Bayesian Persuasion and Costly State Verification*

Subhajit Pramanik †
Souvik Roy ‡
Raghul S. Venkatesh§

January 20, 2026

Abstract

We study a model in which a sender tries to persuade a decision-maker by committing to a signaling strategy (Kamenica and Gentzkow, 2011). The decision-maker observes this and, in turn, commits to an *interim* verification strategy that reveals the true state of the world with an endogenously determined probability. The possibility of verification induces a trade-off for the sender: providing more information decreases the verification probability but also decreases the ex-ante persuasion payoff. The decision-maker trades off learning the state with more precision against the cost of verification. We provide the precise conditions under which introducing verification leads to *(a)* strictly more informative signaling, and *(b)* higher ex-ante welfare for the decision-maker compared to the canonical persuasion game.

JEL codes: D82

Keywords: bayesian persuasion, costly state verification, commitment

*We are thankful to Sourav Bhattacharya, Rahul Deb, Shraman Banerjee, and seminar participants at Shiv Nadar University (New Delhi-NCR) and Centre for Mathematical and Computational Economics at IIT Jodhpur, for useful comments and suggestions.

†Department of Economics and Management, University of Padova, Italy; Email: subhajit.pramanik@phd.unipd.it

‡Applied Statistics Unit, Indian Statistical Institute Kolkata, India; Email: souvik.2004@gmail.com

§Department of Economic Sciences, IISER Bhopal, India; Email: rsvecon@gmail.com

1 Introduction

In many economic settings, an agent would like to persuade a decision-maker about an unknown state (quality, prices, etc.) by strategically committing to disclosing information. In a seminal paper, Kamenica and Gentzkow (2011) consider this problem as a model of persuasion whereby the sender commits to any distribution of signal realizations as a function of the underlying state. With one-sided commitment power, the sender usurps all the benefits from persuasion. However, there are scenarios where the decision-maker can independently verify the underlying state with some precision by directly paying a cost. For example, a judge can order an independent inquiry commission to verify the evidence submitted by a prosecutor; an insurance company could hire experts to verify claims made by firms or individuals; a buyer could consult with an expert to find out the true underlying quality of the product. By appropriately choosing a verification strategy, the decision-maker could *(i)* incentivize the sender to design more informative signals and *(ii)* appropriate a share of the sender's value from persuasion. While recent literature has addressed the question of ex-post information acquisition (Matyskova and Montes, 2023) and ex-ante commitment (Tsakas, Tsakas, and Xefteris, 2021) by the decision-maker, settings with *interim* commitment have not yet been studied in the context of persuasion.

In this paper, we study a model of Bayesian persuasion between a sender (her) and decision-maker (him), with the innovation that the decision-maker is able to commit to a costly verification strategy *after* observing the sender's persuasion strategy. In this setting, we ask, does the sender provide more informative signals to the decision-maker in the presence of two-sided commitment? Under what conditions does the introduction of verification technology improve the decision-maker's welfare compared to canonical persuasion?

To specifically address these questions, we restrict attention to a stylized setup with two states and two actions. We assume that the sender and decision-maker have no common interests: the decision-maker wants to take an action that matches the state, while the sender wants him to take the highest possible action. The sender first publicly commits to a signaling strategy, and after observing this, the decision-maker in turn commits to costly effort to try to discover the true underlying state. The effort measures the probability with which the state is (perfectly) revealed to the decision-maker. Kamenica and Gentzkow (2011) simplifies the sender's optimal persuasion strategy in the absence of decision-maker commitment. In particular, the persuasion problem is equivalent to the sender choosing a distribution of posteriors, subject to the posteriors averaging back to the prior. The sender's payoff is the value function over the posterior belief induced by each signal realization. Given these two observations, they derive the optimal signal from the concavification of the sender's value function. This is the optimal persuasion strategy in the canonical problem (henceforth, optimal strategy). With two-sided sequential commitment, the sender's signaling strategy takes into account the best response of the decision-maker in the continuation game. Since the commitment strategies are undertaken sequentially and the sender is the first mover, the verification probability in equilibrium is a function of the sender's persuasion strategy.

We first characterize the decision-maker's equilibrium verification as a function of the informativeness of the signal. Specifically, any signaling strategy induces an expected state-mismatch loss due to *wrongful belief attribution*. It measures the probability that the decision-maker attributes the state to be when it is indeed the other state. The loss from wrongful belief attribution quantifies the marginal benefits of verification to the decision-maker. Therefore, in equilibrium, the verification probability is increasing in the marginal benefits, which in turn is decreasing in the informativeness of the signal.

When verification fails, the payoff to the sender depends on the persuasion strategy chosen. When verification is successful, the sender gets a payoff equivalent to the expected state. In equilibrium, the sender's value function is a convex combination of these two expected payoffs, weighted by the verification probability. Since the value of persuasion is positive, the sender always prefers the decision-maker to verify with a lower probability. This induces a "*information vs. verification* trade-off for the sender. By providing more informative signals, the sender can lower the state-mismatch losses for the decision-maker, which in turn lowers the verification probability. However, more informative signaling decreases the value of persuasion for the sender. Therefore, an equilibrium persuasion strategy balances information provision and verification incentives.

We show that equilibrium signaling is weakly more informative compared to the regime with no verification. This is because the sender can guarantee herself an expected payoff that is at least equivalent to that under optimal persuasion (i.e., the solution to the canonical Bayesian persuasion problem). Any signaling strategy that is less informative than the optimal strategy would result in greater losses to the decision-maker and therefore be verified with a higher probability. Since less informative strategies also give lower expected payoffs to the sender when verification fails, they cannot be an equilibrium of the two-sided commitment game.

Next, we characterize the precise condition for the sender's strategy to be strictly more informative. Intuitively, this requires the sender's value function to be increasing at the optimal strategy. For this to hold, we require the marginal benefits from a more informative signaling strategy to be higher than the marginal costs. The former consists of two components. The first is the benefit to the sender from a decrease in the probability of verification arising from a more informative signal. The second is the difference between the payoff from the optimal strategy (determined by the concavification) and the payoff when verification is successful (equal to the expected state). The greater this difference, the greater the magnitude of the impact on the change in verification probability from a more informative signal. The marginal costs are given by the loss in payoff from a suboptimal persuasion strategy. In order for the sender to use a suboptimal signal, it must therefore hold that the benefits from decreased verification are higher compared to the expected utility losses due to this change.

Equipped with the precise characterization of the equilibrium verification strategy, we identify a possibility for verification to improve the decision maker's ex-ante welfare. This follows from two observations. When verification is successful, the decision-maker action matches the revealed state. This implies that the only time the decision-maker faces a loss from taking the wrong action is when verification fails with the complementary probability. However, for using the verification technology, the decision-maker pays a cost. Therefore, for the decision-maker to benefit, this cost of verification measured at the equilibrium verification strategy must outweigh the net benefits. This is determined by the probability of failure and the expected loss to the decision-maker under the equilibrium signaling strategy.

Related Literature. Following Kamenica and Gentzkow (2011), henceforth KG, a number of recent papers have extended the canonical model in various directions.[1] None of these papers address the possibility of a strategic decision-maker. The two papers closest to our setup are the ones by Matyskova and Montes (2023) and Tsakas et al. (2021).[2] The former investigates persuasion by allowing for costly information acquisition by a decision-maker once the signals are revealed (ex post). In Tsakas et al. (2021), the decision-maker commits to a resistance strategy ex ante, *before* the sender's persuasion strategy is chosen. We instead focus on *interim commitment* in that the decision-maker's commitment strategy is chosen after the sender's persuasion strategy but before signals are realized. The cost functions used by these two papers are also different from our setup. Specifically, Matyskova and Montes (2023) consider posterior-separable cost functions where the decision-maker can acquire information independent of the signaling strategy of the sender.[3] Tsakas et al. (2021) model a stochastic cost function via a distribution chosen by the receiver. That is, the resistance strategy is akin to money burning by the receiver. In contrast, we allow for two-sided sequential commitments

[1]These include models with heterogeneous priors (Alonso and Câmara, 2016; Galperti, 2019), partial commitment (Lipnowski, Ravid, and Shishkin, 2022; Perez-Richet and Skreta, 2022), privately informed sender (Koessler and Skreta, 2023; Zapechelnyuk, 2023), and communication constraints (Doval and Skreta, 2024; Le Treust and Tomala, 2019). Kamenica (2019) and Bergemann and Morris (2019) provide a comprehensive survey.

[2]Yoder (2022) considers a contracting problem in which the uninformed principal screens the agent whose private type is unobservable. The screening process therefore involves commitment to transfers by the principal.

[3]Cost functions that exhibit a posterior-separable form are extensively used in the literature on rational inattention (e.g., Sims (2003)). Gentzkow and Kamenica (2014) also study a problem of costly persuasion, which extends the basic model by making the signals potentially costly for the sender. They also use the posterior-separable form to model the cost function.

and model the information acquisition technology along the lines of Che and Kartik (2009) and Dur and Swank (2005).

Another paper closely related to our model is the work by Ederer and Min (2022). They consider a model of Bayesian persuasion in which the decision-maker can detect lies with positive probability. Crucially, in their setup, detection probability is exogenous. In the cheap talk literature, some papers have studied the possibility of verification by an uninformed decision-maker. Prominent among them include Balbuzanov (2019), Carroll and Egorov (2019), and Dziuda and Salas (2018). These papers do not consider the problem of commitment since neither the sender nor the decision-maker commits ex ante to their strategies.[4]

2 Model

Consider two players: a sender and a decision-maker (DM). There are two states of the world $\Omega = \{\omega_1, \omega_2\}$, where $\omega_1, \omega_2 \in \mathbb{R}$ and $\omega_1 < \omega_2$. The set of actions available to the DM is denoted by $A = \Omega$ such that the set of actions equals the set of states. The utility function of the players depends on the DM's action a and the state of the world ω. The utility function for the DM is $u(a, \omega) = -(a - \omega)^2$; the sender's utility function is $v(a, \omega) = a$. That is, the sender has monotone preferences over the DM's actions. The DM, on the other hand, prefers to match the action with the state. Players share a common prior $\mu_0 \in \Delta(\Omega)$ where $\mu_0(\omega)$ is the prior probability that the state is $\omega \in \Omega$. Both players are expected utility maximizers.

