Bayesian Persuasion Followed by Receiver's Mechanism Design*

Takuro Yamashita† Shuguang Zhu‡

December 1, 2021

Abstract

In Bayesian persuasion, Receiver simply plays an action after Sender's public signaling. However, in some applications, Receiver could elicit more information by offering a screening contract. This contracting stage may make both Sender and Receiver better off than in Bayesian persuasion: Sender prefers "being further screened". The outcome is most efficient with weak commitment (where Sender privately acquires full information), less so with strong commitment (where Sender jointly designs both public and private signals), and least in Bayesian persuasion. This suggests that economic predictions based on the standard Bayesian persuasion model might be biased toward overly inefficient outcomes.

KEYWORDS: Bayesian persuasion, mechanism design, information acquisition.

---

*We are grateful to Anne-Katrin Roesler, Shintaro Miura, and workshop participants in Communication and Persuasion Workshop (Waseda University) for useful comments. Takuro Yamashita acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 714693), and ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). Shuguang Zhu is supported by "the Fundamental Research Funds for the Central Universities".

†Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France. takuro.yamashita@tse-fr.eu

‡School of Economics, Shanghai University of Finance and Economics, Shanghai, P. R. China. zhu.shuguang@mail.shufe.edu.cn

1 Introduction

Bayesian persuasion / information design models have been applied in many economic contexts, as a representative model where an informed party designs an uninformed party's information structure. Typically, the model assumes the following timing of the game play: (i) The informed party ("Sender") sets up an experiment (at the ex ante stage without knowing the state realization); (ii) the uninformed party ("Receiver") observes a realized signal based on the experiment; and then, (iii) Receiver takes a payoff-relevant action.

The objective of the current paper is to examine an (often implicit) assumption of this framework, namely, that Receiver just takes an action at (iii). To give an idea, imagine an equilibrium of this Bayesian persuasion game where Sender does *not* offer a fully-revealing experiment at (i). This means that, after observing a signal at (ii), Receiver is still uncertain about the state. In this case, even though the standard model assumes that Receiver simply plays an action at (iii), it seems to us quite natural for Receiver to try to *screen* Sender's information further (see the screening / mechanism design literature).

Motivated by this idea, in this paper, we consider a more general model where the third stage (iii) is replaced by (iii') where Receiver *designs a mechanism* (comprising a message space and a mapping from the message space to the action space), followed by (iv') where Sender chooses a message, which determines Receiver's action. As we discuss more formally later, some results depend on the assumption as to how Sender behaves at (iv'). First in Section 3, we consider the case where Sender behaves sequentially rationally at (iv'), that is, given what has happened at (i)-(iii'), Sender inputs the message to the mechanism that is optimal for him. This would correspond to the "weak-commitment" interpretation that Sender can acquire any information at any point (i.e., between (i) and (iv')) costlessly, and thus, he maximizes the pre-acquisition expected payoff at each point. In Section 4, we consider an alternative "strong-commitment" assumption where Sender commits to a joint information structure of public and private(-to-Sender) signals at the beginning, in order to maximize his ex ante payoff. In this case, Sender is unable to further acquire information at (iv'). Which approach is more reasonable may depend on the contexts.

As is discussed in the screening literature, our modeling assumes some commitment power on Receiver's side (to the extent that he should not revise his action after observing Sender's message choice, even if what the mechanism prescribes is not optimal), and in this sense, whether this alternative model is sensible may depend on the applications of interest. For example, imagine that Sender is an investment consultant and Receiver is an institution planning some investment. After the consultant's publishing a generic research report, it is quite common that the investor desires to know more about the investment environment and hence contract with the consultant in order to elicit further advice. In such a context, often the contract between them specifies how the payment from the investor to the consultant is made depending on the investment outcome, consistent with our modeling.

Our first question is whether the equilibrium prediction would change if Receiver is allowed to offer a mechanism. For example, if it does not (under appropriate assumptions on the model parameters, say), then it means that the no-mechanism assumption of the standard framework is innocuous. If it does change, then the next-step question would be how it changes. For example, are Sender and Receiver better-/worse-off in our model relative to the standard case? If Sender is always worse-off, say, then it would be possible to interpret the standard model as the one where Sender *rationally refuses* further elicitation by Receiver (by committing to not acquiring further information). Another question at a more methodological level is whether some well-known useful techniques in the standard model, such as *concavification* (Kamenica and Gentzkow (2011)), are still applicable, perhaps with appropriate modification.

Our results are as follows. First, we find that there are cases where Receiver has a strict incentive of designing a mechanism. In this sense, the (implicit) assumption of the standard model of Receiver's not designing a mechanism can be, potentially, *with* loss of generality. Furthermore, in some cases, not only Receiver but also Sender are strictly better off if Receiver designs a mechanism. Therefore, we cannot justify the standard model by simply arguing that "Sender commits to refusing any additional communication with Receiver", as that would be strictly suboptimal for Sender. Sender sometimes desires to be screened.

On the other hand, there are cases where Receiver has no strict incentive of designing a mechanism. In those cases, the standard no-mechanism assumption is justified. Interestingly, some environments popular in applied papers belong to this class, namely, the environments where Sender's preference does not depend on the state (but only on Receiver's action). This is based on the simple idea that it is basically impossible for Receiver to separate different sender types, and is quite robust: The standard no-mechanism assumption is justified for any prior, and any specification of Receiver's action space and his payoff function.

We then investigate the equilibrium characterization of more specific model in order to obtain further insights. Specifically, motivated by the above investment application, we consider a binary-state environment where Receiver's action comprises a (binary) non-monetary action and an amount of (non-negative) monetary transfer to Sender.

After briefly explaining the equilibrium in the standard Bayesian persuasion model, we first compare it with the weak commitment case, where Sender acquires full information as his private information. We find that, because of the additional channel of communication between Sender and Receiver through a mechanism, the amount of *public* information becomes less than that in the standard model, in the sense of Blackwell ordering. This suggests that the public and private information are treated as "substitutes" by Sender. The *total* communicated information (i.e., how well Receiver is informed at the time of finally taking an action) is improved relative to the standard case. Relatedly, we find that *both* Sender and Receiver are better off: Sender prefers "being further screened".

In the strong commitment model where Sender jointly designs both public and private signals, he sometimes finds it more profitable to have less precise private information. The outcome is most efficient (both in terms of information and welfare) with weak commitment, less so with strong commitment, and least in Bayesian persuasion. Overall, our analysis suggests that economic predictions based on the standard model might be biased toward overly inefficient outcomes, both in terms of information and welfare.

1.1 Related literature

The paper is related to the large literature of Bayesian persuasion. Pioneered by Kamenica and Gentzkow (2011), most of the papers follow the "standard" timing as explained above. In some financial applications, as we argue above, it might make sense to allow Receiver to further screen Sender's information by offering a contract.

For instance, Szydlowski (2021) studies a Bayesian persuasion game in relationship financing where the entrepreneur designs a disclosure policy which sends a signal (about the future cash flows of a project) to the investor, who then decides whether to buy the securities (issued by the entrepreneur) at a fixed price. Azarmsa and Cong (2020) adopts a very similar framework to study the informational hold-up problem, except that after observing the signal, the investor makes a take-it-or-leave-it offer to purchase the securities, and then the entrepreneur decides whether to accept the offer. Ding and Zhu (2019) study the optimal disclosure policy in security issuance where an issuer designs an experiment which generates a signal to persuade an investment bank to underwrite, and then the bank makes its underwriting and retention decisions. Our paper illustrates how Receiver's potentiality to design a mechanism (instead of simply choosing an action) can affect economic outcomes. In this sense, the fundamental idea is based on the vast screening / mechanism design literature (see, for example, Laffont and Martimort (2002) and Börgers (2015), and references therein).

The intersection of mechanism and information design is a smaller but important and growing literature, and our model can be seen as a particular type of "mechanism and information design". In this vein, the most closely related papers to ours is Roesler and Szentes (2017), Condorelli and Szentes (2020), and Deb and Roesler (2021). In these papers, one of the players (say a buyer) can design his private information structure, anticipating that the other player (say a seller) designs an optimal mechanism given that information structure. This can be interpreted as a special case of our strong-commitment model where Sender does not send any public signal.[1] Also in these papers, by construction, Receiver (seller) finds it always better to offer some mechanism than not, while in our case, sometimes he finds it better to *not* offer any non-trivial mechanism (and simply play some action), depending on his belief at that point. This "individual rationality" aspect adds richness to our model. In particular, this makes *joint* design of public and private information crucial.

Our model may also be seen as a model of mechanism design with (pre-mechanism) information acquisition, where Sender (or an "agent" in the mechanism design terminology) may acquire information, and release a part of it publicly. Our model allows flexible information acquisition, but assumes that any information structure is costless. See Mensch (2020) who studies mechanism design where a single agent acquires flexible and costly information. In terms of the grand framework, our model is also related (at a high level) to the cognitive games studied by Pavan and Tirole (2021). The optimality of fully-revealing private signals in our "weak commitment" case is consistent with what they name "unilateral expectation conformity" in self-directed information acquisition.

2 Model

2.1 Basic ingredients

There exist Sender and Receiver. The payoff-relevant state, denoted by $\theta \in \Theta$, follows a common prior $\mu_0 \in \Delta(\Theta)$. Assume Θ is finite. Receiver takes an action $a \in A$, where A may be finite or infinite (e.g., $A = \Delta(\mathcal{A})$ for finite set $\mathcal{A}$; or $A = [0, 1] \times \mathbb{R}_+$ where the first element corresponds to the probability of taking one of the binary non-monetary action, and the second element corresponds to monetary transfer from Receiver to Sender). Sender's ex post payoff is $u_S(\theta, a)$, and Receiver's ex post payoff is $u_R(\theta, a)$.

Both in the standard Bayesian persuasion / information design model and in

[1]In general, as we show, not using public signals can be strictly suboptimal. However, it seems that it is without loss of optimality in the above contexts. We thank Anne-Katrin Roesler for this point.

our model, the game is initiated by Sender's setting up an *experiment* or a *signaling device* denoted by $(\hat{M}, \hat{\sigma})$, where $\hat{M}$ is a measurable space and $\hat{\sigma} : \Theta \to \Delta(\hat{M})$. Receiver observes Sender's choice of $(\hat{M}, \hat{\sigma})$ and also a signal realization $\hat{m} \in \hat{M}$, which makes Receiver update his belief: $\mu_{\hat{m}} \in \Delta(\Theta)$. In this sense, choosing $(\hat{M}, \hat{\sigma})$ is essentially equivalent to choosing a distribution over $\Delta(\Theta)$ (denoted $\lambda \in \Delta(\Delta(\Theta))$) that is *Bayes plausible*:

$$\int_{\mu \in \Delta(\Theta)} \mu \ d\lambda = \mu_0.$$

Thus, in what follows, we assume that Sender chooses a Bayes-plausible $\lambda \in \Delta(\Delta(\Theta))$ without loss of generality.

2.2 Standard Bayesian persuasion model

The standard Bayesian persuasion model assumes that, after Receiver's updating, Receiver simply takes an action $a \in A$. We review the basic results in the standard model here.

Let $a^*(\mu) \in A$ denote Receiver's optimal action if his belief is $\mu \in \Delta(\Theta)$.[2] Thus, Sender's optimal experiment λ^* maximizes:

$$\int_{\mu \in \Delta(\Theta)} \sum_{\theta \in \Theta} u_S(a^*(\mu), \theta)\mu(\theta) \ d\lambda = \int_{\mu \in \Delta(\Theta)} U_S^{BP}(\mu) \ d\lambda$$

subject to Bayes plausibility:

$$\int_{\mu \in \Delta(\Theta)} \mu \ d\lambda = \mu_0,$$

where $U_S^{BP}(\mu) = \mathbb{E}_{\theta \sim \mu}[u_S(a^*(\mu), \theta)]$ denotes Sender's reduced-form payoff given Receiver's optimal action choice ("BP" stands for "Bayesian persuasion").

