Bayesian Persuasion and Moral Hazard*

Raphael Boleslavsky† and Kyungmin Kim‡

March 2018

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 determines the distribution of 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; media bias; grade inflation

*We thank Eduardo Faingold, George Georgiadis, Ilwoo Hwang, Marina Halac, Johannes Hörner, and Alex Wolitzky for various helpful comments.

†University of Miami. Contact: r.boleslavsky@miami.edu

‡University of Miami. Contact: kkim@bus.miami.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.

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 $(\pi(g|s), \pi(g|u))$ 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, $\pi(g|s) = 1$. By increasing $\pi(g|s)$, the school increases both the frequency of g and the good news it conveys. In contrast, by increasing $\pi(g|u)$, the school increases the frequency of grade g but reduces its informativeness. In this example, maintaining the informativeness of grade g constrains $\pi(g|u) \leq 3/7$. Therefore, the optimal grading policy is $\pi(g|s) = 1$ and $\pi(g|u) = 3/7$. 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,

$$\underbrace{\frac{3}{10}\,\pi(g|s) + \frac{7}{10}\,\pi(g|u) - c}_{\text{Work}} \geq \underbrace{\pi(g|u)}_{\text{Shirk}},$$

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

Both incentive and information provision shape the optimal grading policy. On one hand, increases in $\pi(g|s)$ 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, $\pi(g|s) = 1$ is also optimal with moral hazard. On the other hand, increases in $\pi(g|u)$ 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 $\pi(g|u) \leq 1/3$, which is more restrictive than the condition on informativeness.[1] Thus, the optimal grading policy is $\pi(g|s) = 1$ and $\pi(g|u) = 1/3$. 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 has a

[1] Given $\pi(g|s) = 1$, and that the constraint on informativeness is satisfied, work is preferred when $3/10 + (7/10)\pi(g|u) - c \geq \pi(g|u)$.

strict preference to hire a student with grade g.

In what follows, we explore the 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.

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 no effort by the agent and, therefore, cannot be optimal in any non-trivial environment.[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 necessarily the case. Furthermore, we explicitly characterize the optimal signal. In the binary-state environment, we derive conditions under which the optimal signal garbles

[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).

information about only one state, and apply our results to obtain novel economic insights into media censorship and the design of social monitoring systems. In the binary-action environment, we show that the optimal signal is a binary partition, illustrating how it is affected by incentive provision and moral hazard.

One particularly interesting question is the effect 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. Furthermore, we provide a simple example in which transparency increases equilibrium effort, yet the receiver prefers not to have transparency because it leads to worse information about the realized state. 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 2015, Au and Kawai 2017, Gentzkow and Kamenica 2017, Boleslavsky and Cotton 2018), multiple receivers (Alonso and Câmara 2016, Chan et al. 2017), a privately informed receiver (Guo and Shmaya 2017, Kolotilin 2017, Kolotilin et al. 2017), 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 in various economic contexts, including price discrimination (Bergemann et al. 2015), monopoly pricing (Roesler and Szentes 2017), and auctions (Bergemann et al. 2017). To our knowledge, this is the first paper that incorporates moral hazard into the Bayesian persuasion framework.

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. 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.[3] On one hand, this objective can be accommodated in our

[3]In this sense, these papers are related to Hörner and Lambert (2016), who characterize the rating system

analysis by specifying that the sender's payoff is linear 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.

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.

Two recent papers combine elements of information design and moral hazard in novel ways. Barron et al. (2017) consider a canonical principal-agent model in which the agent can engage in "gaming" (adding mean-preserving noise) after observing an intermediate output (which effectively enables the agent to concavify his compensation) and show that if the agent is risk neutral, then a linear contract is optimal. Georgiadis and Szentes (2018) introduce endogenous monitoring into a dynamic principal-agent model and, by applying the ideas in information design and zero-sum games, show that the optimal contract has a simple binary structure (a base salary plus a fixed performance-based bonus).

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 concludes.

---

that maximizes the agent's effort in a dynamic career concerns model with various information sources.

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 endogenously determined. The game unfolds in three stages. In the first stage, the sender designs and publicly reveals a signal structure π, which consists of a message space Σ and a set of conditional probability distributions $\{\pi(\cdot|\omega)\}_{\omega\in\Omega}$ over Σ. As in KG, we impose no structural restriction on the sender's choice of π; that is, we assume that the sender can choose any finite set Σ and any conditional probabilities $\pi(\cdot|\omega)$ over this set. In the second stage, the agent observes the chosen signal structure π and exerts an unobservable effort $e \in [0, 1]$. Given e, the state is drawn according to the probability vector $\eta(e) = (\eta(1|e), ..., \eta(N|e)) \in \Delta(\Omega)$.[4] In the third stage, a message $s \in \Sigma$ is realized according to the given signal structure π. The receiver observes π and s and chooses an action x from a compact set of feasible actions X.

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.[5] 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: $u_A(x)$ is non-negative and bounded, while $c(e)$ is strictly increasing, convex and twice continuously differentiable, with $c(0) = c'(0) = 0$ and $c'(1)$ sufficiently large.[6] All agents are risk neutral and maximize their expected utility.

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

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

In other words, inducing a particular action $x \in X$ is identical to inducing a posterior μ under which the receiver's optimal action is x. As in KG, this reformulation allows us to

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

[5] We assume that the agent's utility does not depend on the state ω 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, as explained shortly. From an economic perspective, this assumption implies that the agent's motivation for effort is purely extrinsic: 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.

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

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.[7]

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

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

where both $\underline{\mu}$ and $\overline{\mu}$ have full support on Ω. In other words, the probability distribution that generates the underlying state ω 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 the agent's 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 π, the agent's effort e (for each possible signal), and a belief system $(\mu_s \in \Delta(\Omega), s \in \Sigma)$ (also for each possible signal), which satisfy the following properties: (i) given any signal π and the receiver's belief system $(\mu_s \in \Delta(\Omega), s \in \Sigma)$, the agent's effort e maximizes his expected utility, (ii) the receiver's belief system $(\mu_s \in \Delta(\Omega), s \in \Sigma)$ is consistent with the agent's effort choice e and the signal π, and (iii) the chosen signal π maximizes the sender's expected utility.

Vector operations. We make use of the following vector operations.[8] For any $y, z \in \mathcal{R}^N$,

$$\begin{aligned}
&\text{Inner product:} && \langle y, z\rangle \equiv y(1)z(1) + ... + y(N)z(N), \\
&\text{Hadamard product:} && y \odot z \equiv (y(1)z(1), ..., y(N)z(N)), \text{ and} \\
&\text{Hadamard division:} && y \oslash z \equiv (y(1)/z(1), ..., y(N)/z(N)).
\end{aligned}$$

[7] Our subsequent arguments can be applied verbatim if the graphs of $v_A(\cdot)$ and $v_S(\cdot)$ are closed.

[8] Clearly, if $z(i) = 0$ for some i, then Hadamard division is undefined. In our analysis, this scenario never arises.

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 $\pi(s|\omega)$ denotes the probability that $s \in \Sigma$ is realized when the state is ω. Let $\widehat{e}$ denote the receiver's conjecture about the agent's effort and $\mu^*(s, \pi, \widehat{e})$ represent the receiver's updated belief given $\widehat{e}$ and π. Then, by Bayes' rule, for any $s \in \Sigma$,

$$(2) \quad \mu^*(s,\pi,\widehat{e}) \equiv \left(\frac{\pi(s|1)\eta(1|\widehat{e})}{\sum_{\omega'} \pi(s|\omega')\eta(\omega'|\widehat{e})}, ..., \frac{\pi(s|N)\eta(N|\widehat{e})}{\sum_{\omega'} \pi(s|\omega')\eta(\omega'|\widehat{e})} \right) = \frac{\pi(s|\cdot) \odot \eta(\widehat{e})}{\langle \pi(s|\cdot), \eta(\widehat{e}) \rangle} \in \Delta(\Omega).$$

Thus, (2) defines the receiver's belief structure.

The agent's payoff depends on his effort, the sender's signal structure, and the receiver's belief structure. In particular, the effort and signal structure together determine the probability of each signal realization, while the belief structure determines the associated reward. To formulate the agent's problem, let $\overline{v}_A(\omega|\pi, \mu'_s) \equiv \sum_s \pi(s|\omega) v_A(\mu'_s)$ denote the agent's expected payoff conditional on generating state ω given the sender's signal π and the receiver's belief structure $\{\mu'_s\}_{s\in\Sigma}$, and let $\overline{v}_A(\pi, \mu'_s) \equiv (\overline{v}_A(1|\pi, \mu'_s), ..., \overline{v}_A(N|\pi, \mu'_s))$ denote the corresponding vector. The agent's problem can be written as

$$\max_{e\in[0,1]} \sum_{\omega} \left(\sum_{s} \pi(s|\omega) v_A(\mu'_s) \right) \eta(\omega|e) - c(e) = \langle \eta(e), \overline{v}_A(\pi, \mu'_s) \rangle - c(e).$$

The first term is linear in e, because $\eta(e) = (1-e)\underline{\mu} + e\overline{\mu}$, while the second term is convex. Therefore, the agent's optimal effort is characterized by the following first-order condition:[9]

$$\langle \eta_e(e), \overline{v}_A(\pi, \mu'_s) \rangle = \langle \overline{\mu} - \underline{\mu}, \overline{v}_A(\pi, \mu'_s) \rangle = c'(e).$$

The receiver's belief structure depends on the conjectured effort, while the agent's optimal

[9]Recall our maintained assumption that $c(0) = c'(0) = 0$, while $c'(1)$ is sufficiently large, which ensures that the agent's optimal effort is always in the interior.

effort depends on the receiver's belief structure. In equilibrium, the conjectured effort should coincide with the agent's actual effort choice. Therefore, the equilibrium belief structure $\{\mu_s\}_{s\in\Sigma}$ 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 unique equilibrium in the subgame between the agent and the receiver given signal π:

$$\mu_s = \mu^*(s, \pi, e) \text{ and } \langle \overline{\mu} - \underline{\mu}, \overline{v}_A(\pi, \mu_s)\rangle = c'(e). \tag{3}$$

Sender's problem. Note that the sender's payoff depends on both the agent's effort level and the receiver's belief structure. As above, let $\overline{v}_S(\omega|\pi, \mu_s) \equiv \sum_s \pi(s|\omega)v_S(\mu_s)$ denote the sender's expected payoff conditional on the state being ω, given the sender's signal structure π and the receiver's belief system $\{\mu_s\}_{s\in\Sigma}$. In addition, let $\overline{v}_S(\pi, \mu_s) \equiv (\overline{v}_S(1|\pi, \mu_s), ..., \overline{v}_S(N|\pi, \mu_s))$ be the corresponding vector. Given the preceding characterization of the subgame, the sender's problem can be written as

$$\max_{\pi,e} \langle \eta(e), \overline{v}_S(\pi, \mu_s)\rangle \text{ subject to the two conditions in (3).}$$

In other words, the sender chooses π and e in order to maximize his expected payoff $\langle \eta(e), \overline{v}_S(\pi, \mu_s)\rangle$ subject to the constraint that e and $\{\mu_s\}_{s\in\Sigma}$ must be an equilibrium effort level and belief structure in the subgame following the sender's choice of π.

For most of the paper, we consider the sender's optimization problem given a particular target effort level. In other words, we treat e as a parameter of the sender's optimization, rather than a choice variable. Given the characterization of optimal π, it suffices to identify the effort level that maximizes the sender's (indirect) expected payoff, which is often straightforward. Note that for some effort levels $e \in [0, 1]$, there may not exist a signal that generates them as part of a subgame equilibrium. We fully characterize the set of implementable efforts in more specific environments in Sections 4 and 5.

