Bayesian Persuasion and Moral Hazard*

(Preliminary and Incomplete)

Raphael Boleslavsky† and Kyungmin Kim‡

February 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.

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 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

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 (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

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

[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).

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, 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

[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.

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 action depends on the (conjectured) effort. In both Rodina (2016) and Rodina and Farragut

[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).

(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 5 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

[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.

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 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, $c(0) = c'(0) = 0$ and $c'(1) < 1$. As shown later, the assumption $c'(1) > 1$ ensures that the principal can never 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,

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

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

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

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

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 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

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

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 2 *Let $\overline{e}$ be the value such that $c'(\overline{e}) = 1$. Then, e is implementable if and only if $e \leq \overline{e}$.*

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

Proof. According to KG,

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

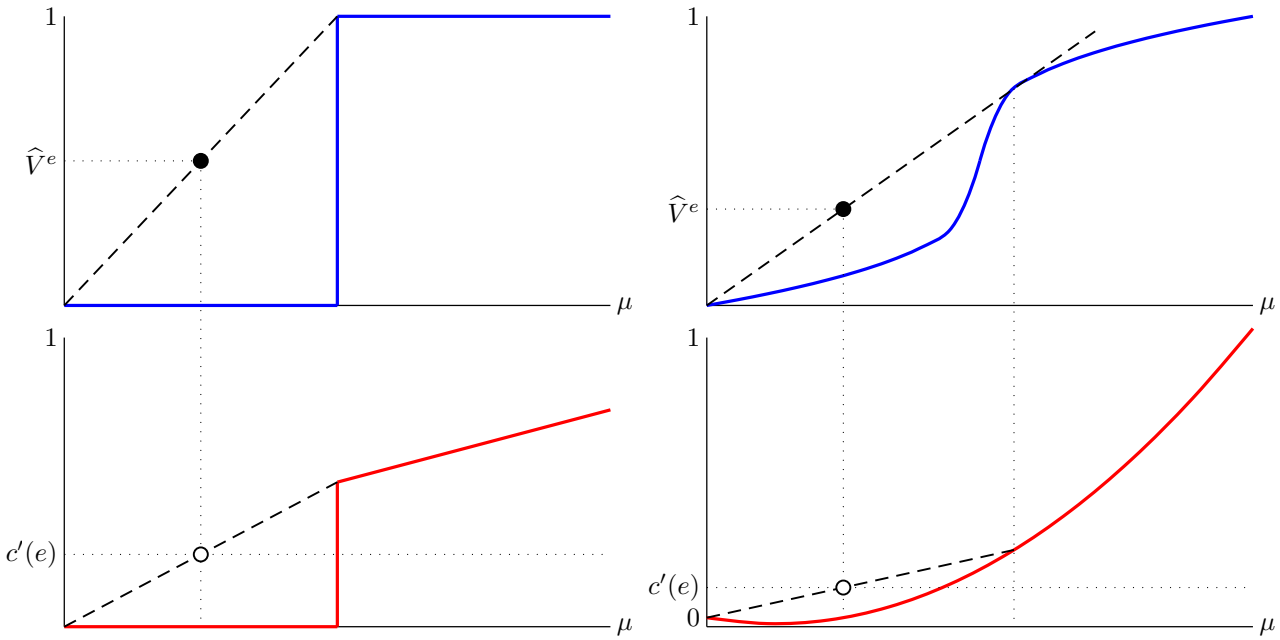

Figure 1: Finding $\underline{e}$ when $v_P(\mu) = v_A(\mu) = \mathcal{I}_{\{\mu \geq 1/2\}}$ and when $v_P(\mu) = (1 + (2\mu - 1)^{1/3})/2$ and $v_A(\mu) = \mu$. The upper panels draw $v_P(\mu)$, while the lower panels show $(\mu - e)v_A(\mu)/(e(1-e))$. The cost function used for the left panel is $c(e) = e^2/4$, while that for the right panel is $c(e) = 5e^2/6$.