[2] If Receiver has multiple optimal actions, let $a^*(\mu)$ be Sender's most preferred action, so that Sender's indirect utility as a function of μ is upper semi-continuous.

It is well-known that the optimal λ^* is characterized by *concavification*:

$$\mathbb{E}_{\mu\sim\lambda^*}[U_S^{BP}(\mu)] = \overline{U}_S^{BP}(\mu_0)$$

where $\overline{U}_S^{BP}$ is the smallest concave function everywhere above U_S^{BP}.

The implicit assumption of this standard model is that Receiver simply takes an action after Sender's public disclosure. However, as pointed out in the introduction, unless Sender fully discloses the state realization in the initial stage, Receiver may have an incentive of further eliciting the information from Sender, as generally observed by the screening / mechanism design literature.

3 Information design followed by Receiver's mechanism design

In this section, we formally introduce our model where Receiver designs a mechanism, followed by Sender's sequentially rational response (i.e., the "weak-commitment" case). The "strong-commitment" case is studied in the next section.

The timing of the game is as follows. First, the seller sets up an experiment generating a public signal, $\lambda \in \Delta(\Delta(\Theta))$. After updating his belief to μ, Receiver designs a *mechanism* (M, α), where M is a message set and $\alpha : M \to A$. Then, Sender sends $m \in M$ (and Receiver plays $a = \alpha(m)$).

The "weak-commitment" assumption means that, given true state θ and mechanism (M, α), Sender chooses $m \in M$ which maximizes:

$$u_S(\theta, \alpha(m)).$$

A possible interpretation is that, right after Receiver's offering a mechanism, Sender can acquire any information about θ costlessly, and then reports a message optimally to the mechanism. Given that this information is private, it is weakly dominant for Sender to fully know θ.

Taking a step backward, for Receiver, it is without loss to focus on a *direct* mechanism with $M = \Theta$ where truth-telling is optimal ("revelation principle"):

$$u_S(\theta, \alpha(\theta)) \geq u_S(\theta, \alpha(\theta')), \forall \theta, \theta'.$$

Taking a step further backward, Sender's optimal choice of λ is characterized by *concavification*. Let $U_S^{RW}(\mu)$ denote Sender's continuation payoff given Receiver's belief μ, and given that Receiver designs an optimal mechanism and Sender responds sequentially rationally ("RW" stands for "Receiver's mechanism design + Weak commitment"). Then, the optimal λ satisfies:

$$\mathbb{E}_{\mu \sim \lambda}[U_S^{RW}(\mu)] = \overline{U}_S^{RW}(\mu_0),$$

where $\overline{U}_S^{RW}(\mu)$ is the concavification of U_S^{RW}.

3.1 General results

Here, we provide two general observations. First, there exists an example where Receiver designs a non-trivial mechanism on the equilibrium path. Moreover, in that example, Sender is better off (always weakly, sometimes strictly) if Receiver can design a mechanism, relative to the standard no-mechanism case. In this sense, the (implicit) assumption of the standard model of Receiver's not designing a mechanism can be, potentially, *with* loss of generality.

Theorem 1. There exists an example where $\overline{U}_S^{RW}(\mu_0) > \overline{U}_S^{BP}(\mu_0)$, which necessarily means that Receiver designs a non-trivial mechanism on the equilibrium path.

Proof. See Example in Section 3.3. □

On the other hand, if Sender's preference does not depend on the state (but only on Receiver's action), then Receiver never designs a mechanism.

Theorem 2. If $u_S(\theta, a) = u_S(\theta', a)(\equiv u_S(a))$ for all $a, \theta \neq \theta'$, and $u_S(a) \neq u_S(a')$ for all $a \neq a'$, then Receiver never has a strict incentive of designing a non-trivial mechanism. Accordingly, $\overline{U}_S^{RW} = \overline{U}_S^{BP}$.

3.2 Binary, Transferable Environment

Here, we consider a model with monetary transfer, binary states and binary non-monetary actions, in order to obtain further insights.

Let $\Theta = \{0, 1\}$ and $A = [0, 1] \times \mathbb{R}_+$, where for $a = (x, p) \in A$, $x \in [0, 1]$ denotes the probability that Receiver's playing "action 1" (and $1 - x$ for "action 0"); and $p \in \mathbb{R}_+$ denotes a (non-negative) monetary transfer from Receiver to Sender.[3]

Sender's ex post payoff is $u_S = v_S(\theta)x + p$, and Receiver's ex post payoff is $u_R = v_R(\theta)x - p$.

Note that, in the standard case without Receiver's mechanism design, given any μ, Receiver always chooses $p = 0$. This makes the problem a binary-state, binary-action Bayesian persuasion problem. Receiver plays $x = 1$ ($x = 0$) if $\mathbb{E}_{\theta\sim\mu}[v_R(\theta)] > 0$ (< 0); and he always chooses $p = 0$. Sender's optimal λ is by concavification. One can easily check that, in the Bayesian persuasion solution, we have: (1) if $\mu_0 \geq \frac{-v_R(0)}{-v_R(0)+v_R(1)}$, Sender's ex ante expected payoff is $\mu_0 v_S(1) + (1 - \mu_0)v_S(0)$, and Receiver's ex ante expected payoff is $\mu_0 v_R(1) + (1 - \mu_0)v_R(0)$; (2) if $\mu_0 < \frac{-v_R(0)}{-v_R(0)+v_R(1)}$, Sender's ex ante expected payoff is $\mu_0\big(\frac{v_R(1)}{-v_R(0)} v_S(0) + v_S(1)\big)$, and Receiver's ex ante expected payoff is 0.

If Receiver designs a mechanism, he offers a direct mechanism $(x(\theta), p(\theta))_{\theta\in\Theta}$ where Sender finds it optimal to report θ truthfully:

$$v_S(\theta)x(\theta) + p(\theta) \geq v_S(\theta)x(\theta') + p(\theta'), \ \forall\theta, \theta'.$$

In what follows, we assume $v_S(1) > v_S(0)$ and $v_S(1) > 0$. This is basically a normalization in the sense that the other cases with (opposite) strict inequalities are just the mirror images of this case.[4] By the standard argument, Sender's incentive compatible constraint implies $x(1) \geq x(0)$.

We also assume that $v_R(1) > 0 > v_R(0)$: (i) the case where $v_R(1)$ and $v_R(0)$ have the same sign is not interesting, because Receiver always plays the same action optimally for any given belief; (ii) the case with $v_R(1) < 0 < v_R(0)$ is not

[3] The non-negativity restriction may be interpreted as a limited liability constraint.

[4] The cases with equality are similar too, although slightly more complicated due to indifference.

interesting either, as Receiver's optimal mechanism subject to incentive compatibility is trivially constant: $x(1) = x(0)$ with $p(1) = p(0) = 0$.

As for Sender, if $v_S(1) > 0 > v_S(0)$, then the first best (for both) is trivially possible by full revelation, and thus, we assume $v_S(1) > v_S(0) > 0$.

Due to the above reasons, we impose the following assumption:

Assumption 1. $v_S(1) > v_S(0) > 0$ and $v_R(1) > 0 > v_R(0)$.

The next lemma characterizes the necessary and sufficient condition (on μ and the payoff parameters) where Receiver offers a non-trivial mechanism.

Lemma 1. Under Assumption 1, if $\mu v_R(1) - (1-\mu)v_S(0) \geq 0$ and $v_R(0) + v_S(0) < 0$, then Receiver's optimal mechanism is:

$$\begin{aligned} (x(1), p(1)) &= (1, 0) \\ (x(0), p(0)) &= (0, v_S(0)); \end{aligned}$$

otherwise, Receiver chooses a constant mechanism (and hence, the standard model is without loss).

As suggested in the proof, we may interpret the condition that $\mu v_R(1) - (1 - \mu)v_S(0) \geq 0$ (equivalently, $\mu \geq \underline{\mu} = \frac{v_S(0)}{v_S(0)+v_R(1)}$ [5]) and $v_R(0) + v_S(0) < 0$ as saying that the necessary transfer for screening is not too costly for Receiver. Otherwise, Receiver would not design a non-trivial mechanism, and thus there would be no difference between U_S^{BP} and U_S^{RW}.

Taking a step backward, Sender's optimal choice of λ is characterized by the same *concavification*, but for the above "modified" payoff. First, if $\mu_0 \geq \underline{\mu}$, then Sender initially reveals no information, while it is *followed by R's mechanism*, to which Sender reports the true state. Thus, Receiver plays $(x, p) = (1, 0)$ if $\theta = 1$; $(x, p) = (0, v_S(0))$ if $\theta = 0$. Sender's ex ante expected payoff is $\overline{U}_S^{RW}(\mu_0) = \mu_0 v_S(1) + (1-\mu_0)v_S(0)$. Accordingly, Receiver's ex ante expected payoff is $\mu_0 v_R(1) - (1 - \mu_0)v_S(0)$.

[5] Clearly, we have $\frac{v_S(0)}{v_S(0)+v_R(1)} < \frac{-v_R(0)}{-v_R(0)+v_R(1)}$.

If $\mu_0 < \underline{\mu}$, then Sender initially reveals $\mu = 0$ with probability $1 - \frac{\mu_0}{\underline{\mu}}$ (implying $(x, p) = (0, 0)$ and Sender earns 0) and $\mu = \underline{\mu}$ with probability $\frac{\mu_0}{\underline{\mu}}$, *followed by Receiver's mechanism*, to which Sender reports the true state.[6] Therefore, Sender's ex ante expected payoff is $\overline{U}_S^{RW}(\mu_0) = \frac{\mu_0}{\underline{\mu}}(\underline{\mu} v_S(1) + (1 - \underline{\mu}) v_S(0))$, while Receiver's ex ante expected payoff is 0.

Based on the above analyses, we have the following result.

Theorem 3. If $v_R(1) > 0 > v_R(0)$, $v_S(1) > v_S(0) > 0$, and $v_R(0) + v_S(0) < 0$, then given any $\mu_0 \in (0, 1)$, Receiver designs a non-trivial mechanism with a strictly positive probability on the equilibrium path. Relative to the standard Bayesian persuasion environment, the possibility of Receiver's mechanism design makes both Sender and Receiver weakly better off given any μ_0, and strictly so for some μ_0. More specifically, for $\mu_0 \in \left(0, \frac{-v_R(0)}{-v_R(0)+v_R(1)}\right)$, Sender is strictly better off; for $\mu_0 \in \left(\frac{v_S(0)}{v_S(0)+v_R(1)}, 1\right)$, Receiver is strictly better off; and in their nonempty intersection, both are strictly better off.

3.3 Example

Here, we revisit more formally the investment consulting example briefly discussed in the introduction. There is an investment consultant ("Sender") and an investor ("Receiver"). The state is binary, $\theta \in \{-1, 1\}$, where $\mu_0 = \Pr(\theta = 1)$. Receiver's action comprises an investment decision $x \in \{0, 1\}$ and a non-negative monetary transfer $p \in \mathbb{R}_+$ to Sender.

