Bayesian Persuasion and Moral Hazard*

(Preliminary and Incomplete)

Raphael Boleslavsky† and Kyungmin Kim‡

October 2017

Abstract

We study optimal Bayesian persuasion when the prior on the underlying state is determined by an agent's unobservable effort. Specifically, we consider a three-player game in which the principal designs a signal, the agent exerts effort, and the decision-maker takes an action that affects the other players' utilities. The principal faces double objectives, persuading the decision-maker and incentivizing the agent. We identify a trade-off between information provision and incentive provision and develop a general method of characterizing an optimal signal. We provide more concrete implications of moral hazard for optimal information design by fully analyzing several natural examples.

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

Keywords: Bayesian persuasion; moral hazard; information design.

*We thank Ilwoo Hwang and Marina Halac for various helpful comments.

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

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

1 INTRODUCTION

We study an optimal information design in the presence of moral hazard. Specifically, we introduce an additional player, the agent, to the Bayesian persuasion framework of Kamenica and Gentzkow (2011) (KG, hereafter). The agent has preferences for the decision-maker's (receiver's) actions, which are partially or fully aligned with those for the principal (sender). He exerts unobservable effort, which determines the prior on the underlying state. In this context, the information designer is concerned not only with information provision (persuasion) for the decision-maker, but also with incentive provision for the agent. We investigate when there is a trade-off between the two objectives and how the information designer optimally resolves the trade-off. To put it different, we endogenize the prior, which is a primitive in KG, through the agent's effort and analyze its implications for Bayesian persuasion.

To understand the underlying problem more clearly, consider the following example, which is borrowed from KG but cast into a different context.[1] A school (principal) wishes to place a student in the labor market (the decision-maker). There are two types of jobs, a low-paying job and a high-paying job. Although the school prefers the latter placement, the student may or may not have acquired skills necessary for the high-paying job. The probability that the student is skilled at the time of placement is given by 0.3. Suppose that a student gets a high-paying job if and only if the market believes that the student is skilled with at least probability 1/2. This arises, for example, if a risk-neutral firm earns utility 1 when hiring a low-skilled (high-skilled) student at a low-paying (high-paying) and utility -1 otherwise.

Kamenica and Gentzkow (2011) show that the school can benefit from designing a sophisticated grading policy. In the current example, the student gets a low-paying job for sure if the school does not reveal any information about the student's skill level. If the school reveals full information, then the student gets a high-paying job if and only if he is indeed skilled and, therefore, with probability 30%. An optimal policy involves a certain amount of obfuscation: the school assigns a good grade (i.e., claims that the student is skilled) with probability 1 if the student is skilled and with probability 3/7 even if the student is not skilled. In this case, given a good grade, the student is believed to be skilled with probability 1/2 and, therefore, placed at a high-paying job. Under this grading policy, a student gets a high-paying job with probability 60%.

We consider the case in which the prior belief, which determines the school's capacity to persuade the market, is determined through the student's effort. To be specific, suppose that the student privately chooses whether to shirk or work hard. In the former case, the student never becomes skilled, while in the latter case, he successfully acquires skills with probability 0.3. Assume that the (risk-neutral) student obtains utility 1 if he gets a high-paying job and 0 if he gets a low-paying job, and his disutility from working is given by 0.2. In what follows, we let 1 (0) denote a skilled

[1]Other natural examples include a credit ratings agency that interacts with both security issues (agent) and investors (decision-maker), a marketing department which deals with both a production department (agent) and consumers (decision-maker), a prosecutor who hires an investigator (agent) and faces a judge (decision-maker), and a news media that transmit information about the government (agent) to the public (decision-maker).

(unskilled) student and $\pi(A|\omega)$ represent the probability that the school assigns a good grade (A) to the student (or, claims that the student is skilled) when his type (skill level) is $\omega = 0, 1$.

To see how moral hazard influences an optimal information design, first consider the policy that is optimal in KG's model (i.e., $\pi(A|1) = 1$ and $\pi(A|0) = 3/7$). That policy, although optimal given prior 0.3, does not provide sufficient incentive for the student to work, because

$$-c + 0.3 \cdot \pi(A|1) + 0.7 \cdot \pi(A|0) = \frac{2}{5} < 1 \cdot \pi(A|0) = \frac{3}{7}.$$

The market rationally expects this and assigns probability 0 to the student having acquired skills, in which case the student never gets a high-paying job. A full information policy ($\pi(A|1) = 1$ and $\pi(A|0) = 0$) performs better, because it at least induces the student to work ($-c + 0.3 \cdot \pi(A|1) = 0.1 > \pi(A|0) = 0$). However, the policy provides too much incentive and, therefore, can be further improved upon.

An optimal grading policy (signal) for the example is $\pi(A|1) = 1$ and $\pi(A|0) = 1/3$.[2] This policy exemplifies how the principal stirkes a balance between information vs. incentive provision. She obfuscates information in the same way as in KG, but provides more precise information. The latter is necessary for the student's incentive, because $-0.2+0.3\cdot\pi(A|1)+0.7\cdot\pi(A|0) = \pi(A|0)$ when $\pi(A|0) = 1/3$ but the right-hand side exceeds as soon as $\pi(A|0) > 1/3$. The overall probability of a high-paying job placement is equal to $0.3 + 0.7 \cdot 1/3 = 8/15 \approx 53.33\%$. Notice that this exceeds the outcome under full information (30%) but falls short of the optimal outcome in the absence of moral hazard (60%). The former shows the value of optimal information design, while the latter represents the cost of moral hazard.

We characterize a principal-optimal signal for the general model in which there are n underlying states, the principal and the agent have arbitrarily preferences regarding the decision-maker's actions, and the agent can choose any effort level. In the absence of moral hazard, KG show that the principal's problem reduces to choosing an optimal one among all Bayes-plausible distributions of posteriors (i.e., the distributions of posteriors such that the expected value of posteriors is equal to the prior) and an optimal distribution of posteriors can be found by a con-convification technique developed by Aumann and Maschler (1995). We explain how to extend these arguments in our model. Moral hazard introduces an additional constraint to the principal's problem, which is that, as in the canonical principal-agent model, a signal (distribution of posteriors) must be such that the agent has an incentive to choose an effort level that the principal intends to induce (and the decision-maker expects). In other words, the principal's problem becomes more stringent,

[2] In this particular example, there is a continuum of optimal signals. For example, suppose that there are three grade levels, A, B, and C. Any signal with the following properties is optimal:

$$\pi(A|1) + \pi(B|1) = 1,\ \pi(A|0) + \pi(B|0) = \frac{1}{3},\ \text{and } \pi(1|A), \pi(1|B) \geq \frac{1}{2}.$$

All of these signals are outcome equivalent: the student works (because $-0.2 + 0.3\pi(A, B|1) + 0.7\pi(A, B|0) = \pi(A, B|0)$) and gets a high-paying job whenever his grade is A or B, which occurs with probability 8/15. This severe multiplicity arises because both payoff and cost structures are discrete, which is not the case in our general model.

in the sense that she now faces an incentive constraint as well as a Bayes-plausibility constraint. Provided that the first-order approach is valid (i.e., the agent's optimal effort is characterized by the first-order condition of the agent's problem), the incentive constraint shrinks to one equality constraint. Con-convification then can be applied jointly over the principal's objective function and the incentive constraint, and it suffices to select the maximal achievable value subject to both Bayes-plausibility constraint and the incentive constraint.

The following two general results highlight distinguishing features of our model relative to KG's. If the principal's utility is concave in the decision-maker's induced posterior, in KG, it is optimal for the principal to reveal no information. In our model, such a policy leads to no effort by the agent and, therefore, cannot be optimal in any non-trivial environment.[3] If there are n possible states, then an optimal outcome can be achieved with at most n signal realizations (or posteriors) in KG. In our model, the number increases by 1, that is, an optimal signal may necessitate $n+1$ realizations (but not more than that). Both economically and geometrically, this is because of the new incentive constraint, which calls for an extra degree of freedom.

We provide a more comprehensive set of results for the binary-state case (and under some natural economic assumptions). We show that the agent's effort is maximized under a fully informative signal and any effort below is also implementable. One corollary of this result is that if the principal's utility is convex in the decision-maker's posterior, then a fully informative signal, which is optimal in KG, continues to be optimal in our model. We also characterize the set of incentive-free effort levels which can be implemented by the optimal policy in KG (i.e., for which the incentive constraint does not bind). From this analysis, it follows that a fully informative signal. Finally, we show that an optimal signal often takes a very simple form: it uses only two signal realizations and introduces noise from one state into the other, so that an optimal distribution of posteriors includes either 0 or 1. We explain why this is the case and when each case arises.

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 examples, multiple sellers (e.g., Boleslavsky and Cotton 2015, Gentzkow and Kamenica 2017, Li and Norman 2015), multiple receivers (e.g., Alonso and Câmara 2016, Chan et al. 2016), a privately informed receiver (e.g., Kolotilin et al. 2015), and dynamic environments (e.g., Ely 2017, Renault et al. 2014). More broadly, optimal information design has been incorporated in various economic contexts, such as price discrimination (e.g., Bergemann et al. 2015), monopoly pricing (e.g., Roesler and Szentes 2017), and auctions (e.g., Bergemann et al. 2017). To our knowledge, this is the first paper that incorporates moral hazard into the general Bayesian persuasion framework.

Two contemporary papers, Rodina (2016) and Rodina and Farragut (2016), are particularly close to this paper. Both papers study the same three-player game as ours. The main difference lies in the principal's objective. In our model, the principal has her own and general preferences over the decision-maker's actions. She is concerned with the agent's effort, because the decision-maker's

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

action depends on the (conjectured) effort. In both Rodina (2016) and Rodina and Farragut (2016), the principal is concerned only with maximizing the agent's effort.[4] This can be interpreted as a special case of our model in which the principal's utility is linear in the decision-maker's posterior belief. On the other hand, they provide a more thorough analysis of the special case than us. In particular, they allow for the general state space and consider multiple specifications with different observability assumptions.

Barron et al. (2016) study another problem that combines information design (Bayesian persuasion) and moral hazard, but in a starkly different way from ours. They analyze a principal-agent model in which the agent can engage in "gaming" (adding mean-preserving noise) after observing an intermediate output. The agent, due to his gaming ability, can always con-convificate his payoffs, which implies that the principal cannot implement a contract that is convex in output. They show that if the agent is risk neutral, then the maximal effort can be implemented by a linear contract and the optimal effort necessarily has a linear concave closure.

The remainder of this paper is organized as follows. Section 2 introduces our baseline model with binary states. Section 3 provides a general characterization of the model. Section 4 considers three representative examples. Section **??** concludes by discussing a few relevant points, including, in particular, how to generalize our analysis beyond the binary-state case.

2 THE MODEL

The game. There are three players, agent (A), principal (P), and decision-maker (D). There is an underlying state $\omega \in \Omega \equiv \{0,1\}$, which is endogenously determined by the agent's effort. The principal designs, and publicly announces, a signal π that relates Ω to a realization space S. The principal is unrestricted in her signal design, in that she can choose any finite set S and any stochastic process from Ω to S. For each $\omega \in \Omega$, we let $\pi_\omega(s)$ denote the probability that s is realized conditional on the agent's type ω. Given π, the agent exerts effort $e \in \mathcal{R}_+$, which stochastically determines the agent's type $\omega \in \Omega \equiv \{0,1\}$ but is unobservable by the other players. More effort increases the probability that the agent becomes type 1. Specifically, we assume that e is identical to the probability of type 1 (i.e., $Pr\{\omega = 1|e\} = e$). The decision-maker observes a signal realization s and chooses an action $a \in A$. The agent's utility u_A depends on the decision-maker's action a and his own effort e.[5] For convenience, we assume that u_A is additively separable and given by $u_A(a,e) = u_A(a) - c(e)$. The principal's utility u_P and the decision-maker's utility u_D depend on the decision-maker's action a and the agent's type ω. All agents maximize their expected utility. Our main objective is the study of an optimal signal design by the principal, and

[4] In this sense, these papers are related to Hörner and Lambert (2016), who characterize the rating system that maximizes the agent's effort in a dynamic career concerns model with various information sources.

[5] We assume that u_A is independent of the agent's type ω for two reasons. First, technically, it ensures that the agent's utility depends only on the decision-maker posterior beliefs even after her deviation from the equilibrium effort. In other words, the subsequent reformulation fails if u_A also depends on ω. Second, economically, it means that the agent exerts effort, not for her own consumption (i.e., not because she enjoys a direct benefit from becoming type 1), but to generate favorable information about her.

thus we focus on a principal-preferred perfect Bayesian equilibrium of this game.

Reformulation. Let μ denote the decision-maker's belief about the state ω (the probability that the decision-maker assigns to $\omega = 1$). For any μ, let $a(\mu)$ denote the set of the decision-maker's optimal (mixed) actions.[6] Then, we can reformulate the agent's and the principal's utility functions as follows:

$$v_A(\mu) \equiv u_A(a(\mu)), \text{ and } v_P(\mu) \equiv E_\mu[u_P(a(\mu), \omega)].$$

In other words, inducing a particular action $a \in A$ is identical to inducing a posterior μ under which the decision-maker's optimal action is a. As in KG, this reformulation allows us to abstract away from details of the decision-maker's actual problem without incurring any loss of generality. Note that $a(\mu)$ is not necessarily a singleton and, therefore, both v_A and v_P are correspondences in general. In what follows, for notational convenience, we treat $a(\mu)$ (and v_A and v_P as well) as a function unless necessary and noted otherwise.

Assumptions. The cost function $c(e)$ is strictly increasing, convex and continuously differentiable. In addition, $c(0) = 0$, $c'(0) < 1$ and $c'(1) > 1$. As shown later, the assumption that $c'(0) < 1$ ensures that the principal can induce non-zero effort in equilibrium, while $c'(1) > 1$ ensures that the principal cannot induce $e = 1$. Both v_A and v_P are upper hemi-continuous and increasing in μ (precisely, $\max\{v_i(\mu)\} \leq \min\{v_i(\mu')\}$ for any $\mu < \mu'$ and both $i = A, P$). The latter monotonicity assumption reflects a natural economic force (that the more optimistic the decision-maker is about the agent's type, the more favorable action he takes to the agent) and allows us to provide sharper characterization results. In addition, the problem becomes trivial, with the agent always choosing $e = 0$, if v_A or v_P is strictly decreasing. Finally, we normalize both the agent's and the principal's utilities, so that $v_A(0) = v_P(0) = 0$ and $v_A(1) = v_P(1) = 1$.

