)]$. Then, we can reformulate the agent's and the sender's payoffs as follows:

$$v_A(\mu) \equiv u_A(x^*(\mu)) \text{ and } v_S(\mu) \equiv \mathbb{E}_\mu[u_S(x^*(\mu),\omega)].$$

In other words, inducing a particular action $x \in X$ is identical to inducing a posterior $\mu$ under which the receiver's optimal action is $x$. As in KG, this reformulation allows us to abstract away from details of the receiver's actual decision problem without incurring any loss of generality. Note that $x^*(\mu)$ is not necessarily a singleton and, therefore, both $v_A$ and $v_S$ are correspondences in general. For ease of exposition, we treat $x^*(\mu)$ (and $v_A(\cdot)$ and $v_S(\cdot)$) as a function unless necessary and noted otherwise.

[6]As is standard, we let $\Delta(Y)$ denote the set of all probability distributions (vectors) over finite set $Y$.

[7]We assume that the agent's utility does not depend on the state $\omega$ for two reasons. From a technical perspective, this assumption enables us to redefine the agent's utility as a function of the receiver's posterior belief alone, as explained shortly. If the agent's utility also depends on the state, then this reformulation is more complicated, because the agent's payoff would depend both on the receiver's public belief (which depends on his conjecture about the agent's effort) and on the agent's private belief (which depends on his actual effort choice). From an economic perspective, this assumption implies that the agent's motivation for effort is purely extrinsic: that is, he exerts effort not because he inherently cares about the realization of the state, but to induce the receiver to take a desirable action. For example, in the context of education, a student exerts effort to increase the probability of getting a job, not because he values education itself.

[8]The assumption on $c'(1)$ is only to ensure that the agent's optimal effort is less than 1. The necessary bound for $c'(1)$ varies across different specifications and will be provided when necessary.

**Linear production technology.** We restrict attention to the following production technology: for two probability vectors $\underline{\mu}$ and $\overline{\mu}(\neq \underline{\mu})$ in $\Delta(\Omega)$,

$$\eta_e = (1-e)\underline{\mu} + e\overline{\mu},$$

where both $\underline{\mu}$ and $\overline{\mu}$ have full support on $\Omega$. In other words, the probability distribution that generates the underlying state $\omega$ is linear in the agent's effort $e$. One may imagine that the state is the realization of a compound lottery in which the agent's effort represents the probability that the state is drawn from $\overline{\mu}$ rather than $\underline{\mu}$. As shown in Section 3.1, this technology guarantees that the agent's optimal effort under any signal is fully characterized by the first-order condition of his optimization (i.e., the first-order approach is valid), which streamlines the formulation of the sender's signal design problem.

**Equilibrium definition.** We study perfect Bayesian equilibria of this game. An equilibrium consists of a signal $\pi$, the agent's effort $e$ (for each possible signal), and a belief system $M \equiv \{\mu_s\}_{s\in\Sigma}$ (also for each possible signal), which satisfy the following properties: (i) given any signal $\pi$ and the receiver's belief system $M$, the agent's effort $e$ maximizes his expected utility, (ii) the receiver's belief system $M$ is consistent with the agent's effort $e$ and the signal $\pi$, and (iii) the chosen signal $\pi$ maximizes the sender's expected utility.

# 3 General Characterization

In this section, we provide a general characterization of the sender's optimal signal. In particular, we show how to extend the geometric characterization in Kamenica and Gentzkow (2011) to allow for the agent's moral hazard.

## 3.1 Formulating the Sender's Problem

We first analyze the equilibrium of the subgame between the receiver and the agent and use it to reformulate the sender's choice as a constrained optimization problem.

**Subgame.** Suppose that the sender has chosen a signal $\pi : \Sigma \times \Omega \to [0, 1]$, where $\pi_\omega(s)$ denotes the probability that $s \in \Sigma$ is realized conditional on $\omega \in \Omega$. The receiver does not observe the agent's effort $e$ but should have a conjecture (belief) about it. Let $\widehat{e}$ denote such a conjecture by the receiver and $\mu_{s|(\pi,\widehat{e})} \in \Delta(\Omega)$ represent the receiver's updated belief following $s \in \Sigma$. Given $\widehat{e}$, (the receiver believes that) state $\omega \in \Omega$ is realized with probability $\eta_{\widehat{e}}(\omega)$. Therefore, by Bayes' rule,

$$\mu_{s|(\pi,\widehat{e})}(\omega) \equiv \frac{\eta_{\widehat{e}}(\omega)\pi_\omega(s)}{\sum_{\omega'} \eta_{\widehat{e}}(\omega')\pi_{\omega'}(s)} \text{ for all } \omega \in \Omega.$$

The agent's payoff depends on his effort, the signal structure $\pi$, and the receiver's belief structure $\{\mu_s\}_{s\in\Sigma}$, where $\mu_s \in \Delta(\Omega)$ denotes (the agent's belief about) the receiver's belief upon observing $s \in \Sigma$. In particular, the effort and signal structure together determine the probability of each signal realization, while the belief structure determines the associated reward. Specifically, given $\pi$ and $\{\mu_s\}_{s\in\Sigma}$, the agent's problem is given by

$$\max_{e\in[0,1]} \sum_\omega \eta_e(\omega) \left(\sum_s \pi_\omega(s) v_A(\mu_s)\right) - c(e).$$

The first term is linear in $e$, because $\eta_e(\omega) = (1-e)\underline{\mu}(\omega) + e\overline{\mu}(\omega)$ for all $\omega \in \Omega$. Since the second term $c(e)$ is strictly convex, the agent's optimal effort is fully characterized by the following first-order condition:

$$\sum_\omega (\overline{\mu}(\omega) - \underline{\mu}(\omega)) \left(\sum_s \pi_\omega(s) v_A(\mu_s)\right) \leq c'(e),$$

with equality holding if $e > 0$.[9]

The receiver's belief structure $\{\mu_{s|(\pi,\widehat{e})}\}$ depends on the conjectured effort $\widehat{e}$, while the agent's optimal effort $e$ depends on the receiver's belief structure $\{\mu_s\}$. In equilibrium, the conjectured effort $\widehat{e}$ should coincide with the agent's actual effort choice $e$. Therefore, the equilibrium belief structure $\{\mu_s\}$ must be based on the equilibrium effort level $e$, and this effort level must be optimal for the agent given the receiver's belief structure. This yields the following equilibrium conditions, which fully summarize the equilibria of the subgame between the agent and the receiver given signal $\pi$:

$$\mu_s = \mu_{s|(\pi,e)} \text{ and } \sum_\omega (\overline{\mu}(\omega) - \underline{\mu}(\omega)) \left(\sum_s \pi_\omega(s) v_A(\mu_s)\right) \leq c'(e), \tag{1}$$

with equality holding if $e > 0$.

**Sender's problem.** Given the preceding characterization of the subgame, the sender's problem can be written as

$$\max_{\pi,e} \sum_\omega \eta_e(\omega) \left(\sum_s \pi_\omega(s) v_S(\mu_s)\right) \text{ subject to the two conditions in (1).}$$

[9] Our maintained assumption that $c'(1)$ is sufficiently large ensures that the optimal effort is always less than 1.

In other words, the sender chooses $\pi$ and $e$ in order to maximize her expected payoff subject to the constraint that $e$ and $\{\mu_s\}_{s\in\Sigma}$ must constitute an equilibrium of the subgame given the sender's choice of $\pi$.

For most of the paper, we study the sender's optimization problem given a particular target effort level $e$.[10] In other words, we treat $e$ as a parameter of the sender's optimization, rather than a choice variable. Following the standard approach in mechanism design, we assume that if multiple equilibria exist for a given signal structure, then the agent and receiver will comply with the sender's choice among them. For the global solution to the sender's problem, it suffices to identify the effort level that maximizes the sender's (indirect) expected payoff, which is complicated in general but reduces to a straightforward optimization in many specific settings.

Note that for some effort levels, there may not exist a signal that generates them as part of a subgame equilibrium. We characterize the set of *implementable* efforts–the set of effort levels that can be induced by some signal $\pi$–in more specific environments in Sections 4 and 5.

The following result allows us to reformulate the sender's problem so that she chooses a distribution of posteriors $\tau \in \Delta(\Delta(\Omega))$ instead of a signal $\pi$, which increases tractability.

**Lemma 1** *Given any positive target effort $e(> 0)$, there exists a signal $\pi$ that satisfies the two conditions in (1) if and only if there exists a distribution of posteriors $\tau \in \Delta(\Delta(\Omega))$ such that*

*(i)* $\mathbb{E}_\tau[\mu] = \eta_e$ *(Bayes-Plausibility), and*

*(ii)* $\mathbb{E}_\tau\left[\mathbb{E}_\mu\left[\frac{\overline{\mu}(\omega)-\underline{\mu}(\omega)}{\eta_e(\omega)}\right] v_A(\mu)\right] = c'(e)$ *(Incentive Compatibility).*

The two conditions in the lemma are restatements of the equilibrium conditions in (1) in the space of posterior beliefs. To understand the link, fix a signal structure $\pi$ and effort $e$. Without loss of generality, assume that for each realization $s \in \Sigma$, the receiver has a distinct Bayesian update $\mu_s \in \Delta(\Omega)$. Then, the probability that the Bayesian update is $\mu_s$ is equal to

$$\tau(\mu_s) = \sum_{\omega'} \eta_e(\omega')\pi_\omega(s).$$

Plugging this into $\mu_s = \mu_{s|(\pi,e)}$ and arranging the terms, for each $\omega \in \Omega$, we get

$$\mu_s(\omega) = \frac{\eta_e(\omega)\pi_\omega(s)}{\sum_{\omega'} \eta_e(\omega')\pi_\omega(s)} = \frac{\eta_e(\omega)\pi_\omega(s)}{\tau(\mu_s)} \Rightarrow \pi_\omega(s) = \tau(\mu_s)\frac{\mu_s(\omega)}{\eta_e(\omega)}. \tag{2}$$

[10] An equilibrium with zero effort always exists in the subgame between agent and receiver. Furthermore, multiple equilibria with distinct, strictly positive effort levels may also exist in this subgame. However, because different equilibria are associated with distinct effort levels, for a particular target effort, the sender's problem is well-defined, despite the possibile equilibrium multiplicity in the subgame.

Clearly, starting with a signal structure and an equilibrium effort, we can compute the resulting distribution of posterior beliefs. Equation (2) shows that the process can be reversed: given any Bayes-plausible distribution of posteriors, one can also derive the underlying signal structure that gives rise to the distribution. To derive (IC), it suffices to combine equation (2) with equation (1):

$$\sum_{\omega}(\overline{\mu}(\omega)-\underline{\mu}(\omega))\left(\sum_{s}\pi_{\omega}(s)v_A(\mu_s)\right)=\sum_{\omega}(\overline{\mu}(\omega)-\underline{\mu}(\omega))\left(\sum_{s}\tau(\mu_s)\frac{\mu_s(\omega)}{\eta_e(\omega)}v_A(\mu_s)\right)$$
$$=\sum_{s}\tau(\mu_s)\left[\sum_{\omega}\left(\mu_s(\omega)\frac{\overline{\mu}(\omega)-\underline{\mu}(\omega)}{\eta_e(\omega)}\right)\right]v_A(\mu_s)=\mathbb{E}_{\tau}\left[\mathbb{E}_{\mu}\left[\frac{\overline{\mu}(\omega)-\underline{\mu}(\omega)}{\eta_e(\omega)}\right]v_A(\mu)\right].$$

Two aspects of (IC) are worth highlighting. First, it immediately implies that a degenerate posterior distribution (equivalently, an uninformative signal) can implement only $e=0$.[11] In other words, dispersion in the posterior belief distribution is necessary to induce the agent's positive effort. As shown later, however, it is not necessarily the case that more dispersion induces more effort (see Section 5). Furthermore, in the absence of (IC), the sender would use an uninformative signal whenever her payoff function is concave in $\mu$. Thus, the desire to motivate the agent forces the sender to introduce distortions in that case (see Section 4).

Second, the likelihood ratio term $\frac{\overline{\mu}(\omega)-\underline{\mu}(\omega)}{\eta_e(\omega)}$ also appears in the standard moral hazard model (where the principal controls the agent's rewards). There, this term arises from the *principal's optimization*, reflecting the benefit/cost of distorting the agent's compensation away from the first best for a particular output (state) realization (Hölmstrom, 1979). In contrast, in our model, these terms emerge from the *agent's optimization*, appearing directly in his first-order condition via the inversion of Bayes' rule. Despite this difference, existing insights in the literature help us to interpret (IC). In the standard model, this likelihood ratio is interpreted as a statistical estimate of the agent's effort from the observed output (state), where a high value suggests high effort. Thus, the principal acts as if he estimates the agent's effort from the output and rewards the agent according to this estimate. In our model, allocating mass to a particular posterior belief has a large impact on (IC) when the covariance between the agent's payoff from the realization ($v_A(\mu)$) and the expected value of the receiver's statistical estimate of effort ($\mathbb{E}_{\mu}\left[\frac{\overline{\mu}(\omega)-\underline{\mu}(\omega)}{\eta_e(\omega)}\right]$) is large. Thus, in our model the agent acts as if his benefit or loss at a particular realization is scaled by the receiver's estimate of his effort, even though his reward depends only on the receiver's action.[12]

[11]For a degenerate distribution to satisfy (BP), the unique posterior belief must be equal to the prior, that is, $\mu=\eta_e$. Then, it violates the IC constraint whenever $e>0$, because

$$\mathbb{E}_{\tau}\left[\mathbb{E}_{\mu}\left[\frac{\overline{\mu}(\omega)-\underline{\mu}(\omega)}{\eta_e(\omega)}\right]v_A(\mu)\right]=\left(\sum_{\omega}\left(\overline{\mu}(\omega)-\underline{\mu}(\omega)\right)\right)v_A(\eta_e)=0\cdot v_A(\eta_e)=0<c'(e)\text{ for any }e>0.$$

[12]In particular, given a Bayes-Plausible distribution of posteriors, the equilibrium effort can be derived from the

## 3.2 Main Characterization

We now characterize the sender's optimal signal structure. Lemma 1 implies that given $e > 0$, the sender's problem can be written as

$$\max_{\tau \in \Delta(\Delta(\Omega))} \mathbb{E}_\tau[v_S(\mu)] \text{ subject to } (BP)\ \mathbb{E}_\tau[\mu] = \eta_e \text{ and } (IC)\ \mathbb{E}_\tau[h(\mu)] = 0, \quad (3)$$

where

$$h(\mu) \equiv \mathbb{E}_\mu \left[ \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)} \right] v_A(\mu) - c'(e) \text{ for each } \mu \in \Delta(\Omega).$$

We let $\tau^e$ denote an optimal solution to this problem (i.e., an optimal distribution of posteriors that implements effort $e$) and $V^e$ denote the corresponding expected utility of the sender (i.e., $V^e \equiv \mathbb{E}_{\tau^e}[v_S(\mu)]$). Crucially, in (3), the objective function and the two constraints are expectations of certain functions of $\mu$ with respect to $\tau$. This property allows us to extend the geometric characterization in Aumann and Maschler (1995) and KG to our problem.