We denote $S = \{s_1, s_2\}$ as the set of signal realizations. A signaling strategy maps each state to a probability distribution over signal realizations. Let Σ denote the set of all possible signaling strategies. A strategy $\pi \in \Sigma$ is a mapping $\pi : \Omega \to \Delta(S)$. We denote by $\pi_\omega(s)$ the probability of the signal s when the realized state is $\omega \in \Omega$. Given the sender's choice of π and the signal realization s, the posterior belief distribution is $\mu_s \equiv \mu_\pi(\cdot \mid s)$. $\mu_s(\omega)$ is the belief that the state is ω when the realized signal is s. We use the notation τ_π to indicate the distribution of posteriors induced by the signal π, where $\mathbf{E}_{\mu \in \tau_\pi} \mu = \mu_0$. Utility maximizing action for the DM is defined as $a^*(\mu) = \text{argmax}_{a \in A} \mathbb{E}_{\omega \sim \mu} u(a, \omega)$. For a given μ, the sender's expected utility is therefore, $\hat{v}(\mu) = \mathbf{E}_{\omega \sim \mu} v(a^*(\mu), \omega)$. Without verification, let the optimal persuasion strategy dictated by KG be π_{op}. Let V_{op} be the expected payoff of the sender under this strategy.

The verification strategy of the DM is a mapping $q : \Sigma \to [0, 1]$. We denote by Φ the set of all verification strategies. The effort choice of the DM is modeled as the outcome probability

[4] More recently, a growing literature has looked into mechanism design with costly verification (see, e.g., Ben-Porath, Dekel, and Lipman (2014); Kartik and Tercieux (2012); Mylovanov and Zapechelnyuk (2017)). Our paper focuses on persuasion outcomes in the presence of costly verification by the decision-maker, which is different from the mechanism design approach.

of successful verification. The cost of verification is a mapping $c : [0,1] \to \mathbb{R}^+$, where q is both the effort and the probability of observing the underlying state and $c(q)$ is the total cost incurred by the DM. The function $c(\cdot)$ is smooth, $c'(\cdot), c''(\cdot) > 0$, and satisfies the Inada conditions $c'(0) = 0$ and $c'(q) \to \infty$ as $q \to 1$. The timing of the two-sided commitment game is as follows:

1. Sender chooses a signaling strategy $\pi \in \Sigma$
2. DM observes the signaling strategy π and chooses a verification strategy $q_\pi \in [0,1]$.
3. Nature chooses ω according to μ_0 and the signal realization s is according to π.
4. DM verifies the state with probability $q(\pi)$.
5. DM takes the action $\bar{a}(\omega) = \omega$ when verification succeeds and action $a^*(\mu_s)$ when it fails.
6. Payoffs are realized.

The extensive-form captures a two-sided commitment game. Specifically, the DM chooses and commits to a probability of verification *after* observing the signaling strategy but *before* the state or the associated signals are realized. This implies that the ex-ante problem for the DM is a choice of q depending on the sender's strategy π.

Discussion of setup Consider a buyer-seller transaction over used cars. The seller can choose signals (e.g., a test drive) that convey positive information about the quality of the car. The buyer could hire an expert to check the other technical components of the car. Crucially, the price that the buyer is willing to pay for an expert depends on what the seller chooses as her commitment strategy. Consequently, if the seller were to perfectly reveal the quality by designing fully informative signals, the buyer would have no incentives to hire the car expert. In the case of insurance markets, suppose a firm wants to claim damages to its building due to an extreme weather incident. The firm submits evidence that provides information about the genuineness of the damage to the insurance company. The firm can exaggerate its damages and claim more than what is commensurate. In such scenarios, insurance companies hire a third-party "*verification specialist*" to thoroughly investigate the claims made by the firm (e.g., through physical verification). This feature is also common in financial markets. In the case of private equity investments, mergers, or acquisitions, the firm planning to invest in another firm engages in due diligence operations. This involves costly audits that try to uncover the true value of the target firm. The extent of scrutiny depends on the financial disclosures of the target firm. When the disclosure is ex ante not informative, a more stringent audit ensues. The DM's commitment strategy in the model captures these types of strategic interactions.

Our formulation of the cost function is different from — or similar to — both Che and Kartik (2009) and Dur and Swank (2005) in certain aspects. In Dur and Swank (2005), the information acquisition results in a *perfect* signal about the underlying state, similar to our setup. However, they model the probability of the state being revealed as an increasing and concave function of an underlying effort choice. In Che and Kartik (2009), the outcome of information acquisition is imperfect in that only a normally distributed signal about the true state is revealed. However, there is no distinction between effort choice and the probability of success. We treat both effort and the probability of successful verification as one and the same, and verification perfectly reveals the state.

Equilibrium Definition Given a pair (π, q), the value functions of the DM and sender, $U(\pi, q)$ and $V(\pi, q)$ respectively, are:

$$U(\pi,q) = q(\pi)\cdot \mathbb{E}_{\omega\sim\mu_0}\Big[u\Big(\bar{a}(\omega),\omega\Big)\Big] + (1-q(\pi))\cdot \mathbb{E}_{\omega\sim\mu_0}\mathbb{E}_{s\sim\pi(\omega)}\Big[u\Big(a^*(\mu_s),\omega\Big)\Big] - c(q(\pi)) \tag{1}$$

$$V(\pi,q) = q(\pi)\cdot \mathbb{E}_{\omega\sim\mu_0}\Big[v(\bar{a}(\omega),\omega)\Big] + (1-q(\pi))\cdot \mathbb{E}_{\omega\sim\mu_0}\mathbb{E}_{s\sim\pi(\omega)}\Big[v(a^*(\mu_s),\omega)\Big] \tag{2}$$

We focus on Perfect Bayesian Equilibrium (henceforth, equilibrium) of the two-sided commitment game. An equilibrium is a pair (π^*, q^*) such that, *(i)* the DM chooses q^* to maximize $U(\pi^*, q)$; and *(ii)* the sender chooses a signaling strategy π^* that maximizes $V(\pi, q^*)$.

- **DM optimality:** q^* maximizes the value function $U(\pi^*, q)$, conditional on observing π^*

$$U(\pi^*,q^*) \quad = \quad \max_{q\in[0,1]} U\Big(\pi^*,q\Big) \tag{3}$$

- **Sender optimality:** the signaling strategy π^* maximizes the value function $V(\pi, q^*)$ conditional on q^*

$$V(\pi^*,q^*) \quad = \quad \max_{\pi\in\Sigma} V\Big(\pi,q^*\Big) \tag{4}$$

- **Value of verification:** the difference between DM's value function under verification and under canonical persuasion

$$\text{value of verification} \quad = \quad U(\pi^*,q^*) - U(\pi_{op},0)$$

3 Equilibrium Characterization

We use the following notational conventions going forward. Let $\mu_s(\omega)$ be the posterior belief that the state is ω when the realized state is s. Let $V_{\mu_0} = \left[\mu_0(\omega_1)\omega_1 + (1-\mu_0(\omega_1))\omega_2\right]$ be the expected payoff to the sender when verification succeeds. The value function of the sender, $V(\pi, q(\pi))$, is a convex combination between V_{μ_0} and the expected payoff from the Bayesian persuasion strategy π, V_π. That is, $V(\pi, q(\pi)) = q(\pi)V_{\mu_0} + (1-q(\pi))V_\pi$. In the case of the optimal strategy π_{op}, let the verification be $q_{op} = q(\pi_{op})$. Therefore, $V(\pi_{op}, q_{op}) = q_{op}V_{\mu_0} + (1-q_{op})V_{op}$. Clearly, given that the ex-post payoffs of the sender v is upper-semicontinuous and V_π is a continuous function defined for all $\pi \in \Pi$, the sender can always find an information structure and an associated distribution of posteriors that guarantees the payoff $V(\pi_{op}, q_{op})$, in the absence of verification. That is, there exists a $\underline{\pi}$ such that, in the absence of verification by the DM, $V_{\underline{\pi}} = V(\pi_{op}, q_{op})$. We refer to $V_{\underline{\pi}}$ as the lower bound that the sender can achieve in the presence of verification. In other words, by sticking to the optimal persuasion strategy, the sender can guarantee an expected payoff at least $V_{\underline{\pi}}$ when there is verification. Figure 1 captures these payoffs.

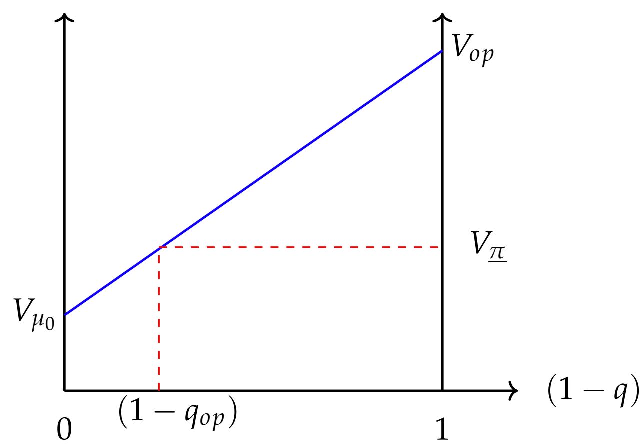

Figure 1: The expected payoff from π_{op} in the presence of verification is on the line joining V_{μ_0} and V_{op}.

Optimal Verification

To characterize the verification strategy $q(\pi)$, we first compute the marginal costs and marginal benefits from verification. This is given by the first-order condition from Equation 1:

$$c'(q) = -\mathbb{E}_{\omega\sim\mu_0}\mathbb{E}_{s\sim\pi(\omega)}\left[u\Big(a(\mu_s), \omega\Big)\right] \tag{5}$$

Intuitively, when both the posteriors induced by π result in the same action as the prior, i.e., $a(\mu_s) = a(\mu_0) = \omega_1$, the ex ante utility of the DM is unchanged. This implies that the verification decision of the DM is unchanged for any π that does not induce ω_2 action. Suppose $\Pi_{\mu_0} = \left\{\pi : a(\mu_{s_1}) = a(\mu_{s_2}) = \omega_1\right\}$ represents the set of all signaling strategies that have this feature.