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 1 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

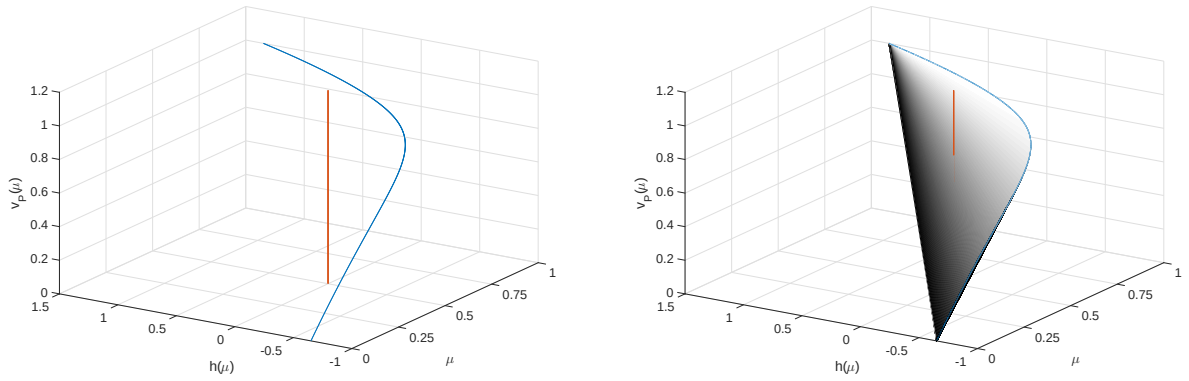

Figure 2: The left panel depicts the curve K, while the right panel depicts its convex hull $co(K)$. In this example, $e = 0.5$.

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 2 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 2. 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 2, 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

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

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^*. Notice that this means that K^* represents all possible convex combinations of the points in K that satisfy both BP constraint ($\sum_s \mu_s \tau(\mu_s) = 0$) and 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.

Theorem 1 *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 2 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, which, as shown in the next section, allows us to derive an optimal solution in a simple fashion in many examples. The above geometric analysis suggests that there is a hyperplane that is tangent to $co(K)$ at $(e,0,V^e)$. This means that there exists a (normalized) direction vector $d = (-\lambda_1, \psi, 1)$ and a scalar λ_0 such that $d \cdot x \leq \lambda_0$ for any $x = (\mu, h^e(\mu), v_P(\mu)) \in co(K)$ and

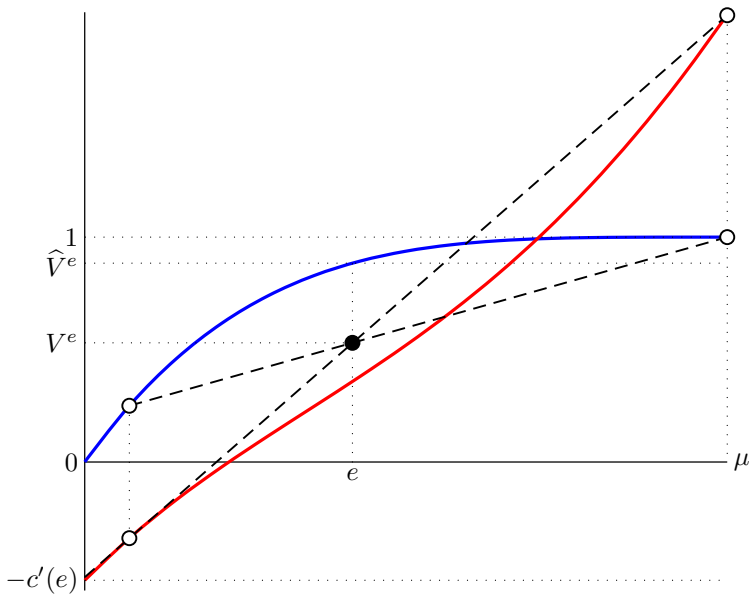

Figure 3: The concave solid curve depicts $v_P(\mu) = 1-(1-\mu)^4$, while the other solid curve depicts $\mathcal{L}(\mu, \psi) = v_P(\mu) + \psi h^e(\mu)$ when $v_A(\mu) = \mu$.

$d \cdot x = \lambda_0$ if $\tau^e(\mu) > 0$. Since $co(K)$ is the convex hull of K, it is necessary and sufficient that the former part holds only for $x \in K$. Arranging the terms, we obtain the following result.[7]

Corollary 1 *If τ^e is an optimal distribution of posteriors that induces e, then 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 holding if $\tau^e(\mu) > 0$. If $e \in (\underline{e}, \overline{e})$, then $\psi > 0$.

Proof. See the appendix for a proof for the last statement. ■

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 3, $v_P(\mu)$ 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 equal to 0 if $v_P(\mu)$ is concave). Figure 3 shows how $\mathcal{L}(\mu, \psi)$ differs from $v_P(\mu)$ how it affects the shape of the straight line. The

[7]Note that Corollary 1 provides only a necessary condition. The underlying reason is identical to that for the method of Lagrange multipliers.