Consider the following curve in $\mathcal{R}^{N+2}$:

$$K^e \equiv \{(\mu, h(\mu), v_S(\mu)) : \mu \in \Delta(\Omega)\}.$$

By construction, each element of $K^e$ corresponds to *ex post* values for the three objects in (3): the first $N$ components are $\mu = (\mu(1), ..., \mu(N))$, which are the ex post values of each component of (BP). The last two components are the ex post values of (IC) and the sender's payoff, respectively.

Next, construct the convex hull of $K^e$, denoted by $co(K^e)$. By definition, $co(K^e)$ consists of all convex combinations of the elements of $K^e$. Therefore, $co(K^e)$ captures all *ex ante* values of (BP), (IC), and sender payoff that can be generated by choosing a probability measure over $\Delta(\Omega)$. Economically, this can be interpreted as the "production possibility set" for the Bayesian persuasion problem, specifying which values of (BP), (IC), and sender payoff are feasible (consistent) with the sender's "technology." The problem is then to find the maximal expected utility of the sender inside of $co(K^e)$, while respecting the two constraints, as formally stated in the following theorem. The result on the cardinality of the support follows from Caratheodory's theorem.[13]

**Theorem 1** *Suppose $e(> 0)$ is implementable. Then, the maximal utility the sender can obtain*

following optimization problem: $\max_e \mathbb{E}_\tau[\mathbb{E}_\mu[\log(\eta_e(\omega))]v_A(\mu)] - c(e)$. Thus, the equilibrium effort is identical to the effort choice of a "virtual agent" who takes the distribution of the posterior belief as given and has payoff $\mathbb{E}_\mu[\log(\eta_e(\omega))]v_A(\mu)] - c(e)$ at each $\mu$. The "virtual" payoff function weighs the agent's true payoff $v_A(\mu)$ by the expected value of the log-likelihood function.

[13] To be precise, it is not a direct application of Caratheodory's theorem, which states that any point on the boundary of a convex set in $\mathcal{R}^{N+2}$ can be composed of at most $N+2$ extreme points (realizations). However, one dimension can be eliminated, because $\Delta(\Omega)$ is an $(N-1)$-dimensional simplex (i.e., $\sum_\omega \mu(\omega) = 1$ for any $\mu \in \Delta(\Omega)$). Therefore, we can reduce the necessary number of realizations by 1.

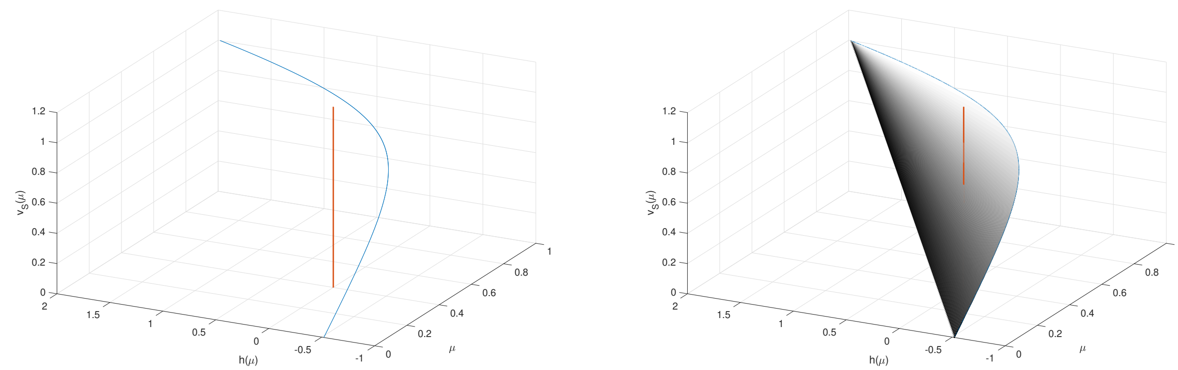

Figure 1: The left panel depicts the curve $K^e$, while the right panel depicts its convex hull $co(K^e)$. In this example, $n = 2$ and $\mu$ represents the probability that $\omega = 2$. In addition, $\eta_e = E[\mu] = 0.4$, $v_A(\mu) = \mu$, and $v_S(\mu) = 1 - (1 - \mu)^4$.

*conditional on targeting $e$ is equal to*

$$V^e = \max\{v : (\eta_e, 0, v) \in co(K^e)\}.$$

*In addition, there exists an optimal distribution of posteriors $\tau^e \in \Delta(\Delta(\Omega))$ whose support contains at most $N + 1$ posteriors (i.e., $|supp(\tau^e)| \leq N + 1$).*

**Proof.** Consider the following subset of $co(K^e)$:

$$H^e \equiv \{(y_1, y_2, y_3) \in co(K^e) : y_1 = \eta_e, y_2 = 0\}.$$

Since $e$ is implementable, $H^e$ is not empty. In addition, by construction, it includes all the points that are convex combinations of the elements of $K^e$ (i.e., $y_1 = \mathbb{E}_\tau[\mu]$, $y_2 = \mathbb{E}_\tau[h(\mu)]$, and $y_3 = \mathbb{E}_\tau[v_S(\mu)]$) and satisfy the two constraints (i.e., $y_1 = \eta_e$ and $y_2 = 0$). Therefore, $\max\{v : (\eta_e, 0, v) \in co(K^e)\}$ is the maximal value to the problem in (3). Because $K^e$ is closed and bounded, $co(K^e)$ is also closed and bounded, and hence, this maximum is attained. For the result on the cardinality of the support of $\tau^e$, see the appendix. ■

Figure 1 illustrates the argument in the binary-state case. Here, the receiver's belief can be represented by a single variable $\mu(2) \in [0, 1]$. In a slight abuse of notation, we replace $\mu(2)$ by $\mu$ and $\eta_e(2)$ by $\eta_e$. The left panel depicts the 3-dimensional curve $K^e \equiv \{(\mu, h(\mu), v_S(\mu)) : \mu \in [0, 1]\}$, while the right panel shows its convex hull $co(K^e)$. The vertical rod in both panels is built upon $(\eta_e, 0, 0)$. In order to find $V^e$, it suffices to move up along the rod and identify the highest point in $co(K^e)$. Clearly, the optimal point $(\eta_e, 0, V^e)$ is on the boundary of $co(K^e) \subset \mathcal{R}^3$. By Caratheodory's theorem for the boundary, $(\eta_e, 0, V^e)$ is a convex combination of no more than three elements of $K^e$.

In the absence of moral hazard, $h(\mu)$ is irrelevant. Therefore, the component of $co(K^e)$ representing $h(\mu)$ can be eliminated, and our geometric argument reduces to the one in KG. In this case, the boundary of $co(K^e)$ reduces to the concave envelope of $v_S(\cdot)$, and the sender's payoff is the value of the concave envelope at the prior belief $\eta_e$. In Figure 1, eliminating the dimension corresponding to $h(\mu)$ projects $co(K^e)$ into the plane of $(\mu, v_S(\mu))$, which is identical to the concave envelope in KG. From a technical perspective, moral hazard introduces an additional dimension (IC) into the sender's optimization problem, constraining the sender's choice of posterior belief distribution along this dimension and increasing the number of required realizations by 1.

There are two other noteworthy points regarding Theorem 1. First, it does not crucially depend on our restriction to the linear production technology (i.e., $\eta_e = e\overline{\mu} + (1-e)\underline{\mu}$). The same logic goes through unchanged as long as the subgame equilibrium effort is fully characterized by the agent's first-order condition (i.e., the first-order approach is valid). As explained above, the linear production technology guarantees this latter property but is not necessary for it. Second, our argument also extends to other settings in which a game between the receiver and the agent (or agents) imposes additional constraints on the sender's problem, provided that these constraints can be written as expectations with respect to $\tau$ (i.e. each constraint can be written as $\mathbb{E}_\tau[F(\mu)] = 0$ for some function $F(\cdot)$).[14] The only difference is that when there are a total of $k$ constraints (including (BP)), the maximal necessary number of posterior belief realizations is $n + k - 1$.

### 3.3 Optimality Conditions

While useful for understanding the structure of the sender's problem, Theorem 1 does not establish explicit necessary and sufficient conditions for optimality. We now develop such conditions using our preceding characterization. The result is based on the observation that $(\eta_e, 0, V^e)$ lies on the boundary of a convex set $co(K^e)$ and, therefore, there exists a supporting hyperplane to $co(K^e)$ at $(\eta_e, 0, V^e)$.

**Theorem 2** *A distribution of posteriors $\tau^e$ is a solution to the sender's problem (3) if and only if it satisfies (BP), (IC), and there exist $\lambda_0 \in \mathcal{R}$, $\psi \in \mathcal{R}$, and $\lambda_1 \in \mathcal{R}^N$ such that*[15]

$$\mathcal{L}(\mu, \psi) \equiv v_S(\mu) + \psi h(\mu) \leq \lambda_0 + \langle \lambda_1, \mu \rangle, \textit{ for all } \mu \in \Delta(\Omega),$$

[14] Whether or not a constraint can be represented as an expectation with respect to $\tau$ depends on the underlying economic problem. For example, the participation constraints in Rosar (2017) are reduced to a single constraint that can be represented as an expectation, while the no-falsification constraint in Perez-Richet and Skreta (2017) cannot be represented in this way.

[15] We use $\langle \cdot, \cdot \rangle$ to denote the inner product between two vectors of the same dimension; that is, for any $y, z \in \mathcal{R}^N$, $\langle y, z \rangle \equiv y(1)z(1) + ... + y(N)z(N)$.

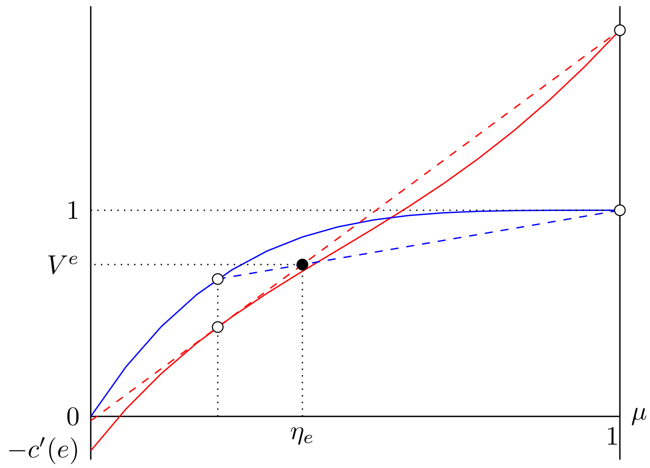


Figure 2: The concave solid curve depicts $v_S(\mu)$, while the other solid curve depicts $\mathcal{L}(\mu, \psi) = v_S(\mu) + \psi h(\mu)$, with the same data as in Figure 1 and $c(e) = e^2/2$.

*with equality for all $\mu$ such that $\tau^e(\mu) > 0$.*

**Proof.** Here, we illustrate only how the existence of the supporting hyperplane leads to the inequality above, relegating the rest of the proof to the appendix. Given the supporting hyperplane, we can find a normalized direction vector $d \equiv (-\lambda, \psi, 1) \in \mathcal{R}^{N+2}$ and a scalar $\lambda_0$ such that $\langle d, y \rangle \leq \lambda_0$ for all $y \in co(K^e)$, with equality for $y = (\eta_e, 0, V^e)$. It follows that $\langle d, z \rangle \leq \lambda_0$ for any vector $z \in K^e$. Expanding this inner product yields $\mathcal{L}(\mu, \psi) \leq \lambda_0 + \langle \lambda_1, \mu \rangle$ for all $\mu \in \Delta(\Omega)$. ■

Figure 2 illustrates Theorem 2 for the binary-state case where $v_S(\mu)$ is an increasing concave function of $\mu = \Pr\{\omega = 2\}$. If $\psi = 0$, then the condition is identical to the corresponding condition in KG. An optimal signal can be found by first drawing a line $\lambda_0 + \lambda_1 \mu$ that supports $v_S(\mu)$ above $\eta_e$. The straight line is, in essence, the concave closure of $v_S(\mu)$, the sender's payoff is the value of the line at $\eta_e$, and the optimal signal is supported on the posterior beliefs at which the supporting line meets the sender's payoff function. In Figure 2, $v_S(\mu)$ is concave, and thus, in the absence of moral hazard the sender's maximal utility is achieved with a degenerate posterior distribution.

Moral hazard requires two changes. First, concavification applies to $\mathcal{L}(\cdot, \psi)$ rather than $v_S(\cdot)$, which are different whenever $\psi \neq 0$ (which is always the case if $v_S(\cdot)$ is concave). Second, (IC) must hold, which also imposes restrictions on the Lagrangian and optimal posterior distribution. Note that (IC) implies $\mathbb{E}_{\tau^e}[\mathcal{L}(\mu, \psi)] = \mathbb{E}_{\tau^e}[v_S(\mu)]$. Graphically, $\mathbb{E}_{\tau^e}[\mathcal{L}(\mu, \psi)]$ is the value of the Lagrangian's supporting line (the dashed red line) evaluated at $\eta_e$. Similarly, $\mathbb{E}_{\tau^e}[v_S(\mu)]$ can be obtained by constructing the corresponding chord of the sender's payoff function (the blue dashed line) and evaluating it at $\eta_e$. If (IC) is satisfied, then the chord of the payoff function and the supporting line of the Lagrangian intersect above $\eta_e$. In some cases, as shown in Section 4.4, satisfying (IC) requires a third realization, in which case the supporting line meets the Lagrangian at three distinct values of $\mu$ and each $\mu$ receives positive probability in the optimal signal. In the binary case of KG, this is inconsequential: $v_S(\cdot)$ may meet the supporting line at more than two points, but an optimal distribution of posteriors can always be supported on only two such points.

# 4 Binary States

In this section, we consider a tractable environment in which there are only two states (i.e., $\Omega = \{1, 2\}$) and, therefore, the receiver's belief can be represented by a scalar. In a slight abuse of notation, we treat $\mu$, $\underline{\mu}$, $\overline{\mu}$, and $\eta_e$ as scalars between $0$ and $1$ and use them to represent the probability that $\omega = 2$. We assume that $\overline{\mu} > \underline{\mu}$, so that effort increases the probability that $\omega = 2$. We first offer some general characterization results and then analyze three representative examples in detail.