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

Proposition 3.1 *Given target effort e, there exists a signal π that satisfies the two conditions in* (3) *if and only if there exists a distribution of posteriors $\tau \in \Delta(\Delta(\Omega))$ such that*

(i) $E_\tau[\mu] = \eta(e)$ *(Bayes-Plausibility), and*

(ii) $E_\tau\left[E_\mu\left[\frac{\eta_e(\omega|e)}{\eta(\omega|e)}\right] v_A(\mu)\right] = c'(e)$ *(Incentive Compatibility).*

The Bayes-Plausibility and incentive compatibility conditions are restatements of the equilibrium conditions in (3) in the space of posterior beliefs. To understand the link, fix a signal structure π and effort level e. Without loss of generality, assume that for each realization $s \in \Sigma$, the receiver has a distinct Bayesian update μ_s. Then, the probability that the Bayesian update is μ_s, denoted by $\tau(\mu_s)$, is simply

$$\tau(\mu_s) = \sum_{\omega} \pi(s|\omega)\eta(\omega|e) = \langle \pi(s|\cdot), \eta(e) \rangle.$$

Plugging this into $\mu_s = \mu^*(s, \pi, e)$ and arranging the terms, we get

$$\mu_s = \frac{\pi(s|\cdot) \odot \eta(e)}{\langle \pi(s|\cdot), \eta(e) \rangle} = \frac{\pi(s|\cdot) \odot \eta(e)}{\tau(\mu_s)} \Rightarrow \pi(s|\cdot) = \tau(\mu_s)(\mu_s \oslash \eta(e)). \tag{4}$$

Clearly, starting with the signal structure and an equilibrium effort level, we can calculate the distribution of posterior beliefs generated by the signal. Equation (4) shows that the process can also be reversed: given any Bayes-plausible distribution of posteriors, one can also derive the underlying signal structure that yields the distribution. To recover the incentive compatibility condition, it suffices to combine equation (4) with equation (3):

$$\begin{aligned}
\langle \overline{\mu} - \underline{\mu}, \overline{v}_A(\pi, \mu_s) \rangle &= \sum_{\omega} \left(\sum_{s} \pi(s|\omega) v_A(\mu_s) \right) (\overline{\mu}(\omega) - \underline{\mu}(\omega)) \\
&= \sum_{\omega} \left(\sum_{s} \tau(\mu_s) \frac{\mu_s(\omega)}{\eta(\omega|e)} v_A(\mu_s) \right) (\overline{\mu}(\omega) - \underline{\mu}(\omega)) \\
&= \sum_{s} \tau(\mu_s) \sum_{\omega} \left(\frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{\eta(\omega|e)} \mu_s(\omega) \right) v_A(\mu_s) \\
&= \sum_{s} \tau(\mu_s) \sum_{\omega} \left(\frac{\eta_e(\omega|e)}{\eta(\omega|e)} \mu_s(\omega) \right) v_A(\mu_s) = E_\tau \left[E_\mu \left[\frac{\eta_e(\omega|e)}{\eta(\omega|e)} \right] v_A(\mu) \right].
\end{aligned}$$

Two aspects of (IC) are worth highlighting. First, (IC) immediately implies that a degenerate posterior distribution can implement only $e = 0$.[10] In other words, dispersion in the posterior belief distribution is necessary to provide the agent with incentives to exert effort. As shown later, however, it is not necessarily the case that more dispersion induces

[10] For a degenerate distribution to satisfy Bayes-plausibility, the unique posterior belief must be equal to the prior, that is, $\mu = \eta(e)$. Then, it is easy to show that it violates the IC constraint whenever $e > 0$:

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

more effort. Furthermore, in the absence of (IC), the sender would use an uninformative signal (degenerate posterior distribution) whenever his payoff function is concave in μ. Thus, the desire to motivate the agent forces the sender to introduce distortions in this case.

Second, the likelihood ratio term $\eta_e(\omega|e)/\eta(\omega|e)$ 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 bigger impact on (IC) when (i) the agent derives a larger benefit from the realization ($v_A(\mu)$ is high) or (ii) the posterior belief realization generates a high expected value for the receiver's statistical estimate of the agent's effort ($E_\mu[\eta_e(\omega)/\eta(\omega)]$ is high). Thus, in our model the agent acts *as if* he is rewarded when the receiver's estimate of his effort is high, even though his reward depends only on the receiver's action.[11]

3.2 MAIN CHARACTERIZATION

We now characterize the sender's optimal signal structure. Proposition 3.1 implies that the sender's problem can be written as

$$\max_{\tau\in\Delta(\Delta(\Omega))} E_\tau[v_S(\mu)], \text{ subject to } (BP)\ \ E_\tau[\mu]=\eta(e), \text{ and } (IC)\ \ E_\tau[h(\mu)]=0, \tag{5}$$

where

$$h(\mu)\equiv E_\mu\left[\frac{\eta_e(\omega|e)}{\eta(\omega|e)}\right]v_A(\mu)-c'(e) \text{ for each } \mu\in\Delta(\Omega). \tag{6}$$

We let τ^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

[11] In particular, given a Bayes-Plausible distribution of posteriors, the equilibrium effort can be derived from the following optimization problem: $\max_e E_\tau[E_\mu[\log(\eta(\omega|e))]v_A(\mu)]-c(e)$. Thus, 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 $E_\mu[\log(\eta(\omega|e))]v_A(\mu)]-c(e)$ at each μ. The "virtual" payoff function weighs the agent's true payoff $v_A(\mu)$ by the expected value of the log-likelihood function. By implication, the virtual agent derives utility from the perception, ex post, that he exerted effort.

(i.e., $V^e \equiv E_{\tau^e}[v_S(\mu)]$). Crucially, in (5), the objective function and the two constraints are expectations of certain functions of μ with respect to τ. 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 (5): the first N components are $(\mu(1), ..., \mu(N))$, 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. Now 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 $\Delta(\Omega)$. Roughly, 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 proposition. The result on the cardinality of the support follows from Caratheodory's theorem.[12]

Proposition 3.2 *In problem* (5)*, the maximal utility the sender can obtain given target effort 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 $co(K^e)$:

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

By construction, H^e includes all the points that are convex combinations of the elements of K^e (i.e., $y_1 = E_\tau[\mu]$, $y_2 = E_\tau[h(\mu)]$, and $y_3 = E_\tau[v_S(\mu)]$) 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 (5). Because K^e is closed, $co(K^e)$ is also closed, and hence, this maximum is attained. For the result on the cardinality of the support of τ^e, see the appendix. ■

[12]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,

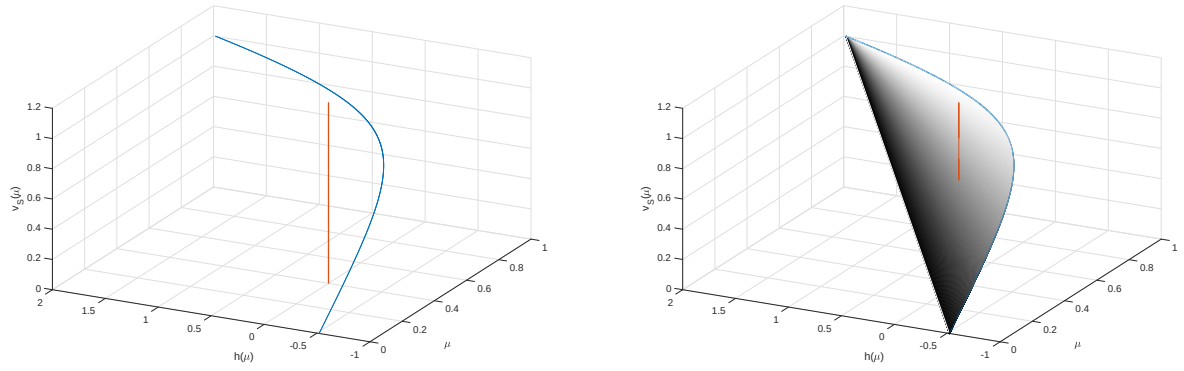

Figure 1: The left panel depicts the curve K^e, while the right panel depicts its convex hull $co(K^e)$. In this example, $n = 2$ and μ represents the probability that $\omega = 2$. In addition, $\eta(e) = E[\mu] = 0.4$, $v_A(\mu) = \mu$, and $u_S(\mu) = 1 - (1 - \mu)^4$.

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 μ and $\eta(2|e)$ 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 (i.e., without (IC)), $h(\mu)$ 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 (possibly) increasing the number of required realizations by 1.

There are two other noteworthy points regarding Proposition 3.2. First, it does not crucially depend on our restriction to the linear production technology (i.e., $\eta(e) \equiv (1 - e)\underline{\mu} + e\overline{\mu}$). The same logic goes through unchanged as long as the subgame equilibrium effort level is fully characterized by the agent's first-order condition (i.e., the first-order approach is

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.

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 τ (i.e. each constraint can be written as $E_\tau[F(\mu)] = 0$ for some function $F(\cdot)$). The only difference is that when there are a total of k constraints (including the BP constraint), 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, Proposition 3.2 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 $(\eta(e), 0, V^e)$.

Proposition 3.3 *A distribution of posteriors τ^e is a solution to the sender's problem* (5) *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*

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

with equality for all μ 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 $d \equiv (-\lambda_1(1), ..., -\lambda_1(N), \psi, 1) \in \mathcal{R}^{N+2}$ and a scalar λ_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 and rearranging yields $\mathcal{L}(\mu, \psi) \leq \lambda_0 + \langle \lambda_1, \mu \rangle$ for all $\mu \in \Delta(\Omega)$. ■

Figure 2 illustrates Proposition 3.3 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.

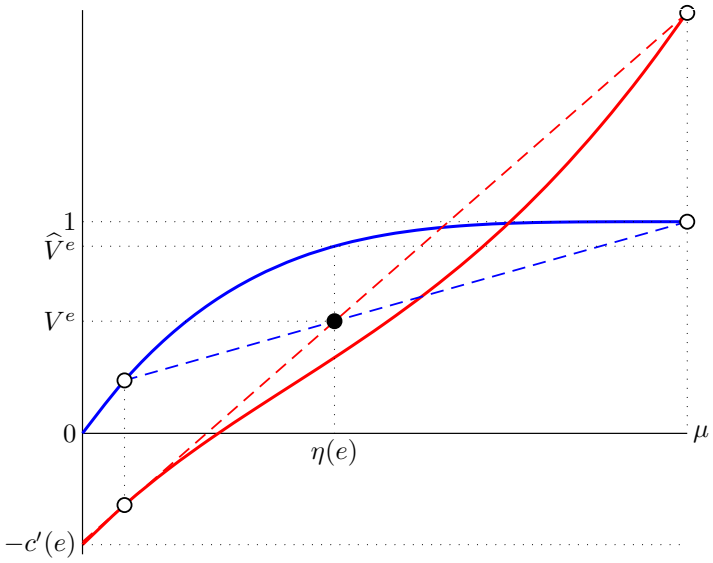

Figure 2: The concave solid curve depicts $u_S(\mu)$, while the other solid curve depicts $\mathcal{L}(\mu, \psi) = v_S(\mu) + \psi h(\mu)$, with the same data as in Figure 1.

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 typically the case and 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 $E_{\tau^e}[\mathcal{L}(\mu, \psi)] = E_{\tau^e}[v_S(\mu)]$. Graphically, $E_{\tau^e}[\mathcal{L}(\mu, \psi)]$ is the value of the Lagrangian's supporting line (the dashed red line) evaluated at $\eta(e)$. Similarly, $E_{\tau^e}[v_S(\mu)]$ 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 μ, each of which 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 μ, $\underline{\mu}$, $\overline{\mu}$, and $\eta(e)$ as scalars between 0 and 1 and use them to represent the probability that $\omega = 2$. We first offer some general characterization results and then provide a more comprehensive analysis of three representative examples.

We maintain the following simplifying assumptions through this section.

(i) Monotonicity: both $v_A(\cdot)$ and $v_S(\cdot)$ are increasing in μ, 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: $c'(1) > \overline{\mu} - \underline{\mu}$.

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 level is less than 1.

4.1 IMPLEMENTABLE AND INCENTIVE-FREE EFFORT LEVELS

We say that a target effort level is implementable if there exists a signal π (equivalently, a distribution of posteriors τ) 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 4.1 *In the binary-state model, let $\overline{e}$ be the value such that $c'(\overline{e}) = \overline{\mu} - \underline{\mu}$. Then, e is implementable if and only if $e \leq \overline{e}$. The maximum effort $\overline{e}$ is incentive compatible if and only if the signal is fully informative.*

Proof. See the appendix. ■

For the intuition, notice that with binary states, equation (3) reduces to

$$(\overline{\mu} - \underline{\mu})(\overline{v}_A(2) - \overline{v}_A(1)) = c'(e),$$

where $\overline{v}_A(\omega)$ is the agent's expected payoff when the realized state is ω. Therefore, the agent's incentive depends exclusively on the difference in the expected rewards in the two possible states, $\overline{v}_A(2) - \overline{v}_A(1)$. Given the assumptions on $v_A(\cdot)$, this difference cannot exceed 1, and achieves 1, if and only if the sender uses a fully informative signal, because it is necessary and sufficient that $\overline{v}_A(2) = 1$ and $\overline{v}_A(1) = 0$.

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 4.1 is that the former result continues to hold, while the latter result fails in the model with moral hazard.

Corollary 4.2 *In the binary-state model, a fully informative signal is optimal when $v_S(\cdot)$ is convex in μ, while a fully uninformative signal is never optimal when $v_S(\cdot)$ is concave.*

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 (concave) 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 level: the sender's ex post payoff is bounded from below by $v_S(\underline{\mu})$ and is strictly greater than $v_S(\underline{\mu})$ with positive probability.

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

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

Obviously, $V^e \leq \widehat{V}^e$ for any $e \leq \overline{e}$. Let $\underline{e}$ be the maximal value of e such that $V^e = \widehat{V}^e$. Our preceding result 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).[13]

The following result shows that the sender (weakly) prefers an effort level $e \in [\underline{e}, \overline{e}]$ to any effort level less than $\underline{e}$ and, therefore, it suffices to consider those effort levels. It also shows that implementing $e > \underline{e}$ requires a distortion from the relaxed problem.