Receiver's ex post payoff is $u_R(\theta, x, p) = \theta x - p$. Thus, Receiver desires to make an investment if and only if $\theta = 1$. Sender's ex post payoff is $u_S(\theta, x, p) = \pi_\theta x + p$, where $0 < \pi_{-1} < \pi_1 < 1$. Thus, Sender always prefers an investment than no investment, but it is more so given $\theta = 1$ (possibly interpreted as his reputation concern for recommending a "right" decision).

[6] In fact, at $\mu = \underline{\mu}$, Receiver is indifferent between the screening mechanism and the constant allocation rule $(x, p) = (0, 0)$. As in the standard Bayesian persuasion model, we assume that Receiver chooses the Sender-preferred action to break the tie.

3.3.1 Standard Bayesian persuasion model

In the standard model, given Receiver's belief $\mu \in [0,1]$ for $\theta = 1$, he plays $(x,p) = (0,0)$ if $\mu < \frac{1}{2}$, while $(x,p) = (1,0)$ if $\mu \geq \frac{1}{2}$. Note that $p \equiv 0$ given any μ.

Given this, Sender's optimal λ^* is characterized by concavification: (i) If $\mu_0 \geq \frac{1}{2}$, then by no revelation, $x \equiv 1$ and:

$$\overline{U}_S^{BP}(\mu_0) = \mu_0\pi_1 + (1-\mu_0)\pi_{-1};$$

(ii) If $\mu_0 < \frac{1}{2}$, then reveal $\mu = 0$ with probability $1 - 2\mu_0$ (which implies $x = 0$ and Sender earns 0), and reveal $\mu = \frac{1}{2}$ with probability $2\mu_0$ (which implies $x = 1$ and Sender earns $\frac{\pi_1+\pi_{-1}}{2}$). Thus, Sender's ex ante expected payoff is:

$$\overline{U}_S^{BP}(\mu_0) = \mu_0(\pi_1 + \pi_{-1}).$$

3.3.2 Receiver's mechanism design

Here, we consider the case where Receiver designs a mechanism, after Receiver observes a signal given by Sender. Let μ denote Receiver's belief. By the previous analysis, if $\mu \geq \underline{\mu} \equiv \frac{\pi_{-1}}{1+\pi_{-1}} (< \frac{1}{2})$, then it is optimal for Receiver to screen Sender's type by offering the menu mechanism comprising two actions, $(x,p) = (1,0)$ and $(x,p) = (0,\pi_{-1})$. Sender chooses $(x,p) = (1,0)$ if $\theta = 1$, while he chooses $(x,p) = (0,\pi_{-1})$ otherwise. In other words, when the investment is more likely to be "right", the investor would like to learn (and pay for) more accurate information from the consultant, so as to avoid potential investment losses. Meanwhile, the consultant is ready to give honest advice, since the loss of his personal benefit in a bad state would be well compensated. Given this, Receiver's expected payoff (given μ) is $\mu - (1-\mu)\pi_{-1}$.

If $\mu < \underline{\mu}$, then Receiver just plays $(x,p) = (0,0)$ (by offering the corresponding constant mechanism). That is, when facing a grim investment prospect, the investor would simply walk away, because the potential gain from the investment cannot cover the consulting fees.

In view of Sender, his payoff is $U_S^{RW}(\mu) = 0$ if $\mu < \underline{\mu}$, while it is $U_S^{RW}(\mu) =$

$\mu\pi_1 + (1-\mu)\pi_{-1}$ if $\mu \geq \underline{\mu}$. Recall that, in the standard model, the seller earns $U_S^{BP}(\mu) = 0$ if $\mu < \frac{1}{2}$ and $U_S^{BP}(\mu) = \mu\pi_1 + (1-\mu)\pi_{-1}$ if $\mu > \frac{1}{2}$. Therefore, for any μ, the seller is weakly better off, and strictly so if $\mu \in (\underline{\mu}, \frac{1}{2})$.

Taking a step backward, Sender's optimal choice of λ is characterized by the concavification for the above "improved" payoff U_S^{RW}. More specifically, if $\mu_0 > \underline{\mu}$, then no *initial* revelation occurs, while it is *followed by R's mechanism*, to which Sender reports the true state. Thus, Receiver plays $(x,p) = (1,0)$ if $\theta = 1$; $(x,p) = (0,\pi_{-1})$ if $\theta = -1$. Sender's ex ante expected payoff is:

$$\overline{U}_S^{RW}(\mu_0) = \mu_0\pi_1 + (1-\mu_0)\pi_{-1}.$$

If $\mu_0 < \underline{\mu}$, then Sender initially reveals $\mu = 0$ with probability $1 - \frac{\mu_0}{\underline{\mu}}$ (implying $(x,p) = (0,0)$ and Sender earns 0) and $\mu = \underline{\mu}$ with probability $\frac{\mu_0}{\underline{\mu}}$, *followed by R's mechanism*, to which Sender reports the true state. Therefore, Sender's ex ante expected payoff is $\overline{U}_S^{RW}(\mu_0) = \frac{\mu_0}{\underline{\mu}}(\underline{\mu}\pi_1 + (1-\underline{\mu})\pi_{-1})$.

4 Joint Information Design

The previous "weak-commitment" assumption may be interpreted as (costlessly-)information-acquiring, sequentially-rational Sender: at each point in time, he can acquire information costlessly, in order to maximize his pre-acquisition expected payoff given the history up to then. In this section, we consider an alternative "strong-commitment" situation where Sender commits to an information structure jointly of public and private(-to-Sender) signals at the beginning of the game, before knowing the state. This corresponds to a "committed Sender" interpretation of the standard Bayesian persuasion model, but given that Receiver's action is to design a mechanism, Sender designs not only public information but also Sender's private information, which is to be extracted through the mechanism.

The main observation is that, sometimes, Sender may want to *limit* private knowledge of the state, if he can commit to it. As we observe in the transferable-utility context below, this is in order to make Receiver's mechanism more favorable (to Sender) in the sense that more information rent is paid to Sender.

Example 1. Revisit the investment example in Section 3.3, and assume $\mu_0 = \frac{1}{2}$ so that $\overline{U}_S^{BP}(\mu_0) = \overline{U}_S^{RW}(\mu_0) = \frac{\pi_1+\pi_{-1}}{2}$.

Consider the following information structure that Sender commits to. The public signal, as in both the standard Bayesian persuasion case and weak-commitment case, is assumed to reveal no information, so that Receiver's posterior belief after observing the public signal, denoted by μ_1, is still $\frac{1}{2}$.

As for the private(-to-Sender) signal, fix small $\varepsilon > 0$, and consider an information structure which assigns probability $\frac{1}{2}-\varepsilon$ for $\mu_2 = 0$, $\frac{1}{2}-\varepsilon^2$ for $\mu_2 = 1$, and $\varepsilon+\varepsilon^2$ for $\mu_2 = \frac{\varepsilon}{1+\varepsilon}$, where μ_2 denotes the posterior belief induced by Sender's private signal. Note that this is Bayes plausible. Of course, the case with $\varepsilon = 0$ corresponds to full information, making the situation identical to the weak-commitment case. If $\varepsilon > 0$ but small, Receiver's optimal direct mechanism is $(x, p) : \{0, \frac{\varepsilon}{1+\varepsilon}, 1\} \to [0, 1] \times \mathbb{R}_+$, where

$$
\begin{aligned}
&(x(1), p(1)) = (1, 0); \\
&(x(0), p(0)) = (x(\frac{\varepsilon}{1+\varepsilon}), p(\frac{\varepsilon}{1+\varepsilon})) = (0, \frac{\varepsilon\pi_1 + \pi_{-1}}{1+\varepsilon}).
\end{aligned}
$$

We omit its derivation, but its optimality should be interpreted as quite natural: Receiver plays action 1 if Sender reports $\mu_2 = 1$, and plays action 0 if Sender reports that "θ is very likely -1" (i.e., $\mu_2 \in \{0, \frac{\varepsilon}{1+\varepsilon}\}$); the payment given $\mu_2 = 1$ is the lowest possible (i.e., 0), and the payment given $\mu_2 \neq 1$ makes Sender's incentive compatibility binding, in case his true posterior belief is $\mu_2 = \frac{\varepsilon}{1+\varepsilon}$.

Sender's ex ante payoff is then:

$$
(\frac{1}{2} - \varepsilon^2)\pi_1 + (\frac{1}{2} + \varepsilon^2)\frac{\varepsilon\pi_1 + \pi_{-1}}{1+\varepsilon},
$$

which is strictly higher than $\frac{\pi_1+\pi_{-1}}{2}$, because:

$$
\begin{aligned}
& (\frac{1}{2} - \varepsilon^2)\pi_1 + (\frac{1}{2} + \varepsilon^2)\frac{\varepsilon\pi_1 + \pi_{-1}}{1+\varepsilon} - \frac{\pi_1 + \pi_{-1}}{2} \\
= \; & \frac{\varepsilon(1-2\varepsilon)}{2(1+\varepsilon)}(\pi_1 - \pi_{-1}) > 0.
\end{aligned}
$$

Therefore, by making his private information *less precise*, Sender can earn strictly higher expected payoff. This is because, by having a pessimistic but imprecise signal, Receiver has to pay higher in order to elicit Sender's information. However, in case Sender's signal turns out to be extreme (in particular, $\mu_2 = 0$), this means that Sender earns higher information rent than the case with fully precise private information.

A question is whether more involved information structure allows further information rent, and how the optimal joint information structure looks like. To answer these questions, we formally introduce the strong-commitment model.

Let s_1 denote a public signal and s_2 denote a private signal for Sender. Sender designs a joint distribution G for (s_1, s_2), which is equivalent to a marginal distribution G_1 solely for s_1, and a conditional distribution $G_2(\cdot|s_1)$ for s_2 given each realized s_1. Thus, the timing of the game is as follows:

1. Sender chooses G_1 and G_2.

2. Receiver observes s_1, while Sender observes (s_1, s_2).

3. Receiver offers a mechanism (M, α), where M is a message set for Sender, and $\alpha : M \to A$.

4. Sender sends a message m, and Receiver implements $a = \alpha(m)$.

After observing (s_1, s_2), Sender forms a posterior belief about θ. For Receiver, after he observes s_1, he forms a posterior belief jointly about (θ, s_2).

By a similar argument as in the previous section, it is without loss to consider the following equivalent formulation:

1. Sender first chooses $\lambda_1 \in \Delta(\Delta(\Theta))$ such that $\int_{\mu_1} \mu_1 d\lambda_1(\mu_1) = \mu_0$.

2. After observing μ_1, Sender chooses $\lambda_2 \in \Delta(\Delta(\Theta))$ such that $\int_{\mu_2} \mu_2 d\lambda_2(\mu_2) = \mu_1$.

3. After observing μ_1 but not μ_2, Receiver chooses a direct mechanism $(\Delta(\Theta), \alpha)$, where $\Delta(\Theta)$ is the message space for Sender, and $\alpha : \Delta(\Theta) \to A$.

4. Based on Sender's reported μ_2, Receiver implements $a = \alpha(\mu_2)$.

An immediate consequence of this strong-commitment modeling is that Sender would never be worse off relative to (i) the standard model, because Sender can always commit (if he desires so) to not having any private information; nor to (ii) the weak-commitment model, because Sender can always commit (if he desires so) to privately knowing θ fully. These two observations imply that Theorem 1 in Section 3 continues to hold even in the strong-commitment model. Also, it is not difficult to show Theorems 2 in the strong-commitment model too.