Subgame. Given a signal π, the agent and the decision-maker play a simple extensive-form game. Let e^* denote an equilibrium effort level and $\mu(s)$ denote the decision-maker's posterior belief following a signal realization s. By Bayes' rule,

$$\mu(s) = \frac{e^*\pi_1(s)}{e^*\pi_1(s) + (1 - e^*)\pi_0(s)}.$$

For e^* to be indeed an equilibrium, it must solve

$$\max_e \sum_s (e\pi_1(s) + (1 - e)\pi_0(s))v_A(\mu(s)) - c(e).$$

Since the first term is linear in e and $c(e)$ is strictly convex, it is necessary and sufficient that

$$\sum_s (\pi_1(s) - \pi_0(s))v_A(\mu(s)) = c'(e^*).$$

[6] To be formal, let $A(\mu) \equiv argmax_{a \in A} E_\mu[u_R(a, \omega)]$, and define $a(\mu) \equiv \Delta(A(\mu))$.

Taken together, an equilibrium in the subgame given π is characterized by an effort level e^* such that

$$\sum_s (\pi_1(s) - \pi_0(s)) v_A \left(\frac{e^* \pi_1(s)}{e^* \pi_1(s) + (1 - e^*)\pi_0(s)} \right) = c'(e^*). \tag{1}$$

Notice that there may exist multiple equilibria. In particular, $e^* = 0$ is always an equilibrium, as long as no signal realization s fully reveals $\omega = 1$. Intuitively, if the decision-maker believes that the agent would not exert effort, then $\mu(s) = 0$ for any s, which in turn justifies $e^* = 0$. This equilibrium multiplicity can be used to restrict the principal's strategy (e.g., by playing $e^* = 0$ unless the principal chooses a signal that satisfies a particular property) but is inconsequential in our analysis, because the principal-preferred equilibrium, which is our focus, involves the optimal choice of equilibrium effort e^* as well.

The principal's problem. Given the characterization of the subgame above, the principal's problem can be written as

$$\max_{\pi, e} \sum_s (e\pi_1(s) + (1 - e)\pi_0(s)) v_P(\mu(s)),$$

subject to

$$\sum_s (\pi_1(s) - \pi_0(s)) v_A(\mu(s)) = c'(e),$$

where

$$\mu(s) = \frac{e\pi_1(s)}{e\pi_1(s) + (1 - e)\pi_0(s)}.$$

The principal's problem can be reformulated as the one in which the sender chooses a distribution of posteriors $\tau \in \Delta(\Delta(\Omega))$, instead of a signal π, as formally stated in the following proposition.

Proposition 2.1 *Given e, there exists a signal π that yields utility v to the principal if and only if there exists a distribution of posteriors $\tau \in \Delta(\Delta(\Omega))$ such that (i) $E_\tau[v_P(\mu)] = v$, (ii) $E_\tau[\mu] = e$, and (iii) $E_\tau[(\mu - e)v_A(\mu)]/(e(1 - e)) = c'(e)$, where $E_\tau[f(\mu)] = \int f(\mu)\tau(\mu)d\mu$.*

Proof. See the appendix. ■

The second requirement that $E_\tau[\mu] = e$ is identical to the one in KG and commonly referred to as the Bayes-plausibility (BP) constraint. The last requirement corresponds to the agent's incentive constraint. To see how equation (1) can be translated into (iii) in the proposition, fix a signal π. Without loss of generality, assume that $\mu(s) \neq \mu(s')$ whenever $s \neq s'$. Then, for any $s \in S$,

$$\tau(\mu(s)) = e\pi_1(s) + (1 - e)\pi_0(s), \text{ and } \mu(s) = \frac{e\pi_1(s)}{e\pi_1(s) + (1 - e)\pi_0(s)}.$$

Solving these two equations yields

$$\pi_1(s) = \frac{\mu(s)\tau(\mu(s))}{e} \text{ and } \pi_0(s) = \frac{(1-\mu(s))\tau(\mu(s))}{1-e}.$$

Plugging these two into equation (1) leads to (iiii).

There are two noteworthy facts about the IC constraint. First, it holds for $e > 0$ only when τ includes at least two posteriors: if τ is degenerate on μ, then $\mu = e$ because $E_\tau[\mu] = e$, in which case $E_\tau[(\mu - e)v_A(\mu)]/(e(1-e)) = 0 < c'(e)$. This is a clear manifestation of the underlying moral hazard problem in our model. In the absence of moral hazard, if v_P is concave in μ, then it is optimal for the principal not to reveal any information. Such an uninformative policy does not provide a proper incentive for the agent and, therefore, can never be optimal in our model. Second, the effect of inducing a particular posterior μ on the IC constraint, summarized by $(\mu - e)v_A(\mu)$, takes an intriguing form: it is decreasing initially, reaches 0 when $\mu = e$, and increases fast thereafter. This pattern is driven by the presence of two channels through which the agent can be incentivized. One (reflected in $v_A(\mu)$) is through differentiating rewards based on a signal realization, and the other (reflected in $\mu - e$) is through controlling the probability of each reward.

3 SEARCHING FOR THE OPTIMAL SOLUTION

In this section, we characterize an optimal solution to the principal's problem. We let τ^e denote an optimal distribution of posteriors that implements effort e and V^e the corresponding expected utility of the principal. In other words, τ^e solves

$$(2) \qquad \max_{\tau \in \Delta(\Delta(\Omega))} E_\tau[v_P(\mu)], \text{ subject to } (BP) \;\; E_\tau[\mu] = e, \text{ and } (IC) \;\; \frac{E_\tau\left[(\mu - e)v_A(\mu)\right]}{e(1-e)} = c'(e),$$

and $V^e \equiv E_{\tau^e}[v_P(\mu)]$. We also define $V^* \equiv \max_e V^e$.

3.1 IMPLEMENTABLE AND INCENTIVE-FREE EFFORT LEVELS

We say that an effort level e is implementable if there exists a signal π (equivalently, a distribution of posteriors τ) that satisfies both BP and IC constraints. The following proposition shows that an effort level is implementable if and only if it is below a certain threshold.

Proposition 3.1 *Let $\overline{e}$ be the value such that $c'(\overline{e}) = 1$. 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. ■

Importantly, the upper bound $\overline{e}$ is achieved by a fully informative signal. To see this clearly, notice that under a fully informative signal, the agent's problem is simply

$$\max_e ev_A(1) + (1-e)v_A(0) - c(e),$$

whose solution is given by $c'(e) = 1$ and, therefore, identical to $\overline{e}$. Intuitively, a fully informative signal maximizes incentive provision in both relevant channels. First, it maximizes dispersion in rewards, because $v_A(1) - v_A(0) \geq v_A(\mu) - v_A(\mu')$ for any $\mu, \mu' \in [0,1]$. Second, it minimizes both type I and type II errors and, therefore, provides a maximal incentive given any rewards.

There may or may not be a conflict between incentive provision and information provision. For example, if $v_P(\mu)$ is convex, then a fully informative signal is optimal in the absence of the IC constraint. Since it also induces maximal effort by the agent, it is clearly an optimal signal. To the contrary, if $v_P(\mu)$ is concave, then an optimal signal is completely uninformative without the IC constraint. However, the signal clearly violates the IC constraint and, in fact, leads to the most undesirable outcome of $e = 0$.

In order to utilize this idea, let $\widehat{V}^e$ denote the maximal attainable value to the principal in the relaxed problem without the IC constraint. In other words,

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

Obviously, $V^e = \widehat{V}^e = 0$ if $e = 0$ and $V^e \leq \widehat{V}^e$ for any $e \leq \overline{e}$. Let $\underline{e}$ be the maximal value such that $V^e = \widehat{V}^e$. The following result shows that, in our search for the optimal solution, it suffices to consider the effort levels between $\underline{e}$ and $\overline{e}$.

Proposition 3.2 *For any $e < \underline{e}$, $V^e \leq V^{\underline{e}} \leq V^*$.*

Proof. According to KG,

$$\widehat{V}^e = \sup\{z | (e,z) \in co(v_P)\},$$

where $co(v_P)$ is the convex hull of the graph of p. Since v_P is increasing in μ, $\widehat{V}^e$ is increasing in e. It then follows that for any $e < \underline{e}$,

$$V^e \leq \widehat{V}^e \leq \widehat{V}^{\underline{e}} = V^{\underline{e}} \leq \max_{e \leq \overline{e}} V^e = V^*.$$

■

Although $\underline{e}$ depends on all relevant functions, it is often straightforward calculate its value. In particular, $\underline{e} = 0$ if v_P is concave, while $\underline{e} = \overline{e}$ if v_P is convex. Figure **??** illustrates how to find $\underline{e}$ when $v_P(\mu)$ is neither concave nor convex. The left panels are for the case where both v_P and v_A have a discrete jump at $1/2$, and the right panels are for the case where $v_P(\mu)$ is initially convex but eventually concave and v_A is linear. In both cases, given e, the con-convification technique in Aumann and Maschler (1995) can be used to find $\widehat{V}^e$ and the corresponding optimal distribution of posteriors. For the distribution to provide a just enough incentive for the agent, it suffices to check whether the IC constraint holds.

In our model, the principal designs a signal first and the agent exerts effort then. Suppose, instead, that the principal designs, or can revise, a signal after the agent chooses e. In this case, the principal 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 that can be induced in this alternative scenario. This shows that it is the principal's commitment power to a signal that enables her to implement $e \in (\underline{e}, \overline{e}]$.

3.2 MAIN CHARACTERIZATION

In order to characterize the maximal value V^e and the optimal distribution of posteriors τ^e, we extend an elegant geometric method by KG. For notational simplicity, define a function $h : [0,1] \to \mathcal{R}$, so that

$$h^e(\mu) \equiv \frac{(\mu - e)v_A(\mu)}{e(1-e)} - c'(e).$$

Notice that the IC constraint reduces to $E_\tau[h^e(\mu)] = 0$.

Define the following curve in $\mathcal{R}^3$:

$$K \equiv \{(\mu, h^e(\mu), v_P(\mu)) : \mu \in [0,1]\}.$$

The left panel of Figure **??** depicts a sample path when $v_P(\mu)$ is concave and $v_A(\mu)$ is linear. Each point in K represents the value of the constraint ($h^e(\mu)$) and the principal's utility ($v_P(\mu)$) when a particular posterior is induced. Clearly, $e(> 0)$ is not even implementable if the principal reveals no information and induces a degenerate posterior e. In the figure, that is reflected in the fact that the vertical line built on $(e, 0, 0)$ does not cross K.

Now construct the convex hull of the curve K, denoted by $co(K)$ and visualized in the right panel of Figure **??**. Then, select the points in the convex hull such that the first coordinate is equal to e and the second coordinate is equal to 0. Formally, define $K^* \equiv \{(x_1, x_2, x_3) \in co(K) : x_1 = e, x_2 = 0\}$. In Figure **??**, K^* corresponds to the intersection of $co(K)$ and the vertical line above $(e, 0, 0)$. Since K^* is a subset of the convex hull of K, for any $(e, 0, z) \in K^*$, there exists a probability vector $(\tau(\mu_1), ..., \tau(\mu_n))$ and a sequence $\{(\mu_s, h^e(\mu_s), v_P(\mu_s)) \in K\}_{s=1}^n$ such that

$$(e, 0, v) = \sum_s \tau(\mu_s)(\mu_s, h^e(\mu_s), v_P(\mu_s)). \tag{3}$$

Conversely, since K^* includes all the points in the intersection of $co(K)$ and the vertical line on $(e, 0, 0)$, any convex combination of the points in K such that $\sum \mu_s \tau(\mu_s) = e$ and $\sum h^e(\mu_s)\tau(\mu_s) = 0$ belongs to K^*. Thus, K^* represents all possible convex combinations of the points in K that satisfy both the BP constraint ($\sum_s \mu_s \tau(\mu_s) = 0$) and the IC constraint ($\sum_s h^e(\mu_s) = 0$). It then follows that the maximal principal utility subject to the two constraints coincides with the maximal third coordinate value of K^*, as formally reported in the following theorem.

Proposition 3.3 *The maximal utility the principal can obtain conditional on inducing e is equal to*

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

If $e \leq \overline{e}$, then there exists an optimal distribution of posteriors $\tau^e \in \Delta(\Delta(\Omega))$ such that its support contains at most three posteriors (i.e., $|supp(\tau^e)| \leq 3$).

Proof. Proposition 3.1 implies that K^* is non-empty if and only if $e \leq \overline{e}$. Since $co(K)$ is closed and bounded, K^* is also closed and bounded. These imply that if $e \leq \overline{e}$, then there exists a distribution of posteriors τ^e that is implementable and yields expected utility V^e to the principal. For the result on the cardinality of the support of τ^e, notice that V^e is the value on the boundary of the convex hull in $\mathcal{R}^3$. The result then follows from Carathéodory's theorem, which states that any point on the convex hull on a 2-dimensional hyperplane can be made of at most three extreme points. ■

Recall that in the absence of moral hazard (i.e., in the model of KG), if there are only two states, then the maximal principal can be achieved with at most two posteriors. In our model, moral hazard introduces the IC constraint and, therefore, an additional dimension. Via Carathéodory's theorem, this translates into the possibility of necessitating one additional signal. As shown in the next section, two posteriors (signals) are still sufficient in many examples, but there are cases where at least three posteriors (signals) are necessary.