Lemma 1. *The ex ante expected marginal benefit to the DM when*

1. $\pi \in \Pi_{\mu_0}$:
$$-\mathbb{E}_{\omega\sim\mu_0}\mathbb{E}_{s\sim\pi(\omega)}\left[u(a(\mu_s),\omega)\right] = \mu_0(\omega_2)(\omega_2-\omega_1)^2 \quad \equiv \underline{\mathcal{Z}}$$

2. $\pi \notin \Pi_{\mu_0}$:
$$-\mathbb{E}_{\omega\sim\mu_0}\mathbb{E}_{s\sim\pi(\omega)}\left[u(a(\mu_s),\omega)\right] = (\omega_2-\omega_1)^2\left[\Pr(s_1)\mu_{s_1}(\omega_2) + \Pr(s_2)\mu_{s_2}(\omega_1)\right] \quad \equiv \mathcal{Z}_\pi$$

When $\pi \notin \Pi_{\mu_0}$, the sender has a positive value from persuasion. The belief $\mu_{s_i}(\omega_j)$ captures the mismatch probability: the posterior belief that the state is $\omega_{j\neq i}$ when the signal realized is s_i. The summation term measures the probability of "*wrongful belief attribution*" to state ω_j when the true state is ω_i. We assume that $G(\cdot) = c'^{-1}(\cdot)$. Consequently, the equilibrium verification probability when $\pi \in \Pi_{\mu_0}$ is $q^*(\pi) = G\left(\underline{\mathcal{Z}}\right)$, and when $\pi \notin \Pi_{\mu_0}$ it is $q^*(\pi) = G\left(\mathcal{Z}_\pi\right)$. The verification probability characterizes an important trade-off for the sender: "*information revelation vs. verification*." A more informative signal decreases verification, puts a lower weight on the payoff V_{μ_0}, and increases the weight on the persuasion payoff V_π.

Since the verification depends on the magnitude of the mismatch for the DM, when the sender's persuasion strategy provides more precise information about the states, the DM verifies less. In order to see this more clearly, let us consider the two strategies: full revelation (π_{fr}) and optimal persuasion (π_{op}). For the full revelation strategy, $\mathcal{Z}_{\pi_{fr}} = 0$ and $q^*(\pi_{fr}) = 0$. From an ex ante perspective, the sender's expected payoff from full revelation is therefore, $V_{\pi_{fr}} = \mu_0(\omega_1)\omega_1 + \mu_0(\omega_2)\omega_2 \equiv V_{\mu_0}$.

For the optimal persuasion strategy, $\mathcal{Z}_{\pi_{op}} = \underline{\mathcal{Z}}$ and $q^*(\pi_{op}) = G\left(\underline{\mathcal{Z}}\right)$. The verification probabilities of the DM under these two strategies are computed by substituting the posteriors into the marginal benefit function given in the second part of Lemma 1. Specifically, when the sender chooses full revelation, there are no incentives for the DM to verify. Instead, if the sender chooses the optimal persuasion strategy, the DM's marginal benefit from verification is the same as in the case of no information revelation (pooling). This is captured in Figure 2. The marginal benefit to the DM is the same for all strategies that induce posteriors on the payoff line joining the origin point O and point A. Consequently, the verification probability is the same across all such strategies. The sender therefore has an intuitive trade-off: a more informative signal decreases verification, puts a lower weight on the payoff V_{μ_0}, and increases the weight on the persuasion payoff V_{π^*}. The following proposition establishes this trade-off.

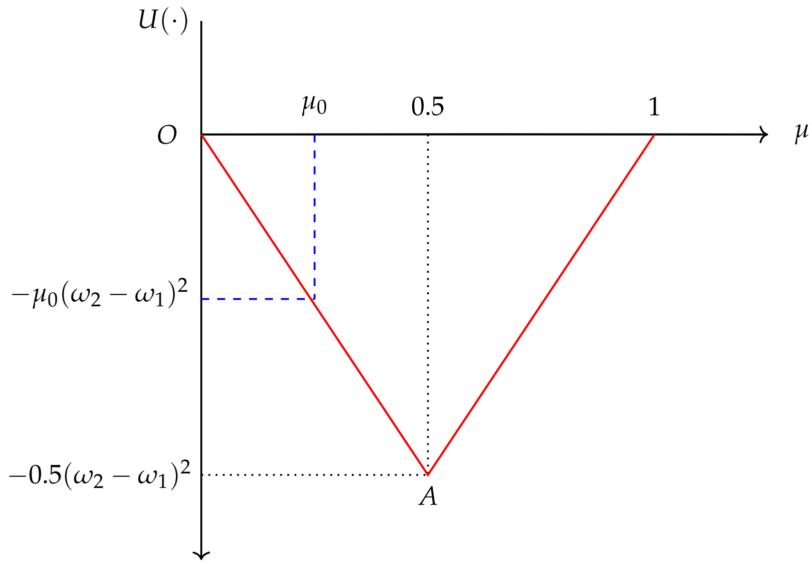

Figure 2: The payoffs to the DM under pooling and optimal persuasion (π_{op}) strategies of the sender are equal.

Proposition 1. *In equilibrium, the sender discloses weakly more information in the presence of verification: $q^* \leq q_{op}$ and $V_{\pi^*} \leq V_{op}$.*

The proof is geometric in nature since there are only two states. The result follows from observing three arguments. First, in any equilibrium with verification, the sender would never choose a strategy that gives a pure persuasion payoff that is less than V_{μ_0}. In Figure 3, any posteriors $\mu \leq \underline{\mu}$ and $\frac{1}{2}$ results in a payoff lower than V_{μ_0}. This implies that for all signaling strategies where $\mu_{s_1}(\omega_2) \in (0, \underline{\mu}]$, the verification probability is increasing in $\mu_{s_1}(\omega_2)$. Further, the persuasion payoff to the sender is decreasing in the interval $(0, \underline{\mu}]$. Consequently, all strategies $\mu_{s_1}(\omega_2) \in (0, \underline{\mu}]$ are dominated by the optimal persuasion strategy π_{op}, which induces the posteriors $\left\{\{1,0\}, \{\frac{1}{2}, \frac{1}{2}\}\right\}$.

Second, consider strategies that induce posteriors $\left\{\mu_{s_1}, \mu_{s_2}\right\}$, where $\mu_{s_1}(\omega_2) \in (0, \underline{\mu})$ and $\mu_{s_2}(\omega_2) \in (\frac{1}{2}, 1)$. This strategy results in greater verification compared to the strategy $\left\{\{1,0\}, \mu_{s_2}\right\}$. This is because the latter is more informative when signal realization is s_1. Consequently, the only feasible set of equilibrium strategies are those that induce the posteriors $\left\{\{1,0\}, \{1-\right.$

$\mu'', \mu''\}\Big\}$, where $\mu'' > \frac{1}{2}$. In Figure 4, these arguments imply that the concavification line joining μ' and point A is dominated by the line from the origin to A. Consequently, the feasible set of equilibrium beliefs is in between the points P and B, and the concavification payoffs are between V_{op} and V_{μ_0}.

Suppose $\mathcal{Z}_{\pi_{op}}$ and $\mathcal{Z}_{\pi^*}$ are the marginal benefits to the sender from the optimal persuasion strategy and the equilibrium strategy with verification, respectively. For the persuasion strategy under verification to be strictly more informative, the marginal benefit from more information revelation must be positive at the optimal persuasion strategy π_{op}.

Proposition 2. *The signaling strategy π^* is more informative than π_{op}, i.e. $\mathcal{Z}_{\pi_{op}} > \mathcal{Z}_{\pi^*}$, if the following holds:*

$$\frac{G'(\mathcal{Z}_{\pi_{op}})}{(1 - G(\mathcal{Z}_{\pi_{op}}))} > \frac{1}{\mathcal{Z}_{\pi_{op}}}$$

The condition in Proposition 2 has an intuitive interpretation: the inverse of the hazard rate of the G function must be strictly greater than the likelihood ratio of the states under the prior. It is a consequence of the "information vs. verification" tradeoff. When there is more precise information revelation, the verification probability decreases. This implies that for the persuasion strategy to be more informative, the marginal benefit from information revelation must be positive at the optimal persuasion strategy π_{op}. To see this, we can rewrite the expected payoff with verification under the optimal strategy as:

$$V(\pi_{op}, q_{op}) = \Pr(s_1)\Big[q_{op}V_{\mu_0} + (1 - q_{op})\omega_1\Big] + \Pr(s_2)\Big[q_{op}V_{\mu_0} + (1 - q_{op})\omega_2\Big]$$

When the prior $\mu_0(\omega_2)$ is sufficiently high, $\Pr(s_2)$ increases to maintain Bayes plausibility. This in turn increases the marginal benefit from verification, $\mathcal{Z}_{\pi_{op}}$. This incentivizes the DM to increase verification, q_{op}, thereby putting a greater weight on the V_{μ_0} payoff. This results in a drop in the expected payoff at $\mu = \frac{1}{2}$, and a corresponding increase at $\mu = 0$ (see points A and B respectively in Figure 5). Further, as the sender provides more information, the expected payoff when s_2 is realized increases from point A, reaching ω_2 when the signal is perfectly informative. In Figure 5, this corresponds to the concave curve from point A. Therefore, the condition in the proposition requires that the verification probability and the marginal change in verification be sufficiently high at the optimal strategy such that the new concavification results in a more informative signal from the sender. In Figure 6, for the same prior, the verification probability and marginal change in this probability at π_{op} are both low. This results in points A and B being far apart which implies that the concavification is unchanged and the equilibrium signal remains π_{op}.

Value of Verification Technology

Are there societal benefits from introducing the verification technology? Clearly, the sender's payoff is lower with verification since the persuasion strategy with verification is (weakly) more informative than the optimal persuasion strategy. This implies that the equilibrium payoff is a convex combination of a (weakly) suboptimal persuasion payoff and the expected utility at the given prior. The DM's expected utility is, however, ambiguous. On the one hand, verification results in a more informative signaling strategy, which is beneficial to the DM. On the other hand, verification is costly and imposes sunk costs on the DM. Therefore, the net benefits of introducing verification are not clear at first sight. Since the introduction of verification improves information revelation under some conditions, it is possible for the DM to benefit from this option.

Proposition 3. *The value of verification is positive if*

$$c\Big(G\big(\mathcal{Z}_{\pi^*}\big)\Big) < \left(\frac{\mu^*}{1-\mu^*} - \Big(1 - G\big(\mathcal{Z}_{\pi^*}\big)\Big)\right)\mathcal{Z}_{\pi^*}$$

The DM benefits as long as the cost of verification in equilibrium is strictly lower than the benefits from verification. The cost of verification is positive when the signaling is not fully informative, i.e., $\mu^* < 1$. The total cost is therefore a function of both μ^* and q^*. The benefit from using verification is the product of the complement of the equilibrium verification and the marginal benefits to the DM at the equilibrium π^*.