Furthermore, validity of *concavification* (with appropriately modified payoff) is easy to see. Given any public signal $\mu_1 \sim \lambda_1$, let $U_S^{RS}(\mu_1)$ denote Sender's expected payoff given that he optimally designs his private information λ_2 given μ_1, and that Receiver optimally designs his incentive compatible direct mechanism ("RS" stands for "Receiver's mechanism design + Strong commitment"). Then, Sender's maximum ex ante payoff is given by $\overline{U}_S^{RS}(\mu_0)$, where $\overline{U}_S^{RS}$ is the smallest concave function everywhere above U_S^{RS}.

4.1 Binary, Transferable Environment

In general, the joint design problem is far more complicated than the weak-commitment case, because Sender now designs a distribution of two (arbitrarily correlated) continuous random variables instead of one, and in principle, for each of the possible joint distributions, we must solve Receiver's mechanism design problem.

Here, in order to provide some key insights by contrasting the strong- and weak-commitment cases, we focus on the binary, transferable environment, and obtain the optimal disclosure policy (and its associated continuation equilibrium) under the same parametric condition as before.[7]

Assumption 2. $v_S(1) > v_S(0) > 0$, $v_R(1) > 0 > v_R(0)$, and $v_S(0) + v_R(0) < 0$.

Assumption 2 is, as in Lemma 1, in order to avoid trivial cases.

[7] We also fully solve the binary-state, three-action case (without monetary transfer) in Online Appendix, and we deliver similar qualitative messages.

As shown in the example above, in general, Sender can be better off by limiting his private information: Having a not-so-informative private signal, Sender earns a relatively high information rent in case the private signal turns out to be extreme. Obviously, if we *fix* the receiver's mechanism choice, making λ_2 assign more probability on the extreme realization would be more beneficial for Sender. However, doing so too much would eventually change Receiver's mechanism (for example, Sender loses the rent if his private signal is fully informative, as in the previous section). Thus, the optimal disclosure policy must hit a right balance.

Based on this logic, in what follows, we "guess" that the optimal information policy makes Receiver indifferent across multiple (in fact, continuously many) mechanisms.[8] We characterize the optimal one within that class. Later, based on a (weak) duality argument, we verify that our guess is indeed correct.

For each $c \in [0, 1]$, we say that Receiver's (direct) mechanism $\alpha = (x, p) : \Delta(\Theta) \to A$ is a *cutoff mechanism* with cutoff c if $x(\mu_2) = 1_{\{\mu_2 > c\}}$. By incentive compatibility and optimality, it is without loss to let the transfer be $p(\mu_2) = (1-x(\mu_2))v_S(c)$. It is worth mentioning that the cutoff mechanism with cutoff $c = 0$ weakly dominates the constant mechanism of "always playing 1" for Receiver: The only difference between two mechanisms is that the cutoff mechanism implements $x = 0$ and $p = v_S(0)$ if Sender reports $\mu_2 = 0$ (hence Receiver earns $-v_S(0)$), while the constant mechanism implements $x = 1$ without payment (hence Receiver earns $v_R(0)$, where $-v_S(0) > v_R(0)$). Sender prefers the cutoff mechanism with $c = 0$ too. Thus, in what follows, we ignore this constant mechanism without loss of generality. On the other hand, the constant mechanism of "always playing 0" is more carefully treated.

Given any public signal realization μ_1, consider Sender's problem of choosing λ_2. As discussed above, one candidate for optimal λ_2 is such that Receiver is made indifferent across multiple cutoff mechanisms. Of course, if μ_1 is too low (in fact, lower than $\underline{\mu} = \frac{v_S(0)}{v_S(0)+v_R(1)}$), then Receiver plays $a = 0$ regardless of λ_2. Even if

[8] This logic of making a mechanism designer indifferent across multiple (continuously many) mechanisms also plays a key role in the buyer's optimal information structure in the monopoly-pricing context (Roesler and Szentes (2017)), hold-up context (Condorelli and Szentes (2020)) and in the robust mechanism design context (Brooks and Du (2021)).

$\mu_1 > \underline{\mu}$, the set of cutoff mechanisms among which Receiver can be indifferent depends on μ_1.

Proposition 1. There exists $\mu_1^* \in [\underline{\mu}, 1]$ such that the following holds.

(i) If $\mu_1 < \underline{\mu}$, then any λ_2 is optimal, and Receiver plays $a = (0, 0)$. Sender (and Receiver) earn expected payoff 0.

(ii) If $\mu_1 \in [\underline{\mu}, \mu_1^*]$, then there exists $\overline{c}(\mu_1)$ such that λ_2 is supported on $[0, \overline{c}(\mu_1)] \cup \{1\}$, and that Receiver is indifferent across all cutoff mechanisms with cutoff $c \in [0, \overline{c}(\mu_1)]$.[9] Moreover, $\overline{c}(\cdot)$ is continuous; strictly increasing; and $\overline{c}(\underline{\mu}) = 0$. Sender's expected payoff is strictly concave, and its supporting hyperplane at $\mu_1 = \underline{\mu}$ goes through the origin.

(iii) If $\mu_1 > \mu_1^*$, then λ_2 is supported on $[0, \overline{c}(\mu_1^*)] \cup \{1\}$, and that Receiver is indifferent across all cutoff mechanisms with cutoff $c \in [0, \overline{c}(\mu_1^*)]$.[10]

The upper bound of the cutoff $\overline{c}(\mu_1)$ is because of "Receiver's individual rationality" condition that he prefers the cutoff mechanism with cutoff $\overline{c}(\mu_1)$ to playing $a = (0, 0)$. As explained above, no such cutoff exists if $\mu_1 < \underline{\mu}$, and the set of possible cutoffs is very limited if μ_1 is small. As μ_1 becomes larger, Receiver becomes able to offer higher-cutoff mechanisms (i.e., $\overline{c}(\mu_1)$ is increasing). Once μ_1 becomes sufficiently large ($\mu_1 > \mu_1^*$), "Receiver's IR" becomes non-binding in Sender's problem, after which $\overline{c}(\mu_1)$ becomes constant.

In the special case of $\mu_1 = \underline{\mu}$, Sender's optimal λ_2 is supported just on $\{0, 1\}$ (because $\overline{c}(\underline{\mu}) = 0$), that is, his private information is fully revealing. Hence, the outcome becomes the same as the weak-commitment case.[11]

Finally, consider the problem of choosing λ_1, the optimal public disclosure. Recall that $U_S^{RS}(\mu_1)$ denotes Sender's expected payoff given μ_1 based on the proposition above. The optimal λ_1 is given by *concavification*: let $\overline{U}_S^{RS}(\cdot)$ be the smallest

[9] On the path of the continuation equilibrium, Receiver chooses the cutoff mechanism with cutoff $\overline{c}(\mu_1)$ with probability one; a standard tie-breaking rule which ensures upper semi-continuity of Sender's indirect utility as a function of λ_2.

[10] Again, Receiver chooses the cutoff mechanism with cutoff $\overline{c}(\mu_1^*)$ with probability one.

[11] This also explains why its supporting hyperplane goes through the origin. Basically, this supporting hyperplane at $\mu_1 = \underline{\mu}$ depicts how Sender's payoff would change for $\mu_1 < \underline{\mu}$, *if both Sender and Receiver behaved the same way* (i.e., Sender choosing fully revealing λ_2 and Receiver choosing the cutoff mechanism with cutoff 0 (even though it is suboptimal for both). As $\mu_1 \to 0$, such hypothetical payoff of Sender linearly converges to 0 (i.e., the origin).

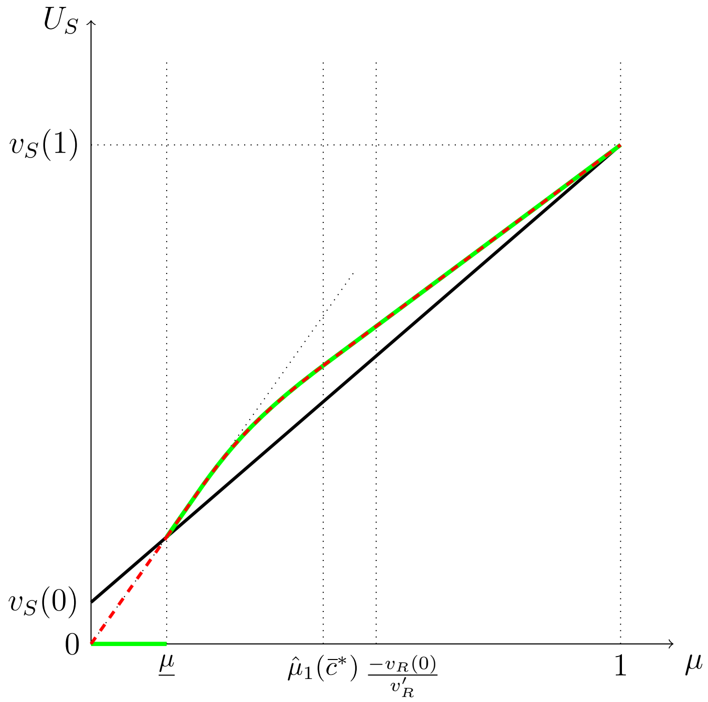

Figure 1: $U_S^{RS}(\cdot)$ (green line) and $\overline{U}_S^{RS}(\cdot)$ (red line)

concave function everywhere above $U_S^{RS}(\cdot)$. Then, Sender's payoff given optimal λ_1 is $\overline{U}_S^{RS}(\mu_0)$. The optimal policy λ_1 is obvious from the above graph.

Theorem 4. If $\mu_0 > \underline{\mu}$, then optimal λ_1 is fully uninformative (i.e., $\mu_1 = \mu_0$ with probability one).

If $\mu_0 < \underline{\mu}$, then optimal λ_1 splits μ_0 to $\mu_1 \in \{0, \underline{\mu}\}$ (with probability $\frac{\mu_0}{\underline{\mu}}$ for $\mu_1 = \underline{\mu}$).

4.2 Comparison of the three regimes

Sender with strong commitment limits his private information in order to make Receiver's mechanism more favorable to Sender. Naturally, this shifts some surplus from Receiver to Sender, at the cost of informational efficiency.

In the following theorems, BP refers to the standard Bayesian persuasion model (i.e., without Receiver's mechanism design), RW refers to the model with Receiver's mechanism design and weak commitment, and RS is with strong commitment. First, from the informational viewpoint, we can rank the three regimes in terms of Blackwell ordering as follows:

Theorem 5. Sender's private information is most informative in RW, less in RS, and not at all in BP.

The public information in BP is more informative than those in RW and RS, which are actually the same as each other.

The total information (i.e., how much Receiver is informed when he takes an action) is most informative in RW, less in RS, and least in BP.

Sender's private information is perhaps the easiest to understand: It is fully informative in RW by definition, and no private information is considered in BP.

The result about the public information implies that the public and private information are "substitute" to each other in this problem. As the private information is expected to be more informative, less public information is released.

For the total informativeness, the private part dominates the public part. Intuitively, this is because Sender can earn information rent based on private information, while public information is released for free; hence, he has less incentive to release more public information.

Next, from the payoff / surplus viewpoint, again the three regimes can be fully ranked, as follows:

Theorem 6. For any given prior μ_0, we have:

Sender's payoff is the highest in RS, less in RW, and lowest in BP;

Receiver's payoff is the highest in RW, less in RS, and lowest in BP;

The total payoff is the highest in RW, less in RS, and lowest in BP.

The payoff comparison for each player is straightforward. For the total payoff, the ranking coincides with that of the total information, which suggests that the ranking of the total payoff is basically based on how informed Receiver's decision-making is.

That the standard Bayesian persuasion case is the least favorable for both of the players strongly suggests that, in applications where Receiver's commitment power (to mechanisms) is reasonable, we should seriously consider the models with

Receiver's mechanism design. As the analysis here suggests, the obtained economic insights could be significantly different.