Convex hull is, in general, hard to construct from a curve in $\mathcal{R}^3$. We provide an alternative characterization of the optimal signal, which, as shown in the next section, allows us to derive an optimal solution in a simple fashion in many examples. Because $co(K)$ is a convex set in $\Re^3$ and $x^* \equiv (e, 0, V^e)$ is on its boundary, a supporting hyperplane for $co(K)$ exists at x^*. Thus, there exists a normalized direction vector $d \equiv (-\lambda_1, \psi, 1)$ and a scalar λ_0 such that $d \cdot x \leq \lambda_0$ for all $x \in co(K)$, with equality for $x = x^*$. Because it is in $co(K)$, vector x^* can be written as a convex combination of vectors in K, as in (3). Therefore, each vector $x_s \equiv (\mu_s, h^e(\mu_s), v_P(\mu_s))$ which receives positive weight in the convex combination, $\tau(\mu_s) > 0$, must also satisfy $d \cdot x_s = \lambda_0$.[7] Rearranging terms, we find the necessary conditions for optimality, presented in the following lemma. The proof that these conditions are also sufficient for optimality is elementary.

Proposition 3.4 *A distribution of posteriors τ^e is a solution to the principal's problem, (2), if and only if it satisfies (BP), (IC), and there exists a vector $(\lambda_0, \lambda_1, \psi)$ such that*

$$\mathcal{L}(\mu, \psi) \equiv v_P(\mu) + \psi h^e(\mu) \leq \lambda_0 + \lambda_1 \mu, \text{ for all } \mu \in [0, 1],$$

with equality for all μ such that $\tau^e(\mu) > 0$. Furthermore, if $e \in (\underline{e}, \overline{e})$, then $\psi > 0$.

In order to understand this condition, notice that if $\psi = 0$, then the condition is identical to the one for KG. An optimal signal can be found by drawing a straight line $\lambda_0 + \lambda_1 \mu$ that stays just above $v_P(\mu)$ and identifying a set of posteriors that span $(e, \widehat{V}^e)$. In Figure **??**, $v_P(\mu)$ is concave, and thus $\widehat{V}^e = v_P(e)$ and the optimal value can be induced with a degenerate posterior e. The only necessary change due to moral hazard is that the same technique is applied over $\mathcal{L}(\mu, \psi) = v_P(\mu) + \psi h^e(\mu)$ for some ψ, which needs not be equal to 0 in general (and is never

[7] Suppose some vector x with $\tau(\mu) > 0$ has $d \cdot x < \lambda_0$. Because any vector $y \in K$ has $d \cdot y \leq \lambda_0$, it follows that $d \cdot x^* < \lambda_0$.

equal to 0 if $v_P(\mu)$ is concave). Figure **??** shows how $\mathcal{L}(\mu, \psi)$ differs from $v_P(\mu)$ how it affects the shape of the straight line. The multiplier ψ is also an unknown variable, but the IC constraint provides an additional equation to solve for τ^e as well as ψ. In Figure **??**, the IC constraint is reflected in the fact that the two dashed lines cross at the optimal point (e, V^e), as it implies that $E_{\tau^e}[v_P(\mu)] = E_{\tau^e}[\mathcal{L}(\mu, \psi)] = E_{\tau^e}[v_P(\mu)] + \psi E_{\tau^e}[h^e(\mu)]$, and thus $E_{\tau^e}[h^e(\mu)] = 0$.

4 EXAMPLES AND APPLICATIONS

In this section, we analyze some representative examples. Each example not only illustrates how to apply the general method developed in the previous section, but also has a natural economic interpretation and, therefore, is of interest by itself.

4.1 CONCAVE-LINEAR CASE

We first consider the case where $v_P(\mu)$ is concave, while $v_A(\mu)$ is linear. This case arises, for example, when the market (decision-maker) offers a competitive wage, the student (agent) is risk neutral and, therefore, maximizes the expected wage, and the school (principal) is mainly concerned with undesirable placement outcomes. For analytical tractability, we assume that $v_P(\mu)$ is twice continuously differentiable.

Since $v_A(\mu) = \mu$, $h^e(\mu)$ simplifies to

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

This implies that the IC constraint becomes identical to

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

This highlights the relationship between dispersion of the distribution of posteriors and the agent's effort. The more dispersed the induced posteriors are, the higher effort the agent chooses. Conversely, the principal can induce a particular effort level as long as she introduces enough dispersion into the distribution of posteriors.

Now fix $e \in (0, \overline{e})$ and consider the function $\mathcal{L}(\mu, \psi)$. Since $v_A(\mu) = \mu$, its second derivative with respect to μ takes the following form:

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

Although $v_P''(\mu) < 0$, $\mathcal{L}_{\mu\mu}$ is not necessary negative because of the second term. In fact, for ψ to be a part of the principal's solution, $\mathcal{L}_{\mu\mu}$ cannot be uniformly negative: if so, the optimal signal is degenerate and, therefore, cannot implement e. Conversely, $\frac{\partial^2 \mathcal{L}(\mu,\psi)}{\partial \mu^2}$ cannot be uniformly positive either: if so, the optimal signal is fully informative and, therefore, provides too much incentive for

the agent. This discussion implies that an optimal value of ψ is such that $\mathcal{L}_{\mu\mu}$ has both positive and negative regions.

It is useful to define the following two types of signals, both of which take a particularly simple form but play a crucial role in subsequent discussions.

Definition 4.1 *A simple inflationary signal (policy) is a binary signal that induces either* 0 *or* $\mu_I(>0)$*. A simple deflationary signal (policy) is a binary signal that induces either* $\mu_D(<0)$ *or* 1*.*

A simple inflationary signal introduces noise into a good signal realization. In other words, it induces a high posterior μ_D with probability 1 if $\omega = 1$ but does so with a positive probability even if $\omega = 0$ (thus, partially inflating the agent's type). A simple deflationary signal does the opposite, inducing a low posterior μ_D with probability 1 if $\omega = 0$ but with a positive probability even if $\omega = 1$.

For both signals, there are two unknowns, one unknown posterior (μ_I or μ_D) and the probability of the posterior being induced (denoted by γ_I and γ_D, respectively). These two unknowns can be obtained from the two equality constraints. For the inflationary one, since $v_A(\mu) = \mu$,

$$(BP)\ \ \mu_I\tau_I = e \text{ and } (IC)\ \ \frac{(\mu_I - e)\mu_I\gamma_I}{e(1-e)} = c'(e).$$

Therefore,

$$\mu_I = e + (1-e)c'(e) \text{ and } \gamma_I = \frac{e}{\mu_I}.$$

It is also easy to show that

$$\mu_D = e(1 - c'(e)) \text{ and } \gamma_D = \frac{1-e}{1-\mu_D}.$$

Note that this implies that implementable simple inflationary and deflationary signals are independent of $v_P(\mu)$.

The following result shows that a full characterization of the optimal signal is available for an important class of concave functions such that $v_P''(\mu)$ is monotone.

Proposition 4.1 *Suppose* $v_P''(\mu) < 0$ *and* $v_A(\mu) = \mu$*. The optimal signal that induces* $e \in (0, \overline{e})$ *is a simple inflationary policy if* $v_P''(\mu)$ *decreases in* μ *and a simple deflationary policy if* $v_P''(\mu)$ *increases in* μ*.*

Proof. If $v_P''(\mu)$ decreases in μ, then $\mathcal{L}_{\mu\mu}$ also decreases in μ. This means that with an optimal ψ, there exists $\overline{\mu} \in (0,1)$ such that $\mathcal{L}_{\mu\mu} \geq 0$ if and only if $\mu \leq \overline{\mu}$. This means that the function $\mathcal{L}(\cdot, \psi)$ is convex until $\overline{\mu}$ and concave after $\overline{\mu}$. By Corollary 3.4, an optimal signal induces 0 or a certain posterior above e as a simple inflationary signal. The logic can be easily modified for the case in which $v_P''(\mu)$ increases in μ. ■

Intuitively, the principal with a concave value function wishes to minimize dispersion of induced posteriors. In the absence of moral hazard, this leads to her revealing no information. In our

model, it induces the principal to use two posteriors, rather than three posteriors. The result that an optimal signal involves extreme posteriors 0 or 1 is due to our focus on well-behaved concave functions. Since $v_P''(\mu)$ is monotone, $\mathcal{L}(\cdot, \psi)$ can have at most one inflection point, and thus the supporting line $(\lambda_0 + \lambda_1\mu)$ crosses either 0 or 1. This property is not guaranteed if $v_P''(\mu)$ is sufficiently irregular that $\mathcal{L}(\cdot, \psi)$ has multiple inflection points.

In order to understand which policy is optimal when, consider the polynomial case in which $v_P(\mu) = 1 - (1-\mu)^\eta$ for some $\eta > 1$. In this case,

$$v_P'''(\mu) = (1 - (1-\mu)^\eta)''' = \eta(\eta-1)(\eta-2)(1-\mu)^{\eta-3}.$$

Therefore, by Proposition 4.1, the optimal signal is inflationary if $\eta \in (1, 2)$ and deflationary if $\eta > 2$. The result certainly depends on the curvature of $v_P(\mu)$. However, risk aversion is not the underlying determinant. Consider the CARA utility function case in which $v_P(\mu) = (1 - e^{-\eta\mu})/(1 - e^{-\eta})$ for some $\eta > 0$. In this case,

$$v_P'''(\mu) = \frac{(1 - e^{-\eta\mu})'''}{1 - e^{-\eta}} = \frac{\eta^3 e^{-\eta\mu}}{1 - e^{-\eta}}.$$

Therefore, the optimal signal is deflationary no matter how close η is to 0 (i.e., the principal is almost risk neutral).

The crucial property is the effect that clockwise variance-preserving rotation has on the principal's expected utility. To see this, again, consider the polynomial case in which $v_P(\mu) = 1-(1-\mu)^\eta$ for some $\eta > 1$. Given $e \in (0, \overline{e})$, there is a continuum of pairs (μ_1, μ_2) such that $\mu_1 < e < \mu_2$ and there exists γ_1 that satisfies

$$(BP)\ \ \gamma_1\mu_1 + (1-\gamma_1)\mu_2 = e \text{ and } (IC)\ \ \frac{var(\mu)}{e(1-e)} - c'(e) = 0.$$

An increase in μ_1 increases μ_2 (because of IC) but decreases γ_1 (because of BP), causing (μ_1, μ_2) to rotate clockwise (see the dashed lines in Figure **??**). In the quadratic case where $v_P(\mu) = 1-(1-\mu)^2$, this rotation has no effect on the principal's expected payoff, because

$$\gamma_1 v_P(\mu_1) + (1-\gamma_1)v_P(\mu_2) = \gamma_1(2\mu_1 - \mu_1^2) + (1-\gamma_1)(2\mu_2 - \mu_2^2) = 2e - var(\mu) + e^2.$$

Indeed, in this quadratic case, any pair of (μ_1, μ_2) that satisfy both BP and IC, including both simple inflationary and deflationary signals, are optimal. If $\eta \in (1, 2)$, then the same rotation always decreases the principal's expected payoff (see the left panel of Figure **??**), which ultimately leads to the optimality of the simple inflationary signal. If $\eta > 2$ (or $v_P(\mu)$ is a CARA function), then the rotation always increases the principal's expected payoff (see the right panel of Figure **??**) and, therefore, the optimal policy is deflationary.

4.2 IDENTICALLY CONCAVE CASE

We now consider the case in which the principal and the agent have an identical concave utility function (i.e., $v_P(\mu) = v_A(\mu) = v(\mu)$). This emerges, for example, when a school's reputation depends on its full placement records. It also captures the case where the principal is altruistic or another self of the (time-inconsistent) agent.[8] As in the previous case, we assume that $v(\mu)$ is twice continuously differentiable.

Differentiating $\mathcal{L}(\mu, \psi)$ with respect to μ twice yields

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

Let $r(\mu) \equiv -v''(\mu)/v'(\mu)$ (Arrow-Prat measure of risk aversion). The equation then reduces to

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

This implies that

$$\mathcal{L}_{\mu\mu} > 0 \Leftrightarrow \frac{e(1-e-\psi)}{2\psi} + \frac{\mu}{2} < \frac{1}{r(\mu)}. \tag{4}$$

As in the concave-linear case, an optimal ψ must be such that $\mathcal{L}$ is neither concave nor convex and has at least one inflection point. From these observation, it is possible to extrapolate the following result.

Proposition 4.2 *Suppose that $v_P(\mu) = v_A(\mu) = v(\mu)$ and $v(\mu)$ is concave.*

- *If $r(\mu)$ increases in μ (Increasing Absolute Risk Aversion), then the optimal signal is a simple deflationary policy.*
- *Suppose that $1/r(\mu) = a + b\mu$ for some a and b (Hyperbolic Absolute Risk Aversion). The optimal signal is a simple inflationary policy if $b < 1/2$ and a simple deflationary policy if $b > 1/2$.*

Proof. If $r(\mu)$ increases in μ, then the left-hand side in equation (4) rises, while the right-hand side falls, as μ increases. This implies there exists $\overline{\mu} \in (0, 1)$ such that $\mathcal{L}_{\mu\mu} \geq 0$ if and only if $\mu \leq \overline{\mu}$. It follows that an optimal distribution of posteriors involves 0 and a certain positive μ. If $1/r(\mu) = a + b\mu$, then the right-hand side rises faster than the left-hand side if and only if $b > 1/2$. This means that $\mathcal{L}$ switches from concave to convex (and the optimal policy is deflationary) if $b > 1/2$ and from convex to concave (and the optimal policy is deflationary) if $b < 1/2$. ■

[8] Recall that we assume that the principal's utility does not depend on the agent's effort. However, this assumption does not affect the characterization of an optimal signal given e, although it does matter for the optimal choice of e. In other words, the principal would choose a lower e if she internalizes the agent's effort, but our main analysis regarding the optimal signal for each e carries through unchanged.