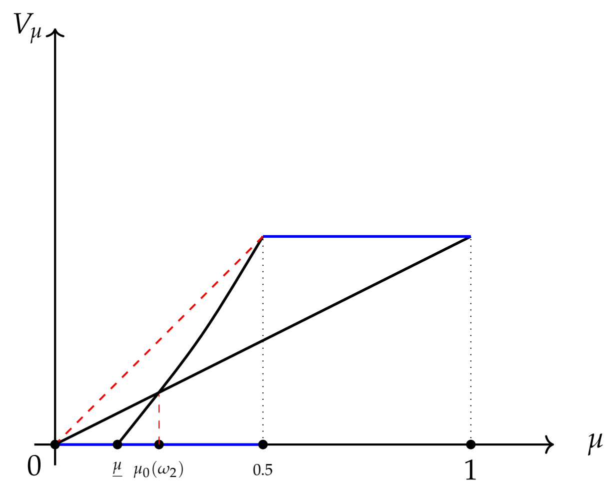

Figure 3: Any signal that induces posteriors $\{\mu, \frac{1}{2}\}$, where $\mu \in [0, \underline{\mu}]$, results in a lower payoff to the sender than the optimal persuasion strategy.

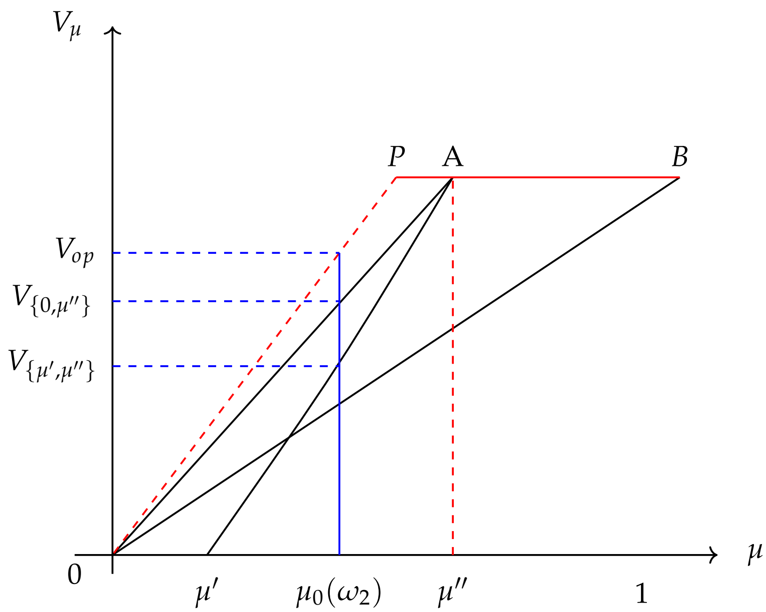

Figure 4: Any posterior on the interior of the line PB is strictly more informative than the optimal strategy.

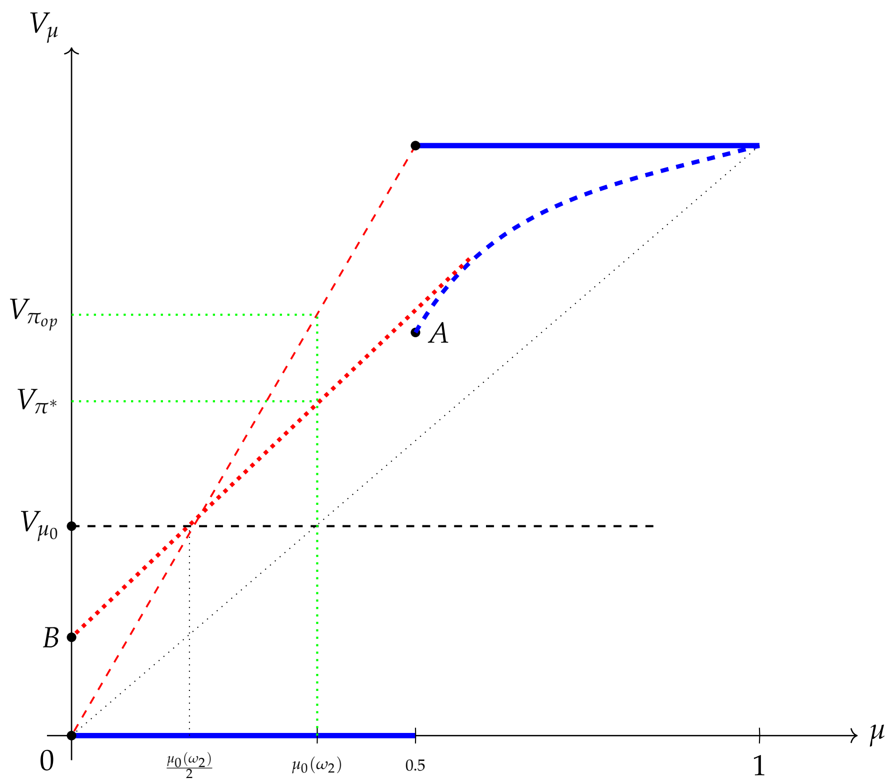

Figure 5: In this case, q_π is such that there is more informative signaling, and the sender's equilibrium payoff is lower than in the canonical persuasion game, $V_{\pi_{op}}$.

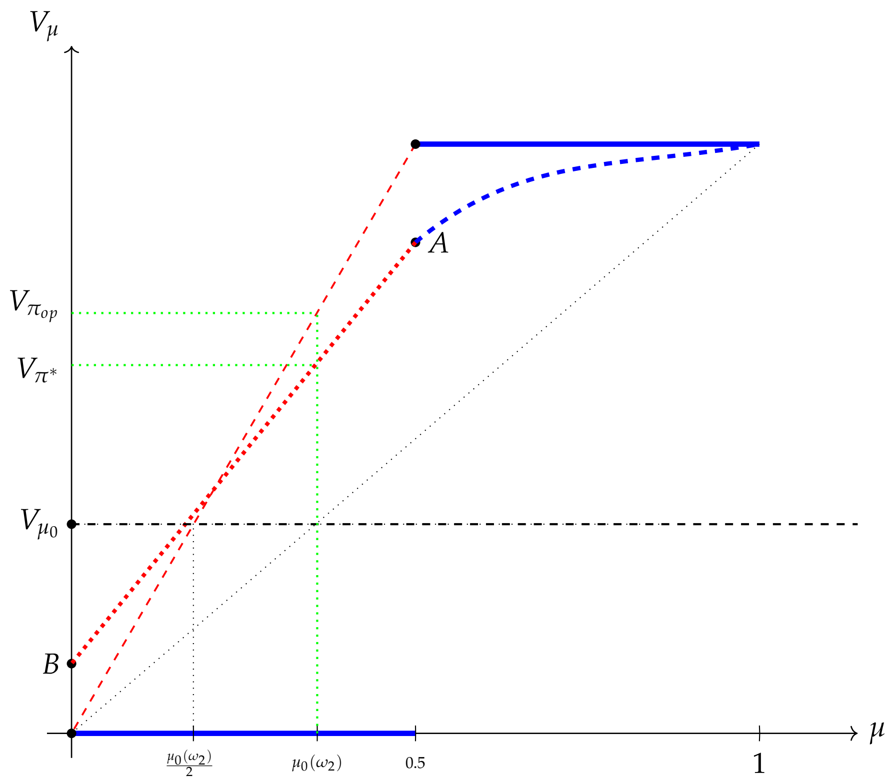

Figure 6: In this case, q_π is such that there is no change in the equilibrium signaling strategy of the sender.

4 Conclusion

In canonical Bayesian persuasion, only the sender benefits from fully committing to a signaling strategy. We look at a model in which the decision-maker, after observing the signal designed by the sender, makes a counter-commitment to verify the state at a cost. We show that interim commitment results in weakly more informative signaling by the sender, decreasing her rents from persuasion. We then find the conditions under which *(i)* there are strictly more informative signals in equilibrium, and *(ii)* the decision-maker benefits from using the verification strategy. The model with interim commitment opens the literature for studying settings where there are competing interests between a sender and a decision-maker. For instance, after observing the prosecutor's (persuader) strategy of evidence provision, the judge (decision-maker) can commit to choosing an independent committee to verify the evidence submitted. Similarly, insurance companies can hire verification specialists to check the claims made by firms. Our analysis shows that since the process of independent verification is costly, there is never full revelation of information by the sender. Instead, in the presence of commitment by the decision-maker, there is at least weakly better signaling and possible gains to the decision-maker.

A Proofs

A.1 Proof of Lemma 1

The ex ante welfare for the DM from a persuasion strategy π of the sender is:

$$
\begin{aligned}
& -\mathbb{E}_{\omega\sim\mu_0}\mathbb{E}_{s\sim\pi(\omega)}\Big[u(a(\mu_s),\omega)\Big] \\
&= \mu_0(\omega_1)\Bigg[\pi(s_1 \mid \omega_1)\mathbb{E}_{\omega\sim\mu_{s_1}}\Big[u(a(\mu_{s_1}),\omega)\Big] + \pi(s_2 \mid \omega_1)\mathbb{E}_{\omega\sim\mu_{s_2}}\Big[u(a(\mu_{s_2}),\omega)\Big]\Bigg] \\
&\quad + \mu_0(\omega_2)\Bigg[\pi(s_1 \mid \omega_2)\mathbb{E}_{\omega\sim\mu_{s_1}}\Big[u(a(\mu_{s_1}),\omega)\Big] + \pi(s_2 \mid \omega_2)\mathbb{E}_{\omega\sim\mu_{s_2}}\Big[u(a(\mu_{s_2}),\omega)\Big]\Bigg] \\
&= \mu_0(\omega_1)\Bigg[\pi(s_1 \mid \omega_1)\Big[\mu_{s_1}(\omega_2)(\omega_2-\omega_1)^2\Big] + \pi(s_2 \mid \omega_1)\Big[\mu_{s_2}(\omega_1)(\omega_2-\omega_1)^2\Big]\Bigg] \\
&\quad + \mu_0(\omega_2)\Bigg[\pi(s_1 \mid \omega_2)\Big[\mu_{s_1}(\omega_2)(\omega_2-\omega_1)^2\Big] + \pi(s_2 \mid \omega_2)\Big[\mu_{s_2}(\omega_1)(\omega_2-\omega_1)^2\Big]\Bigg]
\end{aligned}
$$

Since $\Pr(s_i) = \pi(s_i \mid \omega_1)\mu_0(\omega_1) + \pi(s_i \mid \omega_2)\mu_0(\omega_2)$, the above equation reduces to:

$$
-\mathbb{E}_{\omega\sim\mu_0}\mathbb{E}_{s\sim\pi(\omega)}\Big[u(a(\mu_s),\omega)\Big] \quad = \quad (\omega_2-\omega_1)^2\Bigg[\Pr(s_1)\mu_{s_1}(\omega_2) + \Pr(s_2)\mu_{s_2}(\omega_1)\Bigg]
$$

This follows from noting the fact that when the true state ω is realized according to μ_0, the second expectations term captures the payoff to the DM given the possible signal realizations that are induced by π. Suppose, for example, the signal realized is s, and the posterior is $\mu_s(\omega)$ along with a corresponding action $a(\mu_s) = \omega$. Since each s induces a different action by the DM, it results in a mismatch with probability $\mu_s(\omega'_{\neq\omega})$. The expected utility expression captures this mismatch over all states and every possible signal realization given an underlying state.