Proposition 4.3 *In the sender's problem* (5)*, if* $e < \underline{e}$*, then* $V^e \leq V^{\underline{e}}$*. Furthermore, for any* $e \in (\underline{e}, \overline{e}]$*, the solution to* (5) *has* $\psi > 0$*.*

Proof. See the appendix. ■

We now apply our characterization to study three representative environments. For ease of exposition, we focus on the case in which $\underline{\mu} = 0$ and $\overline{\mu} = 1$. This implies that $\eta(e) = e$.[14] In addition, the function $h(\mu)$ in (IC) reduces to

$$h(\mu) = E_\mu \left[\frac{\eta_e(\omega|e)}{\eta(\omega|e)} \right] v_A(\mu) - c'(e) = \frac{\mu - e}{e(1-e)} v_A(\mu) - c'(e),$$

because $\eta(e) = (1-e, e)$ when written as a vector, and thus, $\eta_e(e) \oslash \eta(e) = (-1/(1-e), 1/e)$.

[13] 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 choice. $\underline{e}$ is the maximal effort attainable under such a scenario. Therefore, it is the sender's power to commit that enables him to implement $e \in (\underline{e}, \overline{e}]$.

[14] 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, Proposition 3.1 and the subsequent characterizations apply to these cases unchanged.

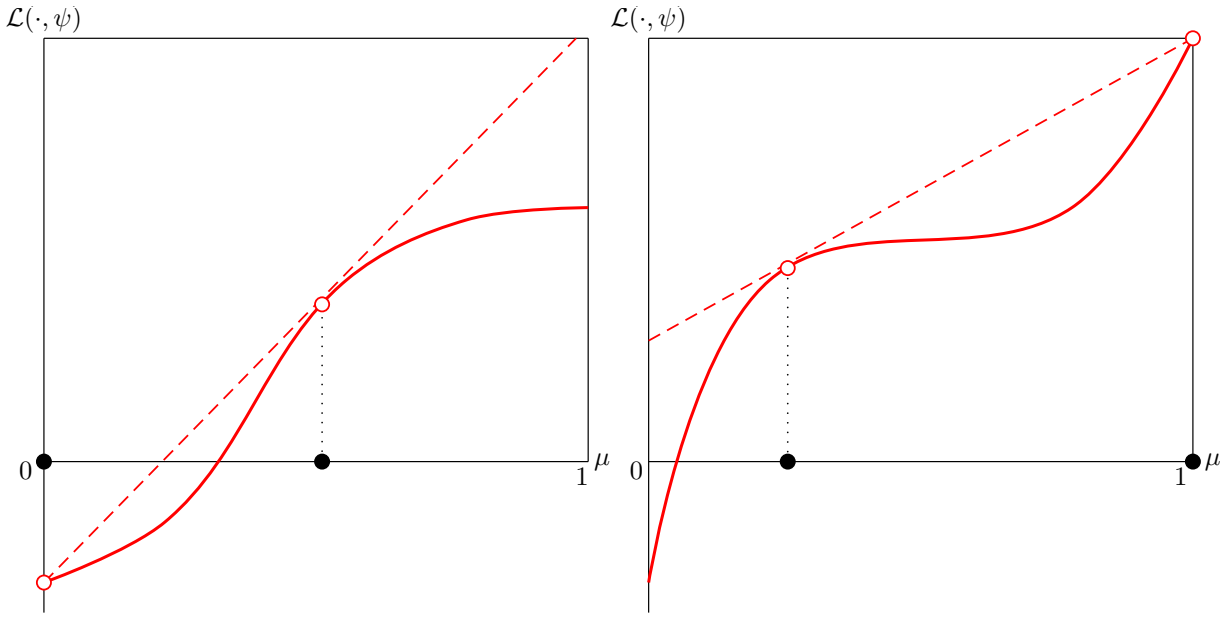

Figure 3: Both panels depict the Lagrangian function $\mathcal{L}(\mu,\psi)$. In the left panel $v_S'''(\cdot)<0$, while in the right panel $v_S'''(\cdot)>0$.

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 $v_S''(\cdot)<0$). This arises, for example, if the sender suffers a convex cost as the receiver's posterior belief falls from the sender's ideal point 1.

Fix $e\in(0,\overline{e})$ and consider the Lagrangian function $\mathcal{L}(\mu,\psi)$ introduced in Proposition 3.3. Because $v_A(\cdot)$ is linear, $h(\cdot)$ is quadratic in μ, and the second derivative of $\mathcal{L}$ with respect to μ takes the following form:

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

Although $v_S''(\cdot)<0$, the Lagrangian is not necessarily concave because of the (positive) second term. 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, $\mathcal{L}_{\mu\mu}$ also cannot be positive everywhere: if the Lagrangian is convex, then the optimal signal is fully informative and implements $\overline{e}$ (see Proposition 4.1). 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 for the case where $\mathcal{L}_{\mu\mu}$ 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 ψ. 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 Proposition 3.3, 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 4.4 *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 $e \in (0, \overline{e})$ 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 $v_S'''(\cdot) > 0$, then the optimal distribution of posteriors that implements $e \in (0, \overline{e})$ is supported on $\{\mu_D, 1\}$ with $\tau^e(\mu_D) = (1-e)/(1-\mu_D)$, where $\mu_D \equiv e(1-c'(e))$.

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

$$h(\mu) = \frac{\mu(\mu - e)}{e(1-e)} - c'(e). \tag{7}$$

Since $E_\tau[\mu] = e$ (due to (BP)),

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

Thus, (IC) constrains the variance of the posterior belief to a specific value, $e(1-e)c'(e)$, while (BP) constrains the mean to e. Therefore, if $v_A(\mu) = \mu$ then 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). These authors 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.[15] 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*

[15]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.

loving and prefers to shift all necessary dispersion to the bottom of the distribution, resulting in a binary distribution whose smaller realization is 0.

Transparency I. To further examine the interplay between information and incentive provision, consider a benchmark model in which the agent's effort is observable. In that case, given any Bayes-Plausible distribution of posteriors τ, the agent's problem reduces to

$$\max_e E_\tau[v_A(\mu)] - c(e) = E_\tau[\mu] - c(e) = e - c(e),$$

whose unique optimal solution is given by $\overline{e}$, as defined in Proposition 4.1. Importantly, this holds for any signal structure, that is, the agent chooses $\overline{e}$ regardless of π (or τ). It follows that it is optimal for the sender, who has concave preferences, to reveal no information. In this example, transparency "crowds out" the sender's concern for incentive provision, severing the link between the signal structure and effort. Therefore, the sender is only concerned with information provision, adopting the optimal policy in KG. This result highlights a potential tradeoff for the receiver: with transparency, the agent exerts effort $\overline{e}$, but the receiver acquires no further information about the state. Depending on the relative importance of these two effects, transparency may harm the receiver.

Application: media bias and government accountability. Our results can be used to gain new insights on the issues of media bias and government accountability.[16] We can identify the government as the agent, the media as the sender, and a representative citizen as the receiver in our model. The government exerts effort e to increase the probability of state $\omega = 2$, which is beneficial for the citizen (for example, reduced corruption or increased economic development), but the citizen observes neither the government's effort nor the realized state. Instead, information about the realized state is communicated to the general public by an independent news media. As in Gehlbach and Sonin (2014) and Gentzkow et al. (2015), the media commits to a reporting strategy $\pi(\cdot|\omega)$, which is observed by the citizen.

Under standard conditions on his decision problem, the citizen's interim payoff function $v_R(\cdot)$ is increasing and convex in his belief μ about the state. If the media shares the citizen's preferences (i.e., $v_S(\cdot)$ is also convex), either because it is altruistic or because it can extract the citizen's total surplus through an access price, then the optimal reporting policy is a

[16] Freedom of the press is essential to democracy, which is built on the fundamental notion that government is accountable to the people. Thus, in its role as a government watchdog, the free press has a "clear, instrumental role in preventing corruption, financial irresponsibility, and underhanded dealings" (Sen 2001, p. 40). However, there is significant concern that the media's economic or political interests may undermine their incentive to report the information needed for proper oversight (see, e.g., Besley and Prat 2006).

fully informative one (Corollary 4.2). It not only provides the most accurate information but also maximizes the government's effort (Proposition 4.1).

Now suppose that the government's payoff is linear in "public opinion" (μ) and the media's payoff is concave. In this case, Proposition 4.4 has interesting implications for the media's reporting strategy. Depending on the media's downside risk aversion, it either uses a binary signal which perfectly reveals the bad state (but not good), or a binary signal which perfectly reveals the good state (but not bad). Specifically, if $v_S'''(\cdot) < 0$, then the media reports good news for sure if $\omega = 2$ and some of the time even if $\omega = 1$. In other words, the media commits to *censoring bad news*, sometimes replacing a "truthful" bad report with a nominally good report, reducing the good report's overall credibility while suppressing unfavorable information. Conversely, if $v_S'''(\cdot) > 0$, then the media reports bad news for sure if $\omega = 1$ and some of the time even if $\omega = 2$. In this case the media commits to *censoring good news* about the state, sometimes replacing truthful good news with nominally bad news. In the presence of moral hazard, this "hostile" reporting policy arises in equilibrium, even though the media and government both prefer to keep public opinion high.

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 $v''(\cdot) < 0$, and the sender also internalizes the agent's effort cost.[17] This perfect alignment of interests could arise, for example, if the sender and the agent are the same entity, 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 fee.

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

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

Since $\underline{e} = 0$ (due to the sender's concave preferences), $\psi > 0$ for any $e > 0$ (see Proposition 4.3). It then follows that

$$\mathcal{L}_{\mu\mu} > 0 \Leftrightarrow \frac{e(1-e-\psi)}{2\psi} + \frac{\mu}{2} < \frac{1}{r(\mu)}, \tag{8}$$

where $r(\mu) \equiv -v''(\mu)/v'(\mu)$ is the Arrow-Pratt measure of risk aversion. As in Section 4.2, at

[17] 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.

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. Using similar logic to Propositions 4.4, we obtain the following result.

Proposition 4.5 *In the binary-state model, suppose that 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 (α, β).*

(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 $\tau(\mu_D) = (1-e)/(1-\mu_D)$.

Transparency II. Suppose that the agent's effort is observable by the receiver. Since the agent and the sender have identical preferences, the problem reduces to:

$$\max_{e,\tau} E_\tau[v(\mu)] - c(e) \quad \text{subject to} \quad E_\tau[\mu] = e.$$

Since $v(\cdot)$ is concave, given any e, it is optimal to reveal no information. Thus, the objective function can be further simplified to $v(e) - c(e)$, for which the optimal effort is characterized by $v'(e) = c'(e)$. Unlike in Section 4.2, the equilibrium observable effort may fall short of $\overline{e}$: it is smaller than $\overline{e}$ if and only if $v'(\overline{e}) < 1$. Therefore, it is not necessarily the case that transparency enhances the agent's effort. In fact, it is possible that transparency reduces both the agent's equilibrium effort and the informativeness of the equilibrium signal, as shown in the following example.

Example. Suppose that $v(\mu) = 1 - (1-\mu)^2$ and $c(e) = ce^2/2$. If the agent's effort is unobservable then, by Proposition 4.5, the optimal signal induces posteriors 0 and μ_U, because $1/r(\mu) = 1 - \mu$ in this case. Using the equilibrium conditions that $E_\tau[\mu] = e$ and $E_\tau[h(\mu)] = 0$, one can find that the sender's optimal signal is such that

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

From this explicit solution, it is immediate that the optimal effort is $2/(3c)$ whenever $c \in (2/3, 1)$. If the agent's effort is observable, then the equilibrium effort is given by

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

It follows that transparency lowers the agent's equilibrium effort if $c \in (2/3, 1)$.

Application: optimal monitoring and prosocial behavior. In social interactions, individuals sometimes undertake costly actions that may generate benefits exclusively for others. One explanation in the literature is that such individuals are motivated by social pressure or social norms that reward those who are believed to have generated such benefits (see, e.g., Bénabou and Tirole 2006, Daughety and Reinganum 2010). Such rewards and punishments are, fundamentally, derived from the beliefs that society holds about an individual's actions and their consequences. In this context, an important question is how such information should be transmitted, on which our results can be used to shed new light.

We interpret e as the amount of effort an agent devotes to a prosocial activity. Effort translates into social benefits stochastically: the agent produces a positive benefit to others ($\omega = 2$, good state) with probability e and no benefit ($\omega = 1$, bad state) with probability $1 - e$. The agent is motivated by "social pressure," which rewards the agent to the extent that society believes a benefit was generated. In this context, the agent's payoff is given by $v(\mu) - c(e)$, where μ denotes the society's belief about the agent's realized social contribution. The problem of designing a monitoring system that maximizes the agent's ex ante welfare then becomes identical to the sender's problem in this subsection.

Proposition 4.5 shows that when the agent's effort is unobservable, then the optimal monitoring system takes an intriguing form. If $\alpha < 1/2$, then it searches for conclusive evidence that no benefit was realized. Such evidence can be found only in the bad state, but even in this state, the system only discovers it with a known probability, which depends on the exhaustiveness of the search. If evidence is uncovered, then it is unveiled, revealing the bad state ($\mu = 0$). If no evidence is discovered, then a null message is generated, which improves beliefs to the extent that the search was exhaustive ($\mu = \mu_U$). Thus, for $\alpha < 1/2$, the monitoring system conducts a limited search for conclusive evidence of the bad state, disclosing anything that it uncovers. Conversely, for $\alpha > 1/2$ the optimal monitoring system conducts a limited search for evidence of the good state, disclosing anything that it uncovers.