4.3 DISCRETE-LINEAR CASE

Now suppose that $v_P(\mu)$ has a discrete jump at $\theta \in (0,1)$ (i.e., $v_P(\mu) = \mathcal{I}_{\{\mu \geq \theta\}}$) and $v_A(\mu)$ is linear (i.e., $v_A(\mu) = \mu$).[9] The latter assumption is not necessary for the subsequent analysis but gives extra tractability. This case corresponds to the case in which the school wishes to maximize the proportion of students who get a wage above a certain threshold. In order to reduce the number of cases to consider, we also assume that $c'(\theta) > 1$, so that $\overline{e} < \theta$.[10]

Unlike in the previous concave cases, $\underline{e} > 0$ in this case. Specifically, in the absence of moral hazard, the principal induces only 0 or θ, unless $e \geq \theta$. This implies that $\underline{e}$ is given by the value such that for some $\gamma > 0$,

$$(BP) \;\; \gamma\theta = \underline{e} \text{ and } (IC) \;\; \frac{(\theta - \underline{e})\gamma}{\underline{e}(1-\underline{e})} - c'(\underline{e}) = 0.$$

Combining the two conditions yields

$$\frac{\theta - \underline{e}}{\theta(1-\underline{e})} = c'(\underline{e}).$$

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

%beginfigure[tbp]

Since $v_P(\mu) = \mathcal{I}_{\{\mu \geq \theta\}}$ and $v_A(\mu) = \mu$,

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

In other words, $\mathcal{L}(\cdot, \psi)$ is a quadratic function but is shifted upward by 1 from θ (see Figure **??**). This means that there are three possibilities: the supporting line $\lambda_0 + \lambda_1\mu$ touches (i) $(0, \mathcal{L}(0,\psi))$ and $(\theta, \mathcal{L}(\theta,\psi))$, (ii) $(0, \mathcal{L}(0,\psi))$ and $(1, \mathcal{L}(1,\psi))$, and (iii) all three points at 0, θ, and 1. However, (i) leads to $\underline{e}$, while (ii) results in $\overline{e}$. Therefore, the only possibility is that ψ is such that all three points lie on the supporting line, as shown in Figure **??**.

Proposition 4.3 *Suppose that $v_P(\mu) = \mathcal{I}_{\{\mu \geq \theta\}}$, $v_A(\mu) = \mu$, and $c'(\theta) > 1$. Then, for any $e \in (\underline{e}, \overline{e})$, an optimal signal $\tau(e)$ involves three posteriors, 0, θ, and 1. The probabilities of each posterior are $\tau^e(\theta) = \frac{e(1-e)(1-c'(e))}{\theta(1-\theta)}$ and $\tau^e(1) = e - \tau^e(\theta)\theta$.*

Proof. The result on the use of three posteriors follows from the discussion above. τ^e can be explicitly calculated from the following three equations:

$$\tau^e(0) + \tau^e(\theta) + \tau^e(1) = 1, \; (BP) \;\; E_{\tau^e}[\mu] = e, \text{ and } (IC) \;\; E_{\tau^e}[h^e(\mu)] = 0.$$

[9] It is straightforward to modify the analysis for the case in which $v_A(\mu)$ is also discrete, as in the example used in the introduction. One disadvantage of the alternative discrete case is that there is a continuum of optimal solutions.

[10] Without this assumption, it may be optimal to induce θ or 1. In KG, this form of dispersion is not relevant, because if the prior is above θ, then an uninformative signal is optimal. In our model, it can be useful and indeed optimal because the principal can induce a certain level of effort at no cost on her side.

■

Among other things, this case demonstrates that the result on the number of necessary posteriors in Theorem 3.3 binds. In other words, although a binary signal (in particular, simple inflationary and deflationary signals) is often sufficient, as shown in all the previous cases, an optimal signal may require three posteriors (signal realizations).

5 ADDITIONAL STATES

Among various simplifying assumptions we have maintained so far, the most restrictive assumption is, arguably, that the set of states Ω has only two elements. In this section, we illustrate how our main characterization (Theorem 3.3) generalizes when there are more than two states and also provide a complete characterization for a prominent example.

5.1 GENERAL CHARACTERIZATION

Setup. Suppose that $\Omega \equiv \{0, 1, ...n\}$, so that there are $n + 1$ states where $n \geq 2$. In this case, the players' beliefs are represented by vector $\mu = (\mu(0), \mu(1), ...\mu(n))$, where $\mu(k)$ denotes the probability that $\omega = k$. Vector μ is an element of the unit n-simplex $\Delta(\Omega)$. The prior distribution of ω is determined by the agent's effort e, according to the function $f : \Omega \times \mathcal{R}_+ \to [0, 1]$, where $f(\omega|e)$ denotes the probability that the state is ω when the agent's effort is e. We let $f(e) \equiv (f(0|e), f(1|e), ..., f(n|e))$ denote the vector of these probabilities and assume that $f(\omega|\cdot)$ is continuously differentiable for all $\omega \in \Omega$. The agent's cost function $c(e)$ and the players' strategies are identical to those of the baseline two-state model.

As in the baseline model, we reformulate the players' underlying utility functions and express the agent's and the principal's payoffs as functions of the decision-maker's posterior belief μ. In other words, we let $v_A(\mu)$ denote the agent's ex post payoff and $v_P(\mu)$ denote the principal's ex post payoff when the decision-maker's posterior belief about the state is $\mu \in \Delta(\Omega)$.

Vector operations. The following three vector operations are useful in what follows.

- Inner product: for any $x, y \in \mathcal{R}^{n+1}$,

$$\langle x, y \rangle \equiv x(0)y(0) + x(1)y(1) + ... + x(n)y(n).$$

- Hadamard (component-wise) product: for any $x, y \in \mathcal{R}^n$,

$$x \odot y \equiv (x(0)y(0), x(1)y(1), ..., x(n)y(n)).$$

- Hadamard (component-wise) division: for any $x, y \in \mathcal{R}^n$,

$$x \oslash y \equiv \left(\frac{x(0)}{y(0)}, \frac{x(1)}{y(1)}, ..., \frac{x(n)}{y(n)} \right).$$

Subgame. Suppose that the principal has chosen a signal, which consists of a finite set S and a function $\pi : S \times \Omega \to [0,1]$, where $\pi(s|\omega)$ is the probability that s is realized when the state is ω. Given an equilibrium effort e^*, by Baye's rule, the decision-maker's belief following a realization s is given by[11]

$$\mu_s = \frac{1}{\langle \pi(s|\cdot), f(e^*) \rangle} \pi(s|\cdot) \odot f(e^*) \in \Delta(\Omega).$$

The agent's problem is then

$$\max_{e \geq 0} \sum_{\omega} \left(\sum_{s} \pi(s|\omega) v_A(\mu_s) \right) f(\omega|e) - c(e) = \sum_{s} \pi(s|e) v_A(\mu_s) - c(e), \tag{5}$$

where $\pi(s|e) \equiv \langle \pi(s|\cdot), f(e) \rangle$. The first-order condition, combined with an equilibrium requirement that e^* is indeed an optimal effort, yields

$$\sum_{s} \pi_e(s|e^*) v_A(\mu_s) - c'(e^*) = 0. \tag{6}$$

As in other principal-agent problems, this condition is not sufficient for the agent's optimal effort choice in general, but it is technically necessary to assume that this first-order condition fully characterizes the agent's optimal effort (the first-order approach).[12]

As in the baseline model, we rewrite the first-order condition in terms of a distribution of posteriors $\tau \in \Delta(\Delta(\Omega))$, instead of a signal π. We use the fact that given e^* (and the consequent prior distribution $f(e^*)$), a Bayes-plausible posterior distribution γ can be induced by a signal π such that[13]

$$\pi(s|\cdot) = \tau(\mu_s) \cdot \mu_s \oslash f(e^*) = \tau(\mu_s) \left(\frac{\mu_s(0)}{f(0|e^*)}, ..., \frac{\mu_s(n)}{f(n|e^*)} \right).$$

Inserting this into the above first-order condition and arranging the terms yield

$$E_\tau \left[\langle f_e(\cdot|e^*) \oslash f(e^*), \mu \rangle v_A(\mu) \right] - c'(e^*) = 0,$$

which can be simplified to $E_\tau[h(\mu, e^*)] = 0$ by letting

$$h(\mu, e) \equiv \langle f_e(\cdot|e) \oslash f(e), \mu \rangle v_A(\mu) - c'(e). \tag{7}$$

[11] To put it different, for all $\omega \in \Omega$ and $s \in S$,

$$\mu_s(\omega) = \frac{\pi(s|\omega) f(\omega|e^*)}{\sum_{\omega'} \pi(s|\omega') f(\omega'|e^*)}.$$

[12] In Section 5.2, we introduce a simple linear technology that ensures the validity of the first-order approach.

[13] Observe that under this signal, $\tau(\mu_s) = \langle \pi(s|\cdot), f(e^*) \rangle$ and, by Bayes' rule, the posterior following realization s is

$$\frac{1}{\langle \pi(s|\cdot), f(e^*) \rangle} \pi(s|\cdot) \odot f(e^*) = \frac{\tau(\mu_s)}{\langle \pi(s|\cdot), f(e^*) \rangle} \langle \mu_s \oslash f(e^*), f(e^*) \rangle = \mu_s.$$

[SOME INTERPRETATION]

The principal's problem. Given the above incentive constraint, the principal's problem can be written as

$$\max_{\tau \in \Delta(\Delta(\Omega))} E_\tau[v_P(\mu)], \text{ subject to } (BP) \;\; E_\tau[\mu] = f(e), \text{ and } (IC) \;\; E_\tau[h(\mu, e)] = 0.$$

This is a clear generalization of the principal's problem in the baseline model (see equation (2)). Given this representation of the principal's problem, it is straightforward to extend the geometric argument used for the baseline model and obtain the following general characterization.

Proposition 5.1 *Suppose that the set of states Ω has $n+1 (\geq 2)$ elements and the first-order approach is valid for the agent's effort choice problem given any signal. For any implementable e,*

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

where

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

and there exists an optimal distribution of posteriors $\tau^e \in \Delta(\Delta(\Omega))$ such that its support contains at most $n+2$ posteriors (i.e., $|supp(\tau^e)| \leq n+2$). Furthermore, a distribution of posteriors τ^e solves the principal's problem if and only if it satisfies (BP), (IC), and there exist $\lambda_0 \in \mathcal{R}$, $\lambda_1 \in \mathcal{R}^n$, and $\psi \in \mathcal{R}$ such that

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

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

Proof. The argument for the result on V^e is identical to that for the baseline model. The result on the use of at most $n+2$ posteriors follows from the fact that $(f(e), 0, V^e)$ is on the boundary of $co(K) \subset \mathcal{R}^{n+1}$ (via Carathéodory's theorem). The necessary condition for τ^e follows from the same logic as in Corollary 3.4. ■

5.2 PERSUADING A DICTATOR, WHILE MOTIVATING THE POLITICIAN

In order to obtain some concrete insights about the general finite-state model, we consider a canonical model in which there are only two actions available to the decision-maker and both the principal and the agent prefer one action to the other (referred to the (general) binary-action model, hereafter). Since the underlying Bayesian persuasion problem is fully studied by Alonso and Câmara (2016), we adopt an analogous political interpretation to theirs: the decision-maker is a dictator (or a representative voter) who solely decides whether to keep the status quo or adopt a new pol-

icy.[14] The agent is a politician, while the principal is a party leader. Both want the new policy to be implemented. Different from Alonso and Câmara (2016), the principal 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.

5.2.1 The Model

Setup. The decision-maker (dictator) decides whether to take action a_0 (keeping the status quo) or a_1 (adopting a new policy). If she selects a_0, then her payoff is $\theta > 0$, independent of the state. If she selects a_1, then her payoff is v_ω in state ω. Without loss of generality, we assume that the states are ordered from the worst to the best to the decision-maker and normalize the payoffs, so that so that

$$v_0 = 0 < v_1 = 1 < v_2 < ... < v_n.$$

The principal's (the party leader) and the agent's (politician) payoffs depend only on the decision-maker's action, and both prefer a_1: both receive payoff 0 if $a = a_0$ and 1 if $a = a_1$.

Define vector $v_{DM} \equiv (0, 1, v_2, ..., v_n)$, which lists the decision-maker's payoffs by state. Then, the decision-maker prefers a_1 to a_0 if and only if $\langle \mu, v_{DM} \rangle \geq \theta$ and, therefore,

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

We refer to the set of beliefs at which the decision-maker selects a_i as $\mathcal{A}_i (\subset \Delta(\Omega))$ 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. We consider the following simple production technology: there exist two vectors $f_0, f_1 \in \Delta(\Omega)$ such that if the agent exerts effort $e \in [0, 1]$, then the state is determined according to

$$f(e) = (1 - e)f_0 + ef_1.$$

In other words, if the agent makes no effort ($e = 0$), then the state is drawn according to f_0, while if the agent chooses maximal effort ($e = 1$), then the state is distributed according to f_1. If an interior effort ($e \in (0, 1)$) is chosen, then the state is determined according to a compound lottery made of f_0 and f_1, with e representing the weight to f_1. Notice that this technology makes the first term in the agent's optimal effort choice problem (equation (5)) linear (because $f_e(e) = f_1 - f_0$) and, therefore, ensures that the first-order condition (equation (6)) is necessary and sufficient for the agent's optimal choice (i.e., the first-order approach is valid).

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

To avoid trivialities, we assume that $f_0 \in \mathcal{A}_0$ and $f_1 \in \mathcal{A}_1$. In other words, if the decision-maker's posterior is equal to f_0 (which arises when the agent makes no effort and the principal reveals no information), then she selects a_0, while if her posterior is equal to f_1 (which corresponds to maximal effort and no additional information), then she takes a_1. For ease of exposition, we make four assumptions on f_0 and f_1.

Assumption 1 *Monotone likelihood ratio property: $f_1(\omega)/f_0(\omega)$ 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 f_0 and f_1 are probability vectors, $f_1(\omega)/f_0(\omega)$ crosses 1 once from below. We refer to states such that $f_0(i) > f_1(i)$ as *bad-news states* and the other states as *good-news states*.

Assumption 2 *Let ω_e denote the largest ω such that $f_0(\omega) > f_1(\omega)$. 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

$$E_{f_0}[v_i|\omega \neq 0] = \frac{\langle f_0, v_{DM} \rangle}{1 - f_0(0)} \geq \theta.$$

This assumption states that even with belief f_0, if state 0 is ruled out, then the resulting belief vector, with the same relative likelihood of all other states, induces a_1. This assumption is not essential, but streamlines the exposition considerably by reducing the number of equilibrium cases.

Assumption 4

$$\sum_{\omega > \omega_e} (f_1(\omega) - f_1(\omega)) < c'(1).$$

This is a technical assumption that corresponds to $c'(1) > 1$ in the baseline model: it ensures that it suffices to consider interior effort levels strictly below 1. The specific form of the left-hand side is clarified shortly.