A.2 Proof of Proposition 1

Consider the equilibrium strategy and the associated payoff for the sender in the case without verification, i.e., π_{op} and V_{op}, respectively. The lower bound $V_{\underline{\pi}}$ gives the expected payoff for this strategy in the presence of verification. The sender can guarantee at least this (lower-bound) payoff. Therefore, we only look at those strategies $\hat{\pi}$ that provide a pure persuasion payoff of at least $V_{\hat{\pi}} \geq V_{\underline{\pi}}$ to the sender (see Figure 7). Suppose the prior $\mu_0(\omega_1) = p$.

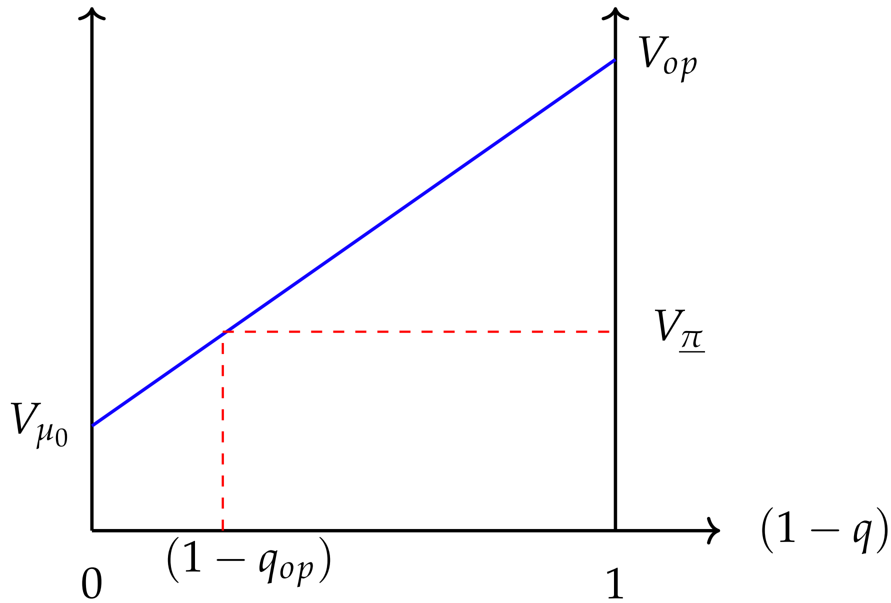

Figure 7: The value function $V(\pi_{op}, q_{op})$ gives the same expected payoff as the pure persuasion strategy $\underline{\pi}$.

Let a generic signaling strategy be represented in terms of two parameters (λ, β), such that,

$$\pi(s_1 \mid \omega_1) = \lambda \qquad \pi(s_2 \mid \omega_1) = 1 - \lambda$$

$$\pi(s_1 \mid \omega_2) = 1 - \beta \qquad \pi(s_2 \mid \omega_2) = \beta$$

We focus on all signaling strategies that give the sender *at least* the lower-bound payoff $V_{\underline{\pi}}$. There is a distribution over posteriors $\left\{\{\underline{\mu}(\omega_1), \underline{\mu}(\omega_2)\}, \{\frac{1}{2}, \frac{1}{2}\}\right\}$ that gives the sender the same payoff as full revelation (see Figure 8). Going forward, for the sake of exposition, we denote $\underline{\mu} = \underline{\mu}(\omega_2)$, and any pair of posterior distributions just by the posterior belief that the state is ω_2.[5] That is, $\left\{\tilde{\mu}', \tilde{\mu}''\right\}$ is equivalent to the distribution over posteriors $\left\{\{1-\tilde{\mu}', \tilde{\mu}'\}, \{1-\tilde{\mu}'', \tilde{\mu}''\}\right\}$.
Given the signaling strategy, the posteriors corresponding with the signal realizations s_1 and s_2 are given by,

$$\mu_{s_1}(\omega_2) = \frac{(1-p)(1-\beta)}{(1-p)(1-\beta) + p\lambda} \qquad \mu_{s_2}(\omega_2) = \frac{(1-p)\beta}{(1-p)\beta + p(1-\lambda)}$$

Where, $\Pr(s_1) = (1-p)(1-\beta) + p\lambda$ and $\Pr(s_2) = (1-p)\beta + p(1-\lambda)$. For any two posteriors $\left\{\mu', \mu''\right\}$ induced by signals (s_1, s_2) respectively, such that $\mu' < 1 - p < \mu''$, the solution to

[5] Further, all beliefs μ in the figures in this proof represent $\mu = \mu(\omega_2)$.

parameters λ and β is given by:

$$\lambda = \frac{\Big(\mu'' - (1-p)\Big)(1-\mu')}{p(\mu''-\mu')} \qquad \beta = \frac{\Big((1-\mu') - p\Big)\mu''}{(1-p)(\mu''-\mu')}$$

$$1-\lambda = \frac{\Big((1-\mu') - p\Big)(1-\mu'')}{p(\mu''-\mu')} \qquad 1-\beta = \frac{\Big(\mu'' - (1-p)\Big)\mu'}{(1-p)(\mu''-\mu')}$$

We focus on all distributions over posteriors of the following form:

$$\textit{(1)} \quad \Big\{\mu', \frac{1}{2}\Big\} \quad \text{for all } \mu' \in [0, \underline{\mu}]$$

$$\textit{(2)} \quad \Big\{\mu', \mu''\Big\} \quad \text{for all } \mu' \in (0, \underline{\mu}) \text{ and } \mu'' \in \Big(\frac{1}{2}, 1\Big]$$

$$\textit{(3)} \quad \Big\{0, \mu''\Big\} \quad \text{for all } \mu'' \in \Big(\frac{1}{2}, 1\Big]$$

The marginal benefit function from Lemma 1 for any generic signaling strategy and induced posteriors $\Big\{\mu', \mu''\Big\}$ is,

$$\mathcal{Z}_\pi(\mu', \mu'') = (\omega_2 - \omega_1)^2 \Bigg[(1-p)(1-\beta) + p(1-\lambda)\Bigg]$$

$$\mathcal{Z}_\pi(\mu', \mu'') = (\omega_2 - \omega_1)^2 \left[\frac{p\mu' - 2\mu'(1-\mu'') + (1-p)(1-\mu'')}{(\mu''-\mu')}\right] \tag{6}$$

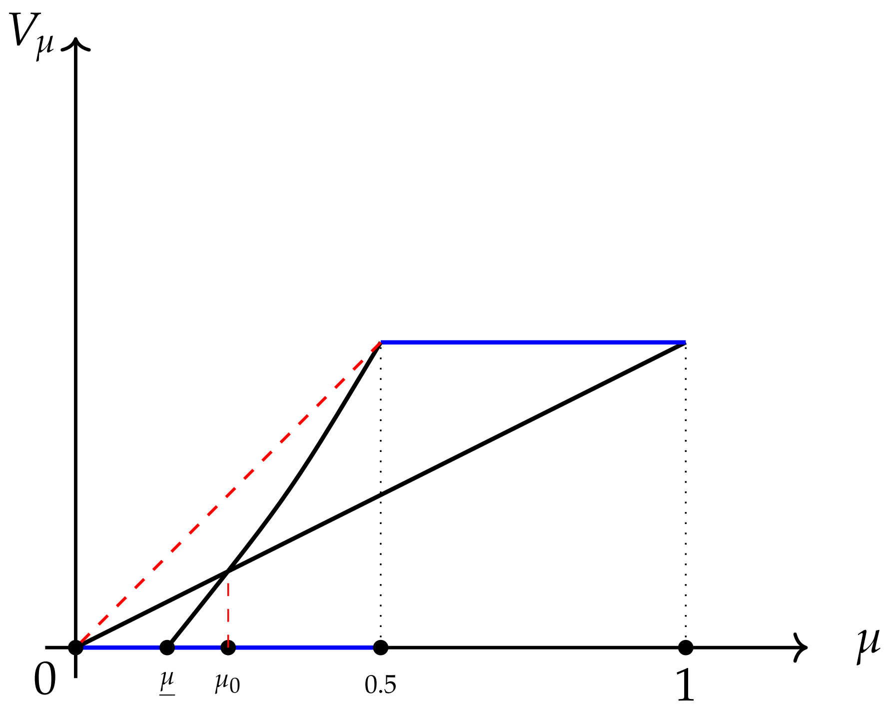

Figure 8: Sender's payoffs over posterior distributions that yield at least the same payoff as the full revelation strategy.

- **Case 1:** Any μ' to the left of $\underline{\mu}$ induces no change in the DM's marginal benefit function (see Figure 2 and Lemma 1). To see this, we compute the marginal benefits under $\left\{\mu', \frac{1}{2}\right\}$ from Equation 6:

$$\mathcal{Z}_\pi\left(\mu', \frac{1}{2}\right) = (1-p)(\omega_2 - \omega_1)^2$$

This is independent of μ' which implies that the verification probability of the DM does not change in this interval. In other words, the DM is *indifferent* among all signaling strategies of the form $\left\{\mu', \frac{1}{2}\right\}$. For the sender, the one that provides the highest ex-ante welfare is:

$$V_{\{\mu', \frac{1}{2}\}} = \Pr(s_1)\omega_1 + \Pr(s_2)\omega_2 \qquad \text{where, } \Pr(s_2) = (1-p)\beta + p(1-\lambda)$$

Since both β and $(1-\lambda)$ are decreasing in μ', $\Pr(s_2)$ is also decreasing in μ'. This implies that the optimal persuasion strategy, $\left\{0, \frac{1}{2}\right\}$ (see Figure 9), gives the sender a strictly higher payoff than any strategy where $\mu' \in (0, \underline{\mu})$. Therefore, $V_{\underline{\pi}} > q'_{\{\mu', \frac{1}{2}\}} V_{\mu_0} + \left(1 - q'_{\{\mu', \frac{1}{2}\}}\right) V_{\{\mu', \frac{1}{2}\}}$ for all $\mu' \in (0, \underline{\mu}]$.