In this environment, transparency may be harmful to society. Suppose that society weighs both the agent's welfare and his social contribution. With transparency, the optimal monitoring system never conveys information and the optimal effort equates the marginal benefit and cost of effort ($v'(e) = c'(e)$). With private effort, the agent is potentially harmed in two ways. First, the monitoring system induces dispersed posteriors, which reduces his payoff because he has concave preferences. Second, as shown in the example above, the agent's equilibrium effort may also be higher, resulting in a larger effort cost. However, if the agent's effort is higher, then social benefits are also more likely to accrue. Thus, if privacy increases the agent's effort, and society places a sufficiently large weight on the benefits it generates, then society may prefer not to observe the agent's effort.

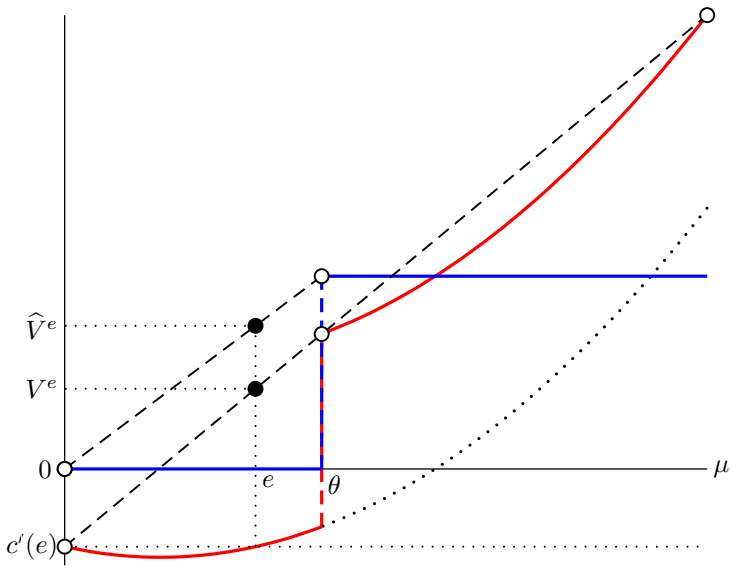

Figure 4: The step function represents $v_S(\mu) = \mathcal{I}_{\{\mu \geq \theta\}}$, while the solid curve depicts $\mathcal{L}(\mu, \psi)$ when $v_A(\mu) = \mu$.

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. Thus, the agent would like to increase the receiver's belief that the state is good, while the sender would like this belief to be "good enough." In order to reduce the number of cases to consider, we assume that $c'(\theta) > 1$, so 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, it is optimal for the sender to use a binary posterior distribution supported on $\{0, \theta\}$ for any $e < \theta$ (see KG). It follows that the unique incentive-free effort level $\underline{e}$ is given by the value that satisfies

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

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 μ, with an upward jump discontinuity of 1 at $\mu = \theta$ (see Figure 4). Because $\mathcal{L}$ is convex over the domains $\mu < \theta$ and $\mu > \theta$, there are three ways

in which the supporting line $\lambda_0 + \lambda_1\mu$ can intersect $\mathcal{L}(\mu, \psi)$ at the solution of the sender's problem. The intersections can occur (i) at $\mu = 0$ and θ, (ii) at $\mu = 0$ and 1, or (iii) at $\mu = 0$, θ, and 1. However, (i) implements $\underline{e}$ (as explained above), while (ii) implements $\overline{e}$ (Proposition 4.1). Therefore, (iii) is the only possibility; that is, at the optimum, the supporting line must intersect $\mathcal{L}(\cdot, \psi)$ at all three points, $\{0, \theta, 1\}$, as shown in Figure 4.

Proposition 4.6 *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 τ^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) $E_{\tau^e}[\mu] = e$, and (iii) (IC) $E_{\tau^e}[h(\mu)] = 0$. ■

This result demonstrates that the result on the maximal number of necessary posteriors in Proposition 3.2 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 4.6 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.

Transparency III. The analysis when the agent's effort is observable is essentially identical to that in Section 4.2: the agent chooses the maximal effort $\overline{e}$, regardless of the signal structure, and the sender chooses a signal as if there is no moral hazard. Because $\overline{e} < \theta$, and the sender's payoff is a step function at θ, the optimal signal is supported on $\{0, \theta\}$.

This result shows in a particularly simple fashion that the receiver may prefer not to observe the agent's effort, despite the fact that the agent's effort under transparency is higher. To be concrete, suppose that the receiver's payoff as a function of his belief μ is given by $\max\{\mu - \theta, 0\}$; for example, the receiver has a binary action and faces the usual tradeoff between Type I and Type II error.[18] In this case, if the agent's effort is observable, then the receiver's expected payoff is equal to 0. In contrast, Proposition 4.6 shows that the receiver obtains a strictly positive expected payoff, because posterior belief $\mu = 1$ is realized with positive probability.

[18]For example, suppose that the receiver has two actions, approve or reject, and his payoffs are given as follows: rejecting when $\omega = 1$ yields payoff θ, approving when $\omega = 2$ gives payoff $1-\theta$, and making a mistake (approving when $\omega = 1$ or rejecting when $\omega = 2$) leads to payoff 0. Note that the agent's payoff is continuous in μ, which cannot be generated by binary actions but can arise if the receiver's action is two-dimensional, one dimension determining the sender's payoff and the other determining the agent's payoff.

5 BINARY ACTIONS

In this section, we analyze another tractable environment in which there are only two actions available to the receiver and both the sender and the agent prefer one action to the other. Since the underlying Bayesian persuasion problem is fully studied by Alonso and Câmara (2016), we adopt an analogous political interpretation to theirs: the receiver is a dictator (or a representative voter) who solely decides whether to keep the status quo or adopt a new policy.[19] The agent is a politician, while the sender is a party leader. Both want the new policy to be implemented. Different from Alonso and Câmara (2016), the sender communicates with the dictator, and the agent can improve the quality of the proposal (the prior belief about the merit of the policy) through effort.

5.1 THE MODEL

Setup. The receiver (dictator) decides whether to take action x_0 (keeping the status quo) or x_1 (adopting a new policy). If she selects x_0, then her payoff is $\theta > 0$, regardless of the state. If she selects x_1, then her payoff is v_ω in state ω. Without loss of generality, the states are ordered from the worst to the best for the receiver and the payoffs are normalized so that $v_1 = 0 < v_2 < ... < v_N$. The sender's (party leader) and the agent's (politician) payoffs depend on the receiver's action, and both prefer x_1: both receive 0 if $x = x_0$ and 1 if $x = x_1$.

Define vector $v_R \equiv (v_1, v_2, ..., v_N)$, which lists the receiver's payoffs by state. Then, the receiver prefers x_1 to x_0 if and only if $\langle \mu, v_R \rangle \geq \theta$ and, therefore,

$$v_A(\mu) = u_S(\mu) = \begin{cases} 1 & \text{if } \langle \mu, v_R \rangle \geq \theta, \text{ and} \\ 0 & \text{if } \langle \mu, v_R \rangle < \theta. \end{cases}$$

We refer to the set of beliefs at which the receiver selects x_i as $\mathcal{X}_i (\subset \Delta(\Omega))$ for $i \in \{0, 1\}$. We also refer to states ω such that $v_\omega < \theta$ as *rejection states*, states such that $v_\omega \geq \theta$ as *acceptance states*, and to the largest rejection state as the *rejection threshold*, ω_r.

Assumptions. To avoid trivialities, we assume that $\underline{\mu} \in \mathcal{X}_0$ and $\overline{\mu} \in \mathcal{X}_1$. In other words, in the absence of additional information, the receiver selects x_0 if she thinks that the agent makes no effort and x_1 if she thinks that the agent makes the maximal effort. For ease of exposition, we make four additional assumptions on $\underline{\mu}$ and $\overline{\mu}$.

[19] Alonso and Câmara (2016) consider a more general problem in which an action is chosen by a set of voters according to a fixed voting rule. Although we restrict attention to the dictator model, our subsequent analysis applies unchanged if the electorate can be represented by a representative voter. Alonso and Câmara (2016) derive conditions under which such a representation is possible.

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

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 from below. We refer to states such that $\underline{\mu}(\omega) > \overline{\mu}(\omega)$ as *bad-news states* and the other states as *good-news states*.

Assumption 2 *Let ω_e denote the largest ω such that $\underline{\mu}(\omega) > \overline{\mu}(\omega)$. Then, $\omega_e \leq \omega_r$.*

This assumption implies that an increase in effort increases the probability of all acceptance states (above ω_r), and in the case of a strict inequality, also increases the probability of some of the (higher) rejection states (between ω_e and ω_r).

Assumption 3

$$E_{\underline{\mu}}[v_\omega | \omega \neq 1] = \frac{\langle \underline{\mu}, v_R \rangle}{1 - \underline{\mu}(1)} \geq \theta.$$

This assumption states that even with belief $\underline{\mu}$, if the worst state (state 1) is ruled out, then the resulting belief vector induces x_1. This assumption is not essential, but streamlines the exposition considerably by reducing the number of equilibrium cases.

Assumption 4

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

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

Binary signals. With binary actions, the number of induced posteriors ($|supp(\tau)|$) does not need to exceed $|X| = 2$. In particular, the set of posterior belief realizations can always be partitioned into two subsets, those in $\mathcal{X}_0$ and those in $\mathcal{X}_1$. Replacing each subset with a single realization at its center of mass does not change the sender's expected payoff because both $\mathcal{X}_0$ and $\mathcal{X}_1$ are convex. Furthermore, because both (BP) and (IC) are linear in μ, the value attained by the constraints is also unaffected by the transformation. Thus, from now on, we focus on binary distributions. Clearly, $e > 0$ can be induced only when each realization induces a different action. In light of these observations, we fix the set of signal realizations with $\Sigma = \{b, g\}$ and let μ^- and μ^+ denote the posteriors corresponding to realizations b and g, respectively. In addition, we take for granted that $\mu^- \in \mathcal{X}_0$ and $\mu^+ \in \mathcal{X}_1$.

5.2 IMPLEMENTABLE EFFORTS

The following lemma is a counterpart to Proposition 4.1 for the binary-state model and fully characterizes the set of implementable efforts in the binary-action model.

Proposition 5.1 *In the binary-action model, e is implementable if and only if $e \leq \overline{e}$, where*

$$\sum_{\omega > \omega_e} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(\overline{e}). \tag{9}$$

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 (the good-news states, $\omega > \omega_e$) and *only* in such states. A binary signal which distinguishes between bad-news states and good-news states has this property: recall that, due to Assumption 3, any posterior belief supported only on the good-news states induces x_1, while, due to Assumption 2, all bad-news states lead to x_0. The left-hand side in equation (9) is the agent's marginal benefit of increasing e under this binary signal.

One intriguing implication of this result is that, in contrast to the binary-state model, a fully informative signal does not necessarily induce the maximal effort $\overline{e}$. The following result provides a necessary and sufficient condition under which it does lead to $\overline{e}$.

Corollary 5.2 *In the binary-action model, a fully informative signal induces $\overline{e}$ if and only if $\omega_e = \omega_r$.*

Proof. The result is immediate from Proposition 5.1 and the following observation: if a signal is fully informative then, by the definition of ω_r and Assumption 1, the receiver selects x_1 if and only if $\omega > \omega_r$. ■

If $\omega_e < \omega_r$, then an increase in effort increases not only the probability of all acceptance states but also the probability of some rejection states (between ω_e and ω_r). When a signal is fully informative, the former is beneficial to the agent, but the latter is not. This undermines the agent's incentive. If instead all states above ω_e are pooled together, then the receiver selects x_1 for all states above ω_e. Therefore, the agent benefits from increasing probabilities of all states above ω_e and has a stronger incentive to exert effort.

5.3 INCENTIVE-FREE EFFORT

In the binary-action model, the sender introduces dispersion in the distribution of posteriors even absent moral hazard, as long as the prior $\eta(e)$ belongs to the rejection region $\mathcal{X}_0$. Thus,

a positive effort can be induced even if the sender ignores the IC constraint. We solve for such an incentive-free effort level in two steps. We first consider a relaxed problem without (IC) and with an exogenously given $e \in [0, \overline{e}]$:

$$\max_{\tau \in \Delta(\Delta(\Omega))} E_\tau[u_S(\mu)] \text{ subject to } (BP) \;\; E_\tau[\mu] = \eta(e).$$

We then find the effort level for which (IC) holds. The following proposition provides a closed-form characterization for the incentive-free effort level.

Proposition 5.3 *In the binary-action model, there exists a unique incentive-free effort level $\underline{e} \in (0, \overline{e})$. Effort $\underline{e}$ is implemented by the following binary posterior distribution:*

$$\mu^- = (1, 0, ..., 0), \;\; \mu^+ = \frac{\eta(\underline{e}) \odot (1 - \underline{r}, 1, ..., 1)}{\langle \eta(\underline{e}), (1 - \underline{r}, 1, ..., 1) \rangle}, \;\; \textit{and } \tau(\mu^+) = 1 - \eta(1|\underline{e})\underline{r},$$

where

$$\underline{r} \equiv \frac{\theta - \langle \mu(\underline{e}), v_R \rangle}{\theta \eta(1|\underline{e})} \in (0, 1).$$

The incentive free effort level satisfies $(\underline{\mu}(1) - \overline{\mu}(1))\underline{r} = c'(\underline{e})$.