5.2.2 Preliminaries

Binary signals. As observed by KG, the number of induced posteriors ($|supp(\tau)|$) does not need to exceed the number of available actions ($|A|$) in Bayesian persuasion. The result holds in our current binary-action model, despite the presence of moral hazard: the set of realizations can be partitioned into two subsets, those in $\mathcal{A}_0$ and those in $\mathcal{A}_1$. It suffices to replace each subset with a single realization at its center of mass, because (BP), (IC), and the objective function are linear in γ and, therefore, independent of the transformation. From now on, we focus on binary distributions. Clearly, $e > 0$ can be induced only when each realization induces each possible action. In light

of these observations, we fix the set of signal realizations with $S = \{0, 1\}$ and let μ^- denote the posterior corresponding to realization 0 and μ^+ denote the posterior corresponding to realization 1. In addition, we take it for granted that $\mu^- \in \mathcal{A}_0$ and $\mu^+ \in \mathcal{A}_1$.

Implementable efforts. The following lemma is a counterpart to Proposition 3.1 for the baseline model and fully characterizes the set of implementable efforts in our binary-action model.

Lemma 5.2 *In the general binary-action model, e is implementable if and only if $e \leq \overline{e}$, where*

$$\sum_{\omega > \omega_e} (f_1(\omega) - f_0(\omega)) = c'(\overline{e}). \tag{8}$$

Proof. Since $f_e(e) = f_1 - f_0$ and the agent's payoff depends only on whether the decision-maker selects a_1 or not, the first-order condition for the agent's effort (equation (6)) can be rewritten as

$$\langle f_1 - f_0, Pr\{a_1|\cdot\}\rangle - c'(e) = 0, \tag{9}$$

where $Pr\{a_1|\cdot\}$ denotes the probabilities that action a_1 is selected depending on the state. By the definition of ω_e (see Assumption 2) and Assumption 1, the marginal benefit of effort (the first term in the above equation) is bounded from above by

$$\langle f_1 - f_0, Pr\{a_1|\cdot\}\rangle \leq \sum_{i > \omega_e} (f_1(\omega) - f_0(\omega)).$$

Since $c'(e)$ is strictly increasing, equation (9) cannot hold for any $e > \overline{e}$. Notice also that Assumption 4 ensures that $\overline{e} < 1$.

We now show that there exists a signal that induces $\overline{e}$. Consider the following binary signal:

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

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 decision-maker selects a_0 following realization 0 and a_1 following realization 1. The former (a_0 after realization 0) follows from the fact that realization 0 reveals that $\omega \leq \omega_e \leq \omega_r$ (i.e., the true state is a rejection state for sure), while the latter (a_1 after realization 1) is due to Assumptions 1 and 3: even when the prior is f_0, the decision-maker is willing to take a_1 as soon as state 0 is excluded, but realization 1 rules out weakly more rejection states from 0 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 decision-maker's incentive to select a_1. As in the baseline model, a convex combination of this signal and one which reveals no information implements any effort below $\overline{e}$. ■

One intriguing implication of this result is that, in contrast to our baseline model with binary states, 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 $\overline{e}$. Note that the signal used in the proof of Lemma 5.2 is the unique binary signal that induces $\overline{e}$, but there may exist non-binary signals that also induce $\overline{e}$.

Corollary 5.3 *In the general binary-action model, a fully informative signal induces $\overline{e}$ if and only if $\omega_e = \omega_r$.*

Proof. The result is immediate from the following two observations: for a signal to induce $\overline{e}$, it is necessary and sufficient that the decision-maker chooses a_1 for any $\omega > \omega_e$. With a fully informative signal, by the definition of ω_r and Assumption 1, the decision-maker selects a_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 makes the agent's incentive not fully maximized, leading to a sub-maximal effort. If instead all states above ω_e are pooled together, then the decision-maker selects a_1 for all states above ω_e. Therefore, the agent benefits from increased probabilities of all states above ω_e. This maximizes the marginal benefit of effort and, therefore, induces the maximal effort.

5.2.3 Incentive-free Effort

In our binary-action model, even in the absence of moral hazard, the principal introduces dispersion in the distribution of posteriors, as long as the prior $f(e)$ belongs to the rejection region $\mathcal{A}_0$. This means that a positive effort can be induced even if the principal ignores the incentive constraint. We solve for such an effort level. To be precise, we first consider a relaxed problem without the incentive constraint and with an exogenously given $e \in [0, \overline{e}]$:

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

We then find the effort level for which the incentive constraint ($E_\tau[h(\mu, e)] = 0$) is satisfied. The following proposition provides a closed-form characterization for such an incentive-free effort level.

Proposition 5.4 *In the general 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:*

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

It is given by the value that satisfies $(f_0(0) - f_1(0))\underline{r} = c'(\underline{e})$, where

$$\underline{r} \equiv \frac{\theta - \langle f(\underline{e}), v_{DM} \rangle}{\theta f(0|\underline{e})} \in (0, 1).$$

Proof. See the appendix. ■

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

$$\pi(1|\omega) = \begin{cases} 1-r & \text{if } \omega = 0, \text{ and} \\ 1 & \text{if } \omega > 0. \end{cases} \tag{10}$$

Clearly, realization 0 reveals state 0, while realization 1 generates Bayesian update

$$\frac{f(e) \odot (1-r, 1, ..., 1)}{\langle f(e), (1-r, 1, ..., 1) \rangle} = \frac{1}{1 - rf(0|e)} \left((1-r) f(0|e), f(1|e), ..., f(n|e) \right).$$

In addition, the signal generates realization 0 with probability $f(0|e)r$ and realization 1 with probability $1 - rf(0|e)$. The optimality of this class of binary signals (in the absence of moral hazard) stems from Assumption 3: since ruling out state 0 with probability 1 moves the decision-maker's belief into the interior of the acceptance region $\mathcal{A}_0$, it suffices to reveal state 0 with a positive probability.

Within this class, the optimal signal is determined by the interaction of two strategic forces. On the one hand, realization 0 leads to rejection. Therefore, the principal wants to reveal state 0 as little as possible (i.e., minimize r). On the other hand, if r is too small, then the decision-maker does not become sufficiently optimistic and will not choose a_1 following realization 1. These mean that the principal would like to reveal state 0 just often enough that the decision-maker is indifferent between actions a_0 and a_1, that is,

$$\langle \mu^+, v_{DM} \rangle = \frac{1}{1 - rf(0|e)} \langle f(e) \odot (1-r, 1, ..., 1), v_{DM} \rangle = \theta.$$

This condition allows us to identify the optimal value of r (using the fact that, since $v_0 = 0$, $\langle f(e) \odot (1-r, 1, ..., 1), v_{DM} \rangle = \langle f(e), v_{DM} \rangle$):

$$r = \frac{\theta - \langle f(e), v_{DM} \rangle}{\theta f(0|e)}. \tag{11}$$

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 the incentive constraint (i.e., solves the relaxed problem without the incentive constraint). An incentive-free effort $\underline{e}$ is the level at which this optimal signal is consistent with the agent's optimal effort choice: given the optimal signal associated with $\underline{e}$, it is optimal for the agent to indeed choose effort level $\underline{e}$, that is, $\underline{e}$ satisfies the incentive constraint. This means that $\underline{e}$ is (uniquely) determined by the following equation (see equation (9) in the proof of Lemma 5.2):

$$\langle f_1 - f_0, Pr\{\cdot|\underline{e}\} \rangle = \langle f_1 - f_0, (1 - \underline{r}, 1, ..., 1) \} \rangle = (f_0(0) - f_1(0)\underline{r} = c'(\underline{e}).$$

Example with three states. Figure 1 provides a graphical illustration of Proposition 5.4 for the case of three states, for which beliefs can be represented by a 2-dimensional simplex.[15] In the figure, effort increases the probability of states 1 and 2 (i.e., $\omega_e = 0$), but only state 2 is an acceptance state (i.e., $\omega_r = 1$). The shaded blue region represents the acceptance region $\mathcal{A}_1 \equiv \{\mu : \langle \mu, v_{DM} \rangle \geq \theta\}$. State distributions f_0 and f_1 are represented by the hollow red dots. Note that increased effort moves the prior $f(e) = (1-e)f_0 + ef_1$ along the red line segment.

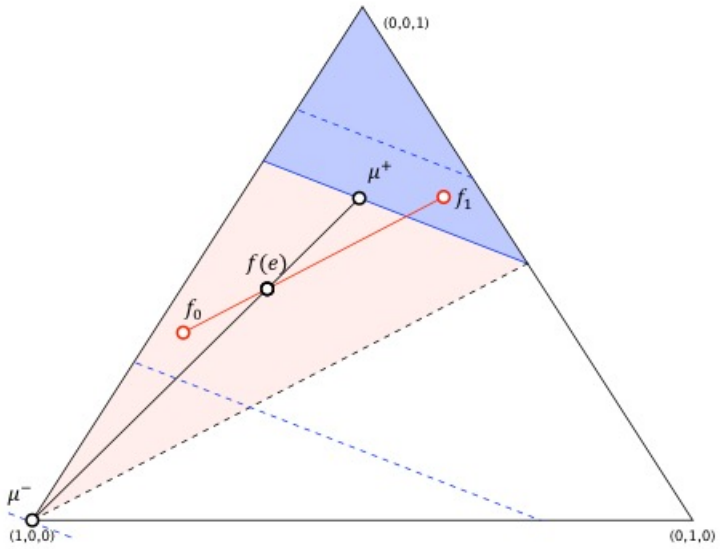

Figure 1: Illustration of Proposition 5.4.

In order to find an incentive-free effort, fix $e(\leq \overline{e})$ (equivalently, prior $f(e)$). We week a hyperplane, $\lambda_0 + \langle \lambda_1, \mu \rangle$, that supports the principal's objective function inside $\Delta(\Omega)$. Since $v_P(\mu) = 1$ if $\mu \in \mathcal{A}_1$ (the blue shaded region), while $v_P(\mu) = 0$ if $\mu \in \mathcal{A}_0$, such a hyperplane must touch $(1,0,0)$, which is the farthest point from $f(e)$ among the elements in $\mathcal{A}_-$, and the lower boundary of $\mathcal{A}_1$ (i.e., the line segment such that $\langle \mu, v_{DM} \rangle = \theta$), which are closest to $f(e)$ among the elements in $\mathcal{A}_+$.[16] In other words, $\lambda_0 + \langle \lambda_1, (1,0,0) \rangle = 0$, and $\lambda_0 + \langle \lambda_1, \mu \rangle = 1$ whenever $\langle \mu, v_{DM} \rangle = \theta$. This implies that $\lambda_0 = 0$ and $\lambda_1 = 1/\theta \cdot v_{DM}$. Jointly imposing the binary-signal restriction (one $\mu^- \in \mathcal{A}_0$ and one $\mu^+ \in \mathcal{A}_+$) and Bayes plausibility ($\tau(\mu^-)\mu^- + \tau(\mu^+)\mu^+ = f(e)$), it follows that $\mu^- = (1,0,0)$ and μ^+ is the unique point that intersects the boundary of $\mathcal{A}_1$ and the extended line that connects $(1,0,0)$ and $f(e)$.

The search for an incentive-free effort level amounts to varying e continuously from 0 to $\overline{e}$ (which varies $f(e)$ along the red line segment) and finding e such that the optimal signal associated with $f(e)$ induces the agent to choose effort e (i.e., also satisfies the incentive constant). Such a point

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

[16] Notice that this is effectively identical to concavificating the graph $\{(\mu, v_P(\mu)) : \mu \in \Delta(\Omega)\}$, which gives rise to the same hyperplane over $\mathcal{A}_0$ and the flat (constant) hyperplane over $\mathcal{A}_1$.

exists: at f_0, 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 realization μ^+. At $f(\overline{e})$, 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 a_1 is taken if and only if $\omega > \omega_e$ (see Lemma 5.2) but, because μ^+ is on the boundary of $\mathcal{A}_1$, action a_1 is selected with a positive probability even when $\omega = 0$.

5.2.4 Optimal Signal

We now study the optimal information design for efforts greater than $\underline{e}$. By definition, moral hazard introduces a distortion in the design and, therefore, the solution to the relaxed problem violates the incentive constraint.

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

$$\sum_{\omega > k} (f_1(\omega) - f_0(\omega)) = c'(e_k). \tag{12}$$

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 (i.e., $\pi(\omega) = 0$ if $\omega \leq k$, while $\pi(\omega) = 1$ if $\omega > k$). This means that e_k is the effort level induced by such a signal. The analysis above for Proposition 5.4 suggests that $\underline{e} < e_0$ (because the optimal signal for $\underline{e}$ reveals state 0 with probability less than 1), while that for Lemma 5.2 implies that $\overline{e} = e_{\omega_e}$. In addition, the left-hand side increases in k as long as $k \leq \omega_e$ (Assumptions 1 and 2) and, therefore, $e_k < e_{k+1}$ for any $k = 0, ..., \omega_e - 1$. For notational convenience, we define $e_{-1} \equiv \underline{e}$.