5 Conclusion

This paper raises the point that, in the standard timing of Bayesian persuasion games, unless the sender's initial public disclosure is fully revealing, the receiver may have an incentive to try to further elicit information from the sender, by offering a mechanism. Importantly, not only the receiver, but the sender may also prefer to be screened by this mechanism. Our analysis also suggests that more information may be communicated from the sender to the receiver if the receiver can offer a mechanism, and in this sense, the economic predictions based on the standard model might be biased toward overly inefficient outcomes, at least in certain environments.

A Omitted proofs

A.1 Proof of Theorem 2

Proof. Fix any μ as Receiver's belief at his mechanism design stage. It suffices to show that only constant mechanisms are incentive compatible.

Direct mechanism $\alpha : \Theta \to A$ is incentive compatible only if:

$$u_S(\alpha(\theta)) \geq u_S(\alpha(\theta'))$$
$$u_S(\alpha(\theta')) \geq u_S(\alpha(\theta)),$$

implying $u_S(\alpha(\theta)) \geq u_S(\alpha(\theta'))$ for all θ, θ'. Sender's no-indifference condition over A implies $\alpha(\theta) = \alpha(\theta')$. Thus, only constant mechanisms are incentive compatible. $\square$

A.2 Proof of Lemma 1

Proof. Given μ, Receiver's optimal mechanism is given by:

$$\begin{aligned}\max_{(x,p):\{0,1\}\to[0,1]\times\mathbb{R}_+} \quad & \mu(v_R(1)x(1)-p(1))+(1-\mu)(v_R(0)x(0)-p(0)) \\ \text{sub. to} \quad & v_S(1)x(1)+p(1)\geq v_S(1)x(0)+p(0) \\ & v_S(0)x(0)+p(0)\geq v_S(0)x(1)+p(1).\end{aligned}$$

In the solution, we must have $x(1)\geq x(0)$, $p(1)=0$, and $p(0)=v_S(0)x(1)-v_S(0)x(0)$, and thus, the problem reduces to:

$$\begin{aligned}\max_{(x,p):\{0,1\}\to[0,1]\times\mathbb{R}_+} \quad & (\mu v_R(1)-(1-\mu)v_S(0))x(1)+(1-\mu)(v_R(0)+v_S(0))x(0) \\ \text{sub. to} \quad & x(1)\geq x(0).\end{aligned}$$

If $\mu v_R(1)-(1-\mu)v_S(0)>0$ and $v_R(0)+v_S(0)<0$, then it is optimal to set $x(1)=1$ and $x(0)=0$. If $\mu v_R(1)-(1-\mu)v_S(0)=0$ and $v_R(0)+v_S(0)<0$, Receiver is indifferent between the screening mechanism (i.e., $x(1)=1$, $x(0)=0$) and the constant allocation rule $x(1)=x(0)=0$. As in the standard Bayesian persuasion model, we assume that Receiver chooses the Sender-preferred action to break the tie, which is the screening mechanism. In the other cases, either $x(1)=x(0)=1$ or $x(1)=x(0)=0$ is optimal. □

A.3 Proof of Proposition 1

Proof. As mentioned in the main text, we characterize the optimal λ_2 for each μ_1 in two steps: first, we consider a particular class of distributions which make Receiver indifferent to continuously many cutoff mechanisms, and then solve the optimal one within that class; second, we prove that the solution we obtain in the first step is indeed optimal among all possible λ_2 such that $\int_{\mu_2}\mu_2 d\lambda_2(\mu_2)=\mu_1$. To simplify the notations, we write $v_S'=v_S(1)-v_S(0)$, $v_R'=v_R(1)-v_R(0)$ and $w(\cdot)=v_S(\cdot)+v_R(\cdot)$.

Step 1

Given any μ_1, we formally define such class of λ_2: (i) for some $\overline{c} \in [0,1]$, any cutoff mechanism with a cutoff in $[0,\overline{c}]$ is optimal for Receiver; (ii) λ_2 is supported on $[0,\overline{c}] \cup \{1\}$ admitting density on $(0,\overline{c})$ and having atoms only at 0 and 1; and (iii) $v_S(c) + v_R(c) \leq 0$ for all $c \in [0,\overline{c}]$ (equivalently, $\overline{c} \leq \frac{-v_S(0)-v_R(0)}{v_S'+v_R'}$). The first condition and the regularity part of the second condition is based on our "indifference" guess above. The reason why λ_2 may be supported not only on $[0,\overline{c}]$ but also on $\{1\}$ is because of the Bayes plausibility requirement (e.g., imagine the case where μ_1 is higher than $\overline{c}$). The third condition also makes sense, because if $v_S(c) + v_R(c) > 0$ for some $c \in [0,\overline{c}]$, then Receiver strictly prefers $a = 1$ if Sender reports exactly c, while the mechanism recommends $a = 0$. By lowering the cutoff, Receiver is strictly better off, because Receiver plays $a = 1$ more often when it is beneficial for him, and he pays less to Sender.

If Receiver chooses a cutoff mechanism with cutoff c, his expected payoff is:

$$\begin{aligned}
& \int_{\mu\leq c} (-v_S(c))d\lambda_2(\mu) + \int_{\mu>c} v_R(\mu)d\lambda_2(\mu) \\
= & \int_{\mu\leq c} (-v_S(c) - v_R(\mu))d\lambda_2(\mu) + v_R(\mu_1) \\
= & \int_{\mu\leq c} (-w(c))d\lambda_2(\mu) + v_R' \int_{\mu\leq c} \lambda_2(\mu)d\mu + v_R(\mu_1)
\end{aligned}$$

and thus, any cutoff mechanism with cutoff $c \in [0,\overline{c}]$ is optimal only if the derivative of this expression with respect to c is 0:

$$(-w(c))\frac{d\lambda_2}{d\mu}(c) - v_S'\lambda_2(c) = 0,$$

where $\lambda_2(c)$ should be interpreted as a cumulative distribution function here; and also, Receiver earns no less than when he simply plays $a = 0$:

$$\int_{\mu\leq c} (-w(c))d\lambda_2(\mu) + v_R' \int_{\mu\leq c} \lambda_2(\mu)d\mu + v_R(\mu_1) \geq 0.$$

The above ordinary differential equation implies:

$$\lambda_2(\mu) = \lambda_2(0)\Big(\frac{-w(0)}{-w(\mu)}\Big)^{\frac{v'_S}{v'_S+v'_R}}$$

for each $\mu \in [0, \overline{c}]$, and $\lambda_2(1) = 1 - \lambda_2(\overline{c})$. This expression indeed makes sense as a cumulative distribution function because $w(\mu) = v_S(\mu) + v_R(\mu) < 0$ for all $\mu \leq \overline{c}$.

Moreover, the Bayes plausibility requires $\mu_1 = \int_{\mu\in[0,1]} \mu d\lambda_2(\mu)$, which implies

$$\lambda_2(0) = \frac{(1-\mu_1)v'_R}{\big(v_S(\overline{c}) + v_R(1)\big)\big(\frac{-w(0)}{-w(\overline{c})}\big)^{\frac{v'_S}{v'_S+v'_R}} - w(0)}(\geq 0).$$

Thus, the form of λ_2 in this class is fully determined up to a single parameter $\overline{c}$. Given any choice of $\overline{c}$, assume that the best continuation equilibrium in view of Sender is to be selected (as standard in Bayesian persuasion); that is, Receiver plays the cutoff mechanism with cutoff $\overline{c}$ for sure, in spite of his indifference to any cutoffs within $[0, \overline{c}]$.[12] Then, Sender's expected payoff (given μ_1) is:

$$\begin{aligned}\underline{U}_S(\overline{c};\mu_1) &= \int_{\mu\leq\overline{c}} v_S(\overline{c})d\lambda_2(\mu) + \int_{\mu>\overline{c}} v_S(\mu)d\lambda_2(\mu) = v'_S\int_{\mu\leq\overline{c}}(\overline{c}-\mu)d\lambda_2(\mu) + v_S(\mu_1)\\ &= (1-\mu_1)v'_S\Big[1 - \frac{(1-\overline{c})v'_R}{v_S(\overline{c}) + v_R(1) + (-w(0))^{\frac{v'_R}{v'_S+v'_R}}(-w(\overline{c}))^{\frac{v'_S}{v'_S+v'_R}}}\Big] + v_S(\mu_1).\end{aligned}$$

Thus, Sender's best expected payoff (given μ_1) within this class of λ_2 is given by:

$$\begin{aligned}\underline{U}^*_S(\mu_1) = \max_{\overline{c}\in[0,\frac{-w(0)}{v'_S+v'_R}]} \quad & \underline{U}_S(\overline{c};\mu_1)\\ \text{sub. to} \quad & (-w(0))\lambda_2(0) + v_R(\mu_1) \geq 0\\ & \lambda_2(\overline{c}) \leq 1,\end{aligned}$$

where the first constraint says that Receiver prefers the cutoff mechanism with

[12]Even if other selection criteria were adopted, Sender could earn the same expected payoff *approximately*, i.e., by a sequence of appropriately perturbed λ_2's. In this sense, our qualitative result would not change.

cutoff 0 (and any cutoff mechanism with cutoff less than or equal to $\overline{c}$) to the constant mechanism of "always playing $x = 0$".

Lemma 2. The constraint $\lambda_2(\overline{c}) \leq 1$ is redundant.

Proof of the lemma. It suffices to show that the constraint $\lambda_2(\overline{c}) \leq 1$ is implied by $(-w(0))\lambda_2(0) + v_R(\mu_1) \geq 0$. Because we have

$$\begin{aligned}\lambda(\overline{c}) \leq 1 \Leftrightarrow \lambda_2(0)\Big(\frac{-w(0)}{-w(\overline{c})}\Big)^{\frac{v'_S}{v'_S+v'_R}} &\leq 1\\ \Leftrightarrow \mu_1 &\geq -\frac{v_S(\overline{c})+v_R(0)}{v'_R} + \frac{w(0)}{v'_R}\Big(\frac{-w(\overline{c})}{-w(0)}\Big)^{\frac{v'_S}{v'_S+v'_R}} := \tilde{\mu}_1(\overline{c}),\end{aligned}$$

$\overline{c}$ satisfies $\lambda_2(\overline{c}) \leq 1$ if and only if $\mu_1 \geq \tilde{\mu}_1(\overline{c})$. Similarly, because we have

$$\begin{aligned}&(-w(0))\lambda_2(0) + v_R(\mu_1) \geq 0\\ \Leftrightarrow &\mu_1 \geq \frac{-v_R(0)}{v'_R} - \frac{v_R(1)}{v'_R}\frac{\big(-w(0)\big)^{\frac{v'_R}{v'_S+v'_R}}\big(-w(\overline{c})\big)^{\frac{v'_S}{v'_S+v'_R}}}{v_S(\overline{c})+v_R(1)} := \hat{\mu}_1(\overline{c}),\end{aligned}$$

$\overline{c}$ satisfies $(-w(0))\lambda_2(0) + v_R(\mu_1) \geq 0$ if and only if $\mu_1 \geq \hat{\mu}_1(\overline{c})$. Since we have

$$\hat{\mu}_1(\overline{c}) - \tilde{\mu}_1(\overline{c}) = \frac{v_S(\overline{c})}{v'_R}\Big[1 + \frac{\big(-w(0)\big)^{\frac{v'_R}{v'_S+v'_R}}\big(-w(\overline{c})\big)^{\frac{v'_S}{v'_S+v'_R}}}{v_S(\overline{c})+v_R(1)}\Big] > 0$$

for any $\overline{c}$, we conclude that the constraint $\lambda(\overline{c}) \leq 1$ is redundant. □

One can easily check that $\hat{\mu}_1(\overline{c})$ is increasing in $\overline{c}$, and $\hat{\mu}_1(0) = \underline{\mu}$. The next lemma shows that Sender's problem has a unique unconstrained maximizer.