Proof. See the online appendix. ■

To understand this result, consider the following class of binary signals, which reveal state 1 with probability $r \in [0, 1]$ but provide no further information:

$$\pi(g|\omega) = \begin{cases} 1 - r & \text{if } \omega = 1, \text{ and} \\ 1 & \text{if } \omega > 1. \end{cases} \tag{10}$$

Clearly, realization b reveals state 1, while realization g generates Bayesian update

$$\mu^+ = \frac{\eta(e) \odot (1 - r, 1, ..., 1)}{\langle \eta(e), (1 - r, 1, ..., 1) \rangle} = \frac{1}{1 - r\eta(1|e)} \left((1 - r)\eta(1|e), \eta(2|e), ..., \eta(n|e) \right).$$

In addition, the signal generates realization b with probability $\tau(\mu^-) = \eta(1|e)r$ and realization g with probability $\tau(\mu^+) = 1 - \eta(1|e)r$. The optimality of this class of binary signals (in the absence of moral hazard) stems from Assumption 3: since ruling out state 1 with probability 1 moves the receiver's belief into the interior of the acceptance region $\mathcal{X}_1$, it suffices to reveal state 1 with a positive probability to induce x_1.

Within this class, the optimal signal is determined by the sender's incentive for information provision. On one hand, realization b leads to rejection. Therefore, the sender wants to minimize r. On the other hand, if r is too small, then the receiver does not become

sufficiently optimistic and will not choose x_1 following realization g. Thus, the sender would like to reveal state 1 just often enough that the receiver is indifferent between x_0 and x_1:

$$\langle \mu^+, v_R \rangle = \frac{1}{1-\eta(1|e)r} \langle \eta(e) \odot (1-r, 1, ..., 1), v_R \rangle = \theta.$$

This condition allows us to identify the optimal value of r (using the fact that, since $v_1 = 0$, $\langle \eta(e) \odot (1-r, 1, ..., 1), v_R \rangle = \langle \eta(e), v_R \rangle$):

$$r = \frac{\theta - \langle \eta(e), v_R \rangle}{\theta \eta(1|e)}. \tag{11}$$

The above discussion suggests that given e, the binary signal of the form in equation (10) with r in equation (11) is optimal in the absence of (IC). For an incentive-free effort, $\underline{e}$, this optimal signal also satisfies (IC). Therefore, $\underline{e}$ is (uniquely) determined by the following equation:

$$\tau(\mu^+) E_{\mu^+} \left[\frac{\eta_e(\omega|\underline{e})}{\eta(\omega|\underline{e})} \right] = c'(\underline{e}) \Rightarrow \langle \overline{\mu} - \underline{\mu}, (1-\underline{r}, 1, ..., 1) \} \rangle = (\underline{\mu}(1) - \overline{\mu}(1)) \underline{r} = c'(\underline{e}).$$

Example with three states. Figure 5 provides a graphical illustration of Proposition 5.3 for the case of three states, where beliefs can be represented by a 2-dimensional simplex,[20] effort increases the probabilities of states 2 and 3, but only state 3 is an acceptance state (i.e., $\omega_e = 1$ and $\omega_r = 2$). The shaded blue region represents the acceptance region $\mathcal{X}_1 \equiv \{\mu : \langle \mu, v_R \rangle \geq \theta\}$. State distributions $\underline{\mu}$ and $\overline{\mu}$ are represented by the hollow red dots, and increased effort moves the prior $\eta(e) = (1-e)\underline{\mu} + e\overline{\mu}$ along the red line segment.

In order to find an incentive-free effort, fix an effort level e. In the solution of the relaxed problem, we seek a hyperplane, $\lambda_0 + \langle \lambda_1, \mu \rangle$, that supports the sender's objective function inside $\Delta(\Omega)$. Since $v_S(\mu) = 1$ if $\mu \in \mathcal{X}_1$ (the blue shaded region), while $v_S(\mu) = 0$ if $\mu \in \mathcal{X}_0$, such a hyperplane must touch the sender's objective at two locations: $(1, 0, 0)$, which is the farthest point from $\eta(e)$ among all vectors in $\mathcal{X}_0$, and the lower boundary of $\mathcal{X}_1$ (defined by $\langle \mu, v_R \rangle = \theta$), which are closest to $\eta(e)$ among all vectors in $\mathcal{X}_1$, as illustrated in the right panel.[21] In other words, $\lambda_0 + \langle \lambda_1, (1, 0, 0) \rangle = 0$, and $\lambda_0 + \langle \lambda_1, \mu \rangle = 1$ whenever $\langle \mu, v_R \rangle = \theta$. It follows that $\lambda_0 = 0$ and $\lambda_1 = (1/\theta) v_R$. Jointly imposing the binary-signal restriction and (BP), we find that that $\mu^- = (1, 0, 0)$, and μ^+ is the unique point that intersects the boundary of $\mathcal{X}_1$ and the extended line that connects $(1, 0, 0)$ and $\eta(e)$.

The search for an incentive-free effort level amounts to varying e continuously from 0 to

[20] We adopt a canonical interpretation of the 2-dimensional simplex, as in Mas-Colell et al. (1995), p. 169.

[21] Notice that this is effectively identical to concavifying the graph $\{(\mu, v_S(\mu)) : \mu \in \Delta(\Omega)\}$, which gives rise to the same hyperplane over the shaded red part of $\mathcal{X}_0$ and the flat (constant) hyperplane over $\mathcal{X}_1$.

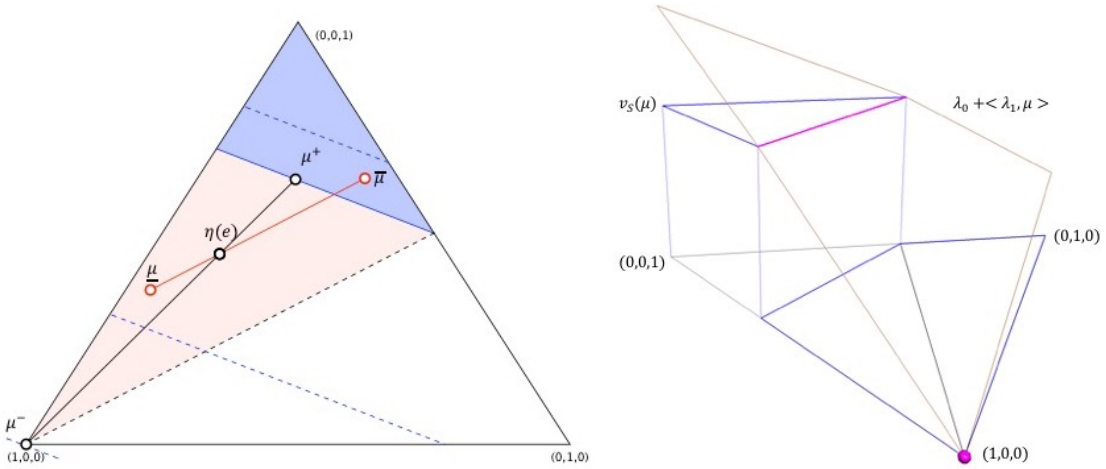

Figure 5: Illustration of Proposition 5.3.

$\overline{e}$ (which varies $\eta(e)$ along the red line segment) and finding e such that the optimal signal associated with $\eta(e)$ also satisfies (IC). Such a point exists: at $\underline{\mu}$, the agent has an incentive to exert positive effort, because the marginal cost is 0 when $e = 0$ (i.e., $c'(0) = 0$), while effort linearly increases the probability of μ^+. At $\eta(\overline{e})$, the signal characterized above does not provide the right incentive for the agent to exert effort $\overline{e}$: recall that the agent chooses $\overline{e}$ only when action x_1 is taken if and only if $\omega > \omega_e$ (see Proposition 5.1) but, because μ^+ is on the boundary of $\mathcal{X}_1$, action x_1 is selected with a positive probability even when $\omega = 1$.

5.4 OPTIMAL SIGNAL

We now study optimal information design for $e \in (\underline{e}, \overline{e})$. In order to facilitate the analysis, as well as simplify the notation, we partition the set $[\underline{e}, \overline{e}]$ as follows: for each $k \leq \omega_e$, let e_k be the unique value that satisfies

$$\sum_{\omega > k} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(e_k). \tag{12}$$

Notice that the left-hand side is the marginal benefit of effort under a signal that separates between states weakly below k and those above k. Therefore, e_k is the effort level induced by such a signal. Proposition 5.3 implies that $\underline{e} < e_1$ (because the optimal signal for $\underline{e}$ reveals state 1 with probability less than 1), while Proposition 5.1 implies that $\overline{e} = e_{\omega_e}$. In addition, Assumptions 1 and 2 ensure that the left-hand side increases in k as long as $k \leq \omega_e$ and, therefore, $e_k < e_{k+1}$ for any $k = 1, ..., \omega_e - 1$. For notational convenience, we define $e_0 \equiv \underline{e}$.

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

Proposition 5.4 *In the binary-action model, for $k = 1, ..., \omega_e$, effort $e \in (e_{k-1}, e_k]$ is optimally implemented by the following binary distribution of posteriors:*

$$\mu^- = \frac{\eta(e) \odot \rho^-}{\langle \eta(e), \rho^- \rangle}, \ \mu^+ = \frac{\eta(e) \odot \rho^+}{\langle \eta(e), \rho^+ \rangle}, \ \textit{and } \tau(\mu^+, e) = \langle \eta(e), \rho^+ \rangle,$$

where $\rho^+ = (1, ..., 1) - \rho^- \in \mathcal{R}^{n+1}$ and

$$\rho^-(\omega) = \begin{cases} 1 & \textit{for} \quad 1 \leq \omega \leq k-1, \\ \frac{c'(e) - c'(e_{k-1})}{\underline{\mu}(k) - \overline{\mu}(k)} & \textit{for} \quad \omega = k, \ \textit{and} \\ 0 & \textit{for} \quad k+1 \leq \omega \leq n. \end{cases}$$

Proof. See the online appendix. ■

The optimal signal varies continuously and systematically, revealing more bad-news states, as e increases from $\underline{e}(= e_0)$ to $\overline{e}(= e_{\omega_e})$. In addition, from e_{k-1} to e_k, the probability of revealing state k (i.e., inducing μ^-, so that the receiver selects x_0) continuously increases from 0 to 1: notice that $c'(e) - c'(e_{k-1})$ increases from 0 to $c'(e_k) - c'(e_{k-1}) = \underline{\mu}(k) - \overline{\mu}(k)$. This is due to the underlying Bayesian persuasion problem: as shown in the analysis for the incentive-free effort, it is optimal for the sender to reveal only the lowest states. On the other hand, moral hazard forces the sender to reveal state 1 with a higher probability and, if $e > e_1$, some other low states as well. This is a distortion from a pure information-provision perspective, but necessary to generate an incentive for the agent to exert more effort than $\underline{e}$.

The receiver benefits from the distortion: in the absence of moral hazard, the receiver strictly prefers x_0 to x_1 following realization b and is indifferent between x_0 and x_1 following realization g (i.e., $\langle \mu^+, v_R \rangle = \theta$) and, therefore, the receiver derives no benefit from communication. With moral hazard, the receiver's belief following realization g is such that $\langle \mu^+, v_R \rangle > \theta$ and, therefore, he strictly prefers x_1 to x_0.

Example with three states. Consider the same three-state example as in Section 5.3, in which $\omega_e = 1$ and, therefore, $\overline{e} = e_1$. To implement $e \in (\underline{e}, \overline{e})$, it is necessary to introduce more dispersion into the distribution of posteriors. However, $\mu^- = (1, 0, 0)$ is already an extremal point and, therefore, μ^+ must move further apart along the extended line that crosses $(1, 0, 0)$ and $\eta(e)$. By implication, the sender must reveal state 1 with a higher probability, moving μ^+ into the interior of $\mathcal{X}_1$, as depicted in the left panel of Figure 6.

To see how our general characterization (Proposition 3.2) applies to this problem, recall that we are seeking a hyperplane $\lambda_0 + \langle \lambda_1, \mu \rangle$ that supports $\mathcal{L}(\mu, \psi)$ inside $\Delta(\Omega)$. Since

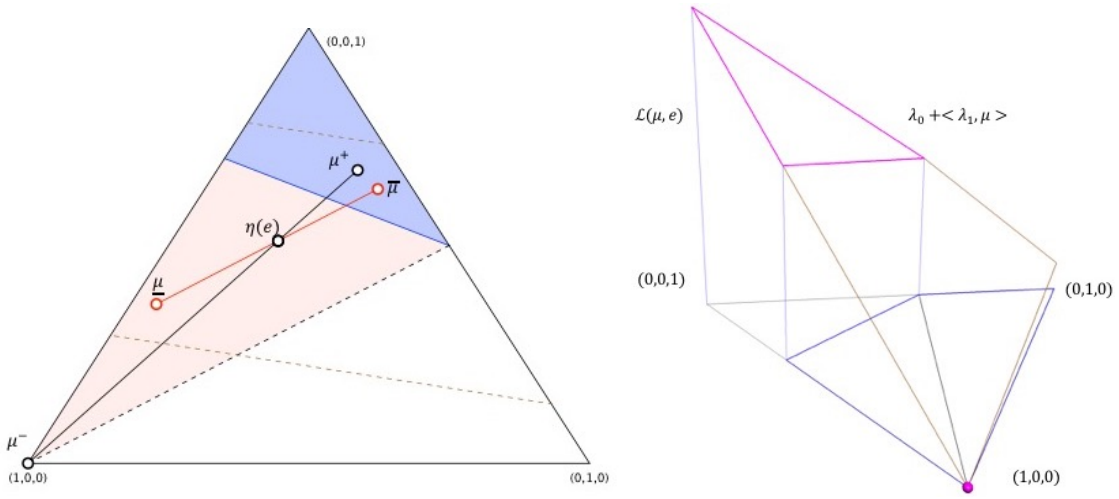

Figure 6: Illustration of Proposition 5.4.