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

Proposition 5.5 *In the general binay-action example, for $k = 0, 1, ..., \omega_e$, effort $e \in (e_{k-1}, e_k]$ is optimally implemented by the following binary distribution of posteriors:*

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

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

$$\rho^-(\omega) = \begin{cases} 1 & \textit{for} \quad 0 \leq \omega \leq k-1, \\ \frac{c'(e) - c'(e_{k-1})}{f_0(k) - f_1(k)} & \textit{for} \quad \omega = k, \ \textit{and} \\ 0 & \textit{for} \quad k+1 \leq \omega \leq n. \end{cases}$$

Proof. See the appendix. ■

Proposition 5.5 suggests that the optimal signal varies continuously and systematically as the target effort e increases from $\underline{e}(= e_{-1})$ to $\overline{e}(= e_{\omega_e})$. The optimal signal reveals more bad-news states

as e increases. In addition, from e_{k-1} to e_k, the probability of revealing state k (i.e., inducing μ^-, so that the decision-maker selects a_0) continuously increases from 0 to 1: notice that $c'(e)-c'(e_{k-1})$ increases from 0 to $c'(e_k)-c'(e_{k-1}) = f_0(k)-f_1(k)$ (see equation (12)). This is due to the underlying Bayesian persuasion problem: as shown in the analysis for the incentive-free effort, it is optimal for the principal to reveal only some lowest states. On the other hand, moral hazard forces the principal to reveal state 0 with a higher probability and, if $e > e_0$, some other low states as well. This is a distortion from a pure information-provision perspective, but is necessary to provide an incentive for the agent to exert more effort than $\underline{e}$. One noteworthy consequence is that the receiver benefits from this distortion: in the absence of moral hazard, the decision-maker strictly prefers a_0 to a_1 following realization 0 and is indifferent between a_0 and a_1 following realization 1 (i.e., $\langle \mu^+, v_{DM} \rangle = \theta$) and, therefore, enjoy no communication benefit. With moral hazard, her belief following realization 1 is such that $\langle \mu^+, v_{DM} \rangle > \theta$ and, therefore, she strictly prefers a_1 to a_0.

Example with three states. Consider the same three-state example as in Section 5.2.3, in which $\omega_e = 0$ and, therefore, $\overline{e} = e_0$. To implement $e \in (\underline{e}, \overline{e})$, it is necessary to introduce more dispersion into the distribution of posteriors (i.e., increase the distance between μ^- and μ^+). 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 $f(e)$. This implies that the principal must reveal state 0 with a higher probability and μ^+ moves into the interior of the acceptance region $\mathcal{A}_1$, as depicted in Figure 2.

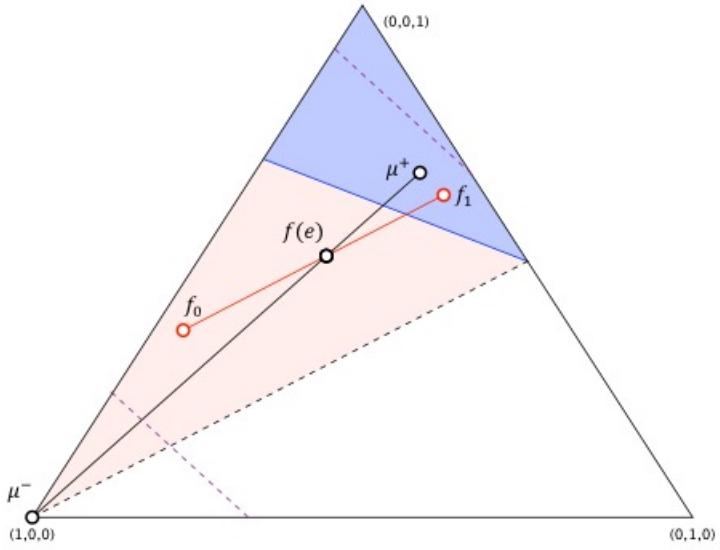

Figure 2: Illustration of Proposition 5.5.

In order to see how our general characterization (Proposition 5.1) applies to this problem, recall that we are seeking a hyperplane $\lambda_0 + \langle \lambda_1, \mu \rangle$ that supports the Lagrangian function $\mathcal{L}(\mu, \psi, e)$

inside $\Delta(\Omega)$. Since $v_A(\mu) = 0$ if $\mu \in \mathcal{A}_0$ and $v_A(\mu) = 1$ if $\mu \in \mathcal{A}_1$, the function $h(\mu, e)$ in equation (7) reduces to

$$h(\mu, e) = \begin{cases} -c'(e) & \text{if } \mu \in \mathcal{A}_0, \text{ and} \\ \langle \mu, (f_1 - f_0) \oslash f(e) \rangle - c'(e), & \text{if } \mu \in \mathcal{A}_1. \end{cases}$$

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

$$\mathcal{L}(\mu, \psi, e) = \begin{cases} -c'(e) & \text{if } \mu \in \mathcal{A}_0, \text{ and} \\ 1 + \psi \left(\langle \mu, (f_1 - f_0) \oslash f(e) \rangle - c'(e) \right), & \text{if } \mu \in \mathcal{A}_1. \end{cases}$$

Notice that $\mathcal{L}$ is linear in $\mu \in \mathcal{A}_1$ and, therefore, itself defines a hyperplane over $\mathcal{A}_1$. This implies that there are the following three possibilities for our target hyperplane $\lambda_0 + \langle \lambda_1, \mu \rangle$, depending on the value of ψ:[17]

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

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

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

In the first case, μ^+ belongs to the lower boundary of $\mathcal{A}_1$. As explained in Section 5.2.3, 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 $\mathcal{A}_1$, which then provides too much an incentive for the agent, as discussed in Section 5.2.2. These mean that the last one is the only relevant case for $e \in (\underline{e}, \overline{e})$. In other words, at the optimal solution, ψ is such that the supporting hyperplane $\lambda_0 + \langle \lambda_1, \mu \rangle$ exactly coincides with the Lagrangian function $\mathcal{L}$. This observation allows us to fully determine the optimal values of λ_0, λ_1, and ψ.[18] The optimal value of μ^+ can also be found from the two equality constraints.[19]

[17] If ψ is sufficiently small, then $\mathcal{L}$ increases slowly over $\mathcal{A}_1$ and, therefore, the supporting hyperplane touches $\mathcal{L}$ at $(1, 0, 0)$ and the lower boundary of $\mathcal{A}_1$, just as in the case for the incentive-free effort. If ψ is sufficiently large, then $\mathcal{L}$ increases fast over $\mathcal{A}_1$, 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.

[18] Specifically, since the supporting hyperplane must meet $\mathcal{L}$ at $(1, 0, 0)$, $\lambda_0 = -c'(e)$. In addition, $\lambda_0 + \langle \lambda_1, \mu \rangle$ must coincide with $\mathcal{L}$ over $\mathcal{A}_+$ and, therefore,

$$-c'(e) + \langle \lambda_1, \mu \rangle = 1 - \psi c'(e) + \int \mu \psi (f_1 - f_0) \oslash f(e).$$

From this equation, it follows that

$$-c'(e) = 1 - \psi c'(e) \Rightarrow \psi = \frac{1 + c'(e)}{c'(e)} \text{ and } \lambda_1 = \psi (f_1 - f_0) \oslash f(e) = \frac{1 + c'(e)}{c'(e)} (f_1 - f_0) \oslash f(e).$$

[19] Specifically, the two constraints reduce to

$$E_\tau[\mu] = (1 - \tau(\mu^+))(1, 0, 0) + \tau(\mu^+)\mu^+ = f(e) \text{ and } E_\tau[h(\mu, e)] = \tau(\mu^+)\langle \mu^+, (f_1 - f_0) \oslash f(e) \rangle - c'(e) = 0.$$

These yield four equations (three from the first and one from the second), which can be used to explicitly calculate the four unknowns, $\tau(\mu^+)$ and $\mu^+ = (\mu^+(0), \mu^+(1), \mu^+(2))$.

6 OBSERVABILITY OF EFFORTS

We have assumed that the agent's effort is not observable to the other two players. Certainly the principal prefers that e is observable (and verifiable by a court), because this would allow her to commit to a menu of signals that conditions on e directly (i.e., the signal is a function $\pi_\omega(s,e)$), allowing her to implement each effort level more efficiently. For example, if v_A is concave (convex), then she can commit to punish the agent by revealing full (no) information, unless the agent chooses a particular effort level.

Suppose instead that effort is observable to the decision maker but unverifiable by a court. In this case, the principal still commits to a signal that does not depend explicitly on effort, $\pi_\omega(s)$, but the decision maker also observes effort when updating beliefs. In this case, the agent's problem is

$$\max_e \sum_s (e\pi_1(s) + (1-e)\pi_0(s))v_A(\mu(s,e)) - c(e), \tag{13}$$

where $\mu(s,e) = e\pi_1(s)/(e\pi_1(s) + (1-e)\pi_0(s))$. In contrast to the baseline model, the decision-maker's posterior belief $\mu(s,e)$ now depends not only on the signal realization s but also on the agent's actual effort choice e, (rather than the decision maker's conjecture). Therefore, if the agent chooses to deviate from the effort level that the principal intends the agent to select, it not only affects the *probability* of a particular signal realization, it also affects the *belief* associated with the realization. In particular, by increasing his effort, the agent can improve beliefs associated with all signal realizations, suggesting that high effort may be easier to sustain when it is observable.

This effect is most clearly seen by differentiating the objective function in (13),

$$\sum_s (\pi_1(s) - \pi_0(s))v_A(\mu(s,e)) + (e\pi_1(s) + (1-e)\pi_0(s))v_A'(\mu(s,e))\frac{\partial \mu}{\partial e} - c'(e) =$$
$$\sum_s (\pi_1(s) - \pi_0(s))v_A(\mu(s,e)) + v_A'(\mu(s,e))\frac{\pi_1(s)\pi_0(s)}{e\pi_1(s) + (1-e)\pi_0(s)} - c'(e).$$

The first term of this expression is identical to the case in which effort is not observed; the agent's effort determines the probability with which the realization is drawn from $\pi_1(\cdot)$ or $\pi_0(\cdot)$. The second term captures the direct effect of the agent's effort on the decision-maker's belief. First, note that if $\pi_1(s)\pi_0(s) > 0$, then $\mu(s,e) \in (0,1)$, and therefore, for such realizations the second term is strictly positive, provided $v_A'(\cdot) > 0$. Second, note that if $\pi_1(s)\pi_0(s) = 0$, then $\mu(s,e) \in \{0,1\}$, that is, the realization reveals the state. In this case, the second term vanishes. Thus, for a fully informative signal the marginal benefit of effort is identical in the observable and unobservable cases, while it is strictly greater in the observable case for all other signals. It is also straightforward to verify that concavity of $v_A(\cdot)$ implies concavity of the agent's objective function, and hence, the first-order approach is valid (consult the proof of Proposition 6.1). We therefore have the following proposition.

Proposition 6.1 *Suppose that the agent's payoff function $v_A(\cdot)$ is strictly increasing and concave. Consider a signal that is incentive compatible for effort level $e < \overline{e}$ assuming effort is not observable. If the principal selects the same signal when effort is observable, then the agent selects a strictly larger effort.*

While an exhaustive analysis of the case with observable effortz is beyond the scope of the current project, such an analysis could be performed using the methods developed here. Indeed, substituting the expressions for $(\pi_1(s), \pi_0(s))$ derived in (), we find that with observable effort, the agent's incentive compatibility constraint can be written as

$$E_\tau[h^*(\mu, e)] = 0, \quad \text{where} \quad h^*(\mu, e) = h(\mu, e) + v_A'(\mu)\frac{\mu(1-\mu)}{e(1-e)}.$$

Thus, our main characterization of the necessary and sufficient conditions also applies to this case, and, in fact, to any Bayesian persuasion model with constraints that can be expressed as the expected value of a function of the posterior belief.

The solution with observable effort is straightforward if $v_A(\mu) = \mu$. In this case the agent's problem reduces to $e - c(e)$, independent of the chosen signal,

$$\sum_s (e\pi_1(s) + (1-e)\pi_0(s))v_A(\mu(s,e)) = \sum_s (e\pi_1(s) + (1-e)\pi_0(s))\mu(s,e) = e.$$

It then follows that the agent always chooses $\overline{e}$ regardless of the principal's signal. Thus, the principal does not need to introduce (additional) dispersion into the signal structure in order to credibly induce effort. Thus, the principal commits to the solution in KG, given effort $\overline{e}$, which may use a less informative signal than in the problem with moral hazard. For example, if the principal's payoff function is concave, then it is optimal for the principal to reveal no information about ω. In this case, the switch from unobservable to observable effort may impose a tradeoff on the decision-maker: effort is higher when it is observable but the principal uses a less-informative signal. Depending on the decision-maker's underlying preferences, on which we do not impose any restrictions, the decision-maker may prefer a setting with unobservable effort, as illustrated by the following example.

Example. Recall the discrete-linear example of Section 4.3, in which $v_P(\mu) = \mathcal{I}\{\mu \geq \theta\}$ and $v_A(\mu) = \mu$. As described previously, this example has a natural interpretation in a lobbying context, where the politician (agent) cares about the public's perception of his policy (μ), and a lobbyist (principal) would like to ensure that the policy is implemented. Under the status quo, the society (decision-maker) payoff is θ, and the payoff of implementing the new policy is $\omega \in \{0, 1\}$. Thus, society will make the reform if and only if $\mu \geq \theta$, and its expected payoff is simply $\max\{\mu, \theta\}$. Suppose that $c'(\theta) > 1$, which implies $\overline{e} < \theta$. From the argument above, with observable effort the agent exerts effort $\overline{e}$, and the principal simply commits to the solution of the relaxed problem (as in KG) anticipating this effort level. Because $\overline{e} < \theta$, the optimal signal is supported on two poste-

riors, $\{0, \theta\}$. Therefore, with observable effort, the decision maker's payoff is simply θ. However, with unobservable effort, the optimal signal is supported on three posteriors, $\{0, \theta, 1\}$. Because $\Pr(\mu > \theta) > 0$, the decision-maker's expected payoff is strictly higher with unobservable effort.

APPENDIX: OMITTED PROOFS

Proof of Proposition 2.1. Given the analysis in the main text, it suffices to show the sufficiency. Let $\tau \in \Delta(\Delta(\Omega))$ be a distribution of posterior distributions that satisfy (i)-(iii). Consider the following signal: let $S \equiv \{\mu \in \Delta(\Omega) : \tau(\mu) > 0\}$. For each $s \in S$,

$$\pi_1(s) = \frac{s}{e}\tau(s) \text{ and } \pi_0(s) = \frac{1-s}{1-e}\tau(s).$$

Notice that

$$\mu(s) = \frac{e\pi_1(s)}{e\pi_1(s) + (1-e)\pi_0(s)} = s.$$

It then follows that

$$\sum_{s\in S}(e\pi_1(s) + (1-e)\pi_0(s))v_P(\mu(s)) = \sum_{s\in S}\tau(s)v_P(s) = E_\tau[v_P(\mu)] = v,$$

and

$$\sum_s(\pi_1(s) - \pi_0(s))v_A(\mu(s)) = \sum_s \frac{(s-e)v_A(s)}{e(1-e)}\tau(s) = \frac{E_\tau[(\mu-e)v_A(\mu)]}{e(1-e)} = c'(e).$$

■

Proof of Proposition 3.1. 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

$$eE[v_A(\mu)|\omega = 1] + (1-e)E[v_A(\mu)|\omega = 0] - c'(e)$$

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

$$\sum_s \pi_1(s)v_A(\mu(s)) - \sum_s \pi_0(s)v_A(\mu(s)) \geq c'(e), \text{ with equality holding if } e < 1.$$

Since v_A is weakly increasing,

$$\sum_s \pi_1(s)v_A(\mu(s)) - \sum_s \pi_0(s)v_A(\mu(s)) \leq v_A(1) - v_A(0) = 1.$$

These imply that e such that $c'(e) > 1$, which is equivalent to $e > \overline{e}$, is not implementable.