We maintain the following assumptions through this section.

(i) Monotonicity: both $v_A(\cdot)$ and $v_S(\cdot)$ are increasing in $\mu$, and $v_A(\underline{\mu}) < v_A(\overline{\mu})$.

(ii) Normalization: $v_A(0) = v_S(0) = 0$ and $v_A(1) = v_S(1) = 1$.

(iii) Interior effort: $c'(1) > \overline{\mu} - \underline{\mu} > 0$.

Assumption (i) implies that the interests of the sender and the agent are aligned: both want the receiver's belief to be as high as possible. Assumption (ii) is purely for convenience. Assumption (iii) ensures that the agent's equilibrium effort is less than 1.

## 4.1 Implementable and Incentive-free Effort Levels

We say that a target effort is implementable if there exists a signal $\pi$ (equivalently, a distribution of posteriors $\tau$) that satisfies both (BP) and (IC). The following proposition shows that an effort level is implementable if and only if it is below a certain threshold.

**Proposition 1** *In the binary-state model, $e$ is implementable if and only if $e \leq \overline{e}$, where $\overline{e}$ is the value that satisfies $c'(\overline{e}) = \overline{\mu} - \underline{\mu}$. The maximal effort $\overline{e}$ is induced if and only if the signal is fully informative.*

**Proof.** See the appendix. ■

For the intuition, notice that with binary states, equation (1) reduces to

$$(\overline{\mu} - \underline{\mu}) \left( \sum_s \pi_2(s) v_S(\mu_s) - \sum_s \pi_1(s) v_S(\mu_s) \right) = c'(e).$$

Therefore, the agent's incentive depends exclusively on the difference in the expected rewards in the two possible states, $\sum_s \pi_2(s)v_S(\mu_s) - \sum_s \pi_1(s)v_S(\mu_s)$. Given the assumptions on $v_A(\cdot)$, this term cannot exceed 1 (because $\pi_1(s), \pi_2(s), v_S(\mu) \in [0,1]$ for all $\omega = 1,2$ and $\mu \in [0,1]$) and achieves 1 if and only if $\sum_s \pi_2(s)v_S(\mu_s) = 1$ and $\sum_s \pi_1(s)v_S(\mu_s) = 0$. For the latter, it is necessary and sufficient that the signal perfectly distinguishes between the two states, so that $v_S(\mu_s) = 1$ whenever $\pi_2(s) > 0$ and $v_S(\mu_s) = 0$ whenever $\pi_1(s) > 0$.

As is well-known, in the absence of moral hazard, a fully informative signal (which maximally disperses the distribution of posteriors) is optimal if $v_S(\cdot)$ is convex, while a fully uninformative signal (which induces a degenerate posterior) is optimal if $v_S(\cdot)$ is concave. An immediate, but important, corollary of Proposition 1 is that the former result continues to hold, while the latter result fails in the case of moral hazard.

**Corollary 1** *In the binary-state model, a fully informative signal is optimal when $v_S(\cdot)$ is convex, while a fully uninformative signal is never optimal.*

**Proof.** See the appendix. ■

If $v_S(\cdot)$ is convex, then a fully informative signal maximizes the sender's expected utility under any fixed prior $\eta_e$. Simultaneously, a fully informative signal maximizes $\eta_e$, which also benefits the sender. For the second result, recall that a fully uninformative signal induces a degenerate posterior distribution and, therefore, results in $e = 0$. The sender can do better, for example, by using a posterior belief distribution supported on $\{\underline{\mu}, 1\}$, which induces the agent to choose a positive effort.

For the general case, let $\widehat{V}^e$ denote the maximal attainable value to the sender in the relaxed problem without (IC):

$$\widehat{V}^e \equiv \max_{\tau \in \Delta(\Delta(\Omega))} \mathbb{E}_\tau[v_S(\mu)] \text{ subject to (BP) } \mathbb{E}_\tau[\mu] = \eta_e.$$

Obviously, $V^e \le \widehat{V}^e$ for any $e \le \overline{e}$. Let $\underline{e}$ be the maximal value of $e$ such that $V^e = \widehat{V}^e$. Corollary 1 implies that $\underline{e} = 0$ if $v_S(\cdot)$ is concave, while $\underline{e} = \overline{e}$ if $v_S(\cdot)$ is convex. If $v_S(\cdot)$ is neither concave nor convex, then $\underline{e}$ can be found by first solving the relaxed problem and then verifying whether the resulting optimal distribution of posteriors satisfies (IC).[16]

The following result shows that the sender prefers $\underline{e}$ to any $e(< \underline{e})$, so it suffices to consider the effort levels in $[\underline{e}, \overline{e}]$. It also shows that implementing $e > \underline{e}$ requires a distortion from the relaxed problem.

[16] An alternative interpretation of $\underline{e}$ is as follows: suppose the sender designs, or can revise, a signal after the agent chooses $e$. In this case, the sender necessarily adopts an optimal signal in the sense of KG and, anticipating this, the agent adjusts his effort. $\underline{e}$ is the maximal effort attainable under such a scenario. This shows that it is the sender's power to commit that enables her to implement $e \in (\underline{e}, \overline{e}]$.

**Proposition 2** *In the sender's problem* (3), *if* $e < \underline{e}$, *then* $V^e \leq V^{\underline{e}}$. *Furthermore, for any* $e \in (\underline{e}, \overline{e}]$, *the solution to* (3) *has* $\psi > 0$.

**Proof.** See the appendix. ■

We apply our characterization to three representative environments. For ease of exposition, we focus on the case in which $\underline{\mu} = 0$ and $\overline{\mu} = 1$. This implies that $\eta_e = e\overline{\mu} + (1-e)\underline{\mu} = e$.[17] In addition, the function $h(\mu)$ in (IC) reduces to

$$\begin{aligned} h(\mu) &= \mathbb{E}_{\mu}\left[\frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)}\right] v_A(\mu) - c'(e) \\ &= \left((1-\mu)\frac{(1-\overline{\mu}) - (1-\underline{\mu})}{1-e} + \mu\frac{\overline{\mu} - \underline{\mu}}{e}\right) v_A(\mu) - c'(e) = \frac{\mu - e}{e(1-e)} v_A(\mu) - c'(e). \end{aligned}$$

Notice that, since $\mathbb{E}_{\tau}[\mu] = e$, (IC) $\mathbb{E}_{\tau}[h(\mu)] = 0$ can be rewritten as

$$\text{Cov}[\mu, v_A(\mu)] = e(1-e)c'(e).$$

Thus, incentive compatibility constrains the covariance between the posterior belief and the agent's payoff to a specific value.

## 4.2 Concave/Linear Preferences

We begin with the case in which the agent's payoff is linear in the receiver's belief (i.e., $v_A(\mu) = \mu$), while the sender's payoff $v_S(\cdot)$ is strictly increasing, concave, and twice differentiable (i.e., $v_S'(\cdot) > 0$ and $v_S''(\cdot) < 0$). In the grading example (where the school is the sender, a student is the agent, and the job market is the receiver), this arises if the student cares only about his expected wage, while the school assigns more weight to underperforming students.

Fix $e \in (0, \overline{e})$ and consider the Lagrangian function $\mathcal{L}(\mu, \psi)$ introduced in Theorem 2. Because $v_A(\cdot)$ is linear, $h(\cdot)$ is quadratic, so the second derivative of $\mathcal{L}$ with respect to $\mu$ is given by

$$\mathcal{L}_{\mu\mu} \equiv \frac{\partial^2 \mathcal{L}(\mu, \psi)}{\partial \mu^2} = v_S''(\mu) + \frac{2\psi}{e(1-e)}.$$

Although $v_S''(\cdot) < 0$, the Lagrangian is not necessarily concave because the second term is positive. In fact, at the sender's optimal solution, $\mathcal{L}_{\mu\mu}$ *cannot* be negative everywhere: if the Lagrangian is concave, then the optimal signal is degenerate and, therefore, cannot implement $e > 0$. Conversely, $\mathcal{L}_{\mu\mu}$ cannot be positive everywhere either: if the Lagrangian is convex, then the optimal signal is

[17] Strictly speaking, $\underline{\mu} = 0$ and $\overline{\mu} = 1$ do not have full support. However, for any $e \in (0, 1)$, the prior belief $\eta_e = e$ does have full support. Therefore, Lemma 1 and the subsequent characterization apply to these cases unchanged.

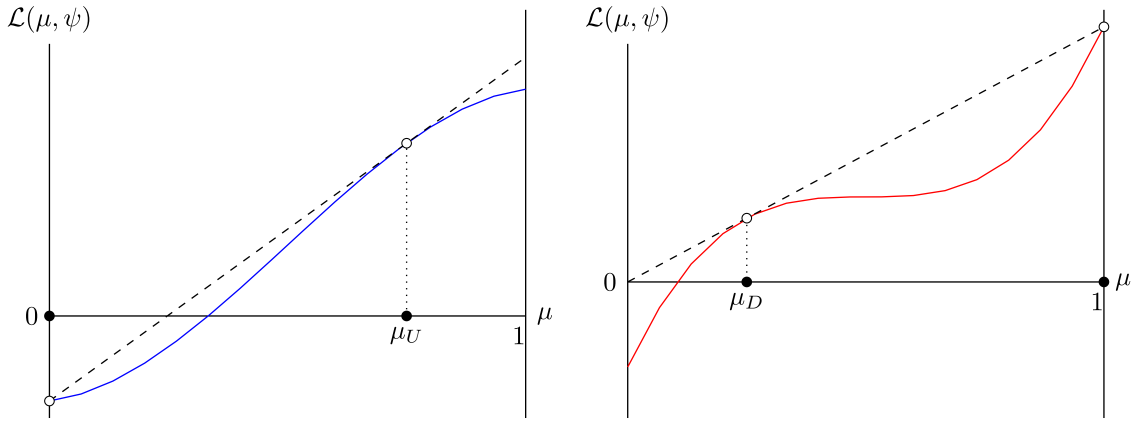


Figure 3: Both panels depict the Lagrangian function $\mathcal{L}(\mu, \psi)$. In the left panel $v_S'''(\cdot) < 0$, while in the right panel $v_S'''(\cdot) > 0$.

fully informative and implements $\bar{e}$ (see Proposition 1). Therefore, $\mathcal{L}_{\mu\mu}$ must have both positive and negative regions, and the Lagrangian must have at least one point of inflection.

A sharp result can be derived if $\mathcal{L}_{\mu\mu}$ is monotone. In this case, $\mathcal{L}$ cannot have more than one inflection point and must have at least one at the optimum. Furthermore, the change in curvature at the inflection point does not depend on $\psi$. Thus, if $v_S'''(\cdot) < 0$ (resp. $v_S'''(\cdot) > 0$), then $\mathcal{L}$ switches once from convex to concave (resp. concave to convex) at the optimum. Applying Theorem 2, it follows that the optimal distribution is supported on two posterior beliefs, one of which must be either $0$ or $1$ (see Figure 3)

**Proposition 3** *Consider the binary-state model with $v_S''(\cdot) < 0$ and $v_A(\mu) = \mu$.*

*(i) If $v_S'''(\cdot) < 0$, then the optimal distribution of posteriors that implements $e \in (0, \bar{e})$ is supported on $\{0, \mu_U\}$ with $\tau^e(\mu_U) = e/\mu_U$, where $\mu_U \equiv e + (1-e)c'(e)$.*

*(ii) If $v_S'''(\cdot) > 0$, then the optimal distribution of posteriors that implements $e \in (0, \bar{e})$ is supported on $\{\mu_D, 1\}$ with $\tau^e(\mu_D) = (1-e)/(1-\mu_D)$, where $\mu_D \equiv e(1 - c'(e))$.*

For an economic intuition, notice that if $v_A(\mu) = \mu$ then $h(\mu)$ simplifies to

$$h(\mu) = \frac{\mu(\mu - e)}{e(1-e)} - c'(e).$$

Since $\mathbb{E}_\tau[\mu] = e$ (due to (BP)),

$$\mathbb{E}_\tau[h(\mu)] = \frac{\mathbb{E}_\tau[\mu(\mu-e)]}{e(1-e)} - c'(e) = \frac{\text{Var}(\mu)}{e(1-e)} - c'(e) = 0 \Leftrightarrow \text{Var}(\mu) = e(1-e)c'(c).$$

In other words, (IC) constrains the variance of the posterior belief to a specific value, while (BP) constrains the mean to $e$. This means that our search for an optimal signal reduces to finding the posterior distribution that maximizes the sender's expected payoff among those with a particular mean and variance.

The sender's problem is identical to the decision problem under uncertainty studied by Menezes et al. (1980). They formalize the notion of *downside risk* (how an individual perceives a small probability of big losses) and show that it is determined by the third derivative of von Neumann-Morgenstern utility function: if $v_S'''(\cdot) > 0$ then the decision-maker (sender) is downside risk averse and always prefers to shift all necessary dispersion to the top of the distribution, while compressing the bottom as much as possible.[18] In our sender's problem, this manifests as the optimality of a binary distribution whose larger realization is 1. Conversely, if $v_S'''(\cdot) < 0$, then the decision-maker (sender) is *downside risk loving* and prefers to shift all necessary dispersion to the bottom of the distribution, resulting in a binary distribution whose smaller realization is 0.

## 4.3 Identical Concave Preferences

We now consider an environment in which the sender and the agent have identical concave preferences. Specifically, $v_S(\mu) = v_A(\mu) = v(\mu)$ with $v''(\cdot) < 0$. This perfect alignment of interests arises if the school and the student have purely common interests in the grading example, or if the sender is a monopolist who designs a signal to maximize the agent's expected payoff, which she then extracts by charging a fixed fee.

The analysis is similar to that of Section 4.2. Given a target effort $e \in (0, \overline{e})$,

$$\mathcal{L}_{\mu\mu} = v''(\mu) + \psi \frac{2v'(\mu) + (\mu - e)v''(\mu)}{e(1-e)}.$$

Since $\underline{e} = 0$ (due to the sender's concave preferences), $\psi > 0$ for any $e > 0$ (see Proposition 2). It then follows that

$$\mathcal{L}_{\mu\mu} > 0 \Leftrightarrow \frac{e(1-e-\psi)}{2\psi} + \frac{\mu}{2} < \frac{1}{r(\mu)},$$

where $r(\mu) \equiv -v''(\mu)/v'(\mu)$ is the Arrow-Pratt measure of risk aversion. As in Section 4.2, at the sender's solution, $\mathcal{L}$ should be neither concave nor convex and have at least one inflection point. Furthermore, if the right-hand side is linear, then the number of inflection points is *at most* one. By the same logic as for Proposition 3, we obtain the following result.