$v_A(\mu) = 0$ if $\mu \in \mathcal{X}_0$, $v_A(\mu) = 1$ if $\mu \in \mathcal{X}_1$, and $E_\mu[\eta_e(\omega|e)/\eta(\omega|e)] = \langle \eta_e(e) \oslash \eta(e), \mu \rangle$,

$$h(\mu) = \begin{cases} -c'(e) & \text{if } \mu \in \mathcal{X}_0, \text{ and} \\ \langle \mu, \eta_e(e) \oslash \eta(e) \rangle - c'(e), & \text{if } \mu \in \mathcal{X}_1. \end{cases}$$

Combining this with the same discrete structure for $v_S(\mu)$ yields

$$\mathcal{L}(\mu, \psi) = \begin{cases} -\psi c'(e) & \text{if } \mu \in \mathcal{X}_0, \text{ and} \\ 1 + \psi \left(\langle \mu, \eta_e(e) \oslash \eta(e) \rangle - c'(e) \right), & \text{if } \mu \in \mathcal{X}_1. \end{cases}$$

Notice that $\mathcal{L}$ is linear in $\mu \in \mathcal{X}_1$, and therefore, it defines a hyperplane over $\mathcal{X}_1$. Thus, there are three possibilities for our supporting hyperplane $\lambda_0 + \langle \lambda_1, \mu \rangle$, depending on ψ:[22]

(i) $\lambda_0 + \langle \lambda_1, \mu \rangle$ meets the lower boundary of $\mathcal{X}_1$ (i.e., $\{\mu \in \mathcal{X}_1 : \langle \mu, v_R \rangle = \theta\}$).

(ii) $\lambda_0 + \langle \lambda_1, \mu \rangle$ meets the upper boundary of $\mathcal{X}_1$ (i.e., $\{\mu \in \mathcal{X}_1 : \mu(1) = 0\}$).

(iii) $\lambda_0 + \langle \lambda_1, \mu \rangle$ coincides with $\mathcal{L}$ over $\mathcal{X}_1$.

In the first case, μ^+ belongs to the lower boundary of $\mathcal{X}_1$. As explained in Section 5.3, this does not provide a sufficient incentive for the agent to choose $e > \underline{e}$. To the contrary, in the second case, μ^+ belongs to the upper boundary of $\mathcal{X}_1$, which induces $\overline{e}$, as discussed in Section 5.2. Therefore, (iii) is the relevant case for $e \in (\underline{e}, \overline{e})$. In other words, at the optimal

[22] If ψ is sufficiently small, then $\mathcal{L}$ increases slowly over $\mathcal{X}_1$ and, therefore, the supporting hyperplane touches $\mathcal{L}$ at $(1, 0, 0)$ and the lower boundary of $\mathcal{X}_1$, just as in the case for the incentive-free effort. If ψ is sufficiently large, then $\mathcal{L}$ increases fast over $\mathcal{X}_1$, in which case the supporting hyperplane stays strictly above $\mathcal{L}$ at any interior value of μ. If ψ is just right, then the last case arises.

solution, ψ is such that the supporting hyperplane $\lambda_0 + \langle \lambda_1, \mu \rangle$ exactly coincides with $\mathcal{L}$ over $\mathcal{X}_1$, as depicted in the right panel of Figure 6. This observation allows us to fully determine the optimal values of λ_0, λ_1, and ψ. The optimal value of μ^+ can also be found from the two equality constraints: $E_\tau[\mu] = \eta(e)$ and $E_\tau[h(\mu)] = 0$.

6 CONCLUSION

In this paper, we study Bayesian persuasion when the prior about the underlying state is generated by an agent's unobservable effort. Thus, the sender is concerned with both information provision and incentive provision in her information design. We show that the sender's problem can be analyzed by extending the concavification technique in Aumann and Maschler (1995) and KG and provide a useful necessary and sufficient condition for an optimal signal (Section 3). We apply the general characterization to two tractable environments, one with binary states (Section 4) and the other with binary receiver actions (Section 5), and derive a number of concrete and economic implications. We also show that transparency, which allows the receiver to observe the agent's effort, may reduce the informativeness of the equilibrium signal and harm the receiver.

Our framework can be used to provide insights into a variety of economic problems. For example, both information and incentive provision are relevant when a credit ratings agency interacts with security issuers and investors, or when a retailer deals with both producers and consumers. Furthermore, our main technique can be applied to Bayesian persuasion problems with other kinds of frictions, provided that the relevant conditions can be written as expectations with respect to the distribution of posterior beliefs.

APPENDIX: OMITTED PROOFS

Continuation of Proof of Proposition 3.2. For the result on the cardinality of the support of τ^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 τ such that $y = (E_\tau[\mu(2)], ..., E_\tau[\mu(N)], E_\tau[h(\mu)], E_\tau[v_S(\mu)])$. $V^e = \max_v \{v : (\eta(2|e), ..., \eta(N|e), 0, v) \in co(\underline{K}^e)\}$. Because the graphs of $h(\cdot|e)$ and $v_S(\cdot)$ are closed, $\underline{K}^e$ is closed, and hence, $co(\underline{K}^e)$ is closed. Thus, $(\eta(2|e), ..., \eta(N|e), 0, V^e)$ is on the boundary of $co(\underline{K}^e)$. By Caratheodry's thoerm, 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, $(\eta(2|e),...,\eta(N|e),0,V^e)$ can be made of at most $N+1$ elements in $\underline{K}^e$. ■

Continuation of Proof of Proposition 3.3. *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 μ such that $\tau^e(\mu)>0$. Suppose that $\mathcal{L}(\mu,\psi)<\lambda_0+\langle\lambda_1,\mu\rangle$ for some μ such that $\tau^e(\mu)>0$. Because $\mathcal{L}(\mu,\psi)\leq\lambda_0+\langle\lambda_1,\mu\rangle$ for all $\mu\in\Delta(\Omega)$, it follows that $E_{\tau^e}[\mathcal{L}(\mu,\psi)]<\lambda_0+\langle\lambda_1,\eta(e)\rangle$. 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 $v_S(\mu)+\psi h(\mu)\leq\lambda_0+\langle\lambda_1,\mu\rangle$ for all $\mu\in\Delta(\Omega)$, then for any τ,

$$E_\tau[v_S(\mu)]+\psi E_\tau[h(\mu)]\leq\lambda_0+E_\tau[\langle\lambda_1,\mu\rangle].$$

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

Proof of Proposition 4.1. We first show that $\overline{e}$ is the upper bound to the set of implementable effort levels. Under any signal π, the agent chooses e to maximize

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

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

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

Since v_A is weakly increasing,

$$(\overline{\mu}-\underline{\mu})\left(\sum_s\pi(s|2)v_A(\mu_s)-\sum_s\pi(s|1)v_A(\mu_s)\right)\leq(\overline{\mu}-\underline{\mu})((\overline{v}_A(2)-\overline{v}_A(1))=\overline{\mu}-\underline{\mu}.$$

These imply that e such that $c'(e)>\overline{\mu}-\underline{\mu}$, which is equivalent to $e>\overline{e}$, 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.

Finally, we show that a fully informative signal is a necessary condition for implementing $\overline{e}$. If $\overline{v}_A(2)-\overline{v}_A(1)=1$, then $\overline{v}_A(2)=1$ and $\overline{v}_A(1)=0$, which can be satisfied only when $v(\mu_s)=1$ for any s such that $\pi(s|2)>0$ and $v(\mu_s)=0$ for any s such that $\pi(s|1)>0$. Thus, the posterior belief distribution is supported on $\{0,1\}$, and hence, it is fully informative. ■

Proof of Corollary 4.2. The first (convex) result is immediate from the fact that $\underline{e} = \overline{e}$ if v_S is convex. For the second (concave) result, first notice that if v_S is concave then $\underline{e} = 0$, which implies that the sender's expected utility is equal to $v_S(\underline{\mu})$ if she adopts a fully uninformative signal. 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. Under this signal, the sender's expected payoff is

$$E_\tau[v_S(\mu)] = \tau(\underline{\mu})v_S(\underline{\mu}) + \tau(1)v_S(1) > v_S(\underline{\mu}),$$

because $\eta(e) > \mu$ only when $\tau(1) > 0$. ■

Proof of Proposition 4.3. 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 $\widehat{\tau}(\mu, e)$ denote the solution to the sender's relaxed problem and $\widehat{V}(e)$ denote the corresponding seller payoff. Applying Proposition 3.3 to the relaxed problem (without (IC) and with $\psi = 0$), there exist $\widehat{\lambda}_0$ and $\widehat{\lambda}_1$ such that

$$v_S(\mu) \leq \widehat{\lambda}_0 + \widehat{\lambda}_1\mu \text{ for all } \mu \in [0,1], \text{ with equality if } \widehat{\tau}(\mu, e) > 0. \tag{13}$$

For later use, define $H(e) \equiv E_{\widehat{\tau}}[h(\mu)]$. Similarly, let $\tau(\mu, e)$ denote the solution to the sender's original (unrelaxed) problem and apply Proposition 3.3. Then, there exist λ_0, λ_1, and ψ such that

$$v_S(\mu) + \psi h(\mu) \leq \lambda_0 + \lambda_1\mu \text{ for all } \mu \in [0,1], \text{ with equality if } \widehat{\tau}(\mu, e) > 0. \tag{14}$$

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

Step 1. We show that if $H(e) \equiv E_{\widehat{\tau}}[h(\mu)] < 0$ then $\psi > 0$. Taking the expectation of equation (14) with respect to τ (not $\widehat{\tau}$), we get

$$E_\tau[v_S(\mu)] + \psi E_\tau[h(\mu)] = \lambda_0 + \lambda_1 E_\tau[\mu] = \lambda_0 + \lambda_1 e,$$

where the last equality is due to (BP). Doing the same with equation (13), we get

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

Combining these two equations, we find that

$$\lambda_0 + \lambda_1 e \leq \widehat{\lambda}_0 + \widehat{\lambda}_1 e. \tag{15}$$

Now taking the expectations of equations (13)) and (14) with respect to $\widehat{\tau}$, we have

$$\begin{aligned} &E_{\widehat{\tau}}[v_S(\mu)] = \widehat{\lambda}_0 + \widehat{\lambda}_1 E_{\widehat{\tau}}[\mu] = \widehat{\lambda}_0 + \widehat{\lambda}_1 e \text{ and} \\ &E_{\widehat{\tau}}[u_S(\mu)] + \psi E_{\widehat{\tau}}[h(\mu)] \leq \lambda_0 + \lambda_1 E_\tau[\mu] = \lambda_0 + \lambda_1 e. \end{aligned}$$

Combining these two conditions, we get

$$\widehat{\lambda}_0 + \widehat{\lambda}_1 e + \psi H(e) \leq \lambda_0 + \lambda_1 e. \tag{16}$$

Subtracting (15) from (16),

$$\widehat{\lambda}_0 + \widehat{\lambda}_1 e + \psi H(e) - (\widehat{\lambda}_0 + \widehat{\lambda}_1 e) \leq \lambda_0 + \lambda_1 e - (\lambda_0 + \lambda_1 e) \Rightarrow \psi H(e) \leq 0,$$

which implies that if $H(e) < 0$ then $\psi \geq 0$. 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 $H(e) < 0$. Suppose $H(e) \equiv E_{\widehat{\tau}}[h(\mu)] > 0$.