Fix $e \in [0, \overline{e}]$, and consider the following distribution of posteriors, which stems from a convex

combination of a fully informative signal and a fully noisy signal:

$$\gamma(0) = c'(e)(1-e), \gamma(e) = 1 - c'(e), \gamma(1) = c'(e)e.$$

This distribution is well-defined, because $c'(e) < c'(\overline{e}) < 1$. It is straightforward to show that this distribution of posteriors satisfies both BP and IC constraints and, therefore, e is implementable. ■

Proof of Proposition 3.4. *Necessity.* As described in the text, the existence of a supporting hyperplane to $co(K)$ at x^* implies that $d \cdot x \leq \lambda_0$ for all $x \in K$, with equality for $(\mu_s, h^e(\mu_s), v_P(\mu_s)$ for which $\tau(\mu_s) > 0$. Hence, $-\lambda_1\mu + \psi h^e(\mu) + v_P(\mu) \leq \lambda_0$ for all $\mu \in [0,1]$, with equality if $\tau(\mu) > 0$. The result follows.

Sufficiency. If $v_P(\mu) + \psi h^e(\mu) \leq \lambda_0 + \lambda_1\mu$ for all μ, then any distribution of posteriors $\tau^e(\mu_s)$ satisfies $\sum \tau^e(\mu_s)(v_P(\mu) + \psi h^e(\mu_s)) \leq \lambda_0 + \lambda_1 \sum \tau^e(\mu_s)\mu_s$. Thus, any (BP) and (IC) distribution of posteriors $\tau^e(\mu)$ satisfies $\sum \tau^e(\mu_s) v_P(\mu) \leq \lambda_0 + \lambda_1 e$. Note that if $\tau(\mu)$ is (BP) and (IC) and $v_P(\mu) + \psi h^e(\mu) = \lambda_0 + \lambda_1\mu$ for all μ such that $\tau^e(\mu) > 0$, then $\tau^e(\mu)$ achieves the upper bound, and it is therefore optimal.

Multiplier $\psi > 0$. Consider a relaxed version of the principal's problem in which IC is absent. Given an effort level, $e \in (\underline{e}, \overline{e})$, let the solution of the relaxed problem be denoted $\widehat{\tau}(\mu, e)$, and the principal's maximum payoff $\widehat{V}(e)$. Also let the value of the left hand side of the incentive constraint achieved by the solution of the relaxed problem be H; that is,

$$H(e) \equiv E_{\widehat{\tau}}[h^e(\mu)].$$

Necessary and sufficient conditions for the relaxed problem are a special case of those in the preceding part of Proposition 3.4, where (IC) is removed and $\psi = 0$. Hence, there exists $(\widehat{\lambda}_0, \widehat{\lambda}_1)$ such that

$$v_P(\mu) \leq \widehat{\lambda}_0 + \widehat{\lambda}_1\mu \tag{14}$$

for all $\mu \in [0,1]$, with equality if $\widehat{\tau}(\mu, e) > 0$. Note that $\widehat{\tau}(\mu, e)$ satisfies (BP).

Given the same e, let the solution of the principal's problem be $\tau(\mu, e)$ and the principal's maximum payoff $V(e)$. From the preceding part of Proposition 3.4, there exist $(\lambda_0, \lambda_1, \psi)$ such that

$$v_p(\mu) + \psi h^e(\mu) \leq \lambda_0 + \lambda_1\mu \tag{15}$$

for all $\mu \in [0,1]$, with equality for all μ such that $\tau(\mu, e) > 0$. Note that $\tau(\mu, e)$ satisfies (BP) and (IC).

Step 1. We show that if $H(e) < 0$, then $\psi > 0$. Take the expected value with respect to $\tau(\mu, e)$ on

both sides of (15). Because (15) holds with equality for all μ such that $\tau(\mu, e) > 0$, we find that

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

where the last equality follows from (BP) and (IC). Next, take the expectation with respect to $\tau(\mu, e)$ in (14):

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

where the last equality follows from (BP). Combining these two conditions, we find

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

Take the expected value with respect to $\widehat{\tau}(\mu, e)$ on both sides of (14). Because (14) holds with equality for all μ such that $\widehat{\tau}(\mu, e) > 0$, we find that

$$E_{\widehat{\tau}}[v_P(\mu)] = \widehat{\lambda}_0 + \widehat{\lambda}_1 E_{\widehat{\tau}}[\mu] = \widehat{\lambda}_0 + \widehat{\lambda}_1 e,$$

where the last equality follows from (BP). Next, take the expected value with respect to $\widehat{\tau}(\mu, e)$ on both sides of (15).

$$E_{\widehat{\tau}}[v_P(\mu)] + \psi E_{\widehat{\tau}}[h^e(\mu)] \leq \lambda_0 + \lambda_1 E_\tau[\mu] = \lambda_0 + \lambda_1 e,$$

where the last equality follows from (BP). Combining these two conditions, we find

$$\widehat{\lambda}_0 + \widehat{\lambda}_1 e + \psi H(e) \leq \lambda_0 + \lambda_1 e. \tag{17}$$

Subtracting (16) from (17),

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

Hence, $H(e) < 0 \Rightarrow \psi \geq 0$. Finally, note that if $\psi = 0$, then the solution to the principal's problem is also a solution of the relaxed problem, and hence, $\widehat{V}(e) = V(e)$, contradicting $e > \underline{e}$.

Step 2. We show that if $e \in (\underline{e}, \overline{e})$, then $H(e) < 0$. We will prove the contrapositive; if $H(e) > 0$, then $e \leq \underline{e}$. Thus, suppose that for some e, $\widehat{\tau}(\mu, e)$ solves the relaxed problem and satisfies $H(e) \equiv E_{\widehat{\tau}}[h^e(\mu)] > 0$.

Case I: Suppose that $\widehat{\tau}(\mu)$ is not the only solution to the relaxed problem, and that there exists another solution, $\widehat{\tau}'(\mu)$ such that $E_{\widehat{\tau}'}[h^e(\mu)] < 0$. 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, $E_{\tau^w}[v_P(\mu)] = wE_{\widehat{\tau}}[v_P(\mu)] + (1-w)E_{\widehat{\tau}'}[v_P(\mu)] = \widehat{V}(e)$, and therefore $\tau^w(e)$ is also optimal in the relaxed problem. Next, note that $E_{\tau^w}[h^e(\mu)] = wE_{\widehat{\tau}}[h^e(\mu)] + (1 - w)E_{\widehat{\tau}'}[h^e(\mu)]$, and hence, there exists some $w^* \in$

$(0,1)$ such that $E_{\tau^w}[h^e(\mu)] = 0$. Thus, $\tau^w(\mu)$ satisfies (IC) at e'. Therefore, $\tau^w(\mu)$ also satisfies the necessary conditions for the principal's (unrelaxed) problem, and thus $V(e) = \widehat{V}(e)$. Hence $e \leq \underline{e}$.

Case II: Suppose that for all solution(s) of the relaxed problem, $H(e) > 0$; that is, if $\widehat{\tau}(\mu, e)$ solves the relaxed problem, then $E_{\widehat{\tau}}[h^e(\mu)] > 0$. Note first that all solution(s) to the relaxed problem that satisfy this condition must be supported on at least two posteriors. To see this, note that any solution of the relaxed problem that concentrates all mass on a single realization must concentrate all mass on $\mu = e$ in order to satisfy (BP), and $h^e(e) = -c'(e) < 0$. Note second that the support must contain at least one realization smaller than e and at least one larger than e in order to satisfy (BP). That is, there exist $\mu_L < e < \mu_H$ such that $\widehat{\tau}(\mu_L, e) > 0$ and $\widehat{\tau}(\mu_H, e) > 0$. Necessary and sufficient conditions for the relaxed problem imply that

$$
\begin{aligned}
(18) \qquad & v_P(\mu_L) = \widehat{\lambda}_0 + \widehat{\lambda}_1 \mu_L \geq v_P(\mu) && \text{for all } \mu \in [0,1] \\
& v_P(\mu_H) = \widehat{\lambda}_0 + \widehat{\lambda}_1 \mu_H \geq v_P(\mu) && \text{for all } \mu \in [0,1].
\end{aligned}
$$

Next, consider $\widetilde{e} \in [e, \mu_H]$, and note that the following distribution of posteriors, supported on $\{\mu_L, \mu_H\}$, satisfies (BP):

$$
\widetilde{\tau}(\mu_H) = \frac{\widetilde{e} - \mu_L}{\mu_H - \mu_L} \qquad \widetilde{\tau}(\mu_L) = \frac{\mu_H - \widetilde{e}}{\mu_H - \mu_L}
$$

Combined with (18), $\widetilde{\tau}$ satisfies the necessary and sufficient conditions for optimality in the relaxed problem. Next, note that

$$
E_{\widetilde{\tau}}[h(\mu, \widetilde{e})] = \frac{1}{\widetilde{e}(1-\widetilde{e})} \Big(\frac{\mu_H - \widetilde{e}}{\mu_H - \mu_L} (\mu_L - \widetilde{e}) v_A(\mu_L) + \frac{\widetilde{e} - \mu_L}{\mu_H - \mu_L} (\mu_H - \widetilde{e}) v_A(\mu_H) \Big) - c'(\widetilde{e}),
$$

which is a continuous function of $\widetilde{e}$. By assumption, all solutions of the relaxed problem at e generate $E_{\widetilde{\tau}}[h(\mu, e)] > 0$. Furthermore, it is clear from direct substitutions that for $\widetilde{e} = \mu_H$, $E_{\widetilde{\tau}}[h(\mu, \mu_H)] < 0$. Hence, an $e^* \in (e, \mu_H)$ exists for which the solution to the relaxed problem satisfies (IC). Thus, $V(e^*) = \widehat{V}(e^*)$, and hence, $\underline{e} \geq e^* > e$. ■

Proof of Proposition 5.4. We prove the result in three steps. First, we define a useful function $r(e)$, which determines the probability of revealing state 0, and establish some basic properties of the function. Second, we construct a specific binary signal and show that it is the unique binary solution to the relaxed problem in the absence of (IC). Finally, we show that there exists a unique incentive-effort level $\underline{e}$.

Step 1. Let

$$
r(e) \equiv \frac{\theta - \langle f(e), v_{DM} \rangle}{\theta f(0|e)}.
$$

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

$$\begin{aligned} & \frac{\theta - \langle f(e), v_{DM} \rangle}{\theta f(0|e)} \leq 1 \iff \theta - \langle f(e), v_{DM} \rangle \leq \theta f(0|e) \\ \iff \quad & \theta(1 - f(0|e)) = \theta \langle f(e), (0,1,...,1) \rangle \leq \langle f(e), v_{DM} \rangle \\ \iff \quad & \frac{\langle f(e), v_{DM} \rangle}{f(0|e)} = \frac{\langle f(e), v_{DM} \rangle}{\langle f(e), (0,1,...,1) \rangle} \geq \theta. \end{aligned}$$

To prove the preceding inequality, recall Assumption 3:

$$\frac{\langle f_0, v_{DM} \rangle}{1 - f_0(0)} = \frac{\langle f_0, v_{DM} \rangle}{\langle f_0, (0,1,...,1) \rangle} \geq \theta.$$

and the maintained assumption that $\langle f_1, v_{DM} \rangle > \theta$. Combining these with the facts that $f(e) = (1-e)f_0 + ef_1$ and $1 = \langle f_1, (1,...,1) \rangle \geq \langle f_1, (0,1,...,1) \rangle$, the desired result is obtained as follows:

$$\begin{aligned} \langle f(e), v_{DM} \rangle &= \langle (1-e)f_0 + ef_1, v_{DM} \rangle = (1-e)\langle f_0, v_{DM} \rangle + e\langle f_1, v_{DM} \rangle \\ &\geq (1-e)\theta \langle f_0, (0,1,...,1) \rangle + e\theta \langle f_1, (1,1,...,1) \rangle \\ &\geq (1-e)\theta \langle f_0, (0,1,...,1) \rangle + e\theta \langle f_1, (0,1,...,1) \rangle \\ &= \theta \langle (1-e)f_0 + ef_1, (0,1,...,1) \rangle = \theta \langle f(e), (0,1,...,1) \rangle. \end{aligned}$$

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

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

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

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

To verify the optimality of the signal, we apply a necessary and sufficient condition in Theorem 5.1, which applies unchanged to the relaxed problem with $\psi = 0$. In other words, we show that there exists a hyperplane that supports $v_P(\mu)$ inside $\Delta(\Omega)$. We explicitly construct such a hyperplane. Let $\lambda_0 = 0$ and $\lambda_1 = 1/\theta \cdot v_{DM}$, so that