[18] Menezes et al. (1980) define a *mean-variance preserving transformation*, which combines a mean-preserving spread over high values with a mean-preserving contraction over low values, while preserving the mean and variance of the distribution. They show that if an individual is downside risk averse, then he always prefers a mean-variance preserving transformation, while the opposite is true if an individual is downside risk loving.

**Proposition 4** *Consider the binary-state model in which the sender and the agent have identical HARA (Hyperbolic Absolute Risk Aversion) preferences: for any $\mu \in [0,1]$, $v_S(\mu) = v_A(\mu) = v(\mu)$ and $1/r(\mu) = \alpha\mu + \beta$ for some $(\alpha, \beta)$.*

*(i) If $\alpha < 1/2$, then the optimal distribution of posteriors that implements $e \in (0, \overline{e})$ is supported on $\{0, \mu_U\}$, with $\tau(\mu_U) = e/\mu_U$.*

*(ii) If $\alpha > 1/2$, then the optimal distribution of posteriors that implements $e \in (0, \overline{e})$ is supported on $\{\mu_D, 1\}$, with $\tau(\mu_D) = (1-e)/(1-\mu_D)$.*

## 4.4 Discrete/Linear Preferences

In our final example, $v_S(\cdot)$ is a step function at $\theta \in (0,1)$ ($v_S(\mu) = \mathcal{I}(\mu \geq \theta)$), while $v_A(\cdot)$ is linear. Then, the agent has a constant marginal benefit to increase the receiver's belief, while the sender would like this belief to be just "good enough." In the grading example, this arises if the school is concerned only with the fraction of graduates earning wages above a threshold. In order to reduce the number of cases to consider, we assume that $\overline{e} < \theta$.

Unlike in the previous cases, $v_S(\cdot)$ is neither convex nor concave, which allows for the possibility that $\underline{e} \in (0, \overline{e})$. In the absence of moral hazard, for any $\eta_e < \theta$, it is optimal for the sender to use a binary posterior distribution supported on $\{0, \theta\}$. It follows that the unique incentive-free effort level $\underline{e}$ is given by the value that satisfies

$$\mathbb{E}_\tau[h(\mu)] = \frac{\theta(\theta - \underline{e})}{\underline{e}(1-\underline{e})}\tau(\theta) = \frac{\theta - \underline{e}}{1-\underline{e}} - c'(\underline{e}) = 0.$$

From now on, we restrict attention to $e \in (\underline{e}, \overline{e})$.

Given the preferences of the sender and the agent,

$$\mathcal{L}(\mu, \psi) = \begin{cases} \psi\left(\frac{(\mu-e)\mu}{e(1-e)} - c'(e)\right) & \text{if } \mu < \theta, \\ 1 + \psi\left(\frac{(\mu-e)\mu}{e(1-e)} - c'(e)\right) & \text{if } \mu \geq \theta. \end{cases}$$

Thus, $\mathcal{L}$ is a quadratic function of $\mu$, with an upward jump discontinuity of 1 at $\mu = \theta$ (see Figure 4). Because $\mathcal{L}$ is convex over $[0, \theta)$ and over $[\theta, 1]$, there are three ways in which the supporting line $\lambda_0 + \lambda_1\mu$ can intersect $\mathcal{L}(\mu, \psi)$. The intersections can occur (i) at $\mu = 0$ and $\theta$, (ii) at $\mu = 0$ and 1, or (iii) at $\mu = 0$, $\theta$, and 1. However, (i) implements $\underline{e}$ (as explained above), while (ii) implements $\overline{e}$ (Proposition 1). Therefore, (iii) is the only possibility; that is, at the sender's optimal solution, the supporting line must intersect $\mathcal{L}(\cdot, \psi)$ at all three points, $\{0, \theta, 1\}$, as shown in Figure 4.

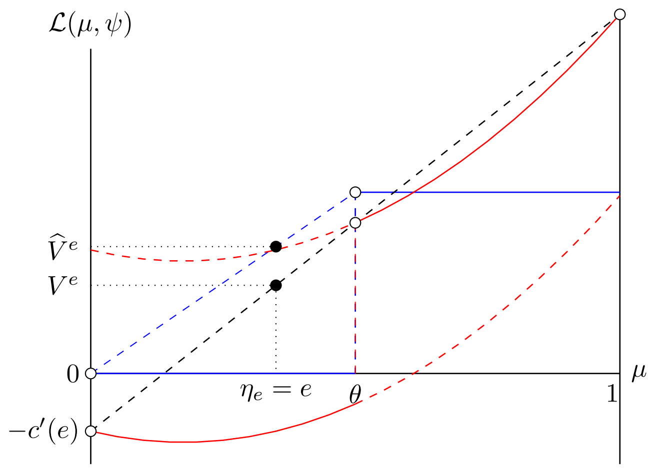


Figure 4: The step function represents $v_S(\mu) = \mathcal{I}_{\{\mu \geq \theta\}}$, while the solid curve depicts $\mathcal{L}(\mu, \psi)$ when $v_A(\mu) = \mu$. In this figure, $\theta = 0.5$, $e = 0.35$, and $c(e) = e^2$.

**Proposition 5** *In the binary-state model with $v_S(\mu) = \mathcal{I}\{\mu \geq \theta\}$, $v_A(\mu) = \mu$, and $c'(\theta) > 1$, for any $e \in (\underline{e}, \overline{e})$, the optimal distribution of posteriors $\tau^e$ is supported on $\{0, \theta, 1\}$, where*

$$\tau^e(0) = (1-e)(1-\tau^e(\theta)),\ \tau^e(\theta) = \frac{e(1-e)(1-c'(e))}{\theta(1-\theta)},\ \textit{and}\ \tau^e(1) = e - \tau^e(\theta)\theta.$$

**Proof.** The result on the use of three posteriors follows from the discussion above. The probabilities of each realization can be calculated from the following three equations: (i) $\tau^e(0) + \tau^e(\theta) + \tau^e(1) = 1$, (ii) (BP) $\mathbb{E}_{\tau^e}[\mu] = e$, and (iii) (IC) $\mathbb{E}_{\tau^e}[h(\mu)] = 0$. ■

This result demonstrates that the result on the maximal number of necessary posteriors in Theorem 1 is binding in the binary-state model. As shown in the previous subsections, it is often the case that a binary signal is optimal with binary states. However, as Proposition 5 shows, three distinct posteriors (equivalently, three signal realizations) may be necessary in our model, which is never the case in the absence of moral hazard.

# 5 Binary Actions

In this section, we analyze another tractable environment in which there are only two actions available to the receiver. The sender has aligned interests with the receiver, but the agent prefers one action to the other.[19] As an example, suppose that the the receiver is a representative voter who solely decides whether to keep the status quo or adopt a new policy. The agent is a politician who

[19]In an earlier version, we analyzed the case in which the sender's preferences are aligned with those of the *agent*. The results of this section can be adapted to that case in a straightforward manner.

exerts effort to improve the new policy but stands to benefit only if the new policy is implemented. Information about the realized state is communicated to the general public by an independent and impartial news media, which seeks to maximize the payoff of the representative voter. As in Gehlbach and Sonin (2014) and Gentzkow et al. (2015), the news media commits to a reporting strategy $\{\pi_\omega(\cdot)\}_{\omega\in\Omega}$, which is observed by the representative voter and the politician.[20] With this interpretation, the media's role is not only to inform citizens, but also to incentivize the politician.[21]

## 5.1 The Model

**Setup.** The receiver (representative voter) decides whether to take action $x_0$ (status quo) or $x_1$ (reform). If she selects $x_0$, then her payoff is $\theta(>0)$, regardless of the state. If she selects $x_1$, then she obtains a benefit of $v_\omega$ in state $\omega$. Without loss of generality, the states are ordered so that $v_1 = 0 < v_2 < ... < v_N$ and $v_N > \theta$. The sender's (news media) payoff is identical to that of the receiver (representative voter). The agent's (politician's) payoff depends only on the receiver's action, and he prefers $x_1$: he receives 0 if $x = x_0$ and 1 if $x = x_1$.

The receiver chooses $x_1$ if and only if $\mathbb{E}_\mu[v_\omega] = \sum_\omega \mu(\omega)v_\omega \geq \theta$. Therefore, the sender's and the agent's payoffs as functions of the posterior belief $\mu$ are respectively given by

$$v_S(\mu) = \max\{\mathbb{E}_\mu[v_\omega], \theta\} \text{ and } v_A(\mu) = \begin{cases} 1, & \text{if } \mathbb{E}_\mu[v_\omega] \geq \theta, \\ 0, & \text{if } \mathbb{E}_\mu[v_\omega] < \theta. \end{cases}$$

For $i \in \{0,1\}$, we let $\mathcal{X}_i(\subset \Delta(\Omega))$ denote the set of beliefs with which the receiver selects $x_i$. In addition, we refer to states $\omega$ such that $v_\omega < \theta$ as *rejection states* and states such that $v_\omega \geq \theta$ as *acceptance states*. Finally, we let $\omega_r$ denote the largest rejection state.

**Assumptions.** We assume that $\underline{\mu}, \overline{\mu} \in \mathcal{X}_0$. This ensures that if the sender provides no information, then the receiver never chooses $x_1$, regardless of the agent's effort; that is, information provision is necessary for persuasion.[22] For ease of exposition, we make four additional assumptions on $\underline{\mu}$ and $\overline{\mu}$.

**Assumption 1** *Monotone likelihood ratio property: $\overline{\mu}(\omega)/\underline{\mu}(\omega)$ is strictly increasing in $\omega$.*

This assumption ensures that higher states are more likely to be realized when the agent exerts greater effort. Notice that, since both $\underline{\mu}$ and $\overline{\mu}$ are probability vectors, $\overline{\mu}(\omega)/\underline{\mu}(\omega)$ crosses 1 once

[20] Alonso and Câmara (2016) consider a related Bayesian Persuasion problem in which the politician's effort is exogenous and she communicates directly with the representative voter.

[21] It has long been recognized that the news media has an essential function as government watchdog, with a "clear, instrumental role in preventing corruption, financial irresponsibility, and underhanded dealings" (Sen, 2001, p. 40).

[22] If $\underline{\mu} \notin \mathcal{X}_0$ then the problem becomes trivial because the receiver always prefers to take $x_1$. The assumption $\overline{\mu} \in \mathcal{X}_0$ is sufficient, but not necessary, for the results in this section but simplifies the technical analysis.

from below. Let $\omega_e$ denote the largest $\omega$ such that $\underline{\mu}(\omega) > \overline{\mu}(\omega)$. As the agent exerts more effort, states $\omega \leq \omega_e$ are less likely to be realized, while states $\omega > \omega_e$ are more likely to be realized.[23] We therefore refer to $\omega \leq \omega_e$ as *effort-negative* and $\omega > \omega_e$ as *effort-positive*.

**Assumption 2** *All acceptance states are effort-positive:* $\omega_e \leq \omega_r$.

This assumption implies that an increase in effort increases the probability of all acceptance states (above $\omega_r$). In the case of strict inequality, additional effort also increases the probability of some rejection states (between $\omega_e$ and $\omega_r$).

**Assumption 3** *With prior* $\underline{\mu}$*, ruling out effort-negative states induces acceptance:*

$$\frac{\sum_{\omega>\omega_e} \underline{\mu}(\omega) v_\omega}{\sum_{\omega>\omega_e} \underline{\mu}(\omega)} \geq \theta \iff \frac{\big(0, ..., 0, \underline{\mu}(\omega_e+1), ..., \underline{\mu}(N)\big)}{\sum_{\omega>\omega_e} \underline{\mu}(\omega)} \in \mathcal{X}_1.$$

This assumption states that starting with the worst prior belief $\underline{\mu}$, if all effort-negative states are ruled out, then the resulting posterior belief induces the receiver to select $x_1$.[24] This assumption is not essential, but streamlines the exposition considerably by reducing the number of equilibrium cases.

**Assumption 4** *The agent's marginal cost is sufficiently high:*

$$\sum_{\omega>\omega_e} \left(\overline{\mu}(\omega) - \overline{\mu}(\omega)\right) < c'(1).$$

This is a technical assumption that corresponds to $c'(1) > \overline{\mu} - \underline{\mu}$ in the binary-state model, ensuring that the agent's effort is always less than one. The specific form is clarified shortly.

## 5.2 Incentive-free and Implementable Efforts

Because the sender's payoff is convex in the belief, in the absence of moral hazard, it is optimal for the sender to fully reveal the realized state.[25] Clearly, if a signal is fully informative, then the receiver takes action $x_1$ if and only if $\omega > \omega_r$. Since the agent receives 1 if the receiver takes $x_1$

[23] The probability that state $\omega$ is realized given effort $e$ is $e\overline{\mu}(\omega) + (1-e)\underline{\mu}(\omega)$. Thus, an increase in effort reduces the probability of state $\omega$ if and only if $\underline{\mu}(\omega) > \overline{\mu}(\omega)$, or equivalently, $\omega \leq \omega_e$.

[24] From Assumption 2, the effort-negative states may be a strict subset of the rejection states, and thus eliminating them leaves a posterior belief that is supported on all acceptance states and some rejection states. The essence of the assumption is that even under the worst prior belief $\underline{\mu}$, eliminating the effort-negative states places enough weight on the acceptance states to induce $x_1$.

[25] In fact, it suffices to distinguish between rejection and acceptance states. Therefore, instead of a fully revealing signal, a binary signal that reveals whether $\omega \leq \omega_e$ or $\omega > \omega_e$ can be employed.

and $0$ otherwise, his problem can be written as

$$\max_{e} \sum_{\omega > \omega_r} (e\overline{\mu}(\omega) + (1-e)\underline{\omega}) - c(e).$$

Solving this problem leads to the following characterization of the incentive-free effort level $\underline{e}$.

**Lemma 2** *In the binary-action model, there exists a unique incentive-free effort level $\underline{e} \in (0, \overline{e})$, which is characterized by*

$$\sum_{\omega > \omega_r} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(\underline{e}).$$

*Effort $\underline{e}$ is implemented by a fully informative signal.*

Recall that in the binary-state model of Section 4, a fully informative signal implements the maximal effort $\overline{e}$, not the incentive-free effort $\underline{e}$. The following proposition characterizes the maximal implementable effort $\overline{e}$ in the current binary-action model and also show that any $e < \overline{e}$ is implementable.