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 3, 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

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 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 *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 1, 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, 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 4). In the quadratic case where $v_P(\mu) = 1 -$

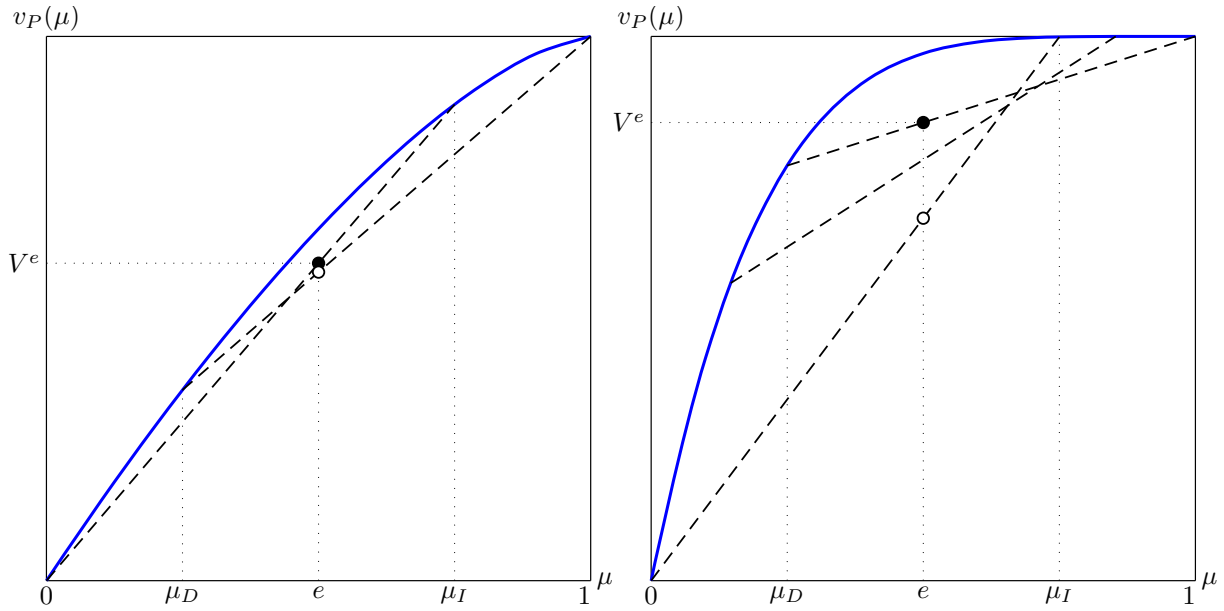

Figure 4: The left panel depicts the case in which $v_P(\mu) = 1 - (1-\mu)^{1.5}$ (i.e., $\eta = 1.5$), while the right panel shows the case in which $v_P(\mu) = 1 - (1-\mu)^5$ (i.e., $\eta = 5$).

$(1-\mu)^2$, 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 4), 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 4) 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

[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.

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

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

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 5 *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 (2) 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 $b < 1/2$. ■

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

[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.

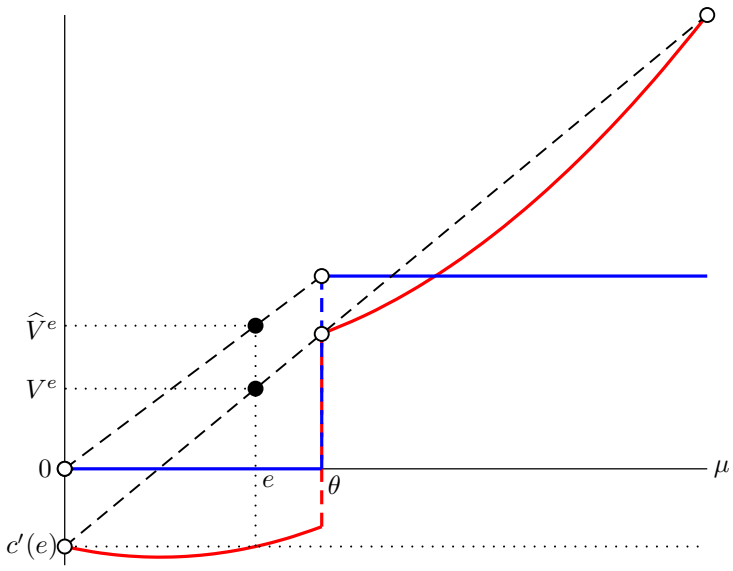

Figure 5: The step function represents $v_P(\mu) = \mathcal{I}_{\{\mu \geq \theta\}}$, while the solid curve depicts $\mathcal{L}(\mu, \psi)$ when $v_A(\mu) = \mu$.