- **Case 2:** Consider signaling strategies that induce posteriors $\left\{\mu', \mu''\right\}$ such that $\mu' \in (0, \underline{\mu})$ and $\mu'' \in (\frac{1}{2}, 1]$ (see Figure 10). From the DM's perspective, the verification probability decreases as μ' decreases from $\underline{\mu}$ to 0. To see this, we compute the derivative of the marginal benefit function with respect to μ', keeping μ'' as fixed.

$$\frac{d\mathcal{Z}_\pi\left(\mu', \mu''\right)}{d\mu'} = (\omega_2 - \omega_1)^2 \left[\frac{(2\mu'' - 1)(p - (1 - \mu''))}{(\mu'' - \mu')^2}\right] \quad > \quad 0 \qquad \text{since } p, \mu'' > \frac{1}{2}$$

That is, the marginal benefit and therefore the verification probability is lower under the strategy $\{0, \mu''\}$ compared to information structures where $\mu' > 0$ (see Figure 11). Let $\hat{V}_{\{\mu',\mu''\}}$ and $\hat{V}_{\{0,\mu''\}}$ be the value functions under the respective information structures. Let the associated verification probabilities be $\hat{q}_{\{\mu',\mu''\}}$ and $\hat{q}_{\{0,\mu''\}}$. As argued in Case 1, both β and $(1-\lambda)$ are decreasing in μ', fixing μ''. Consequently, $\Pr(s_2)$ is also decreasing in μ' and therefore, $\hat{V}_{\{0,\mu''\}} > \hat{V}_{\{\mu',\mu''\}}$. Since $\hat{q}_{\{0,\mu''\}} < \hat{q}_{\{\mu',\mu''\}}$, the sender's welfare is higher under the signal $\left\{0, \mu''\right\}$ (see Figure 12).

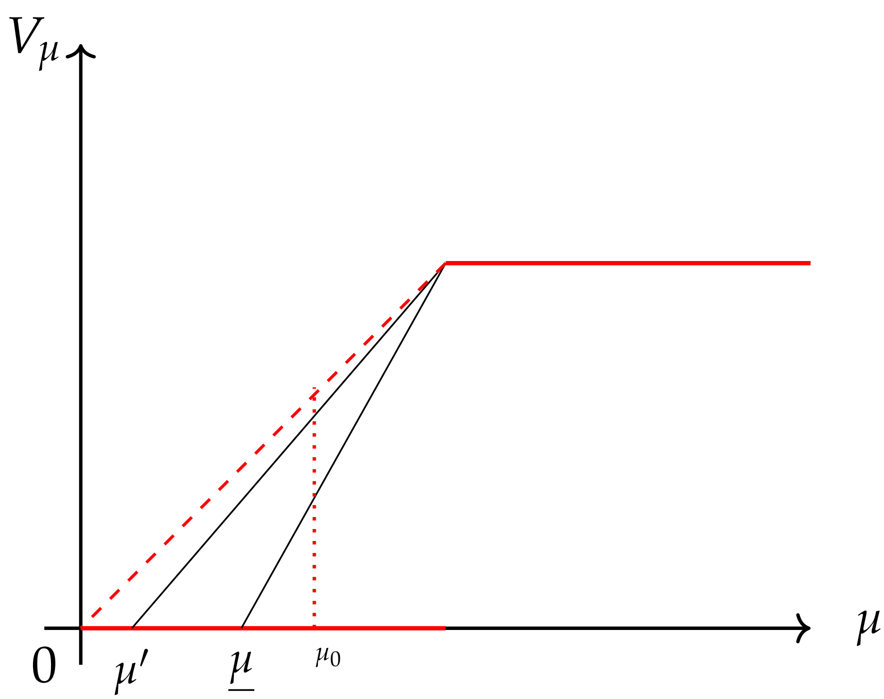

Figure 9: Sender's persuasion payoff under $\{\mu', \frac{1}{2}\}$ is always lower than the optimal persuasion strategy for all $\mu' \in (0, \underline{\mu}]$

- **Case 3:** Consider all information structures $\left\{0, \mu''\right\}$, such that $\mu'' \in [\frac{1}{2}, 1]$. The marginal benefit function is given by:

$$\mathcal{Z}_\pi\left(0,\mu''\right) = (\omega_2-\omega_1)^2\left[\frac{(1-p)(1-\mu'')}{\mu''}\right]$$

Clearly, $\frac{d\mathcal{Z}_\pi(0,\mu'')}{d\mu''} < 0$. This implies that from the perspective of the DM, the verification probability $q'' \equiv q_{\{0,\mu''\}}$ is decreasing as μ'' increases (see Figure 13). Further, starting at $\left\{0, \frac{1}{2}\right\}$, the verification probability and the sender's payoff from persuasion are both decreasing as μ'' increases from $\frac{1}{2}$ to 1. Beginning from $\left\{\underline{\mu}, \frac{1}{2}\right\}$, where the expected payoff to the sender is V_{μ_0} , the welfare of the sender increases as μ' decreases from $\underline{\mu}$ to 0. Further, under full revelation, $\lambda = \beta = 1$, and the expected payoff of the sender is $p\omega_1 + (1-p)\omega_2 \equiv V_{\mu_0}$. That is, the welfare to the sender increases as $\mu' \to 0$ and then decreases as μ'' goes from $\frac{1}{2}$ to 1, resulting in a payoff of V_{μ_0} under full revelation. Therefore, at some intermediate information structure between $\left\{0, \frac{1}{2}\right\}$ and $\left\{0, 1\right\}$, given by $\pi^* = \left\{0, \mu^*\right\}$, the sender's expected payoff attains a maximum. Since $\mu^* \in \left[\frac{1}{2}, 1\right)$, the persuasion strategy is (weakly) more informative.

The equilibrium strategy for the sender is in between the *optimal* Bayesian Persuasion strategy, $\pi_{op} = \{0, 0.5\}$, and the full revelation signal, $\pi_{fr} = \{0, 1\}$ (see Figure 14). As the sender continues to provide information such that $\pi^* = \{0, \mu^*\}$ where $\mu^* \in (0.5, 1]$, verification continues to decrease. This implies,

$$q_{op}V_{\mu_0} + (1-q_{op})V_{op} \gtrless q^*V_{\mu_0} + (1-q^*)V_{\pi^*}$$

In the range of signaling strategies where $\mu^* \in (0.5, 1]$, we know that the sender's expected payoff is V_{μ_0} as $\mu^* \to 1$. By continuity, there must exist a $\pi^* \equiv \{0, \hat{\mu}^*\}$ such that the expected payoff of the sender reaches a maximum. Given the concavity assumptions, this strategy π^* is unique.

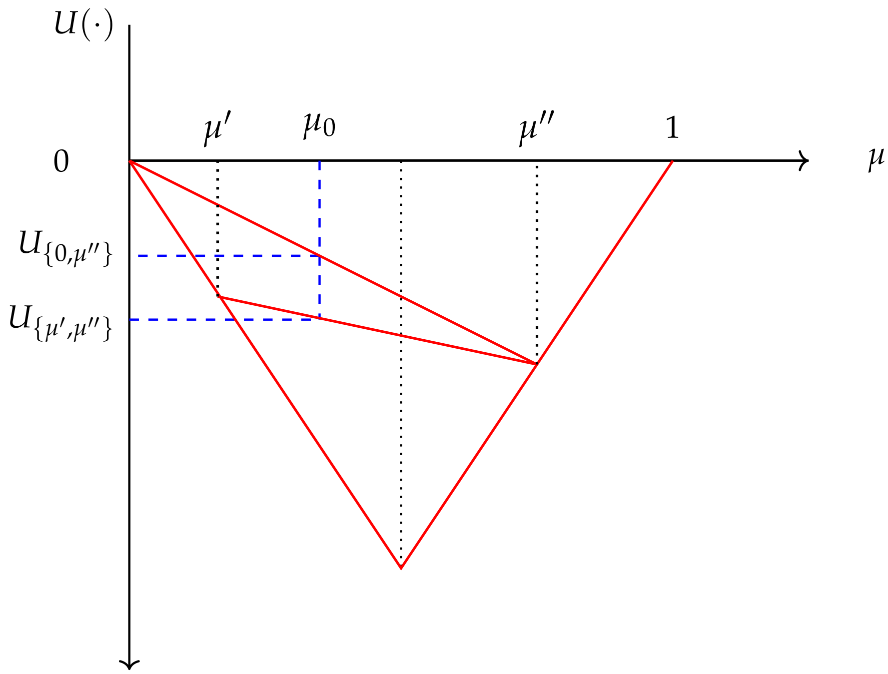

Figure 10: DM's payoff for $\left\{\mu', \mu''\right\}$ for all $\mu' \in (0, \underline{\mu})$ and $\mu'' \in (\frac{1}{2}, 1]$

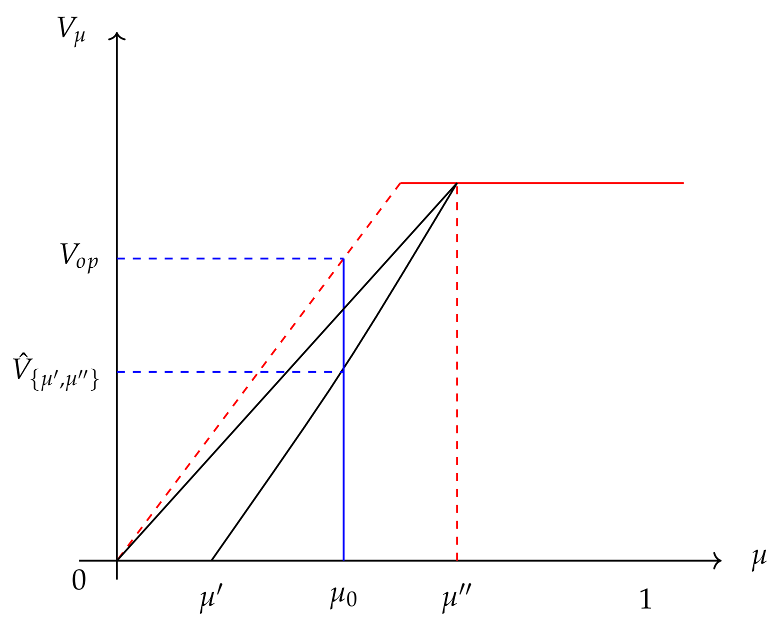

Figure 11: Verification is lower under $\left\{0, \mu''\right\}$ compared to the signal $\left\{\mu', \mu''\right\}$