**Proposition 6** *In the binary-action model, $e$ is implementable if and only if $e \leq \overline{e}$, where $\overline{e}$ is the value that satisfies*

$$\sum_{\omega > \omega_e} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(\overline{e}).$$

**Proof.** See the appendix. ■

In the binary-action model, the agent's incentive to exert effort stems from the possibility that he can switch the receiver's action from $x_0$ to $x_1$. Therefore, this incentive is strongest when $x_1$ is selected in *all* states that are more likely to be realized as the agent puts in more effort (i.e., all effort-positive states, $\omega > \omega_e$) and *only* in such states. The binary signal which perfectly distinguishes between effort-positive and effort-negative states (not between acceptance and rejection states) has this property.

An immediate corollary is that a fully informative signal implements the maximal effort $\overline{e}$ *if and only if* $\omega_e = \omega_r$. This reveals the main tension between information and incentive provision in the binary-action model. Because the sender's payoff is convex in the receiver's belief, the sender would like to reveal all information in the absence of moral hazard. However, whenever $\omega_r \neq \omega_e$, the sender can increase the agent's effort by reducing signal informativeness, transmitting the good message in all effort-positive states, not just the acceptance states. Therefore, the tension between incentive and information provision in the binary-action model operates in the *opposite direction* from the binary-state model. In the binary-state model, effort is maximized by a fully informative signal, while the solution in the absence of moral hazard may not be fully informative (at least in the cases of interest). In contrast, in the binary-action model, effort is maximized by a signal that is not fully informative, while the solution in the absence of moral hazard is fully informative. By implication, in the binary-action model, inducing effort higher than $\underline{e}$ requires the principal to *reduce* signal informativeness.

## 5.3 Optimal Signal

We proceed to characterize the optimal *binary* signal that implements $e \in (\underline{e}, \overline{e})$. This incurs no loss of generality for optimal information design, because with binary actions, the number of induced posteriors ($|supp(\tau)|$) does not need to exceed $|X| = 2$. The set of posterior belief realizations can always be partitioned into two subsets, those in $\mathcal{X}_0$ and those in $\mathcal{X}_1$. Since $\mathcal{X}_0$ and $\mathcal{X}_1$ are convex, and within $\mathcal{X}_i$, the sender's payoff is linear in the belief, replacing each subset with a single realization at its center of mass does not change the sender's expected payoff. Furthermore, because both (BP) and (IC) are linear in $\mu$, the value attained by the constraints is also unaffected by this transformation.

It is convenient to partition $[\underline{e}, \overline{e}]$ as follows: for each $\omega_e \leq k \leq \omega_r$, let $e_k$ denote the unique value that satisfies

$$\sum_{\omega > k} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(e_k).$$

The left-hand side is the marginal benefit of effort under a signal that distinguish between $\omega \leq k$ and $\omega > k$. Therefore, $e_k$ is the effort level induced by such a signal. Lemma 2 implies that $e_{\omega_r} = \underline{e}$ (because the optimal signal for $\underline{e}$ reveals all rejection states), while Proposition 6 implies that $e_{\omega_e} = \overline{e}$ (because the effort-maximizing signal reveals all effort-negative states). In addition, Assumption 1 ensures that the left-hand side decreases in $k$ as long as $k \geq \omega_e$. Thus, for such $k$, the sequence $e_k$ is decreasing.

The following proposition provides a closed-form characterization for the optimal signal that corresponds to each $e \in [\underline{e}, \overline{e})$.

**Proposition 7** *In the binary-action model, for $k \in \{\omega_e + 1, ..., \omega_r\}$, effort $e \in [e_k, e_{k-1})$ is optimally implemented by the following binary signal: $\Sigma = \{b, g\}$, and*

$$\pi_\omega(g) = \begin{cases} 0, & \text{if } \omega < k-1, \\ s, & \text{if } \omega = k, \\ 1, & \text{if } \omega > k \end{cases} \qquad \text{and } \pi_\omega(b) = 1 - \pi_\omega(g) \text{ for all } \omega \in \Omega,$$

*where* $s \in [0,1)$ *is the value that satisfies*

$$s = \frac{c'(e) - c'(e_k)}{\overline{\mu}(k) - \underline{\mu}(k)}.$$

**Proof.** See the appendix. ■

The optimal (binary) signal for target effort $e \in [e_k, e_{k-1})$ for $k \in \{\omega_e+1, ..., \omega_r\}$ is a monotone partition, which transmits message $g$ in all states greater than $k$, message $b$ in all states less than $k$, and a mixture of $g$ and $b$ in state $k$. Because message $g$ rules out all effort-negative states, it induces the receiver to select $x_1$ (see Assumption 3). Conversely, upon observing message $b$, the receiver infers that the realized $\omega$ is a rejection state, so he selects $x_0$. The optimal signal varies continuously and systematically with the target effort. In particular, as the target effort increases within $[e_k, e_{k-1})$, the probability of transmitting message $g$ in state $k$ increases, from 0 at $e = e_k$ to 1 at $e = e_{k-1}$. Thus, providing the agent with incentives requires the sender to transmit $g$ (and select action $x_1$) in an interval of effort-positive rejection states ($k \leq \omega \leq \omega_r$), where she would prefer to transmit $b$ (and select action $x_0$) if the agent's incentives were irrelevant. Because such states are effort-positive, transmitting $g$ increases the marginal benefit of effort for the agent, but because these are also rejection states, transmitting message $g$ (and selecting $x_1$) also imposes a cost on the sender.

In general, determining whether to transmit $g$ in an effort-positive rejection state $\omega$ requires the sender to compare the beneficial impact on incentive provision to the harmful payoff consequences. Message $g$ is optimally transmitted in states where the impact on incentives is large and the impact on the sender's payoff is small. In our context, the monotone likelihood ratio property (Assumption 1) ensures that transmitting message $g$ in a higher state simultaneously generates a larger impact on incentives and a smaller payoff cost for the sender. By implication, if it is optimal for the sender to transmit $g$ in some state $\omega$, then it is also optimal to do so in all larger states, generating the monotone partition structure of the optimal signal.

# 6 Discussion

This section addresses two economic questions that are particularly relevant to our model, one on the effects of transparency regarding the agent's effort choice and the other on the effects of moral hazard on the informativeness of the optimal signal.

## 6.1 The Effects of Transparency

In our model, the agent's effort is unobservable, so the sender can incentivize the agent's effort only by adopting an informative signal. If the agent's effort is *observable* by the receiver, then this is no longer the case. The following example shows that transparency can completely resolve the incentive-provision problem.

**Example 1** *Consider a binary-state environment in which* $\underline{\mu} = 0$, $\overline{\mu} = 1$, *and* $v_A(\mu) = \mu$ *(as in Sections 4.2 and 4.4). If the agent's effort is observable by the receiver, then the agent's problem reduces to*

$$\max_e \mathbb{E}_\tau[v_A(\mu)] - c(e) = \mathbb{E}_\tau[\mu] - c(e) = e - c(e),$$

*whose unique optimal solution is given by* $\overline{e}$, *as defined in Proposition 1. Importantly, this holds for any signal structure; that is, the agent chooses* $\overline{e}$ *regardless of* $\pi$ *(or* $\tau$*).*

In the environment of Example 1, transparency is necessarily beneficial to the sender, as she can effectively take the (maximal) prior $\eta_e = \overline{e}$ as given and focuses on information provision (as in KG). Somewhat paradoxically, transparency–the receiver's observability of the agent's effort–can be harmful to the receiver. Not being concerned with incentive provision, the sender would adopt the optimal policy in KG. Therefore, if the sender has concave preferences (i.e., $v_S(\cdot)$ is concave) as in Section 4.2, then the sender would reveal no information. Although the agent's effort is higher, the receiver obtains less information about the state, which may make him worse off overall. Similarly, if the sender has discrete preferences (i.e., $v_S(\mu) = \mathcal{I}(\mu \geq \theta)$) as in Section 4.4, then she would choose a binary signal that induces only posteriors $0$ and $\theta$. Again, although the agent chooses the maximal effort $\overline{e}$, the receiver's payoff can be lower when he can observe the agent's effort than when he cannot.[26]

Outside of Example 1, even the effect of transparency on the agent's equilibrium effort is not clear-cut, as demonstrated in the following example.

**Example 2** *Suppose, as in Section 4.3, the sender and the agent have common concave preferences, and the sender also internalizes the agent's effort cost.*[27] *In particular, assume that* $\underline{\mu} = 0$, $\overline{\mu} = 1$, $v_S(\mu) = v_A(\mu) = 1 - (1 - \mu)^2$, *and* $c(e) = ce^2/2$ *for some* $c \in (2/3, 1)$.

[26] To be specific, suppose the sender has discrete preferences as in Section 4.4 and the receiver's payoff as a function of his belief $\mu$ is given by $\max\{\mu - \theta, 0\}$; for example, the receiver has a binary action and faces the usual trade-off between Type I and Type II errors. Then, the receiver's expected payoff is equal to 0 with transparency, while his expected payoff is strictly positive without transparency (because, as shown in Section 4.4, posterior belief $\mu = 1$ is realized with positive probability).

[27] To be precise, the sender's underlying utility function $u_S$ now depends on the agent's effort but not on the state, that is, $u_S(x, e) = u_A(x) - c(e)$. The assumption that the sender internalizes the agent's effort cost plays no role in the characterization of the optimal signal. However, it does affect the sender's optimal effort choice.

*If the agent's effort is unobservable then, by Proposition 4, the optimal signal induces posteriors* $0$ *and* $\mu_U$ *(because* $1/r(\mu) = 1 - \mu$ *in this case). Using the equilibrium conditions that* $\mathbb{E}_\tau[\mu] = e$ *and* $\mathbb{E}_\tau[h(\mu)] = 0$*, one can find that the sender's optimal signal is such that*

$$\mu_U = e + ce(1-e),\ \tau(\mu_U) = \frac{e}{e+ce(1-e)},\ \textit{and}\ V^e = (2 - e - ce(1-e))e - \frac{c}{2}e^2.$$

*From this explicit solution, it is immediate that the optimal effort is* $2/(3c)$*.*

*Now suppose the agent's effort is observable by the receiver. Since the agent and the sender have identical preferences, the problem reduces to:*

$$\max_{e,\tau} \mathbb{E}_\tau[v_A(\mu)] - c(e) \quad \textit{subject to} \quad \mathbb{E}_\tau[\mu] = e.$$

*Since* $v_A(\cdot)$ *is concave, given any* $e$*, it is optimal to reveal no information. Thus, the objective function can be further simplified to* $v_A(e) - c(e) = 1 - (1-\mu)^2 - ce^2/2$*, for which the equilibrium effort is given by*

$$v'(e) = 2 - 2e = c'(e) = ce \Rightarrow e = \frac{2}{2+c}.$$

*Importantly, this equilibrium effort* $2/(2+c)$ *exceeds* $2/(3c)$ *for any* $c \in (2/3, 1)$*.*

*In this example, transparency reduces both the agent's effort and the signal informativeness. Therefore, it is unambiguously harmful to the receiver.*

## 6.2 Moral Hazard and Signal Informativeness

Our motivating example in the introduction and Corollary 1 in Section 4 suggest that in the presence of moral hazard (i.e., if the agent must be incentivized to exert effort) the sender should employ a more informative signal than when the prior is exogenously given. Although intuitive, this is not generally true: recall that in our binary-action model of Section 5, a fully informative signal is always optimal if the prior is exogenously given, but it cannot be used (i.e., some information must be withheld) to implement $e \in (\underline{e}, \overline{e}]$. Still, the intuitive result–that moral hazard forces the sender to provide more information–does hold for a class of binary-state environments, as we demonstrate now.

We begin by defining the "Lagrangian Single Inflection" (LSI) property.

**Definition 1** *The payoff functions* $(v_S(\cdot), v_A(\cdot))$ *have the LSI-A (LSI-B) property if for any fixed* $\psi \geq 0$*,* $\mathcal{L}_{\mu\mu}(\cdot, \psi)$ *crosses zero at most once, and the crossing is from above (below).*

This definition formalizes the property of the Lagrangian that was exploited in Sections 4.2 and 4.3. As before, when it holds, the Lagrangian has at most one inflection point. Furthermore, LSI-A(B) guarantees that at the inflection point the switch is from convex to concave (concave to convex). Though this definition is formulated in terms of the Lagrangian, it is easy to check in a variety of settings beyond those already discussed.[28]

In general, the LSI property does not imply that the sender's payoff function is globally concave, so it is possible that $\underline{e} > 0$. Furthermore, except in the case where the sender's payoff is globally convex, $\underline{e} < \overline{e}$. Exploiting the logic underlying Propositions 3 and 4, we have the following generalization.

**Lemma 3** *Consider the binary-state model.*

*(i) If $(v_S(\cdot), v_A(\cdot))$ satisfy LSI-A, then the optimal distribution of posteriors that implements $e \in (\underline{e}, \overline{e})$ is supported on $\{0, \mu_U\}$ with $\tau^e(\mu_U) = e/\mu_U$, where $\mu_U \equiv e + (1-e)c'(e)$.*

*(ii) If $(v_S(\cdot), v_A(\cdot))$ satisfy LSI-B, then the optimal distribution of posteriors that implements $e \in (\underline{e}, \overline{e})$ is supported on $\{\mu_D, 1\}$ with $\tau^e(\mu_D) = (1-e)/(1-\mu_D)$, where $\mu_D \equiv e(1-c'(e))$.*

The following result shows that if the LSI property holds (whether $A$ or $B$), the sender necessarily adopts a more informative signal in our model than she would do in KG.

**Proposition 8** *In the binary-state model, if the LSI property holds, then the optimal signal associated with target effort $e \in (\underline{e}, \overline{e})$ is more informative in the Blackwell sense than the optimal signal in the benchmark model with exogenous prior $\eta_e$.*

**Proof.** See the appendix. ■

Intuitively, for $e > \underline{e}$, the optimal signal in the benchmark with an exogenous prior violates incentive compatibility, and thus, to incentivize effort the sender must adjust the signal in some way. With binary states and LSI, the set of adjustments that could be optimal is limited—it is only ever optimal for the sender to garble information about one of the states. The essence of the proof is to show that the adjustment required by incentive compatibility always results in an increase in signal informativeness.