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

Case 2: Suppose that for all solution(s) of the relaxed problem, $H(e) > 0$. 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, $H(e) = -c'(e) < 0$. In addition, there must exist $\mu_L < e < \mu_H$ such that $\widehat{\tau}(\mu_L, e) > 0$ and $\widehat{\tau}(\mu_H, e) > 0$: otherwise, (BP) cannot hold. Now, applying Proposition 3.3 to the relaxed problem, we have

$$\widehat{\lambda}_0 + \widehat{\lambda}_1 \mu \geq v_S(\mu), \text{ with equality if } \mu = \mu_L, \mu_H. \tag{17}$$

Now 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 (17) applies regardless of $\tilde{e}$, $\tilde{t}$ is an optimal solution to the sender's relaxed problem. Next, observe that

$$H(\tilde{e}) = \frac{1}{\widetilde{e}(1-\widetilde{e})}\Big(\frac{\mu_H - \widetilde{e}}{\mu_H - \mu_L}(\mu_L - \widetilde{e})v_A(\mu_L) + \frac{\widetilde{e} - \mu_L}{\mu_H - \mu_L}(\mu_H - \widetilde{e})v_A(\mu_H)\Big) - c'(\widetilde{e}).$$

By assumption, with any solution to the relaxed problem, $H(e) > 0$. To the contrary, it is clear from direct substitutions that if $\widetilde{e} = \mu_H$ then $H(\widetilde{e}) < 0$. Since $H(\tilde{e})$ is a continuous function, there must exist $e^* \in (e, \mu_H)$ such that the solution to the relaxed problem satisfies (IC), that is, τ 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 5.1. Since $\eta_e(e) = \overline{\mu} - \underline{\mu}$ and the agent's payoff depends only on

whether the receiver selects x_1 or not, the agent's first-order condition (3) can be written as

$$\langle \overline{\mu} - \underline{\mu}, Pr\{x_1|\cdot\}\rangle - c'(e) = 0, \tag{18}$$

where $Pr\{x_1|\cdot\}$ denotes the probabilities that action x_1 is selected depending on the state. By Assumptions 1 and 2, the first term in (18) is bounded from above by

$$\langle \overline{\mu} - \underline{\mu}, Pr\{x_1|\cdot\}\rangle \leq \sum_{i>\omega_e} (\overline{\mu}(\omega) - \underline{\mu}(\omega)).$$

Since $c'(e)$ is strictly increasing, equation (18) cannot hold for any $e > \overline{e}$. Notice also that Assumption 4 ensures that $\overline{e} < 1$.

We now show that $\overline{e}$ can be induced by the following binary signal:

$$\pi(g|\omega) = 1 - \pi(b|\omega) = \begin{cases} 0 & \text{if } \omega \leq \omega_e, \text{ and} \\ 1 & \text{if } \omega > \omega_e. \end{cases}$$

This signal distinguishes between bad-news states (that become less likely as e increases) and good-news states (that become more likely as e increases) and correctly reports which subset contains the true state. By construction, this signal attains the upper bound of the marginal benefit of effort above and, therefore, induces $\overline{e}$ as long as the receiver selects x_0 following realization b and x_1 following realization g. The former (x_0 after realization b) follows from the fact that realization b reveals that $\omega \leq \omega_e \leq \omega_r$ (i.e., the true state is a rejection state for sure), while the latter (x_1 after realization g) is due to Assumptions 1 and 3: even when the prior is $\underline{\mu}$, the receiver is willing to take x_1 as soon as state 1 is excluded, but realization g rules out weakly more rejection states from 1 to ω_e (without ruling out any acceptance states). In addition, an increase in e makes higher states to be realized with higher states, which further strengthens the receiver's incentive to select x_1. As in the binary-state model, a convex combination of this signal and one which reveals no information implements any effort below $\overline{e}$. ■

REFERENCES

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

Au, Pak Hung and Keiichi Kawai, "Competitive Information Disclosure by Multiple Senders," 2017.

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

Barron, Daniel, George Georgiadis, and Jeroen Swinkels, "Optimal contracts with a risk-taking agent," *mimeo*, 2017.

Bénabou, Roland and Jean Tirole, "Incentives and prosocial behavior," *American economic review*, 2006, *96* (5), 1652–1678.

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

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

Besley, Timothy and Andrea Prat, "Handcuffs for the grabbing hand? The role of the media in political accountability," *American Economic Review*, 2006, *96* (3), 720–736.

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

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

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang, "Pivotal persuasion," *mimeo*, 2017.

Daughety, Andrew F and Jennifer F Reinganum, "Public goods, social pressure, and the choice between privacy and publicity," *American Economic Journal: Microeconomics*, 2010, *2* (2), 191–221.

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

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

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

_ , 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.

Georgiadis, George and Balazs Szentes, "Optimal Monitoring Design," *mimeo*, 2018.

Guo, Yingni and Eran Shmaya, "The interval structure of optimal disclosure," *mimeo*, 2017.

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

Hörner, Johannes and Nicolas Lambert, "Motivational ratings," *mimeo*, 2016.

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

Kolotilin, Anton, "Optimal information disclosure: a linear programming approach," *Theoretical Economic, forthcoming*, 2017.

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

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

Li, Fei and Peter Norman, "On Bayesian persuasion with multiple senders," *mimeo*, 2015.

Mas-Colell, Andreu, Michael Dennis Whinston, Jerry R Green et al., *Microeconomic theory*, Vol. 1, Oxford university press New York, 1995.

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

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

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

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

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

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

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

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

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

Online Appendix

Proof of Proposition 5.3. We prove the result in three steps. First, we define a useful function $r(e)$, which determines the probability of revealing state 1, and establish its basic properties. Second, we show that a specific signal is a unique binary solution to the relaxed problem without (IC). Finally, we show that there exists a unique incentive-effort level $\underline{e}$.

Step 1. Let $\tilde{e} \in [0, \overline{e}]$ be the maximal value such that $\langle \eta(\tilde{e}), v_R \rangle \leq \theta$. Then, let

$$r(e) \equiv \frac{\theta - \langle \eta(e), v_R \rangle}{\theta \eta(1|e)} \text{ for each } e \in [0, \tilde{e}].$$

We show that $r(e) \in [0,1]$ for all $e \leq \overline{e}$. $r(e) \geq 0$ follows from the fact that $\langle \eta(e), v_R \rangle \leq \langle \eta(\tilde{e}), v_R \rangle \leq \theta$. To show $r(e) \leq 1$, note that

$$\begin{aligned} \frac{\theta - \langle \eta(e), v_R \rangle}{\theta \eta(1|e)} \leq 1 \quad &\Longleftrightarrow \quad \theta - \theta \eta(1|e) \leq \langle \eta(e), v_R \rangle \\ &\Longleftrightarrow \quad \theta(1 - \eta(1|e)) = \theta \langle \eta(e), (0, 1, ..., 1) \rangle \leq \langle \eta(e), v_R \rangle \\ &\Longleftrightarrow \quad \frac{\langle \eta(e), v_R \rangle}{\eta(1|e)} = \frac{\langle \eta(e), v_R \rangle}{\langle \eta(e), (0, 1, ..., 1) \rangle} \geq \theta. \end{aligned}$$

For the inequality, recall the maintained assumption that $\langle \overline{\mu}, v_R \rangle > \theta$ and Assumption 3:

$$\frac{\langle \underline{\mu}, v_R \rangle}{1 - \underline{\mu}(1)} = \frac{\langle \underline{\mu}, v_R \rangle}{\langle \underline{\mu}, (0, 1, ..., 1) \rangle} \geq \theta$$

Combining these with the facts that $\eta(e) = (1-e)\underline{\mu} + e\overline{\mu}$ and $1 = \langle \overline{\mu}, (1, ..., 1) \rangle \geq \langle \overline{\mu}, (0, 1, ..., 1) \rangle$, the desired result is obtained as follows:

$$\begin{aligned} \langle \eta(e), v_R \rangle &= \langle (1-e)\underline{\mu} + e\overline{\mu}, v_R \rangle = (1-e)\langle \underline{\mu}, v_R \rangle + e \langle \overline{\mu}, v_R \rangle \\ &\geq (1-e)\theta \langle \underline{\mu}, (0, 1, ..., 1) \rangle + e\theta \langle \overline{\mu}, (1, 1, ..., 1) \rangle \\ &\geq (1-e)\theta \langle \underline{\mu}, (0, 1, ..., 1) \rangle + e\theta \langle \overline{\mu}, (0, 1, ..., 1) \rangle \\ &= \theta \langle (1-e)\underline{\mu} + e\overline{\mu}, (0, 1, ..., 1) \rangle = \theta \langle \eta(e), (0, 1, ..., 1) \rangle. \end{aligned}$$

Step 2. We show that for any $e \in [0, \widetilde{e})$, the binary solution of the relaxed problem is

$$\mu^- = (1, 0, ..., 0), \ \mu^+(e) = \frac{\eta(e) \odot (1 - r(e), 1, ..., 1)}{\langle \eta(e), (1 - r(e), 1, ..., 1) \rangle}, \text{ and } \tau(\mu^+(e), e) = 1 - \eta(1|e) r(e).$$

We first show that this signal satisfies (BP). Since $\tau(\mu^+(e), e) = \langle \eta(e), (1 - r(e), 1, ..., 1) \rangle$,

$$\begin{aligned} &\tau(\mu^+(e), e)\mu^+(e) + (1 - \tau(\mu^+(e), e))\mu^- \\ = \ &\eta(e) \odot (1 - r(e), 1, ..., 1) + \eta(1|e) r(e) (1, 0, ..., 0) \\ = \ &\eta(e) \odot (1 - r(e), 1, ..., 1) + \eta(e) \odot (r(e), 1, ..., 1) \\ = \ &\langle \eta(e), (1, ..., 1) \rangle = \eta(e). \end{aligned}$$

To verify the optimality of the signal, we apply a necessary and sufficient condition in Proposition 3.3 with $\psi = 0$. In other words, we show that there exists a hyperplane that supports $v_S(\mu)$ inside $\Delta(\Omega)$. Let $\lambda_0 = 0$ and $\lambda_1 = 1/\theta \cdot v_R$, so that

$$\lambda_0 + \langle \lambda_1, \mu \rangle = \lambda_0 + \langle \frac{1}{\theta} v_R, \mu \rangle = \frac{1}{\theta} \langle \mu, v_R \rangle.$$

For $\mu \in \mathcal{X}_0$, $v_S(\mu) = 0$, and therefore $v_S(\mu) \leq \lambda_0 + \langle \lambda_1, \mu \rangle = \frac{1}{\theta} \langle v_R, \mu \rangle$, with equality holding if and only if $\mu = \mu^- = (1, 0, ..., 0)$. For $\mu \in \mathcal{X}_1$, $v_S(\mu) = 1$, while $\lambda_0 + \langle \lambda_1, \mu \rangle = 1/\theta \cdot \langle v_R, \mu \rangle \geq 1$ with equality holding if and only if $\langle \mu, v_R \rangle = \theta$. It suffices to show that $\mu^+(e)$ belongs to the point where $u_S(\mu) = \lambda_0 + \langle \lambda_1, \mu \rangle$. This follows from

$$\begin{aligned} \langle \mu^+(e), v_R \rangle &= \langle \frac{\eta(e) \odot (1 - r(e), 1, ..., 1)}{\langle \eta(e), (1 - r(e), 1, ..., 1) \rangle}, v_R \rangle \\ &= \frac{\langle \eta(e), v_R \rangle}{1 - \eta(1|e) r(e)} = \frac{\langle \eta(e), v_R \rangle}{1 - \frac{\theta - \langle \eta(e), v_R \rangle}{\theta}} = \frac{\theta \langle \eta(e), v_R \rangle}{\langle \eta(e), v_R \rangle} = \theta. \end{aligned}$$

Note that we use the fact that, since $v_R(0) = 0$, $\langle \eta(e), (1 - r(e), 1, ..., 1) \rangle = \eta(e)$ in the second inequality and apply the definition of $r(e)$, given in Step 1, in the third equality.

Step 3. We now prove that there exists a unique value of $\underline{e} \in (0, \overline{e})$ such that the optimal binary distribution defined above satisfies (IC). Note that for the optimal binary signal,

$$\begin{aligned} E_\tau[h(\mu)] &= \tau(\mu^+(e), e) \langle \mu^+(e), (\overline{\mu} - \underline{\mu}) \oslash \eta(e) \rangle - c'(e) \\ &= \langle \tau(\mu^+(e), e) \mu^+(e), (\overline{\mu} - \underline{\mu}) \oslash \eta(e) \rangle - c'(e) \\ &= \langle \eta(e) \odot (1 - r(e), 1, ..., 1), (\overline{\mu} - \underline{\mu}) \oslash \eta(e) \rangle - c'(e) \\ &= \langle (1 - r(e), ..., 1), \overline{\mu} - \underline{\mu} \rangle - c'(e) = r(e)(\underline{\mu}(1) - \overline{\mu}(1)) - c'(e), \end{aligned}$$

where the last equality is due to $\langle (1, ..., 1), \overline{\mu} - \underline{\mu} \rangle = 0$. $E_\tau[h(\mu)]$ is positive if $e = 0$ (because $r(0) > 0$ while $c'(0) = 0$) and negative if $e = \tilde{e}$ (because if $\tilde{e} < \overline{e}$ then $r(\tilde{e}) = 0$, while if $\tilde{e} = \overline{e}$ then $r(\tilde{e})(\underline{\mu}(1) - \overline{\mu}(1)) < \sum_{\omega > \omega_e} (\overline{\mu}(\omega) - \underline{\mu}(\omega)) = c'(\overline{e})$). Since $E_\tau[h(\mu)]$ is continuous in e, there exists e such that $E_\tau[h(\mu)] = 0$, that is, (IC) holds. For uniqueness, notice that $c'(e)$ increases in e. Therefore, it is sufficient that $r(e)$ decreases in e, which we establish below.