Lemma 3. Sender's objective function $\underline{U}_S(\overline{c};\mu_1)$ has a unique maximizer, named $\overline{c}^*$, over $[0, \frac{-w(0)}{v'_S+v'_R}]$, which is given by

$$\Big(\frac{-w(\overline{c}^*)}{-w(0)}\Big)^{\frac{v'_R}{v'_S+v'_R}} = \frac{v_S(1)+v_R(\overline{c}^*)}{w(1)}. \tag{1}$$

Proof of the lemma. One can easily check that the first order derivative of $\underline{U}_S(\overline{c};\mu_1)$ with respect to $\overline{c}$ has the same sign as

$$w(1)-\big(v_R(\overline{c})+v_S(1)\big)\big(-w(0)\big)^{\frac{v_R'}{v_S'+v_R'}}\big(-w(\overline{c})\big)^{-\frac{v_R'}{v_S'+v_R'}}:=h(\overline{c}).$$

Because we have

$$\frac{dh(\overline{c})}{d\overline{c}}=-v_R'v_S'(1-\overline{c})\big(-w(0)\big)^{\frac{v_R'}{v_S'+v_R'}}\big(-w(\overline{c})\big)^{-\frac{v_R'}{v_S'+v_R'}-1}<0,$$
$$h(0)=v_R'>0,\qquad \lim_{\overline{c}\nearrow\frac{-w(0)}{v_S'+v_R'}}h(\overline{c})=-\infty,$$

there exists a unique $\overline{c}^*\in[0,\frac{-w(0)}{v_S'+v_R'}]$ such that $h(\overline{c}^*)=0$. It follows that $\underline{U}_S(\overline{c};\mu_1)$ is increasing in $\overline{c}$ over $(0,\overline{c}^*)$, and decreasing over $(\overline{c}^*,\frac{-w(0)}{v_S'+v_R'})$. Thus, $\overline{c}^*$ is the unique maximizer of $\underline{U}_S(\overline{c};\mu_1)$, given by $h(\overline{c}^*)=0$, which is equivalent to Equation (1). □

It is worth noting that $\overline{c}^*$ is independent of μ_1. Based on these results, we get the optimal $\overline{c}$ for each μ_1: When $\mu_1\geq\hat{\mu}_1(\overline{c}^*):=\mu_1^*$, the unconstrained maximizer automatically satisfies $(-w(0))\lambda_2(0)+v_R(\mu_1)\geq 0$, thus we have

$$\underline{U}_S^*(\mu_1)=\underline{U}_S(\overline{c}^*;\mu_1)=\frac{-w(\overline{c}^*)}{1-\overline{c}^*}(1-\mu_1)+v_S(\mu_1).$$

When $\mu_1\in[\underline{\mu},\mu_1^*)$, we get a binding constraint $(-w(0))\lambda_2(0)+v_R(\mu_1)=0$ (or equivalently, $\mu_1=\hat{\mu}_1(\overline{c})$), which pins down the optimal $\overline{c}(\mu_1)$, and we have

$$\underline{U}_S^*(\mu_1)=\underline{U}_S(\overline{c}(\mu_1);\mu_1)=v_S'(1-\mu_1)-\frac{v_S'(1-\overline{c}(\mu_1))v_R(1)}{v_S(\overline{c}(\mu_1))+v_R(1)}+v_S(\mu_1).$$

When $\mu_1<\underline{\mu}$, there is no candidate within this class of λ_2 such that $(-w(0))\lambda_2(0)+v_R(\mu_1)\geq 0$. In other words, Receiver always prefers playing $a=(0,0)$ to any cutoff mechanisms, and Sender's choice of $\overline{c}$ doesn't matter.

Step 2

Now we prove that what we get in Step 1 is indeed optimal among all Bayes plausible λ_2. First, consider any information policy $\lambda = (\lambda_1, \lambda_2)$ and its associated continuation equilibrium given each realized μ_1, in particular, Receiver's mechanism $\alpha = (x, p) : \Delta(\Theta) \to [0, 1] \times \mathbb{R}_+$. By the standard argument, $x(\mu_2)$ must be non-decreasing, and $p(\mu_2)$ is given by the envelope formula:

$$x(\mu_2)v_S(\mu_2) + p(\mu_2) = x(1)v_S(1) + p(1) - \int_{z=\mu_2}^{1} x(z)v_S' dz,$$

where $p(1) = 0$ by optimality. Accordingly, Receiver's payoff for each given μ_2 is:

$$\begin{aligned} & x(\mu_2)v_R(\mu_2) - p(\mu_2) \\ = \; & x(\mu_2)v_R(\mu_2) - [x(1)v_S(1) - \int_{z=\mu_2}^{1} x(z)v_S' dz - x(\mu_2)v_S(\mu_2)] \\ = \; & x(\mu_2)w(\mu_2) - x(1)v_S(1) + \int_{z=\mu_2}^{1} x(z)v_S' dz \end{aligned}$$

Case (i). $\mu_1 \geq \mu_1^*$.

Consider the following problem:

$$\begin{aligned} \max_{\lambda_2 \in \Delta([0,1]), x(\cdot)} \quad & \int_{\mu_2 \in [0,1]} [x(1)v_S(1) - \int_{z=\mu_2}^{1} x(z)v_S' dz] d\lambda_2(\mu_2) \\ \text{sub. to} \quad & \int_{\mu_2 \in [0,1]} [x(\mu_2)w(\mu_2) - x(1)v_S(1) + \int_{z=\mu_2}^{1} x(z)v_S' dz] d\lambda_2(\mu_2) \\ & \qquad \geq v_R(\mu_1) - \int_{\mu_2 \leq y} (v_S(y) + v_R(\mu_2)) d\lambda_2(\mu_2), \; \forall y \in [0, \bar{c}^*] \\ & \int_{\mu_2 \in [0,1]} \mu_2 d\lambda_2(\mu_2) = \mu_1, \quad x : \text{ non-decreasing.} \end{aligned}$$

This is to maximize Sender's expected payoff by choosing a distribution λ_2 that is Bayes plausible with respect to μ_1 (the second constraint), letting Receiver's mechanism $\alpha = (x, p) : \Delta(\Theta) \to [0, 1] \times \mathbb{R}_+$. The first constraint requires that Receiver prefers this mechanism to any cutoff mechanism with cutoff in $[0, \bar{c}^*]$. To the extent that only a subset of Receiver's incentive conditions is taken care of, the value of the above problem must be (weakly) higher than Sender's expected payoff given μ_1 under the optimal λ_2.

By a standard (weak) duality argument, the following min-max problem attains a (weakly) even higher value:

$$\begin{aligned}
\min_{\phi\geq 0}\Big[\max_{\lambda_2\in\Delta([0,1]),x(\cdot)} & \int_{\mu_2\in[0,1]} [x(1)v_S(1) - \int_{z=\mu_2}^{1} x(z)v_S' dz] d\lambda_2(\mu_2) \\
& + \int_{y=0}^{\overline{c}^*} \Big[\int_{\mu_2\in[0,1]} [x(\mu_2)w(\mu_2) - x(1)v_S(1) + \int_{z=\mu_2}^{1} x(z)v_S' dz] d\lambda_2(\mu_2) \\
& -v_R(\mu_1) + \int_{\mu_2\leq y} (v_S(y) + v_R(\mu_2)) d\lambda_2(\mu_2) \Big] \phi(y) dy \\
\text{sub. to} & \int_{\mu_2\in[0,1]} \mu_2 d\lambda_2(\mu_2) = \mu_1, \quad x: \text{ non-decreasing} \Big]
\end{aligned}$$

Consider the following candidate solution to this min-max problem:

$$\phi(y) = \frac{v_R(\overline{c}^*) + v_S(1)}{(-w(y))(1-\overline{c}^*)} \Big(\frac{-w(0)}{-w(y)}\Big)^{\frac{v_R'}{v_S'+v_R'}}$$

for $y \in [0, \overline{c}^*]$, and $\phi(y) = 0$ otherwise. Notice that

$$\Phi(y) = \int_{z\in[0,y]} \phi(z) dz = \frac{v_R(\overline{c}^*) + v_S(1)}{v_R'(1-\overline{c}^*)} \Big[\Big(\frac{-w(0)}{-w(y)}\Big)^{\frac{v_R'}{v_S'+v_R'}} - 1\Big]$$

satisfies $\Phi(\overline{c}^*) = 1$ (due to Equation (1)), thus ϕ may be interpreted as a probability density, and Φ denotes its cumulative distribution function. Then, the objective of the min-max problem becomes:

$$-v_R(\mu_1) + \int_{\mu_2=0}^{1} x(\mu_2)w(\mu_2) d\lambda_2(\mu_2) + \int_{y=0}^{\overline{c}^*} \Big[\int_{\mu_2=0}^{y} (v_S(y) + v_R(\mu_2)) d\lambda_2(\mu_2) \Big] \phi(y) dy.$$

Thus, the optimal x is given by $x(\mu_2) = 1_{\{\mu_2 > \mu^*\}}$, where μ^* is such that $w(\mu^*) = 0$.

Since $\overline{c}^* < \mu^* < 1$, the objective further becomes:

$$v_S(\mu_1) - \int_{\mu_2=0}^{\mu^*} w(\mu_2)d\lambda_2(\mu_2) + \int_{y=0}^{\overline{c}^*} \Big[\int_{\mu_2=0}^{y} (v_S(y) + v_R(\mu_2))d\lambda_2(\mu_2) \Big] \phi(y)dy.$$

Applying integration by parts (twice), the third item (that is, the double integral) can be written as:

$$\int_{y=0}^{\overline{c}^*} (1 - \Phi(y))w(y)d\lambda_2(y) + v'_S \int_{y=0}^{\overline{c}^*} J(y)d\lambda_2(y),$$

where

$$\begin{aligned} J(y) &= \int_{y'=y}^{\overline{c}^*} \big(1 - \Phi(y')\big)dy' \\ &= \frac{w(1)(\overline{c}^* - y)}{v'_R(1 - \overline{c}^*)} + \frac{v_R(\overline{c}^*) + v_S(1)}{v'_S v'_R(1 - \overline{c}^*)} (-w(0))^{\frac{v'_R}{v'_S + v'_R}} \Big[(-w(\overline{c}^*))^{\frac{v'_S}{v'_S + v'_R}} - (-w(y))^{\frac{v'_S}{v'_S + v'_R}} \Big] \end{aligned}$$

and therefore, the objective finally becomes:

$$\begin{aligned} & v_S(\mu_1) - \int_{\mu_2=0}^{\mu^*} w(\mu_2)d\lambda_2(\mu_2) + \int_{\mu_2=0}^{\overline{c}^*} \big[(1 - \Phi(\mu_2))w(\mu_2) + v'_S J(\mu_2)\big] d\lambda_2(\mu_2) \\ &= v_S(\mu_1) + \int_{\mu_2=0}^{1} K(\mu_2)d\lambda_2(\mu_2), \end{aligned}$$

where

$$K(\mu_2) = 1_{\{\mu_2 \le \overline{c}^*\}} \big(v'_S J(\mu_2) - \Phi(\mu_2)w(\mu_2)\big) + 1_{\{\mu_2 \in (\overline{c}^*, \mu^*]\}} \big(-w(\mu_2)\big).$$

Lemma 4. $K(\mu_2) = \frac{-w(\overline{c}^*)}{1 - \overline{c}^*}(1 - \mu_2)$ for $\mu_2 \in [0, \overline{c}^*] \cup \{1\}$, and $K(\mu_2) \le \frac{-w(\overline{c}^*)}{1 - \overline{c}^*}(1 - \mu_2)$ otherwise.