# 7 Conclusion

In this paper, we studied Bayesian persuasion when the prior about the underlying state is generated by an agent's unobservable effort, so the sender is concerned with both information provision and

[28]For example, the LSI property holds in the following environments: (i) If $S$ (the sender) and $A$ (the agent) have identical IARA preferences, then LSI-A holds. (ii) If $v_i(\mu) = (1-\exp(-\rho_i\mu)/(1-\exp(-\rho_i))$ (CARA) for both $i = S, A$ and $\rho_A \geq \rho_S > 0$, then LSI-A holds. (iii) If $v_i(\mu) = \mu^{1-\rho_i}$ (CRRA) for both $i = S, A$ and $\rho_A \leq \rho_S < 1$, then LSI-B holds. (iv) Suppose for both $i = S, A$, $v_i(\mu) = \frac{(\mu+\underline{\mu})^{1-\rho_i} - \underline{\mu}^{1-\rho_i}}{(1+\underline{\mu})^{1-\rho_i} - \underline{\mu}^{1-\rho_i}}$, where $\underline{\mu} > 0$ (HARA). If $\rho_A > 2$ and $\rho_S \leq \rho_A$, then LSI-A holds. If $\rho_A < 2$ and $\rho_S \geq \rho_A$, then LSI-B holds.

incentive provision. We showed that the sender's problem can be analyzed by extending the concavification technique in Aumann and Maschler (1995) and KG, and provided a useful necessary and sufficient condition for an optimal signal (Section 3). We applied the general characterization to two tractable environments, one with binary states (Section 4) and the other with binary receiver actions (Section 5), and derived a number of concrete and economic implications. We also showed that transparency, which allows the receiver to observe the agent's effort, may reduce the informativeness of the equilibrium signal and harm the receiver, and provided a sufficient condition under which moral hazard forces the sender to provide more information than she would do otherwise (Section 6).

# Appendix: Omitted Proofs

**Continuation of Proof of Theorem 1.** For the result on the cardinality of the support of $\tau^e$, we present a slightly different but equivalent construction. Note that the dimension of (BP) can be reduced by 1. In particular, since $\mu, \eta_e \in \Delta(\Omega)$, if (BP) is satisfied for any $N-1$ components, then it is also satisfied for the remaining component. Therefore, consider the following curve in $\mathcal{R}^{N+1}$:

$$\underline{K}^e \equiv \{(\mu(2), ..., \mu(N), h(\mu), v_S(\mu)) : \mu \in \Delta(\Omega)\},$$

and let $co(\underline{K}^e)$ denote its convex hull. As in the text, $y \in co(\underline{K}^e)$ if and only if there exists a distribution of posteriors $\tau$ such that $y = (\mathbb{E}_\tau[\mu(2)], ..., \mathbb{E}_\tau[\mu(N)], \mathbb{E}_\tau[h(\mu)], \mathbb{E}_\tau[v_S(\mu)])$. $V^e = \max_v\{v : (\eta_e(2), ..., \eta_e(N), 0, v) \in co(\underline{K}^e)\}$. Because the graphs of $h(\cdot)$ and $v_S(\cdot)$ are closed, $\underline{K}^e$ is closed, and hence, $co(\underline{K}^e)$ is closed. Thus, $(\eta_e(2), ..., \eta_e(N), 0, V^e)$ is on the boundary of $co(\underline{K}^e)$. By Caratheodry's theorem, any vector that belongs to the boundary of the convex hull of a set $Y$ in $\mathcal{R}^{N+1}$ can be written as a convex combination of no more than $N+1$ elements of $Y$. Hence, $(\eta_e(2), ..., \eta_e(N), 0, V^e)$ can be made of at most $N+1$ elements in $\underline{K}^e$. ■

**Continuation of Proof of Theorem 2.** *Necessity.* Given the partial proof in the main text, it suffices to show that $\mathcal{L}(\mu, \psi) = \lambda_0 + \langle \lambda_1, \mu \rangle$ for all $\mu$ such that $\tau^e(\mu) > 0$. Suppose that $\mathcal{L}(\mu', \psi) < \lambda_0 + \langle \lambda_1, \mu' \rangle$ for some $\mu'$ such that $\tau^e(\mu') > 0$. Because $\mathcal{L}(\mu, \psi) \le \lambda_0 + \langle \lambda_1, \mu \rangle$ for all $\mu \in \Delta(\Omega)$, it follows that $\mathbb{E}_{\tau^e}[\mathcal{L}(\mu, \psi)] < \lambda_0 + \langle \lambda_1, \mathbb{E}_{\tau^e}[\mu] \rangle = \lambda_0 + \langle \lambda_1, \eta_e \rangle$. Using (IC) on the left hand side, $V^e < \lambda_0 + \langle \lambda_1, \eta_e \rangle$. Rearranging the terms, $\langle d, (\eta_e, 0, V^e) \rangle < \lambda_0$, which contradicts $\langle d, (\eta_e, 0, V^e) \rangle = \lambda_0$.

*Sufficiency.* If $v_S(\mu) + \psi h(\mu) \le \lambda_0 + \langle \lambda_1, \mu \rangle$ for all $\mu \in \Delta(\Omega)$, then for any $\tau$,

$$\mathbb{E}_\tau[v_S(\mu)] + \psi \mathbb{E}_\tau[h(\mu)] \le \lambda_0 + \mathbb{E}_\tau[\langle \lambda_1, \mu \rangle].$$

If $\tau$ satisfies both (BP) and (IC), then $\mathbb{E}_\tau[v_S(\mu)] \leq \lambda_0 + \langle \lambda_1, \eta_e \rangle$. If $\tau^e$ is such that $\mathcal{L}(\mu, \psi) = \lambda_0 + \langle \lambda_1, \mu \rangle$ for any $\tau^e(\mu) > 0$, then $\mathbb{E}_{\tau^e}[v_S(\mu)] = \lambda_0 + \langle \lambda_1, \eta_e \rangle$, that is, $\tau^e$ achieves the upper bound of the sender's expected utility. Thus, it is optimal. ■

**Proof of Proposition 1.** We first show that $\overline{e}$ is the upper bound to the set of implementable effort levels. Under any signal $\pi$, the agent chooses $e$ to maximize

$$(e(1-\overline{\mu}) + (1-e)(1-\underline{\mu}))\mathbb{E}[v_A(\mu)|\omega = 1] + (e\overline{\mu} + (1-e)\underline{\mu})\mathbb{E}[v_A(\mu)|\omega = 2] - c(e).$$

The first two terms are linear in $e$, while $c(e)$ is strictly convex. Therefore, the optimal effort level is determined by

$$(\overline{\mu} - \underline{\mu})(\mathbb{E}[v_A(\mu)|\omega = 2] - \mathbb{E}[v_A(\mu)|\omega = 1]) = (\overline{\mu} - \underline{\mu}) \sum_s (\pi_2(s) - \pi_1(s)) v_A(\mu_s) = c'(e).$$

Since $v_A$ is weakly increasing,

$$(\overline{\mu} - \underline{\mu}) \left( \sum_s \pi_2(s) v_A(\mu_s) - \sum_s \pi_1(s) v_A(\mu_s) \right) \leq (\overline{\mu} - \underline{\mu}) \left( v_A(1) - v_A(0) \right) = \overline{\mu} - \underline{\mu}.$$

It follows that $e$ such that $c'(e) > \overline{\mu} - \underline{\mu}$, which is equivalent to any $e > \overline{e}$, is not implementable.

Fix $e \in [0, \overline{e}]$, and consider the following distribution of posteriors, which can be interpreted as a convex combination of a fully informative signal and a fully noisy signal:

$$\tau(0) = (1 - \eta_e)\frac{c'(e)}{\overline{\mu} - \underline{\mu}}, \tau(\eta_e) = 1 - \frac{c'(e)}{\overline{\mu} - \underline{\mu}}, \tau(1) = \eta_e \frac{c'(e)}{\overline{\mu} - \underline{\mu}}.$$

This distribution is well-defined, because $\eta_e \in [\underline{\mu}, \overline{\mu}]$ and $c'(e) \leq c'(\overline{e}) < \overline{\mu} - \underline{\mu}$. It is easy to verify that this distribution satisfies both (BP) and (IC) and, therefore, $e$ is implementable.

The final result that $\overline{e}$ is implementable if and only if the signal is fully informative follows from the argument given in the main text. ■

**Proof of Corollary 1.** The first (convex) result is immediate from the fact that $\underline{e} = \overline{e}$ if $v_S$ is convex. For the second (concave) result, notice that if $v_S$ is concave then $\underline{e} = 0$, so if the sender adopts an uninformative signal then the sender's expected utility is equal to $v_S(\underline{\mu})$. Consider an alternative signal that induces either $\underline{\mu}$ or 1. Under the assumption that $v_A(\underline{\mu}) < v_A(1)$, the agent chooses a strictly positive effort level. Therefore, the sender's expected payoff satisfies

$$\mathbb{E}_\tau[v_S(\mu)] = \tau(\underline{\mu}) v_S(\underline{\mu}) + \tau(1) v_S(1) > v_S(\underline{\mu}).$$

This proves that an uninformative signal is never optimal for the sender. ■

**Proof of Proposition 2.** The result that $V^{\underline{e}} \geq V^e$ for any $e \leq \underline{e}$ is immediate from the fact that $\widehat{V}^e$ is increasing in $e$ (due to monotone $v_S(\mu)$), $V^e \leq \widehat{V}^e$ for any $e$, and $V^{\underline{e}} = \widehat{V}^{\underline{e}}$.

We now prove that at the optimal solution, $\psi > 0$ whenever $e > \underline{e}$. For $e \in (\underline{e}, \overline{e})$, let $\widehat{\tau}$ denote the solution to the sender's relaxed problem and $\widehat{V}^e$ denote the corresponding seller payoff. Applying Theorem 2 to the relaxed problem (without (IC) and with $\psi = 0$), there exist $\widehat{\lambda}_0$ and $\widehat{\lambda}_1$ such that

$$v_S(\mu) \leq \widehat{\lambda}_0 + \widehat{\lambda}_1\mu \text{ for all } \mu \in [0,1], \text{ with equality if } \widehat{\tau}(\mu) > 0. \tag{4}$$

Similarly, let $\tau$ denote the solution to the sender's original (unrelaxed) problem and apply Theorem 2. Then, there exist $\lambda_0$, $\lambda_1$, and $\psi$ such that

$$v_S(\mu) + \psi h(\mu) \leq \lambda_0 + \lambda_1\mu \text{ for all } \mu \in [0,1], \text{ with equality if } \tau(\mu) > 0. \tag{5}$$

Note that, whereas $\widehat{\tau}(\mu)$ satisfies only (BP), $\tau(\mu)$ satisfies both (BP) and (IC).

**Step 1:** We show that if $\mathbb{E}_{\widehat{\tau}}[h(\mu)] < 0$ then $\psi > 0$.

Taking the expectation of (4) with respect to $\tau$, we get

$$\mathbb{E}_\tau[v_S(\mu)] \leq \widehat{\lambda}_0 + \widehat{\lambda}_1\mathbb{E}_\tau[\mu] = \widehat{\lambda}_0 + \widehat{\lambda}_1 e.$$

Doing the same with (5) and using the fact that $\tau$ is a solution, we get

$$\mathbb{E}_\tau[v_S(\mu)] + \psi\mathbb{E}_\tau[h(\mu)] = \lambda_0 + \lambda_1\mathbb{E}_\tau[\mu] = \lambda_0 + \lambda_1 e.$$

Since $\tau$ satisfies (IC), $\mathbb{E}_\tau[h(\mu)] = 0$. Then, it follows

$$\lambda_0 + \lambda_1 e = \mathbb{E}_\tau[v_S(\mu)] \leq \widehat{\lambda}_0 + \widehat{\lambda}_1 e.$$

Similarly, taking the expectations of (4) and (5) with respect to $\widehat{\tau}$, we obtain

$$\mathbb{E}_{\widehat{\tau}}[v_S(\mu)] = \widehat{\lambda}_0 + \widehat{\lambda}_1 e \text{ and } \mathbb{E}_{\widehat{\tau}}[u_S(\mu)] + \psi\mathbb{E}_{\widehat{\tau}}[h(\mu)] \leq \lambda_0 + \lambda_1 e,$$

leading to

$$\widehat{\lambda}_0 + \widehat{\lambda}_1 e + \psi\mathbb{E}_{\widehat{\tau}}[h(\mu)] = \mathbb{E}_{\widehat{\tau}}[u_S(\mu)] + \psi\mathbb{E}_{\widehat{\tau}}[h(\mu)] \leq \lambda_0 + \lambda_1 e.$$

Combining the two derived inequalities, it follows that

$$\psi\mathbb{E}_{\widehat{\tau}}[h(\mu)] \leq (\lambda_0 + \lambda_1 e) - \left(\widehat{\lambda}_0 + \widehat{\lambda}_1 e\right) \leq 0.$$

Clearly, $\mathbb{E}_{\widehat{\tau}}[h(\mu)] < 0$ implies $\psi \geq 0$. Strict inequality follows from the fact that if $\psi = 0$ then $\tau = \widehat{\tau}$ and, therefore, $\widehat{V}^e = V^e$, which contradicts $e > \underline{e}$.

**Step 2:** We now show that if $e \in (\underline{e}, \overline{e})$ then $\mathbb{E}_{\widehat{\tau}}[h(\mu)] < 0$.

If $\mathbb{E}_{\widehat{\tau}}[h(\mu)] = 0$, then $\widehat{\tau} = \tau$, which cannot be the case for $e \in (\underline{e}, \overline{e})$. Therefore, it suffices to show that it cannot be that $\mathbb{E}_{\widehat{\tau}}[h(\mu)] > 0$. Toward a contradiction, suppose $\mathbb{E}_{\widehat{\tau}}[h(\mu)] > 0$.

*Case 1:* Suppose that there exists another solution $\widehat{\tau}'$ such that $\mathbb{E}_{\widehat{\tau}'}[h(\mu)] < 0$. Because (BP) is convex, for any $w \in [0, 1]$ the posterior distribution $\tau^w = w\widehat{\tau} + (1-w)\widehat{\tau}'$ is feasible. Furthermore, $\mathbb{E}_{\tau^w}[v_S(\mu)] = w\mathbb{E}_{\widehat{\tau}}[v_S(\mu)] + (1-w)\mathbb{E}_{\widehat{\tau}'}[v_S(\mu)] = \widehat{V}^e$, which means that $\tau^w$ is also optimal in the relaxed problem. Next, note that $\mathbb{E}_{\tau^w}[h(\mu)] = w\mathbb{E}_{\widehat{\tau}}[h(\mu)] + (1-w)\mathbb{E}_{\widehat{\tau}'}[h(\mu)]$, and hence, there exists some $w^* \in (0, 1)$ such that $\mathbb{E}_{\tau^{w^*}}[h(\mu)] = 0$, that is, (IC) is satisfied. This implies that $\tau^{w^*}$ is an optimal solution to the sender's original problem, which contradicts $e > \underline{e}$.