Observe that

$$\begin{aligned} & r'(e) = \frac{-\langle \overline{\mu} - \underline{\mu}, v_R \rangle \theta \eta(1|e) - (\theta - \langle \eta(e), v_R \rangle) \theta (\overline{\mu}(1) - \underline{\mu}(1))}{(\theta \eta(1|e))^2} \leq 0 \\ \iff & \frac{\langle \overline{\mu} - \underline{\mu}, v_R \rangle}{\underline{\mu}(1) - \overline{\mu}(1)} \geq \frac{\theta - \langle \eta(e), v_R \rangle}{\eta(1|e)}. \end{aligned}$$

We prove this inequality by establishing the following two inequalities:

$$\frac{\langle \overline{\mu} - \underline{\mu}, v_R \rangle}{\underline{\mu}(1) - \overline{\mu}(1)} \geq \theta \text{ and } \quad \theta \geq \frac{\theta - \langle \eta(e), v_R \rangle}{\eta(1|e)}.$$

The second inequality is straightforward from the fact that $r(e) \leq 1$ (see Step 1):

$$\theta - \frac{\theta - \langle \eta(e), v_R \rangle}{\eta(1|e)} = \theta - \theta r(e) = \theta(1 - r(e)) \geq 0.$$

In order to establish the first inequality, notice that it can rewritten as

$$\langle \overline{\mu} - \underline{\mu}, v_R \rangle = \langle \overline{\mu} - \underline{\mu}, (0, v_1, ..., v_n) \rangle \geq \theta(\underline{\mu}(1) - \overline{\mu}(1)) = \theta \langle \overline{\mu} - \underline{\mu}, (0, 1, ..., 1) \rangle,$$

which is equivalent to

$$\langle \overline{\mu} - \underline{\mu}, (0, v_1 - \theta, ..., v_n - \theta) \rangle = \sum_{\omega \geq 1} (\overline{\mu}(\omega) - \underline{\mu}(\omega))(v_\omega - \theta) \geq 0.$$

The expression can be decomposed into three pieces as follows:

$$\sum_{\omega=1}^{\omega_e} (\overline{\mu}(\omega) - \underline{\mu}(\omega))(v_\omega - \theta) + \sum_{\omega=\omega_e+1}^{\omega_r} (\overline{\mu}(\omega) - \underline{\mu}(\omega))(v_\omega - \theta) + \sum_{\omega=\omega_r+1}^{n} (\overline{\mu}(\omega) - \underline{\mu}(\omega))(v_\omega - \theta) \geq 0.$$

Recall that $\overline{\mu}(\omega) - \underline{\mu}(\omega) < 0$ if and only if $\omega \leq \omega_e$, while $v_\omega < \theta$ if and only if $\omega \leq \omega_r$. Therefore, the first and the third terms are positive, while the second term is negative. We show that the sum of the second and the third terms is positive.

If $\omega_r = \omega_e$, then the second term is vacuous and, therefore, the result is straightforward. For the case where $\omega_e < \omega_r$, note that, by Assumption 1 (MLRP), $1 - \underline{\mu}(\omega)/\overline{\mu}(\omega)$ increases in ω. Combining this with the fact that $v_\omega - \theta < 0$ if and only if $\omega \leq \omega_r$,

$$\begin{aligned}
& \sum_{\omega=\omega_e+1}^{\omega_r} (\overline{\mu}(\omega) - \underline{\mu}(\omega))(v_\omega - \theta) + \sum_{\omega=\omega_r+1}^{n} (\overline{\mu}(\omega) - \underline{\mu}(\omega))(v_\omega - \theta) \\
= & \sum_{\omega=\omega_e+1}^{\omega_r} \left(1 - \frac{\underline{\mu}(\omega)}{\overline{\mu}(\omega)}\right)(v_\omega - \theta)\overline{\mu}(\omega) + \sum_{\omega=\omega_r+1}^{n} \left(1 - \frac{\underline{\mu}(\omega)}{\overline{\mu}(\omega)}\right)(v_\omega - \theta)\overline{\mu}(\omega) \\
\geq & \sum_{\omega=\omega_e+1}^{\omega_r} \left(1 - \frac{\underline{\mu}(\omega_r+1)}{\overline{\mu}(\omega_r+1)}\right)(v_\omega - \theta)\overline{\mu}(\omega) + \sum_{\omega=\omega_r+1}^{n} \left(1 - \frac{\underline{\mu}(\omega_r+1)}{\overline{\mu}(\omega_r+1)}\right)(v_\omega - \theta)\overline{\mu}(\omega) \\
= & \left(1 - \frac{\underline{\mu}(\omega_r+1)}{\overline{\mu}(\omega_r+1)}\right) \sum_{\omega=\omega_e+1}^{n} (v_\omega - \theta)\overline{\mu}(\omega) \\
\geq & \left(1 - \frac{\underline{\mu}(\omega_r+1)}{\overline{\mu}(\omega_r+1)}\right) \sum_{\omega=0}^{n} (v_\omega - \theta)\overline{\mu}(\omega) = \left(1 - \frac{\underline{\mu}(\omega_r+1)}{\overline{\mu}(\omega_r+1)}\right) \langle v_R - \theta, \overline{\mu} \rangle > 0,
\end{aligned}$$

where the second last inequality is because $v_\omega \leq \theta$ when $\omega \leq \omega_e$. ■

Proof of Proposition 5.4. We prove the result in three steps. First, we show that the proposed distribution of posteriors satisfies the two constraints. Second, we specify the multiplier ψ, explicitly construct a hyperplane $\lambda_0 + \langle \lambda_1, \mu \rangle$, and show that the hyperplane is above the Lagrangian function everywhere (i.e., $\lambda_0 + \langle \lambda_1, \mu \rangle \geq \mathcal{L}(\mu, \psi)$ for all $\mu \in \Delta(\Omega)$.

Third, we show that the hyperplane meets the Lagrangian function at the support of the distribution τ, that is, $\lambda_0 + \langle \lambda_1, \mu \rangle = \mathcal{L}(\mu, \psi)$ if $\mu = \mu^-$ or $\mu = \mu^+$. The last two steps establish the optimality of the proposed binary signal via Proposition 5.4.

Step 1. We first show that the proposed distribution satisfies (BP) and (IC). For (BP),

$$E_\tau[\mu] = \tau(\mu^+, e)\mu^+ + \tau(\mu^-, e)\mu^- = \eta(e) \odot \rho^+ + \eta(e) \odot \rho^- = \eta(e) \odot (\rho^+ + \rho^-) = \eta(e).$$

To verify (IC),

$$\begin{aligned}
E_\tau[h(\mu)] =& \tau(\mu^+, e)\langle \mu^+, (\overline{\mu} - \underline{\mu}) \oslash \eta(e) \rangle - c'(e) \\
=& \langle \eta(e) \odot \rho^+, (\overline{\mu} - \underline{\mu}) \oslash \eta(e) \rangle - c'(e) \\
=& \rho^+ \odot (\overline{\mu} - \underline{\mu}) - c'(e) \\
=& \sum_{\omega \geq k} (\overline{\mu}(i) - \underline{\mu}(i)) - \frac{c'(e) - c'(e_{k-1})}{\underline{\mu}(k) - \overline{\mu}(k)} (\overline{\mu}(k) - \underline{\mu}(k)) - c'(e) \\
=& \sum_{\omega > k-1} (\overline{\mu}(i) - \underline{\mu}(i)) - c'(e_{k-1}) = 0.
\end{aligned}$$

The last equality is due to the definition of e_{k-1} (see equation (12)).

Step 2. We now verify the optimality of the proposed solution by constructing a supporting hyperplane that meets $\mathcal{L}(\mu, \psi)$ at μ^- and μ^+. Let

$$\psi = -\frac{f(k|e)}{\overline{\mu}(k) - \underline{\mu}(k)}, \ \lambda_0 = 1 - \psi c'(e), \text{ and } \lambda_1(\omega) = \begin{cases} -1 & \text{if } \omega \leq k, \\ \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} & \text{if } \omega \geq k+1. \end{cases}$$

Note that $\overline{\mu}(k) - \underline{\mu}(k) < 0$ (because $k \leq \omega_e$) and, therefore, $\psi > 0$. In addition, with this specification,

$$\begin{aligned}
\lambda_0 + \langle \lambda_1, \mu \rangle &= 1 - \psi c'(e) - \sum_{\omega=0}^{k} \mu(\omega) + \sum_{\omega=k+1}^{n} \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \mu(\omega) \\
&= 1 - \psi c'(e) - \left(1 - \sum_{\omega=k+1}^{n} \mu(\omega) \right) + \sum_{\omega=k+1}^{n} \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \mu(\omega) \\
&= \sum_{\omega=k+1}^{n} \left(1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e).
\end{aligned}$$

The following fact is useful in what follows.

Lemma 6.1 *For any $k \leq \omega_e$,*

$$1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} = 1 - \frac{f(k|e)}{\overline{\mu}(k) - \underline{\mu}(k)} \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \begin{cases} < 0 \text{ if } \omega < k \\ > 0 \text{ if } \omega > k. \end{cases}$$

Proof. Since $\overline{\mu}(k) - \underline{\mu}(k) < 0$,

$$1 - \frac{f(k|e)}{\overline{\mu}(k) - \underline{\mu}(k)} \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} < 0 \iff \frac{\overline{\mu}(k) - \underline{\mu}(k)}{f(k|e)} - \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} > 0.$$

Arranging the terms with the fact that $f(\omega|e) = (1-e)\underline{\mu}(\omega) + e\overline{\mu}(\omega)$,

$$\frac{\overline{\mu}(k) - \underline{\mu}(k)}{f(k|e)} - \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} = \frac{\overline{\mu}(\omega)\overline{\mu}(k)}{f(k|e)f(\omega|e)} \left(\frac{\underline{\mu}(\omega)}{\overline{\mu}(\omega)} - \frac{\underline{\mu}(k)}{\overline{\mu}(k)} \right).$$

The desired result follows from the fact that, by Assumption 1 (MLRP), this expression is positive if $\omega < k$ and negative if $\omega > k$. ■

We first establish that $\lambda_0 + \langle \lambda_1, \mu \rangle \geq \mathcal{L}(\mu, \psi)$ for all $\mu \in \Delta(\Omega)$. Consider $\mu \in \mathcal{X}_0$. For such μ, $\mathcal{L}(\mu, \psi) = -\psi c'(e)$. Therefore,

$$\begin{aligned} \lambda_0 + \langle \lambda_1, \mu \rangle - \mathcal{L}(\mu, \psi) &= \sum_{\omega=k+1}^{n} \left(1 - \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) - (-\psi c'(e)). \\ &= \sum_{\omega=k+1}^{n} \left(1 - \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) \geq 0, \end{aligned}$$

where the inequality follows from the above lemma. Next, consider $\mu \in \mathcal{X}_1$. For such μ,

$$\begin{aligned} \mathcal{L}(\mu, \psi) &= 1 + \psi \left(\langle (\overline{\mu} - \underline{\mu}) \oslash \eta(e), \mu \rangle - c'(e) \right) = 1 + \psi \left(\sum_{\omega} \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \mu(\omega) - c'(e) \right) \\ &= \sum_{\omega} \left(1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) \\ &= \sum_{\omega=1}^{k} \left(1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) + \sum_{\omega=k+1}^{n} \left(1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) \\ &= \sum_{\omega=1}^{k-1} \left(1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) + \sum_{\omega=k+1}^{n} \left(1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e), \end{aligned}$$

where the last equality is due to the fact that $1 + \psi(\overline{\mu}(k) - \underline{\mu}(k))/f(k|e) = 0$. Therefore,

$$\lambda_0 + \langle \lambda_1, \mu \rangle - \mathcal{L}(\mu, \psi) = -\sum_{\omega=1}^{k-1} \left(1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)} \right) \mu(\omega) \geq 0,$$

where the inequality is, again, due to the above lemma.

Step 3. We now show that $\lambda_0 + \langle \lambda_1, \mu \rangle$ coincides with $\mathcal{L}(\mu, \psi)$ when $\mu = \mu^-$ and $\mu = \mu^+$. Since $\mu^- \in \mathcal{X}_0$, as shown above,

$$\lambda_0 + \langle \lambda_1, \mu^- \rangle - \mathcal{L}(\mu^-, \psi, e) = \sum_{\omega=k+1}^{n} \left(1 - \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)}\right) \mu^-(\omega).$$

The desired result that $\lambda_0 + \langle \lambda_1, \mu^- \rangle - \mathcal{L}(\mu^-, \psi) = 0$ follows from the fact that $\mu^-(\omega) = 0$ for all $\omega \geq k+1$: recall that $\mu^- = 1/\langle \eta(e), \rho^- \rangle \cdot \eta(e) \odot \rho^-$ where $\rho^-(\omega) = 0$ for $\omega \geq k+1$. $\mu^+ \in \mathcal{X}_1$ and, therefore,

$$\lambda_0 + \langle \lambda_1, \mu^+ \rangle - \mathcal{L}(\mu^+, \psi, e) = -\sum_{\omega=1}^{k-1} \left(1 + \psi \frac{\overline{\mu}(\omega) - \underline{\mu}(\omega)}{f(\omega|e)}\right) \mu^+(\omega).$$

Similarly to the above, the desired result follows because $\mu^+(\omega) = 0$ for all $\omega \geq k$: recall that $\mu^+ = 1/\langle \eta(e), \rho^+ \rangle \cdot \eta(e) \odot \rho^+$ where $\rho^+(\omega) = 0$ for $\omega \leq k$. ■