Proof of the lemma. For $\mu_2 \in (0, \overline{c}^*)$, by differentiation, we have

$$K'(\mu_2) = -v'_S - \phi(\mu_2)w(\mu_2) - v'_R\Phi(\mu_2) = \frac{w(\overline{c}^*)}{1 - \overline{c}^*}.$$

Because $K(\mu_2)$ is continuous at $\mu_2 \in \{0, \overline{c}^*\}$, and

$$
\begin{aligned}
K(0) &= v_S' J(0) = v_S' \int_{y'=0}^{\overline{c}^*} \big(1 - \Phi(y')\big) dy' \\
&= \frac{w(1) v_S' \overline{c}^*}{v_R'(1-\overline{c}^*)} + \frac{v_R(\overline{c}^*) + v_S(1)}{v_R'(1-\overline{c}^*)} \Big[(-w(0))^{\frac{v_R'}{v_S'+v_R'}} (-w(\overline{c}^*))^{\frac{v_S'}{v_S'+v_R'}} + w(0) \Big] \\
&= \frac{w(1) v_S' \overline{c}^*}{v_R'(1-\overline{c}^*)} + \frac{v_R(\overline{c}^*) + v_S(1)}{v_R'(1-\overline{c}^*)} \Big[\frac{-w(\overline{c}^*) w(1)}{v_S(1) + v_R(\overline{c}^*)} + w(0) \Big] = \frac{-w(\overline{c}^*)}{1-\overline{c}^*},
\end{aligned}
$$

we obtain $K(\mu_2) = \frac{-w(\overline{c}^*)}{1-\overline{c}^*}(1-\mu_2)$ for $\mu_2 \in [0, \overline{c}^*]$.

For $\mu_2 \in (\overline{c}^*, \mu^*]$, because $K(\mu_2)$ is linear, and continuous at $\mu_2 = \overline{c}^*$, it suffices to show that $K(\mu^*) \leq \frac{-w(\overline{c}^*)}{1-\overline{c}^*}(1-\mu^*)$. This is indeed the case, because $K(\mu^*) = -w(\mu^*) = 0$, while $\frac{-w(\overline{c}^*)}{1-\overline{c}^*}(1-\mu^*) > 0$. For $\mu_2 \in (\mu^*, 1)$, we have $K(\mu_2) = 0 < \frac{-w(\overline{c}^*)}{1-\overline{c}^*}(1-\mu_2)$. Finally, for $\mu_2 = 1$, we have $K(\mu_2) = 0$. □

Fixed the above candidate solution ϕ, the inner maximization problem of the min-max problem becomes:

$$
\begin{aligned}
\max_{\lambda_2 \in \Delta([0,1])} \quad & v_S(\mu_1) + \int_{\mu_2 \in [0,1]} K(\mu_2) d\lambda_2(\mu_2) \\
\text{sub. to} \quad & \int_{\mu_2 \in [0,1]} \mu_2 d\lambda_2(\mu_2) = \mu_1.
\end{aligned}
$$

This is a standard Bayesian persuasion problem of choosing λ_2 given μ_1. By Lemma 4, its value is at most $v_S(\mu_1) + \frac{-w(\overline{c}^*)}{1-\overline{c}^*}(1-\mu_1)$, which is attainable by any λ_2 supported only on $[0, \overline{c}^*] \cup \{1\}$, and serves as the upper bound of Sender's payoff given $\mu_1 \in [\mu_1^*, 1]$. On the other hand, this value coincides with $\underline{U}_S^*(\mu_1)$ in Step 1 (by the specific λ_2 with parameter $\overline{c} = \overline{c}^*$). We therefore complete the verification.

Case (ii). $\underline{\mu} \leq \mu_1 < \mu_1^*$.

Consider the following problem:

$$\begin{aligned}
\max_{\lambda_2 \in \Delta([0,1]), x(\cdot)} \quad & \int_{\mu_2 \in [0,1]} [x(1)v_S(1) - \int_{z=\mu_2}^1 x(z)v_S' dz] d\lambda_2(\mu_2) \\
\text{sub. to} \quad & \int_{\mu_2 \in [0,1]} [x(\mu_2)w(\mu_2) - x(1)v_S(1) + \int_{z=\mu_2}^1 x(z)v_S' dz] d\lambda_2(\mu_2) \\
& \qquad \geq v_R(\mu_1) - \int_{\mu_2 \leq y} (v_S(y) + v_R(\mu_2)) d\lambda_2(\mu_2), \ \forall y \in [0, \overline{c}(\mu_1)] \\
& \int_{\mu_2 \in [0,1]} [x(\mu_2)w(\mu_2) - x(1)v_S(1) + \int_{z=\mu_2}^1 x(z)v_S' dz] d\lambda_2(\mu_2) \geq 0 \\
& \int_{\mu_2 \in [0,1]} \mu_2 d\lambda_2(\mu_2) = \mu_1, \quad x: \text{ non-decreasing.}
\end{aligned}$$

where $\overline{c}(\mu_1)$ is given by $\mu_1 = \hat{\mu}_1(\overline{c}(\mu_1))$; or equivalently,

$$(-w(0))^{\frac{v_R'}{v_R'+v_S'}} (-w(\overline{c}(\mu_1)))^{\frac{v_S'}{v_R'+v_S'}} = \frac{-v_R(\mu_1)\big(v_R(1) + v_S(\overline{c}(\mu_1))\big)}{v_R(1)}. \tag{2}$$

Unlike Case (i), the incentive constraints are only for $y \in [0, \overline{c}(\mu_1)]$; however, we now have an "individual rationality" constraint where Receiver does not deviate to "always playing $a = 0$". Now the Lagrangian is:

$$\begin{aligned}
L \;=\; & \int_{\mu_2 \in [0,1]} [x(1)v_S(1) - \int_{z=\mu_2}^1 x(z)v_S' dz] d\lambda_2(\mu_2) \\
& + \gamma \int_{\mu_2 \in [0,1]} [x(\mu_2)w(\mu_2) - x(1)v_S(1) + \int_{z=\mu_2}^1 x(z)v_S' dz] d\lambda_2(\mu_2) \\
& + \int_{y=0}^{\overline{c}(\mu_1)} \Big[\int_{\mu_2 \in [0,1]} [x(\mu_2)w(\mu_2) - x(1)v_S(1) + \int_{z=\mu_2}^1 x(z)v_S' dz] d\lambda_2(\mu_2) \\
& - v_R(\mu_1) + \int_{\mu_2 \leq y} (v_S(y) + v_R(\mu_2)) d\lambda_2(\mu_2) \Big] \phi(y) dy,
\end{aligned}$$

where $\gamma \geq 0$ is the multiplier for Receiver's IR constraint.

Consider the following candidate solution to $\min_{\phi, \gamma \geq 0} \max_{\lambda_2 \in \Delta([0,1]), x(\cdot)} L$ (where

λ_2 is Bayes plausible and x is non-decreasing) :

$$\phi(y) = \frac{w(1)}{(1-\overline{c}(\mu_1))}\big(-w(\overline{c}(\mu_1))\big)^{\frac{v'_R}{v'_S+v'_R}}(-w(y))^{-\frac{v'_R}{v'_S+v'_R}-1}$$

for $y \in [0, \overline{c}(\mu_1)]$, and $\phi(y) = 0$ otherwise; $\gamma = 1 - \Phi(\overline{c}(\mu_1)) = 1 - \int_{y=0}^{\overline{c}(\mu_1)} \phi(y)dy$. The following lemma guarantees that this is indeed a feasible solution.

Lemma 5. $\Phi(\overline{c}(\mu_1)) \leq 1$.

Proof of the lemma. First, by integrating ϕ, we have

$$\Phi(y) = \int_{y'=0}^{y} \phi(y')dy' = \frac{w(1)}{v'_R(1-\overline{c}(\mu_1))}\Big[\Big(\frac{-w(\overline{c}(\mu_1))}{-w(y)}\Big)^{\frac{v'_R}{v'_S+v'_R}} - \Big(\frac{-w(\overline{c}(\mu_1))}{-w(0)}\Big)^{\frac{v'_R}{v'_S+v'_R}}\Big].$$

From $\overline{c}(\mu_1) < \overline{c}^*$ and Equation (1), we have

$$\Big(\frac{-w(\overline{c}(\mu_1))}{-w(0)}\Big)^{\frac{v'_R}{v'_S+v'_R}} > \frac{v_S(1) + v_R(\overline{c}(\mu_1))}{w(1)}.$$

Thus,

$$\begin{aligned} 1 - \Phi(\overline{c}(\mu_1)) &= 1 - \frac{w(1)}{v'_R(1-\overline{c}(\mu_1))}\Big[1 - \Big(\frac{-w(\overline{c}(\mu_1))}{-w(0)}\Big)^{\frac{v'_R}{v'_S+v'_R}}\Big] \\ &> 1 - \frac{w(1)}{v'_R(1-\overline{c}(\mu_1))}\Big[1 - \frac{v_S(1) + v_R(\overline{c}(\mu_1))}{w(1)}\Big] = 0. \end{aligned}$$

□

Then, by the similar procedure as in Case (i) (that is, verifying that it is optimal to set $x(z) = 1_{\{z>\mu^*\}}$, and applying integration by parts twice), we obtain:

$$L = w(\mu_1) - \Phi(\overline{c}(\mu_1))v_R(\mu_1) + \int_{\mu_2=0}^{1} H(\mu_2)d\lambda_2(\mu_2),$$

where

$$H(\mu_2) = 1_{\{\mu_2 \in (\overline{c}(\mu_1), \mu^*]\}} \big(- w(\mu_2) \big)$$
$$+ 1_{\{\mu_2 \le \overline{c}(\mu_1)\}} \Big(v_S' \int_{z=\mu_2}^{\overline{c}(\mu_1)} (\Phi(\overline{c}(\mu_1)) - \Phi(z)) dz + w(\mu_2)(\Phi(\overline{c}(\mu_1)) - \Phi(\mu_2)) - w(\mu_2) \Big).$$

Following the same procedure as in Lemma 4, we have $H(\mu_2) = \frac{-w(\overline{c}(\mu_1))}{1-\overline{c}(\mu_1)}(1-\mu_2)$ for $\mu_2 \in [0, \overline{c}(\mu_1)] \cup \{1\}$, and $H(\mu_2) \le \frac{-w(\overline{c}(\mu_1))}{1-\overline{c}(\mu_1)}(1-\mu_2)$ otherwise. Then the optimal λ_2 is any distribution that is supported on $[0, \overline{c}(\mu_1)] \cup \{1\}$ (and has mean μ_1 by Bayes plausibility). It follows that $w(\mu_1) - \Phi(\overline{c}(\mu_1)) v_R(\mu_1) - \frac{w(\overline{c}(\mu_1))}{1-\overline{c}(\mu_1)}(1-\mu_1)$ is an upper bound of Sender's payoff given $\mu_1 \in [\underline{\mu}, \mu_1^*)$. Applying the (weak) duality argument and Equation (2), we have

$$\begin{aligned} 0 \le \min_{\phi, \gamma \ge 0} & \Big[\max_{\lambda_2 \in \Delta([0,1]), x(\cdot)} L, \text{ s.t. } \mathbb{E}_{\lambda_2}[\mu_2] = \mu_1, \ x: \text{ non-decreasing} \Big] - \underline{U}_S(\overline{c}(\mu_1); \mu_1) \\ \le & \Big[w(\mu_1) - \frac{v_R(\mu_1) w(1)}{v_R'(1 - \overline{c}(\mu_1))} \Big(1 - \frac{w(\overline{c}(\mu_1)) v_R(1)}{v_R(\mu_1)(v_R(1) + v_S(\overline{c}(\mu_1)))} \Big) - \frac{w(\overline{c}(\mu_1))}{1 - \overline{c}(\mu_1)} (1 - \mu_1) \Big] \\ & - \Big[v_S'(1 - \mu_1) - \frac{v_S'(1 - \overline{c}(\mu_1)) v_R(1)}{v_S(\overline{c}(\mu_1)) + v_R(1)} + v_S(\mu_1) \Big] = 0, \end{aligned}$$

which means the specific λ_2 with parameter $\overline{c} = \overline{c}(\mu_1)$ in Step 1 is indeed optimal.