*Case 2:* Suppose that for all solution(s) of the relaxed problem, $\mathbb{E}_{\widehat{\tau}}[h(\mu)] > 0$. Note that that any solution to the relaxed problem must induce at least two posteriors: if a signal is degenerate, then it must put all mass on $\mu = e$ and, therefore, $\mathbb{E}_{\widehat{\tau}}[h(\mu)] = -c'(e) < 0$. In addition, there must exist $\mu_L < e < \mu_H$ such that $\widehat{\tau}(\mu_L) > 0$ and $\widehat{\tau}(\mu_H) > 0$: otherwise, (BP) cannot hold. Now, applying Theorem 2 to the relaxed problem, we have

$$\widehat{\lambda}_0 + \widehat{\lambda}_1 \mu \geq v_S(\mu) \text{ for all } \mu, \text{ with equality if } \mu = \mu_L, \mu_H. \tag{6}$$

For each $\widetilde{e} \in [e, \mu_H]$, consider the following distribution, which clearly satisfies (BP):

$$\widetilde{\tau}(\mu_H) = \frac{\widetilde{e} - \mu_L}{\mu_H - \mu_L} \text{ and } \widetilde{\tau}(\mu_L) = \frac{\mu_H - \widetilde{e}}{\mu_H - \mu_L}.$$

Since (6) applies regardless of $\tilde{e}$, $\tilde{\tau}$ is an optimal solution to the sender's relaxed problem. Next, observe that with $\widetilde{e}$,

$$\mathbb{E}_{\widehat{\tau}}[h(\mu)] = \frac{1}{\widetilde{e}(1-\widetilde{e})}\Big(\frac{\mu_H - \widetilde{e}}{\mu_H - \mu_L}(\mu_L - \widetilde{e})v_A(\mu_L) + \frac{\widetilde{e} - \mu_L}{\mu_H - \mu_L}(\mu_H - \widetilde{e})v_A(\mu_H)\Big) - c'(\widetilde{e}).$$

By assumption, with any solution to the relaxed problem, $\mathbb{E}_{\widehat{\tau}}[h(\mu)] > 0$. To the contrary, it is clear from direct substitutions that if $\widetilde{e} = \mu_H$ then $\mathbb{E}_{\widehat{\tau}}[h(\mu)] < 0$. Since $\mathbb{E}_{\widehat{\tau}}[h(\mu)]$ continuously changes in $\widetilde{e}$, there must exist $e^* \in (e, \mu_H)$ such that the solution to the relaxed problem satisfies (IC), that is, $\widetilde{\tau}$ is a solution to the sender's original problem. This implies that $V(e^*) = \widehat{V}(e^*)$, which contradicts $e^* > e > \underline{e}$. ■

**Proof of Proposition 6.** We prove Proposition 6 in three steps.

**Step 1:** $e > \overline{e}$ cannot be implemented.

Fix any signal structure $\pi$. Let $\Sigma^+$ denote the subset of $\Sigma$ such that in the equilibrium of the subgame between the agent and the receiver, $\mu_s \in \mathcal{X}_1$ if and only if $s \in \Sigma^+$; that is, $\Sigma^+$ includes all signal realizations that induce the receiver to take action $x_1$ given $\pi$ and the agent's equilibrium effort $e$. In a slight abuse of notation, let $\pi_\omega(\Sigma')$ denote the probability that the signal realization belongs to $\Sigma'$ (i.e., $\pi_\omega(\Sigma') = \sum_{s\in\Sigma^+} \pi_\omega(s)$). Then, the agent's expected payoff of exerting effort $e$ is given by

$$\sum_{\omega} \eta_e(\omega)\pi_\omega(\Sigma') - c(e) = \sum_{\omega}(e\overline{\mu}(\omega) + (1-e)\underline{\mu}(\omega))\pi_\omega(\Sigma') - c(e).$$

Therefore, the agent's (interior) optimal effort satisfies

$$\sum_{\omega}(\overline{\mu}(\omega) - \underline{\mu}(\omega))\pi_\omega(\Sigma') = c'(e).$$

Since $\pi_\omega(\Sigma') \in [0, 1]$ for all $\omega \in \Omega$, the left-hand side is maximized when $\pi_\omega(\Sigma') = 0$ if $\overline{\mu}(\omega) < \underline{\mu}(\omega)$ and $\pi_\omega(\Sigma') = 1$ if $\overline{\mu}(\omega) > \underline{\mu}(\omega)$. Since $\overline{\mu}(\omega) \geq \underline{\mu}(\omega)$ if and only if $\omega > \omega_e$ (by Assumption 1 and the definition of $\omega_e$), we have

$$c'(e) = \sum_{\omega}(\overline{\mu}(\omega) - \underline{\mu}(\omega))\pi_\omega(\Sigma') \leq \sum_{\omega>\omega_e}(\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(\overline{e}).$$

Since $c'(\cdot)$ is strictly increasing, it is necessarily the case that $e \leq \overline{e}$.

**Step 2:** Any $e \in (\underline{e}, \overline{e}]$ is implementable.

Consider the following class of binary signals: $\Sigma = \{b, g\}$ and for some $k \in \{\omega_e + 1, ..., \omega_r\}$ and $s \in (0, 1]$,

$$\pi_\omega^{k,s}(g) = \begin{cases} 0, & \text{if } \omega < k, \\ s, & \text{if } \omega = k, \\ 1, & \text{if } \omega > k, \end{cases} \quad \text{and } \pi_\omega^{k,s}(b) = 1 - \pi_\omega^{k,s}(g) \text{ for all } \omega \in \Omega.$$

By the structure of the binary signal, we have

$$\mu_b = \frac{(\eta_e(1), ..., \eta_e(k-1), (1-s)\eta_e(k), 0, ..., 0)}{\sum_{\omega<k} \eta_e(\omega) + (1-s)\eta_e(k)} \text{ and } \mu_g = \frac{(0, ..., 0, s\eta_e(k), \eta_e(k+1), ..., \eta_e(N))}{s\eta_e(k) + \sum_{\omega>k} \eta_e(\omega)}.$$

We first show that $\mu_b \in \mathcal{X}_0$ and $\mu_g \in \mathcal{X}_1$. The former is immediate from the fact that $\mu_b$ assigns positive values only to rejection states ($\omega \leq k \leq \omega_r$), so $\mathbb{E}_{\mu_b}[v_\omega] \leq v_{\omega_r} < \theta$ regardless of $e$. To show $\mu_g \in \mathcal{X}_1$, observe that Assumption 1 implies that $\mu_g$ increases in $e$ in the sense of first-order stochastic dominance. Therefore, for any $e > 0$, $\mu_g$ stochastically dominates

$$\mu_g^0 \equiv \frac{(0, ..., 0, s\eta_0(k), \eta_0(k+1), ..., \eta_0(N))}{s\eta_0(k) + \sum_{\omega>k} \eta_0(\omega)} = \frac{\big(0, ..., 0, s\underline{\mu}(k), \underline{\mu}(k+1), ..., \underline{\mu}(N)\big)}{s\underline{\mu}(k) + \sum_{\omega>k} \underline{\mu}(\omega)}.$$

In turn, because $k \geq \omega_e + 1$ and $s \in [0, 1)$, $\mu_g^0$ stochastically dominates

$$\underline{\mu}_{\omega_e} \equiv \frac{\big(0, ..., 0, \underline{\mu}(\omega_e + 1), ..., \underline{\mu}(N)\big)}{\sum_{\omega>\omega_e} \underline{\mu}(\omega)}.$$

By Assumption 3, $\mathbb{E}_{\underline{\mu}_{\omega_e}}[v_\omega] > \theta$. Since $\mu_g$ stochastically dominates $\underline{\mu}_{\omega_e}$ and $v_\omega$ is increasing in $\omega$, it follows that $\mathbb{E}_{\mu_g}[v_\omega] > \theta$, so $\mu_g \in \mathcal{X}_1$.

Given the signal structure $\pi^{k,s}$, $\mu_b \in \mathcal{X}_0$, and $\mu_g \in \mathcal{X}_1$, the agent's problem reduces to

$$\max_e s\eta_e(k) + \sum_{\omega>k} \eta_e(\omega) - c(e).$$

Since $\eta_e(\omega) = e\overline{\mu}(\omega) + (1 - e)\underline{\mu}(\omega)$ for all $\omega \in \Omega$, the agent's optimal effort $\overline{e}$ satisfies

$$s(\overline{\mu}(k) - \underline{\mu}(k)) + \sum_{\omega>k} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(\overline{e}).$$

The left-hand side continuously decreases from $\sum_{\omega>\omega_e}(\overline{\mu}(\omega) - \underline{\mu}(\omega))$ to $\sum_{\omega>\omega_r}(\overline{\mu}(\omega) - \underline{\mu}(\omega))$ as $s$ decreases from 1 to 0 and $k$ increases by 1 (and $s$ jumps up to 1) once $s$ reaches 0. Since $c'(\cdot)$ is strictly increasing and continuous, this means that the equilibrium effort $e$ continuously decreases from $\overline{e}$ to $\underline{e}$.

**Step 3:** any $e \in [0, \underline{e}]$ is implementable.

Consider the following class of signal structures: for $\alpha \in [0, 1]$, $\Sigma = \{b, \omega_r + 1, ..., N\}$,

$$\pi_\omega^\alpha(b) = \begin{cases} 1, & \text{if } \omega \leq \omega_r, \\ 1 - \alpha, & \text{if } \omega > \omega_r, \end{cases} \quad \text{and } \pi_\omega^\alpha(\omega') = \begin{cases} \alpha, & \text{if } \omega' = \omega, \\ 0, & \text{if } \omega' \neq \omega. \end{cases}$$

In other words, $\pi^\alpha$ only reveals acceptance states ($\omega > \omega_r$) with probability $\alpha$. By construction, $\pi^\alpha$ implements 0 if $\alpha = 0$ (as the signal is completely uninformative) and $\underline{e}$ if $\alpha = 1$. If $s \neq b$ then $\mu_s$ is the degenerate distribution at $\omega > \omega_r$, so $\mu_s \in \mathcal{X}_1$.

Now we show that $\mu_b \notin \mathcal{X}_1$. Suppose $\mu_b \in \mathcal{X}_1$. Then, the receiver always takes $x_1$, which eliminates the agent's incentive to exert effort. In that case, $\eta_e = \underline{\mu}$ and

$$\mu_b = \frac{\big(\underline{\mu}(1), ..., \underline{\mu}(\omega_r), (1 - \alpha)\underline{\mu}(\omega_r + 1), ..., (1 - \alpha)\underline{\mu}(N)\big)}{\sum_{\omega\leq\omega_r} \underline{\mu}(\omega) + (1 - \alpha) \sum_{\omega>\omega_r} \underline{\mu}(\omega)}$$

By its structure, $\mu_b$ is stochastically dominated by $\underline{\mu}$. Therefore, $\mathbb{E}_{\mu_b}[v_\omega] < \mathbb{E}_{\underline{\mu}}[v_\omega] < \theta$, contradicting $\mu_b \in \mathcal{X}_1$.

Since $\mu_b \notin \mathcal{X}_1$, the agent's problem is given by

$$\max_e \alpha \sum_{\omega > \omega_r} \eta_e(\omega) - c(e) = \alpha \sum_{\omega > \omega_r} \left( e\overline{\mu}(\omega) + (1-e)\underline{\mu}(\omega) \right) - c(e).$$

Therefore, the agent's optimal effort $e$ satisfies

$$\alpha \sum_{\omega > \omega_r} \left( \overline{\mu}(\omega) - \underline{\mu}(\omega) \right) = c'(e). \tag{7}$$

Since $c'(\cdot)$ is strictly increasing and continuous, $e$ continuously increases from $0$ to $\underline{e}$ as $\alpha$ rises from $0$ to $1$.

It remains to show that for each $\alpha$, $\mu_b$ is indeed well-defined in $\mathcal{X}_0$. Note that

$$\mu_b = \beta \eta_e + (1-\beta) \left( \eta_e(1), ..., \eta_e(\omega_r), 0, ..., 0 \right),$$

where

$$\beta = \frac{\alpha}{\sum_{\omega \le \omega_r} \eta_e(\omega) + (1-\alpha) \sum_{\omega > \omega_r} \eta_e(\omega)} \in [0,1].$$

Both $\eta_e$ and $(\eta_e(1), ..., \eta_e(\omega_r), 0, ..., 0)$ are stochastically dominated by $\overline{\mu}$ and, therefore, belong to $\mathcal{X}_0$. It then follows that $\mu_b$ (their convex combination) also belongs to $\mathcal{X}_0$.

■

**Proof of Proposition 7.** Notice that the proposed signal is identical to that used in Step 2 in the proof of Proposition 6. There, we already proved that $\mu_b \in \mathcal{X}_0$ and $\mu_s \in \mathcal{X}_1$. The induced distribution of posteriors triviality satisfies (BP). To show that it also satisfies (IC), recall that the agent's equilibrium effort $e$ is characterized by

$$s(\overline{\mu}(k) - \underline{\mu}(k)) + \sum_{\omega > k} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(\overline{e}).$$

By definition, $\sum_{\omega > k} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(e_k)$. Therefore, (IC) holds when

$$s(\overline{\mu}(k) - \underline{\mu}(k)) + c'(e_k) = c'(\overline{e}) \iff s = \frac{c'(e) - c'(e_k)}{\overline{\mu}(k) - \underline{\mu}(k)}.$$

The following claim is useful in what follows.

**Claim 1** *For any $k \in \{\omega_e+1, ..., \omega_r\}$ and $\omega \in \{1, ..., N\}$, we have*

$$\omega \gtreqless k \Rightarrow v_\omega - \theta \gtreqless (v_k - \theta)\frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)}\frac{\eta_e(k)}{\overline{\mu}(k) - \underline{\mu}(k)}.$$

**Proof.** Fix $k \in \{\omega_e+1, ..., \omega_r\}$. Note that since $\omega_e < k \leq \omega_r$, we have $\overline{\mu}(k) > \underline{\mu}(k)$ and $v_k < \theta$.

Consider $\omega > k$. Since $k > \omega_e$, $\overline{\mu}(\omega) > \underline{\mu}(\omega)$, so the desired inequality can be rewritten as

$$\begin{aligned}
& (v_\omega - \theta)\frac{\eta_e(\omega)}{\overline{\mu}(\omega) - \underline{\mu}(\omega)} > (v_k - \theta)\frac{\eta_e(k)}{\overline{\mu}(k) - \underline{\mu}(k)} \\
\iff \quad & (\theta - v_\omega)\frac{e(\overline{\mu}(\omega) - \underline{\mu}(\omega)) + \underline{\mu}(\omega)}{\overline{\mu}(\omega) - \underline{\mu}(\omega)} < (\theta - v_k)\frac{e(\overline{\mu}(k) - \underline{\mu}(k)) + \underline{\mu}(k)}{\overline{\mu}(\omega) - \underline{\mu}(\omega)} \\
\iff \quad & (\theta - v_\omega)\left(e + \frac{1}{\frac{\overline{\mu}(\omega)}{\underline{\mu}(\omega)} - 1}\right) < (\theta - v_k)\left(e + \frac{1}{\frac{\overline{\mu}(k)}{\underline{\mu}(k)} - 1}\right).
\end{aligned}$$

This holds for any $\omega > k$, because $v_\omega > v_k$ and $\overline{\mu}(\omega)/\underline{\mu}(\omega) > \overline{\mu}(k)/\underline{\mu}(k)$ (Assumption 1).