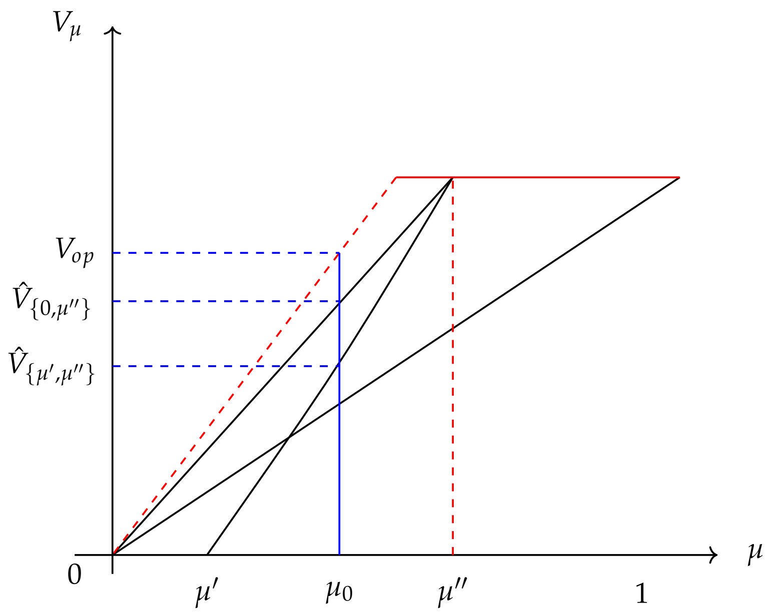

Figure 12: Since $\hat{V}_{\{0,\mu''\}} > \hat{V}_{\{\mu',\mu''\}}$ and $\hat{q}_{\{0,\mu''\}} < \hat{q}_{\{\mu',\mu''\}}$, it implies $\left\{0,\mu''\right\} \succ \left\{\mu',\mu''\right\}$ for the sender

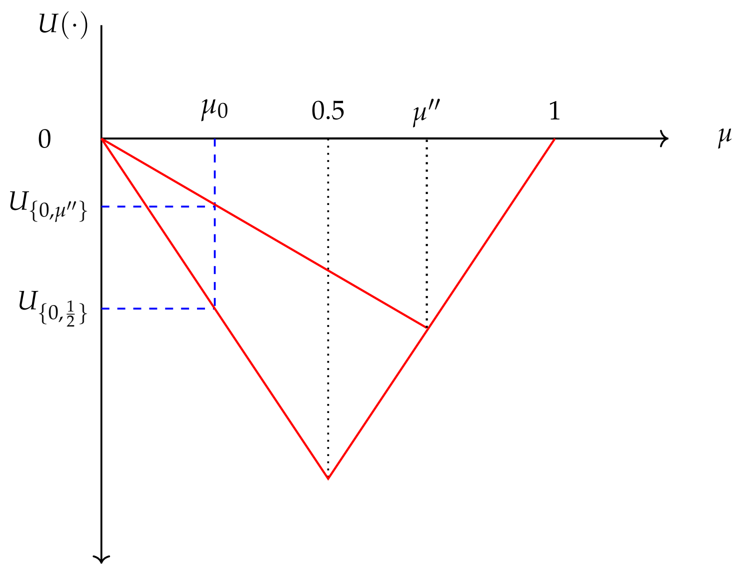

Figure 13: DM's payoff for $\left\{0,\mu''\right\}$ for all $\mu'' \in [\frac{1}{2},1]$

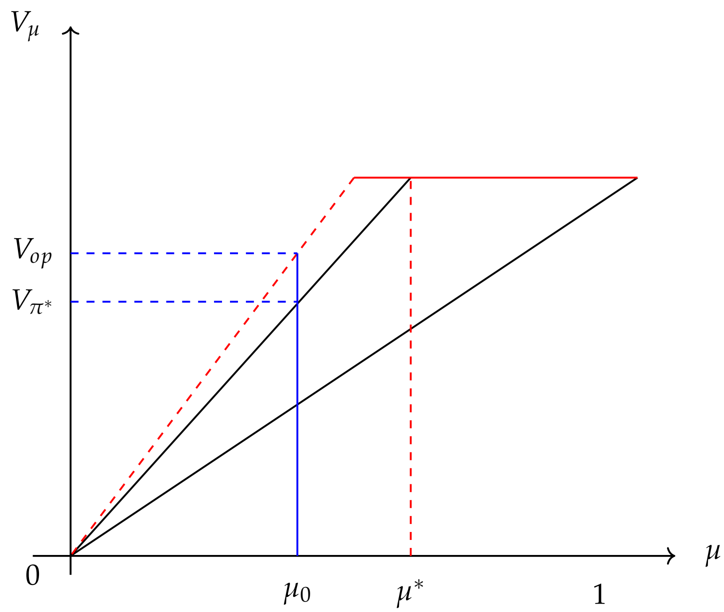

Figure 14: In the equilibrium information structure $\pi^* \equiv \{0, \mu^*\}$, the sender's expected payoff is such that $q^* V_{\mu_0} + (1 - q^*) V_{\pi^*} > V_{\underline{\pi}}$

A.3 Proof of Proposition 2

Consider the following signaling strategy:

$$\pi(s_1 \mid \omega_1) = \lambda \qquad \pi(s_2 \mid \omega_1) = 1 - \lambda$$

$$\pi(s_1 \mid \omega_2) = 0 \qquad \pi(s_2 \mid \omega_2) = 1$$

We know that under the optimal persuasion strategy of the sender ($\lambda = \lambda_{op}$), the payoff to the sender when the DM verifies the state with probability q_{op} is:

$$\underline{V} = q_{op} V_{\mu_0} + (1 - q_{op}) V_{op}$$

Since $V_{\mu_0} < \underline{V} < V_{op}$ whenever $0 < q_{op} < 1$, it follows that there exists a signaling strategy $\tilde{\pi}$ and an associated probability $\tilde{\lambda}$, such that $V_{\tilde{\pi}} = \underline{V}$. Further, the payoff of the sender decreases from V_{op} to $\underline{V}$ as the messaging strategy becomes more revealing, i.e., λ increases. Therefore, it follows that $\tilde{\lambda} > \lambda_{op}$ (Figure 15). To characterize the equilibrium signaling strategy under verification, we require that there exists a $\lambda^* \in (\lambda_{op}, \tilde{\lambda})$ such that,

- the posterior belief distribution under λ^*, given by μ^*, must satisfy $V_{\mu^*} > \underline{V}$;

- the verification probability at λ^*, given by q^*, must satisfy $q^* \leq \tilde{q}$, where,

$$\tilde{q} = q_{op} \cdot \frac{V_{\mu^*} - \underline{V}}{V_{op} - \underline{V}}$$

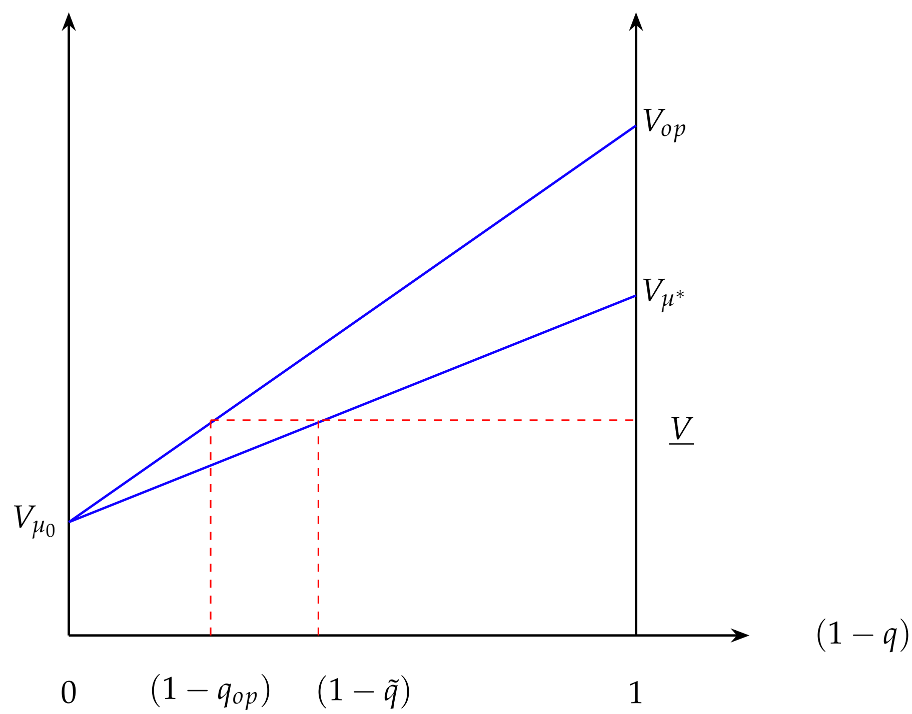

Figure 15: Convex Combination of Sender's *ex-ante* Payoff

For a generic signaling strategy, the verification probability q is determined by the equation:

$$c'(q) = p(1-\lambda)(\omega_2 - \omega_1)^2 \tag{7}$$

Let q_λ be the solution to the above equation. Further, let any generic information structure be represented in terms of the variable λ. That is, we denote the signaling strategy as π_λ and the posterior belief distribution as μ_λ. The expected payoff for the sender under π_λ is:

$$V_{\pi_\lambda} = q_\lambda V_{\mu_0} + (1-q_\lambda)V_{bp}(\pi_\lambda)$$

Given that the payoffs of the sender are linear, we write down the expression for $V_{bp}(\pi_\lambda)$ as:

$$V_{bp}(\pi_\lambda) = \Pr(s_1)\cdot\omega_1 + \Pr(s_2)\cdot\omega_2 = [p\lambda]\omega_1 + [p(1-\lambda) + (1-p)]\omega_2$$

$$V'_{bp}(\pi_\lambda) = \frac{dV_{bp}(\pi_\lambda)}{d\lambda} = -p(\omega_2 - \omega_1)$$

Since $\omega_2 >> \omega_1$, we know that $\frac{dV_{bp}(\pi_\lambda)}{d\lambda} < 0$. In order to show that $\lambda^* > \lambda_{op}$, we have to take the *foc* of V_{π_λ} with respect to λ and evaluate this condition at $\lambda = \lambda_{op}$ to show that $V'_{\pi_\lambda}\Big|_{\lambda=\lambda_{op}} > 0$. Taking the *foc* with respect to λ, we get:

$$V'_{\pi_\lambda} = q'_\lambda V_{\mu_0} + (1 - q_\lambda)V'_{bp}(\pi_\lambda) - q'_\lambda V_{bp}(\pi_\lambda)$$
$$V'_{\pi_\lambda} = q'_\lambda \Big[V_{\mu_0} - V_{bp}(\pi_\lambda)\Big] + (1 - q_\lambda)V'_{bp}(\pi_\lambda)$$