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 $e \in (\underline{e}, \overline{e})$.

[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.

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 5). 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 5.

Proposition 6 *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:

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

■

Among other things, this case demonstrates that the result on the number of necessary posteriors in Theorem 1 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 Discussion

5.1 Generalization

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. If $|\Omega| > 2$, then the players' beliefs can no longer be represented by a one-dimensional variable and, therefore, we cannot provide as many clear and concrete results as in the previous two sections. Nevertheless, it is possible to generalize our main characterization (Theorem 1), as shown below.

Suppose that there are $n(\geq 0)$ states (i.e., $|\Omega| = n$). In this case, the players' beliefs μ are represented by an element in a $(n-1)$-simplex $\Delta(\Omega) \equiv \{(x_1, ..., x_n) \in [0,1]^n : \sum_{k=1}^n x_k = 1\}$.

The state ω is determined 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 assume that $f(\omega|\cdot)$ is continuously differentiable. It is convenient to assume that $f(\cdot|e)$ has full support given any e. A signal, chosen by the principal, 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.

As in the baseline model, first consider the subgame between the agent and the decision-maker given a signal. Given an equilibrium effort e^*, the decision-maker's belief following a signal realization s is given by

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

Given this, the agent's problem is

$$\max_e \sum_\omega \sum_s v_A(\mu_s)\pi(s|\omega)f(\omega|e) - c(e) = \sum_s v_A(\mu_s)\pi(s|e) - c(e),$$

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

$$\sum_s v_A(\mu_s)\pi_e(s|e^*) - c'(e^*) = 0.$$

As in other principal-agent problems, this condition is not sufficient for the agent's optimal effort choice in general, but it is necessary to assume the property in order to proceed further.[11]

As in the baseline model, we rewrite the first-order condition in terms of a distribution of posteriors τ, instead of a signal π. As reported in KG, given e^* (and the consequent prior distribution $f(\cdot|e^*)$, a Bayes-plausible distribution of posteriors can be induced by a signal π such that $\pi(s|\omega) = \mu_s(\omega)\tau(\mu_s)/f(\omega|e^*)$. Inserting this into the above first-order condition and arranging the terms yield

$$E_\tau\left[\left(\sum_\omega \mu(\omega)\frac{f_e(\omega|e^*)}{f(\omega|e^*)}\right)v_A(\mu)\right] - c'(e^*) = 0,$$

[11] One specification under which the first-order approach is valid is when $f(\omega|e)$ takes the following form: there exist two probability assignment functions $f_0, f_1 : \Omega \to [0,1]$ such that

$$f(\omega|e) = ef_1(\omega) + (1-e)f_0(\omega) \text{ for all } \omega \in \Omega \text{ and } e \in [0,1].$$

In this case, as in our baseline model, the first term in the agent's objective function is linear in e and, therefore, the first-order condition is necessary and sufficient for the agent's optimal effort.

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

$$h^e(\mu) \equiv \left(\sum_\omega \mu(\omega) \frac{f_e(\omega|e)}{f(\omega|e)} \right) v_A(\mu) - c'(e).$$

This is a generalization of the IC constraint in the baseline model. Given this constraint, it is straightforward to generalize the geometric argument for the baseline model.

Theorem 2 *Suppose that the set of states Ω has $n(\geq 0)$ elements and the first-order approach is valid. For any implementable e,*

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

where

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

and there exists an optimal distribution of posteriors $\tau^e \in \Delta(\Delta(\Omega))$ such that its support contains at most $n+1$ posteriors (i.e., $|supp(\tau^e)| \leq n+1$). In addition, there exist $\lambda_0 \in \mathcal{R}$, $\lambda_1 \in \mathcal{R}^n$, and $\psi \in \mathcal{R}$ such that

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

with equality holding if $\tau^e(\mu) > 0$.

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 $n+1$ posteriors follows from the fact that $f(\cdot|e), 0, V^e)$ 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 1. ■

5.2 Observable Efforts

We have assumed that the agent's effort is not observable to the other two players. Certainly, the principal prefers to observe e, because if so, she can condition a signal on e as well (i.e., if a signal is a function $\pi(s|\omega, e)$) and, therefore, implements each effort level more efficiently. For example, if v_A is concave (convex), then she can reveal full (no) information, unless the agent chooses a particular effort level.