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

For $\mu \in \mathcal{A}_0$, $v_P(\mu) = 0$, and therefore $v_P(\mu) \leq \lambda_0 + \langle \lambda_1, \mu \rangle = \frac{1}{\theta} \langle v_{DM}, \mu \rangle$, with equality holding if

and only if $\mu = \mu^- = (1, 0, ..., 0)$. For $\mu \in \mathcal{A}_1$, $v_P(\mu) = 1$, while $\lambda_0 + \langle \lambda_1, \mu \rangle = 1/\theta \cdot \langle v_{DM}, \mu \rangle \geq 1$ with equality holding if and only if $\langle \mu, v_{DM} \rangle = \theta$. It suffices to show that $\mu^+(e)$ belongs to the point where $v_P(\mu) = \lambda_0 + \langle \lambda_1, \mu \rangle$. This follows from

$$\begin{aligned} \langle \mu^+(e), v_{DM} \rangle &= \langle \frac{f(e) \odot (1 - r(e), 1, ..., 1)}{\langle f(e), (1 - r(e), 1, ..., 1) \rangle}, v_{DM} \rangle \\ &= \frac{\langle f(e), v_{DM} \rangle}{1 - f(0|e) r(e)} = \frac{\langle f(e), v_{DM} \rangle}{1 - \frac{\theta - \langle f(e), v_{DM} \rangle}{\theta}} = \frac{\theta \langle f(e), v_{DM} \rangle}{\langle f(e), v_{DM} \rangle} = \theta. \end{aligned}$$

Note that we use the fact that, since $v_{DM}(0) = 0$, $\langle f(e), (1 - r(e), 1, ..., 1) \rangle = f(e)$ 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,

$$\begin{aligned} E_\tau[h(\mu, e)] &= \tau(\mu^+(e), e) \langle \mu^+(e), (f_1 - f_0) \oslash f(e) \rangle - c'(e) \\ &= \langle \tau(\mu^+(e), e) \mu^+(e), (f_1 - f_0) \oslash f(e) \rangle - c'(e) \\ &= \langle f(e) \odot (1 - r(e), 1, ..., 1), (f_1 - f_0) \oslash f(e) \rangle - c'(e) \\ &= \langle (1 - r(e), ..., 1), f_1 - f_0 \rangle - c'(e) \\ &= r(e)(f_0(0) - f_1(0)) - c'(e), \end{aligned}$$

where the last equality is due to the fact that $\langle (1, ..., 1), f_1 - f_0 \rangle = 0$. $E_\tau[h(\mu, e)]$ is positive if $e = 0$ (because $r(0) > 0$ while $c'(0) = 0$) and negative if $e = \overline{e}$ (because $r(\overline{e})(f_0(0) - f_1(0)) < \sum_{\omega > \omega_e} (f_1(\omega) - f_0(\omega)) = c'(\overline{e})$). Since $E_\tau[h(\mu, e)]$ is continuous in e (because both $r(e)$ and $c'(e)$ are continuous), there exists e such that $E_\tau[h(\mu, e)] = 0$, 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

$$\begin{aligned} & r'(e) = \frac{-\langle f_1 - f_0, v_{DM} \rangle \theta f(0|e) - (\theta - \langle f(e), v_{DM} \rangle) \theta (f_1(0) - f_0(0))}{(\theta f(0|e))^2} \leq 0 \\ \iff \quad & \frac{\langle f_1 - f_0, v_{DM} \rangle}{f_0(0) - f_1(0)} \geq \frac{\theta - \langle f(e), v_{DM} \rangle}{f(0|e)}. \end{aligned}$$

We prove this inequality by establishing the following two inequalities:

$$\frac{\langle f_1 - f_0, v_{DM} \rangle}{f_0(0) - f_1(0)} \geq \theta \text{ and } \quad \theta \geq \frac{\theta - \langle f(e), v_{DM} \rangle}{f(0|e)}.$$

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

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

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

$$\langle f_1 - f_0, v_{DM}\rangle = \langle f_1 - f_0, (0, v_1, ..., v_n)\rangle \geq \theta(f_0(0) - f_1(0)) = \theta\langle f_1 - f_0, (0, 1, ..., 1)\rangle,$$

which is equivalent to

$$\langle f_1 - f_0, (0, v_1 - \theta, ..., v_n - \theta)\rangle = \sum_{\omega \geq 1}(f_1(\omega) - f_0(\omega))(v_\omega - \theta) \geq 0.$$

The expression can be decomposed into three pieces as follows:

$$\sum_{\omega=1}^{\omega_e}(f_1(\omega) - f_0(\omega))(v_\omega - \theta) + \sum_{\omega=\omega_e+1}^{\omega_r}(f_1(\omega) - f_0(\omega))(v_\omega - \theta) + \sum_{\omega=\omega_r+1}^{n}(f_1(\omega) - f_0(\omega))(v_\omega - \theta) \geq 0.$$

Recall that $f_1(\omega) - f_0(\omega) < 0$ 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, which is sufficient for the inequality. 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), $1 - f_0(\omega)/f_1(\omega)$ increases in ω. Combining this with the fact that $v_\omega - \theta < 0$ if and only if $\omega \leq \omega_r$,

$$\begin{aligned}
& \sum_{\omega=\omega_e+1}^{\omega_r}(f_1(\omega) - f_0(\omega))(v_\omega - \theta) + \sum_{\omega=\omega_r+1}^{n}(f_1(\omega) - f_0(\omega))(v_\omega - \theta) \\
= & \sum_{\omega=\omega_e+1}^{\omega_r}\left(1 - \frac{f_0(\omega)}{f_1(\omega)}\right)(v_\omega - \theta)f_1(\omega) + \sum_{\omega=\omega_r+1}^{n}\left(1 - \frac{f_0(\omega)}{f_1(\omega)}\right)(v_\omega - \theta)f_1(\omega) \\
\geq & \sum_{\omega=\omega_e+1}^{\omega_r}\left(1 - \frac{f_0(\omega_r+1)}{f_1(\omega_r+1)}\right)(v_\omega - \theta)f_1(\omega) + \sum_{\omega=\omega_r+1}^{n}\left(1 - \frac{f_0(\omega_r+1)}{f_1(\omega_r+1)}\right)(v_\omega - \theta)f_1(\omega) \\
= & \left(1 - \frac{f_0(\omega_r+1)}{f_1(\omega_r+1)}\right)\sum_{\omega=\omega_e+1}^{n}(v_\omega - \theta)f_1(\omega) \\
\geq & \left(1 - \frac{f_0(\omega_r+1)}{f_1(\omega_r+1)}\right)\sum_{\omega=0}^{n}(v_\omega - \theta)f_1(\omega) = \left(1 - \frac{f_0(\omega_r+1)}{f_1(\omega_r+1)}\right)\langle v_{DM} - \theta, f_1\rangle > 0,
\end{aligned}$$

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

Proof of Proposition 5.5. 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., $\lambda_0 + \langle\lambda_1, \mu\rangle \geq \mathcal{L}(\mu, \psi, e)$ 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, $\lambda_0 + \langle\lambda_1, \mu\rangle\mathcal{L}(\mu, \psi, e)$ if $\mu = \mu^-$ or $\mu = \mu^+$. The last two steps establish the optimality of the proposed binary signal via Theorem 5.5.

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

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

To verify (IC),

$$\begin{aligned} E_\tau[h(\mu, e)] =& \tau(\mu^+, e)\langle \mu^+, (f_1 - f_0) \oslash f(e)\rangle - c'(e) \\ =& \langle f(e) \odot \rho^+, (f_1 - f_0) \oslash f(e)\rangle - c'(e) \\ =& \rho^+ \odot (f_1 - f_0) - c'(e) \\ =& \sum_{\omega \geq k} (f_1(i) - f_0(i)) - \frac{c'(e) - c'(e_{k-1})}{f_0(k) - f_1(k)}(f_1(k) - f_0(k)) - c'(e) \\ =& \sum_{\omega > k-1} (f_1(i) - f_0(i)) - c'(e_{k-1}) = 0. \end{aligned}$$

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 $\mathcal{L}(\mu, \psi, e)$ at μ^- and μ^+. Let

$$\psi = -\frac{f(k|e)}{f_1(k) - f_0(k)},\ \lambda_0 = 1 - \psi c'(e),\ \text{and}\ \lambda_1(\omega) = \begin{cases} -1 & \text{if } \omega \leq k, \\ \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} & \text{if } \omega \geq k+1. \end{cases}$$

Note that $f_1(k) - f_0(k) < 0$ (because $k \leq \omega_e$) and, therefore, $\psi > 0$. In addition, with this specification,

$$\begin{aligned} \lambda_0 + \langle \lambda_1, \mu\rangle &= 1 - \psi c'(e) - \sum_{\omega=0}^{k} \mu(\omega) + \sum_{\omega=k+1}^{n} \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \mu(\omega) \\ &= 1 - \psi c'(e) - \left(1 - \sum_{\omega=k+1}^{n} \mu(\omega)\right) + \sum_{\omega=k+1}^{n} \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \mu(\omega) \\ &= \sum_{\omega=k+1}^{n} \left(1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)}\right) \mu(\omega) - \psi c'(e). \end{aligned}$$

The following fact is useful in what follows.

Lemma 6.2 *For any $k \leq \omega_e$,*

$$1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} = 1 - \frac{f(k|e)}{f_1(k) - f_0(k)} \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \begin{cases} < 0 \text{ if } \omega < k \\ > 0 \text{ if } \omega > k. \end{cases}$$

Proof. Since $f_1(k) - f_0(k) < 0$,

$$1 - \frac{f(k|e)}{f_1(k) - f_0(k)} \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} < 0 \iff \frac{f_1(k) - f_0(k)}{f(k|e)} - \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} > 0.$$

Arranging the terms with the fact that $f(\omega|e) = (1-e)f_0(\omega) + ef_1(\omega)$,

$$\frac{f_1(k) - f_0(k)}{f(k|e)} - \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} = \frac{f_1(\omega)f_1(k)}{f(k|e)f(\omega|e)} \left(\frac{f_0(\omega)}{f_1(\omega)} - \frac{f_0(k)}{f_1(k)} \right).$$

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

$$\begin{aligned} \lambda_0 + \langle \lambda_1, \mu \rangle - \mathcal{L}(\mu, \psi, e) &= \sum_{\omega=k+1}^{n} \left(1 - \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) - (-\psi c'(e)). \\ &= \sum_{\omega=k+1}^{n} \left(1 - \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \right) \mu(\omega) \geq 0, \end{aligned}$$

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

$$\begin{aligned} \mathcal{L}(\mu, \psi, e) &= 1 + \psi \left(\langle (f_1 - f_0) \oslash f(e), \mu \rangle - c'(e) \right) = 1 + \psi \left(\sum_{\omega} \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \mu(\omega) - c'(e) \right) \\ &= \sum_{\omega} \left(1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) \\ &= \sum_{\omega=1}^{k} \left(1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) + \sum_{\omega=k+1}^{n} \left(1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) \\ &= \sum_{\omega=1}^{k-1} \left(1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e) + \sum_{\omega=k+1}^{n} \left(1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \right) \mu(\omega) - \psi c'(e), \end{aligned}$$

where the last equality is due to the fact that $1 + \psi(f_1(k) - f_0(k))/f(k|e) = 0$. Therefore,

$$\lambda_0 + \langle \lambda_1, \mu \rangle - \mathcal{L}(\mu, \psi, e) = -\sum_{\omega=1}^{k-1} \left(1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)} \right) \mu(\omega) \geq 0,$$

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 $\mathcal{L}(\mu, \psi, e)$ when $\mu = \mu^-$ and $\mu = \mu^+$. Since $\mu^- \in \mathcal{A}_0$, as shown above,

$$\lambda_0 + \langle \lambda_1, \mu^- \rangle - \mathcal{L}(\mu^-, \psi, e) = \sum_{\omega=k+1}^{n} \left(1 - \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)}\right) \mu^-(\omega).$$

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 $\mu^- = 1/\langle f(e), \rho^- \rangle \cdot f(e) \odot \rho^-$ where $\rho^-(\omega) = 0$ for $\omega \geq k+1$. $\mu^+ \in \mathcal{A}_1$ and, therefore,

$$\lambda_0 + \langle \lambda_1, \mu^+ \rangle - \mathcal{L}(\mu^+, \psi, e) = -\sum_{\omega=1}^{k-1} \left(1 + \psi \frac{f_1(\omega) - f_0(\omega)}{f(\omega|e)}\right) \mu^+(\omega).$$

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

Proof of Proposition 6.1. First, we show that if $v_A''(\cdot) \leq 0$, then the agent's objective function in (13) is concave. The second derivative of the agent's payoff is

$$\begin{aligned}
&\sum_s (\pi_1(s) - \pi_0(s)) v_A'(\mu(s,e)) \frac{\pi_1(s)\pi_0(s)}{(e\pi_1(s) + (1-e)\pi_0(s))^2} + \\
&\quad v_A''(\mu(s,e)) \frac{(\pi_1(s)\pi_0(s))^2}{(e\pi_1(s) + (1-e)\pi_0(s))^3} - v_A'(\mu(s,e)) \frac{(\pi_1(s) - \pi_0(s))\pi_1(s)\pi_0(s)}{(e\pi_1(s) + (1-e)\pi_0(s))^2} - c''(e) = \\
&\sum_s v_A''(\mu(s,e)) \frac{(\pi_1(s)\pi_0(s))^2}{(e\pi_1(s) + (1-e)\pi_0(s))^3} - c''(e),
\end{aligned}$$

which is strictly negative if $v_A''(\cdot) \leq 0$. Hence, the agent's optimal effort choice is either the unique critical point, or a corner solution if no critical point exists.

If a signal is incentive compatible for effort level $e < \overline{e}$ assuming unobservable effort, then the signal cannot be fully-informative. Therefore, as argued in the text, the marginal benefit of effort is strictly higher at every e with observable effort. The result follows. ■

REFERENCES

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

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

Barron, Daniel, George Georgiadis, and Jeroen Swinkels, "Risk-taking and simple contracts," *mimeo*, 2016.

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.

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

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang, "Pivotal persuasion," *mimeo*, 2016.

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

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

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, Ming Li, Tymofiy Mylovanov, and Andriy Zapechelnyuk, "Persuasion of a privately informed receiver," *mimeo*, 2015.

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.

Renault, Jérôme, Eilon Solan, and Nicolas Vieille, "Optimal dynamic information provision," *arXiv preprint arXiv:1407.5649*, 2014.

Rodina, David, "Information design and career concerns," *mimeo*, 2016.

_ **and John Farragut**, "Inducing effort through grade," *mimeo*, 2016.

Roesler, Anne-Katrin and Balázs Szentes, "Buyer-optimal learning and monopoly pricing," *mimeo*, 2017.