Now consider $\omega_e < \omega < k$. In this case, it suffices to reverse the inequalities in the previous case; that is, the desired result can be rewritten as

$$(\theta - v_\omega)\left(e + \frac{1}{\frac{\overline{\mu}(\omega)}{\underline{\mu}(\omega)} - 1}\right) > (\theta - v_k)\left(e + \frac{1}{\frac{\overline{\mu}(k)}{\underline{\mu}(k)} - 1}\right).$$

As in the previous case, this holds for any $\omega_e < \omega < k$, because $v_\omega < v_k$ and $\overline{\mu}(\omega)/\underline{\mu}(\omega) < \overline{\mu}(k)/\underline{\mu}(k)$.

Finally, consider $\omega \leq \omega_e < k$. In this case, $\overline{\mu}(\omega) \leq \underline{\mu}(\omega)$, so

$$v_\omega - \theta < 0 \leq (v_k - \theta)\frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)}\frac{\eta_e(k)}{\overline{\mu}(k) - \underline{\mu}(k)}.$$

■

We now verify the optimality of the proposed by applying Theorem 2 in three steps.

**Step 1:** constructing the hyperplane that meets $\mathcal{L}(\mu, \psi)$ at $\mu_b$ and $\mu_g$.

Since $v_S(\mu) = \max\{\mathbb{E}_\mu[v_\omega], \theta\}$, we have

$$\mathcal{L}(\mu, \psi) = \begin{cases} \theta - \psi c'(e), & \text{if } \mu \in \mathcal{X}_0, \\ \mathbb{E}_\mu[v_\omega] + \psi\left(\mathbb{E}_\mu\left[\frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)}\right] - c'(e)\right), & \text{if } \mu \in \mathcal{X}_1. \end{cases}$$

Consider the following multipliers:

$$\psi = -(v_k - \theta)\frac{\eta_e(k)}{\overline{\mu}(k) - \underline{\mu}(k)},\ \lambda_0 = -\psi c'(e),\ \text{and}\ \lambda_1(\omega) = \begin{cases} \theta, & \text{if } \omega \le k, \\ v_\omega + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)}, & \text{if } \omega > k. \end{cases}$$

Then, we have

$$\begin{aligned}
\lambda_0 + \langle \lambda_1, \mu \rangle &= -\psi c'(e) + \sum_{\omega \le k} \mu(\omega)\theta + \sum_{\omega > k} \mu(\omega)\left( v_\omega + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)} \right) \\
&= \theta - \psi c'(e) + \sum_{\omega > k} \mu(\omega)\left( v_\omega - \theta + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)} \right) \\
&= \theta - \psi c'(e) + \sum_{\omega > k} \mu(\omega)\left( v_\omega - \theta - (v_k - \theta)\frac{\eta_e(k)}{\overline{\mu}(k) - \underline{\mu}(k)} \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)} \right)
\end{aligned}$$

**Step 2:** $\lambda_0 + \langle \lambda_1, \mu \rangle \ge \mathcal{L}(\mu, \psi)$ for all $\mu \in \mathcal{X}_0$, with equality holding if $\mu = \mu_b$.

If $\mu \in \mathcal{X}_0$ then $\mathcal{L}(\mu, \psi) = \theta - \psi c'(e)$. Therefore,

$$\lambda_0 + \langle \lambda_1, \mu \rangle \ge \mathcal{L}(\mu, \psi) \iff \sum_{\omega > k} \mu(\omega)\left( v_\omega - \theta - (v_k - \theta)\frac{\eta_e(k)}{\overline{\mu}(k) - \underline{\mu}(k)} \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)} \right) \ge 0.$$

By Claim 1, this latter inequality always holds. If $\mu = \mu_b$ then it holds with equality, because $\mu_b$ assigns 0 to all states strictly above $k$ (i.e., $\mu_b(\omega) = 0$ for all $\omega > k$).

**Step 3:** $\lambda_0 + \langle \lambda_1, \mu \rangle \ge \mathcal{L}(\mu, \psi)$ for all $\mu \in \mathcal{X}_1$, with equality holding if $\mu = \mu_g$.

For $\mu \in \mathcal{X}_1$, we have

$$\begin{aligned}
\mathcal{L}(\mu, \psi) &= \sum_{\omega} \mu(\omega) v_\omega + \sum_{\omega} \mu(\omega)\psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)} - \psi c'(e) \\
&= \theta - \psi c'(e) + \sum_{\omega} \mu(\omega)\left( v_\omega - \theta - (v_k - \theta)\frac{\eta_e(k)}{\overline{\mu}(k) - \underline{\mu}(k)} \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)} \right).
\end{aligned}$$

Therefore, $\lambda_0 + \langle \lambda_1, \mu \rangle \ge \mathcal{L}(\mu, \psi)$ is equivalent to

$$\sum_{\omega < k} \mu(\omega)\left( v_\omega - \theta - (v_k - \theta)\frac{\eta_e(k)}{\overline{\mu}(k) - \underline{\mu}(k)} \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta_e(\omega)} \right) \le 0.$$

Again, this inequality follows from Claim 1. As in Step 2, if $\mu = \mu_g$ then this holds with equality, because $\mu_g$ assigns 0 to all states strictly below $k$ (i.e., $\mu_g(\omega) = 0$ for all $\omega < k$). ■

**Proof of Proposition 8.** In the proof of Proposition 2 (Step 2), we showed that if $e \in (\underline{e}, \overline{e})$, then in the solution of the relaxed problem where (IC) is ignored, $\mathbb{E}_\tau[h(\mu)] < 0$. Here we show that restoring incentive compatibility requires increasing $\mu_U$ in the case of LSI-A, and reducing $\mu_D$ in the case of LSI-B. By implication, the distribution of posteriors in the sender's problem with incentive compatibility is a mean preserving spread of the distribution of the posteriors in the relaxed problem, and it is therefore more informative in the Blackwell sense.

*LSI-A*. If the underlying environment has the LSI-A property, then the sender's optimal distribution of posteriors in the relaxed problem is $\{0, \hat{\mu}\}$ with $\tau(\hat{\mu}) = \frac{e}{\hat{\mu}}$, and the solution in the full problem is $\{0, \mu^*\}$ with $\tau(\mu^*) = \frac{e}{\mu^*}$. Per Step 2 of Proposition 2, we know that the solution of the relaxed problem generates a negative value of the IC constraint, i.e.

$$\mathbb{E}_\tau[h(\mu)] = \frac{e}{\hat{\mu}}\Big(\frac{\hat{\mu} - e}{e(1-e)} v_A(\hat{\mu})\Big) < c'(e).$$

Next, note that for $\mu \geq e$

$$\frac{e}{\mu}\Big(\frac{\mu - e}{e(1-e)} v_A(\mu)\Big) = \frac{1}{1-e}\left(1 - \frac{e}{\mu}\right) v_A(\mu)$$

is a product of two positive, increasing functions of $\mu$, and it is therefore increasing. Furthermore, the solution of the sender's full problem satisfies IC. Combining these observations, it follows that $\mu^* > \hat{\mu}$. Because both the solution of the relaxed problem and the sender's full problem have the same mean, and are supported on $\{0, \hat{\mu}\}$ and $\{0, \mu^*\}$ respectively, the solution of the full problem is a mean preserving spread of the solution to the relaxed problem, and the signal is more informative in the sense of Blackwell.

*LSI-B*. If the underlying environment has the LSI-A property, then the sender's optimal distribution of posteriors in the relaxed problem is $\{\hat{\mu}, 1\}$ with $\tau(\hat{\mu}) = \frac{1-e}{1-\hat{\mu}}$, and the solution in the full problem is $\{\mu^*, 1\}$ with $\tau(\mu^*) = \frac{1-e}{1-\mu^*}$. Per Step 2 of Proposition 2, we know that the solution of the relaxed problem generates a negative value of the IC constraint, i.e.

$$\mathbb{E}_\tau[h(\mu)] = \frac{1-e}{1-\hat{\mu}}\Big(\frac{\hat{\mu} - e}{e(1-e)} v_A(\hat{\mu})\Big) + \frac{e - \hat{\mu}}{1-\hat{\mu}}\Big(\frac{1}{e}\Big) < c'(e).$$

Next, note that for $\mu \leq e$

$$\frac{1-e}{1-\mu}\Big(\frac{\mu - e}{e(1-e)} v_A(\mu)\Big) + \frac{e-\mu}{1-\mu}\Big(\frac{1}{e}\Big) = \Big(\frac{e-\mu}{e(1-\mu)}\Big)(1 - v_A(\mu)).$$

is a product of two positive, decreasing functions of $\mu$, and it is therefore decreasing. Furthermore, the solution of the sender's full problem satisfies IC. Combining these observations, it follows that

$\mu^* < \hat{\mu}$. Because both the solution of the relaxed problem and the sender's full problem have the same mean, and are supported on $\{\hat{\mu}, 1\}$ and $\{\mu^*, 1\}$ respectively, the solution of the full problem is a mean preserving spread of the solution to the relaxed problem, and the signal is more informative in the sense of Blackwell. ■

# References


**Alonso, Ricardo and Odilon Câmara**, "Persuading voters," *The American Economic Review*, 2016, *106* (11), 3590–3605. 3, 4, 22

**Au, Pak Hung and Keiichi Kawai**, "Competitive information disclosure by multiple senders," *Games and Economic Behavior*, 2020, *119*, 56–78. 4

**Aumann, Robert J and Michael Maschler**, *Repeated games with incomplete information*, MIT press, 1995. 3, 11, 30

**Barron, Daniel, George Georgiadis, and Jeroen Swinkels**, "Optimal contracts with a risk-taking agent," *Theoretical Economics*, 2020, *15* (2), 715–761. 5

**Bergemann, Dirk, Benjamin Brooks, and Stephen Morris**, "The limits of price discrimination," *The American Economic Review*, 2015, *105* (3), 921–957. 4

**_ , _ , and _** , "First-price auctions with general information structures: implications for bidding and revenue," *Econometrica*, 2017, *85* (1), 107–143. 4

**Bloedel, Alexander W and Ilya R Segal**, "Persuasion with rational inattention," *Available at SSRN 3164033*, 2018. 4

**Boleslavsky, Raphael and Christopher Cotton**, "Grading standards and education quality," *American Economic Journal: Microeconomics*, 2015, *7* (2), 248–279. 4, 5

**_ and _** , "Limited Capacity in Project Selection: Competition Through Evidence Production," *Economic Theory*, 2018, *65* (2), 385–421. 4

**Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang**, "Pivotal persuasion," *Journal of Economic theory*, 2019, *180*, 178–202. 4

**Doval, Laura and Vasiliki Skreta**, "Constrained Information Design: Toolkit," 2018. 3

**Ely, Jeffrey C**, "Beeps," *The American Economic Review*, 2017, *107* (1), 31–53. 4

**Gehlbach, Scott and Konstantin Sonin**, "Government control of the media," *Journal of Public Economics*, 2014, *118*, 163 – 171. 22

**Gentzkow, Matthew and Emir Kamenica**, "Competition in persuasion," *The Review of Economic Studies*, 2017, *84* (1), 300–322. 4

— , **Jesse M Shapiro, and Daniel F Stone**, "Media bias in the marketplace: Theory," in "Handbook of media economics," Vol. 1, Elsevier, 2015, pp. 623–645. 22

**Georgiadis, George and Balazs Szentes**, "Optimal Monitoring Design," *Econometrica*, 2020, p. forthcoming. 5

**Guo, Yingni and Eran Shmaya**, "The interval structure of optimal disclosure," *Econometrica*, 2019, *87* (2), 653–675. 4

**Habibi, Amir**, "Motivation and information design," *Journal of Economic Behavior & Organization*, 2020, *169*, 1–18. 4

**Hölmstrom, Bengt**, "Moral hazard and observability," *The Bell journal of economics*, 1979, pp. 74–91. 10

**Hörner, Johannes and Nicolas S Lambert**, "Motivational Ratings," *The Review of Economic Studies*, 11 2020. rdaa070. 4

**Kamenica, Emir and Matthew Gentzkow**, "Bayesian persuasion," *The American Economic Review*, 2011, *101* (6), 2590–2615. 1, 3, 4, 7

**Kolotilin, Anton**, "Optimal information disclosure: A linear programming approach," *Theoretical Economics*, 2018, *13* (2), 607–635. 4

— , **Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li**, "Persuasion of a privately informed receiver," *Econometrica*, 2017, *85* (6), 1949–1964. 4

**Le Treust, Maël and Tristan Tomala**, "Persuasion with limited communication capacity," *Journal of Economic Theory*, 2019, *184*, 104940. 3

**Levy, Gilat**, "Decision making in committees: Transparency, reputation, and voting rules," *American economic review*, 2007, *97* (1), 150–168. 4

**Li, Fei and Peter Norman**, "On Bayesian persuasion with multiple senders," *Economics Letters*, 2018, *170*, 66–70. 4

**Menezes, C., C. Geiss, and J. Tressler**, "Increasing Downside Risk," *The American Economic Review*, 1980, *70* (5), 921–932. 19

**Perez-Richet, Eduardo and Vasiliki Skreta**, "Test design under falsification," *mimeo*, 2017. 4, 13

**Prat, Andrea**, "The Wrong Kind of Transparency," *The American Economic Review*, 2005, *95* (3), 862–877. 4

**Renault, Jérôme, Eilon Solan, and Nicolas Vieille**, "Optimal dynamic information provision," *Games and Economic Behavior*, 2017, *104*, 329–349. 4

**Rodina, David**, "Information design and career concerns," *mimeo*, 2017. 4

— **and John Farragut**, "Inducing effort through grade," *mimeo*, 2017. 4

**Roesler, Anne-Katrin and Balázs Szentes**, "Buyer-optimal learning and monopoly pricing," *American Economic Review*, 2017, *107* (7), 2072–80. 4

**Rosar, Frank**, "Test design under voluntary participation," *Games and Economic Behavior*, 2017, *104*, 632–655. 5, 13

**Sen, Amartya**, *Development as freedom*, Oxford Paperbacks, 2001. 22

**Zapechelnyuk, Andriy**, "Optimal quality certification," *American Economic Review: Insights*, 2020, *2* (2), 161–76. 4