For the decision-maker's problem, cnsider the concave-linear case in Section 4.1 and now suppose that the agent's effort e is observable to the decision-maker. In this case, the agent's

problem is given by

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

where $\mu(s,e) = e\pi_1(s)/(e\pi_1(s)+(1-e)\pi_0(s))$. The difference from the baseline model is that the decision-maker's posterior belief $\mu(s,e)$ now depends not only on a signal realization s but also on actual effort e. When $v_A(\mu) = \mu$, independent of a signal π, the problem reduces to $e - c(e)$, because

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

It then follows that the agent chooses $\overline{e}$ and, since v_P is concave, it is optimal for the principal to reveal no further information about ω.

This example demonstrates that it is not so clear cut whether the decision-maker would prefer to observe e. It would increase the agent's effort, but the principal may choose to provide less information. Depending on the decision-maker's underlying preferences, on which we do not impose any restrictions, the decision-maker may prefer not to observe the agent's effort, which is in good contrast to a conventional wisdom in the principal-agent problem.

Appendix: Omitted Proofs

Proof of Proposition 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 2. 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

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

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 Corollary 1. We define another programming problem as follows:

$$\tilde{V}^e = \max_{\tau \in \Delta(\Delta(\Omega))} E_\tau[v_P(\mu)],$$

subject to

$$(BP)\ \ E_\tau[\mu] = e \text{ and } (IC')\ \ E_\tau[h^e(\mu)] \geq 0.$$

This problem is more relaxed than the original problem but more stringent than the problem without IC. Therefore, $V^e \leq \tilde{V}^e \leq \widehat{V}^e$ for any $e \leq \overline{e}$. Now let

$$\underline{e}' \equiv sup\{e : \tilde{V}^e = \widehat{V}^e\}.$$

Certainly, $\underline{e}' \geq \underline{e}$, because $\tilde{V}^e \geq V^e$ for any e. If $\underline{e}' = \underline{e}$, then the desired result (that $\psi > 0$ for any $e \in (\underline{e}, \overline{e})$) immediately follows from Kuhn-Tucker theorem.

Suppose that $\underline{e}' > \underline{e}$. Combining with the fact that $V^{\overline{e}} < \widehat{V}^{\overline{e}}$ (otherwise, $\underline{e} = \overline{e}$), this implies that there exists $\widehat{\tau}$ such that $E_{\widehat{\tau}}[v_P(\mu)] = \widehat{V}^{\overline{e}'}$, $E_{\widehat{\tau}}[\mu] = \overline{e}'$, and $E_{\widehat{\tau}}[h^e(\mu)] < 0$. Meanwhile, since both BP and IC' are closed, there exists $\tilde{\tau}$ such that $E_{\tilde{\tau}}[v_P(\mu)] = \widehat{V}^{\overline{e}'}$, $E_{\tilde{\tau}}[\mu] = \overline{e}'$, and $E_{\tilde{\tau}}[h^e(\mu)] > 0$ (note that $E_{\tilde{\tau}}[h^e(\mu)] = 0$ contradicts $\overline{e}' > \overline{e}$). Now define $\tau^w = w\widehat{\tau} + (1-w)\tilde{\tau}$. For any $w \in [0,1]$,

$$E_{\tau^w}[v_P(\mu)] = wE_{\widehat{\tau}}[v_P(\mu)] + (1-w)E_{\tilde{\tau}}[v_P(\mu)] = \widehat{V}^{\overline{e}'},$$

$$E_{\tau^w}[\mu] = wE_{\widehat{\tau}}[\mu] + (1-w)E_{\tilde{\tau}}[\mu] = \overline{e}',$$

and

$$E_{\widehat{\tau}}[h^e(\mu)] \leq E_{\tau^w}[h^e(\mu)] = wE_{\widehat{\tau}}[h^e(\mu)] + (1-w)E_{\tilde{\tau}}[h^e(\mu)] \leq E_{\tilde{\tau}}[h^e(\mu)].$$

As w increases from 0 to 1, $E_{\tau^w}[h^e(\mu)]$ continuously rises from $E_{\widehat{\tau}}[h^e(\mu)](< 0)$ to $E_{\tilde{\tau}}[h^e(\mu)](> 0)$. This means that there exists w^* such that $E_{\tau^{w^*}}[h^e(\mu)] = 0$. Since τ^{w^*} satisfies the original two constraints but achieves $\widehat{V}^{\overline{e}'}$, $\overline{e} \geq \overline{e}'$, which is a contradiction. ■

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.

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.