The expected payoff for the sender under the prior is $V_{\mu_0} = p\omega_1 + (1-p)\omega_2$. Using this, we get the following:

$$V_{\mu_0} - V_{bp}(\pi_\lambda) = p(1-\lambda)\omega_1 - p(1-\lambda)\omega_2 = -p(1-\lambda)(\omega_2 - \omega_1)$$

$$V'_{\pi_\lambda} = -pq'_\lambda(1-\lambda)(\omega_2 - \omega_1) + (1 - q_\lambda)V'_{bp}(\pi_\lambda)$$

Substituting for $V'_{bp}(\pi_\lambda)$, we get:

$$V'_{\pi_\lambda} = -pq'_\lambda(1-\lambda)(\omega_2 - \omega_1) - p(1 - q_\lambda)(\omega_2 - \omega_1)$$

Simplifying the terms inside the big square brackets:

$$V'_{\pi_\lambda} = -p(\omega_2 - \omega_1)\Big[q'_\lambda(1-\lambda) + (1 - q_\lambda)\Big] \tag{8}$$

Further, we know that:

$$V''_{\pi_\lambda} = -p(\omega_2 - \omega_1) \cdot \Big[q''_\lambda(1-\lambda) - 2q'_\lambda\Big] \quad < \quad 0$$

We have already established earlier that $q'_\lambda < 0$. Therefore, for the right-hand side of Equation 8 to be positive when evaluated at λ_{op}, the following must hold:

$$V'_{\pi_\lambda}\big|_{\lambda=\lambda_{op}} > 0 \quad \textbf{if} \quad -q'_{op} > \frac{(1 - q_{op})}{(1 - \lambda_{op})} \tag{9}$$

Where q_{op} is given by Equation 7:

$$q_{op} = c'^{-1}\left[p(1-\lambda_{op})(\omega_2 - \omega_1)^2\right]$$

We know that at the optimal persuasion strategy λ_{op}:

$$\mu_{s_2}(\omega_1) = \frac{p(1-\lambda_{op})}{p\cdot(1-\lambda_{op}) + (1-p)} = \frac{1}{2} \quad \Longrightarrow \quad \lambda_{op} = \frac{2p-1}{p}$$

$$q_{op} = c'^{-1}\left[(1-p)(\omega_2-\omega_1)^2\right]$$

where $\mathcal{Z}_{\pi_{op}} = (1-p)(\omega_2-\omega_1)^2$ measures the ex ante expected marginal benefits to the DM from the optimal persuasion strategy. Since c' is increasing and convex, the inverse of this function is increasing and concave, i.e., $G' > 0, G'' < 0$. Simplifying the equation for q'_{op} gives:

$$-q'_{op} > \frac{p\left(1 - G\left(\mathcal{Z}_{\pi_{op}}\right)\right)}{1-p} \tag{10}$$

We know that for a generic λ, the verification probability is given by $q = G\left(p(1-\lambda)(\omega_2-\omega_1)^2\right)$.

$$\Longrightarrow \qquad q'_{\lambda} = -p(\omega_2-\omega_1)^2 G'\left(p(1-\lambda)(\omega_2-\omega_1)^2\right) \tag{11}$$

Evaluating the above at $\lambda = \lambda_{op}$:

$$q'_{op} = -p(\omega_2-\omega_1)^2 G'\left(\mathcal{Z}_{\pi_{op}}\right)$$

Substituting this into Equation 10 and simplifying yields:

$$\frac{G'(\mathcal{Z}_{\pi_{op}})}{(1-G(\mathcal{Z}_{\pi_{op}}))} > \frac{1}{\mathcal{Z}_{\pi_{op}}} \tag{12}$$

The left-hand side measures the hazard rate of the c'^{-1} function. This completes the proof. **QED**

A.4 Proof of Proposition 3

When Equation 12 holds, the DM benefits informationally from the use of verification. However, since it is also costly for the DM to verify, the overall benefits of using verification depend on the net benefits and costs. We know that under μ^* and λ^*:

$$\mu^*_{s_2}(\omega_1) \equiv 1-\mu^* = \frac{p(1-\lambda^*)}{p\cdot(1-\lambda^*)+(1-p)} \quad \Longrightarrow \quad \lambda^* = 1 - \frac{1-p}{p}\cdot\frac{1-\mu^*}{\mu^*}$$

$$\implies \quad \lambda^* = \frac{p-(1-\mu^*)}{p\mu^*} \qquad 1-\lambda^* = \frac{(1-p)(1-\mu^*)}{p\mu^*}$$

where $\mu^* > \frac{1}{2}$. We compare the ex ante expected utilities for the DM under two information structures: $\pi_{op} = \{0, \frac{1}{2}\}$ and $\pi^* = \{0, \mu^*\}$.[6] In addition, there is a verification cost associated with the persuasion strategy π^*, given by $c(q^*)$, where:

$$q^* = {c'}^{-1}\left[\frac{(1-p)(1-\mu^*)}{\mu^*}(\omega_2-\omega_1)^2\right] = G\Big(\mathcal{Z}_{\pi^*}\Big)$$

As before, $\mathcal{Z}_{\pi^*} = \frac{(1-p)(1-\mu^*)}{\mu^*}(\omega_2-\omega_1)^2$ is the marginal benefit to the DM from the equilibrium signaling strategy π^*. Further, we know that the payoff to the DM when signal s_1 is realized is just 0 since it perfectly informs the underlying state. Therefore, the only time the DM faces an expectation is when the signal s_2 is realized. We know, given λ_{op} and λ^*, that:

$$\Pr(s_2|\lambda_{op}) = 2(1-p) \qquad \Pr(s_2|\lambda^*) = \frac{1-p}{\mu^*}$$

The expected payoffs in the presence of verification is, therefore:

$$\mathbb{E}U(\pi^*, q^*) = -(1-q^*)\Big[\frac{(1-p)(1-\mu^*)}{\mu^*}(\omega_2-\omega_1)^2\Big] - c(q^*)$$

Since $\mu_{s_2}(\omega_1) = \frac{1}{2}$, the expected payoffs to the DM in the absence of verification is $\mathbb{E}U(\pi_{op}, q_{op}) = -(1-p)(\omega_2-\omega_1)^2$. Given these two expressions, verification benefits the DM if:

$$(1-q^*)\Big[\frac{(1-p)(1-\mu^*)}{\mu^*}(\omega_2-\omega_1)^2\Big] + c(q^*) < (1-p)(\omega_2-\omega_1)^2$$

$$\implies \quad c(q^*) < (1-p)(\omega_2-\omega_1)^2\left[1 - \frac{(1-q^*)(1-\mu^*)}{\mu^*}\right] = \Big[\frac{(1-p)(1-\mu^*)}{\mu^*}(\omega_2-\omega_1)^2$$

$$\implies \quad c\Big(G\big(\mathcal{Z}_{\pi^*}\big)\Big) < \left(\frac{\mu^*}{1-\mu^*} - \Big(1 - G\big(\mathcal{Z}_{\pi^*}\big)\Big)\right)\mathcal{Z}_{\pi^*} \tag{13}$$

This equation characterizes the condition for the value of verification to be positive. **QED**

[6]The numbers in the brackets denote the posterior probability that the state is ω_2 under s_1 and s_2, respectively.

References

ALONSO, R. AND O. CÂMARA (2016): "Bayesian persuasion with heterogeneous priors," *Journal of Economic Theory*, 165, 672–706.

BALBUZANOV, I. (2019): "Lies and consequences: The effect of lie detection on communication outcomes," *International Journal of Game Theory*, 48, 1203–1240.

BEN-PORATH, E., E. DEKEL, AND B. L. LIPMAN (2014): "Optimal allocation with costly verification," *American Economic Review*, 104, 3779–3813.

BERGEMANN, D. AND S. MORRIS (2019): "Information design: A unified perspective," *Journal of Economic Literature*, 57, 44–95.

CARROLL, G. AND G. EGOROV (2019): "Strategic communication with minimal verification," *Econometrica*, 87, 1867–1892.

CHE, Y.-K. AND N. KARTIK (2009): "Opinions as incentives," *Journal of Political Economy*, 117, 815–860.

DOVAL, L. AND V. SKRETA (2024): "Constrained information design," *Mathematics of Operations Research*, 49, 78–106.

DUR, R. AND O. H. SWANK (2005): "Producing and manipulating information," *The Economic Journal*, 115, 185–199.

DZIUDA, W. AND C. SALAS (2018): "Communication with detectable deceit," *Available at SSRN 3234695*.

EDERER, F. AND W. MIN (2022): "Bayesian Persuasion with Lie Detection," Tech. rep., National Bureau of Economic Research.

GALPERTI, S. (2019): "Persuasion: The art of changing worldviews," *American Economic Review*, 109, 996–1031.

GENTZKOW, M. AND E. KAMENICA (2014): "Costly persuasion," *American Economic Review*, 104, 457–462.

KAMENICA, E. (2019): "Bayesian persuasion and information design," *Annual Review of Economics*, 11, 249–272.

KAMENICA, E. AND M. GENTZKOW (2011): "Bayesian persuasion," *American Economic Review*, 101, 2590–2615.

Kartik, N. and O. Tercieux (2012): "Implementation with evidence," *Theoretical Economics*, 7, 323–355.

Koessler, F. and V. Skreta (2023): "Informed information design," *Journal of Political Economy*, 131, 3186–3232.

Le Treust, M. and T. Tomala (2019): "Persuasion with limited communication capacity," *Journal of Economic Theory*, 184, 104940.

Lipnowski, E., D. Ravid, and D. Shishkin (2022): "Persuasion via weak institutions," *Journal of Political Economy*, 130, 2705–2730.

Matyskova, L. and A. Montes (2023): "Bayesian persuasion with costly information acquisition," *Journal of Economic Theory*, 211, 105678.

Mylovanov, T. and A. Zapechelnyuk (2017): "Optimal allocation with ex post verification and limited penalties," *American Economic Review*, 107, 2666–2694.

Perez-Richet, E. and V. Skreta (2022): "Test design under falsification," *Econometrica*, 90, 1109–1142.

Sims, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50, 665–690.

Tsakas, E., N. Tsakas, and D. Xefteris (2021): "Resisting persuasion," *Economic Theory*, 72, 723–742.

Yoder, N. (2022): "Designing incentives for heterogeneous researchers," *Journal of Political Economy*, 130, 2018–2054.

Zapechelnyuk, A. (2023): "On the equivalence of information design by uninformed and informed principals," *Economic Theory*, 1–17.