Case (iii). $\mu_1 < \underline{\mu}$.

Consider the following problem:

$$\begin{aligned} \max_{\lambda_2 \in \Delta([0,1]), x(\cdot)} \quad & \int_{\mu_2 \in [0,1]} [x(1) v_S(1) - \int_{z=\mu_2}^{1} x(z) v_S' dz] d\lambda_2(\mu_2) \\ \text{sub. to} \quad & \int_{\mu_2 \in [0,1]} [x(\mu_2) w(\mu_2) - x(1) v_S(1) + \int_{z=\mu_2}^{1} x(z) v_S' dz] d\lambda_2(\mu_2) \ge 0 \\ & \int_{\mu_2 \in [0,1]} \mu_2 d\lambda_2(\mu_2) = \mu_1, \quad x: \text{ non-decreasing}. \end{aligned}$$

Here, we only consider Receiver's individual rationality constraint. The corresponding Lagrangian becomes:

$$\begin{aligned} L &= \int_{\mu_2\in[0,1]}[x(1)v_S(1) - \int_{z=\mu_2}^1 x(z)v_S' dz]d\lambda_2(\mu_2) \\ &\quad +\gamma\int_{\mu_2\in[0,1]}[x(\mu_2)w(\mu_2) - x(1)v_S(1) + \int_{z=\mu_2}^1 x(z)v_S' dz]d\lambda_2(\mu_2). \end{aligned}$$

The goal is to show that, with an appropriate γ, the value of the Lagrangian cannot be strictly positive. This verifies that the optimal policy for Sender is just to let Receiver always play $a = 0$. Taking integration by parts (twice), we obtain:

$$L = x(1)\big(\gamma v_R(\mu_1) + v_S(\mu_1)\big) + \int_{\mu_2=0}^1 \Big[(\gamma v_R' + v_S')\Lambda(\mu_2) - \gamma\lambda_2(\mu_2)w(\mu_2)\Big]dx(\mu_2),$$

where $\Lambda(\mu_2) = \int_{y=0}^{\mu_2}\lambda_2(y)dy$. Immediately, we have

$$\begin{aligned} L &\le x(1)\big(\gamma v_R(\mu_1) + v_S(\mu_1)\big) + x(1)\max_{z\in[0,1]}\big\{(\gamma v_R' + v_S')\Lambda(z) - \gamma\lambda_2(z)w(z)\big\} \\ &= x(1)\max_{z\in[0,1]}\big\{\underbrace{\gamma v_R(\mu_1) + v_S(\mu_1) + (\gamma v_R' + v_S')\Lambda(z) - \gamma\lambda_2(z)w(z)}_{:=l(z,\lambda_2)}\big\}. \end{aligned}$$

Fixed some $\gamma \ge \frac{v_S(\mu_1)}{v_S(0)-\mu_1(v_S(0)+v_R(1))}$ (> 1), to prove L is always non-positive, it suffices to show that $l(z,\lambda_2) \le 0$ for any z and any Bayes-plausible λ_2.

(I) $w(z) \le 0$**.** Because λ_2 is a cumulative distribution function over $[0,1]$ with mean μ_1, for any $z \in [0,1]$, we have $\Lambda(z) + (1-z)\lambda_2(z) \le \Lambda(1) = 1-\mu_1$, which means $\lambda_2(z) \le \frac{1-\mu_1-\Lambda(z)}{1-z}$. Since $\Lambda(z)$ is convex, we have $\Lambda(z) \le (1-z)\Lambda(0) +$

$z\Lambda(1) = z(1-\mu_1)$. Then, we have

$$\begin{aligned}
l(z,\lambda_2) \leq & \gamma v_R(\mu_1) + v_S(\mu_1) + (\gamma v'_R + v'_S)\Lambda(z) - \gamma \frac{1-\mu_1-\Lambda(z)}{1-z} w(z) \\
= & \gamma v_R(\mu_1) + v_S(\mu_1) - \gamma \frac{1-\mu_1}{1-z} w(z) + \Lambda(z)\Big(v'_S + \gamma \frac{v_R(1)+v_S(z)}{1-z}\Big) \\
\leq & \gamma v_R(\mu_1) + v_S(\mu_1) - \gamma \frac{1-\mu_1}{1-z} w(z) + z(1-\mu_1)\Big(v'_S + \gamma \frac{v_R(1)+v_S(z)}{1-z}\Big) \\
= & \gamma v_R(\mu_1) + v_S(\mu_1) + (1-\mu_1)\big[-(\gamma-1)v'_S z - \gamma w(0)\big] \\
\leq & \gamma v_R(\mu_1) + v_S(\mu_1) - \gamma(1-\mu_1)w(0) \\
= & v_S(\mu_1) - \gamma\big[v_S(0) - \mu_1(v_S(0)+v_R(1))\big] \leq 0.
\end{aligned}$$

(II) $w(z) > 0$**.** Because $\Lambda(z) \leq \lambda_2(z)z$ for any $z \in [0,1]$, we have

$$l(z,\lambda_2) \leq \gamma v_R(\mu_1) + v_S(\mu_1) + (\gamma v'_R + v'_S)\Lambda(z) - \gamma \frac{\Lambda(z)}{z} w(z),$$

which is a linear function with respect to $\Lambda(z)$. Since $\Lambda(z) \in [0, z(1-\mu_1)]$, to prove $l(z,\lambda_2) \leq 0$, it suffices to show that the right hand side is non-positive at $\Lambda(z) = 0$ and $\Lambda(z) = z(1-\mu_1)$. In fact, we have

$$\begin{aligned}
& \gamma v_R(\mu_1) + v_S(\mu_1) + (\gamma v'_R + v'_S)z(1-\mu_1) - \gamma \frac{z(1-\mu_1)}{z} w(z) \\
& = \gamma v_R(\mu_1) + v_S(\mu_1) + (1-\mu_1)\big[-(\gamma-1)v'_S z - \gamma w(0)\big] \\
& \leq \gamma v_R(\mu_1) + v_S(\mu_1) - \gamma(1-\mu_1)w(0) \leq 0;
\end{aligned}$$

and from $v_R(\mu_1) \leq v_R(0) + v'_R \frac{v_S(0)}{v_S(0)+v_R(1)} = \frac{w(0)v_R(1)}{v_S(0)+v_R(1)} < 0$, we have

$$\begin{aligned}
& \gamma v_R(\mu_1) + v_S(\mu_1) + (\gamma v'_R + v'_S)\cdot 0 - \gamma \frac{0}{z} w(z) \\
\leq & \frac{v_S(\mu_1)}{v_S(0) - \mu_1(v_S(0)+v_R(1))} v_R(\mu_1) + v_S(\mu_1) = \frac{v_S(\mu_1)w(0)(1-\mu_1)}{v_S(0) - \mu_1(v_S(0)+v_R(1))} < 0.
\end{aligned}$$

Shape of $\underline{U}_S^*(\mu_1)$

In Case (i), $\underline{U}_S^*(\mu_1)$ is linear in μ_1: when $\mu_1 \geq \hat{\mu}_1(\bar{c}^*)$,

$$\underline{U}_S^*(\mu_1) = \frac{-w(\bar{c}^*)}{1-\bar{c}^*}(1-\mu_1) + v_S(\mu_1).$$

In Case (iii), $\underline{U}_S^*(\mu_1)$ is also linear in μ_1: when $\mu_1 < \underline{\mu}$,

$$\underline{U}_S^*(\mu_1) = 0.$$

In Case (ii), $\underline{U}_S^*(\mu_1)$ is increasing and concave. To prove this, we first take the derivative of Equation (2) with respect to μ_1, and get:

$$\frac{d\bar{c}(\mu_1)}{d\mu_1} = \frac{w(\bar{c}(\mu_1))(v_R(1) + v_S(\bar{c}(\mu_1)))}{v_S' v_R(\mu_1)(1-\bar{c}(\mu_1))}.$$

Then, we check the first-order and second-order derivatives of $\underline{U}_S^*(\mu_1)$ with respect to μ_1 : (We write $\bar{c}$ instead of $\bar{c}(\mu_1)$ to simplify the notation.)

$$\begin{aligned}\frac{d\underline{U}_S^*(\mu_1)}{d\mu_1} &= v_S'\left[-1 - \frac{-v_R(1)(v_R(1)+v_S(\bar{c})) - v_R(1)(1-\bar{c})v_S'}{(v_R(1)+v_S(\bar{c}))^2}\frac{d\bar{c}}{d\mu_1}\right] + v_S' \\ &= \frac{v_R(1)w(1)w(\bar{c})}{v_R(\mu_1)(1-\bar{c})(v_R(1)+v_S(\bar{c}))} \quad (>0),\end{aligned}$$

and

$$\frac{d^2\underline{U}_S^*(\mu_1)}{d(\mu_1)^2} = \frac{v_R(1)w(1)}{\big[v_R(\mu_1)(1-\bar{c})(v_R(1)+v_S(\bar{c}))\big]^2}\Big[(v_R'+v_S')\frac{d\bar{c}}{d\mu_1}v_R(\mu_1)(1-\bar{c})(v_R(1)+v_S(\bar{c}))$$

$$-w(\bar{c})v_R'(1-\bar{c})(v_R(1)+v_S(\bar{c}))+w(\bar{c})v_R(\mu_1)\frac{d\bar{c}}{d\mu_1}(v_R(1)+v_S(\bar{c}))-w(\bar{c})v_R(\mu_1)(1-\bar{c})v_S'\frac{d\bar{c}}{d\mu_1}\Big]$$

$$= \frac{v_R(1)w(1)}{\big[v_R(\mu_1)(1-\bar{c})(v_R(1)+v_S(\bar{c}))\big]^2}\Big[(v_R'+v_S')\frac{w(\bar{c})(v_R(1)+v_S(\bar{c}))}{v_S'}(v_R(1)+v_S(\bar{c}))$$

$$-w(\bar{c})v_R'(1-\bar{c})(v_R(1)+v_S(\bar{c}))+w(\bar{c})\frac{w(\bar{c})(v_R(1)+v_S(\bar{c}))^2}{v_S'(1-\bar{c})}-w(\bar{c})^2(v_R(1)+v_S(\bar{c}))\Big]$$

$$= \frac{v_R(1)w(1)w(\bar{c})(v_R(1)+v_S(\bar{c}))^2}{\big[v_R(\mu_1)(1-\bar{c})(v_R(1)+v_S(\bar{c}))\big]^2}\cdot\Big[\frac{v_R'+v_S'}{v_S'}+\frac{w(\bar{c})}{v_S'(1-\bar{c})}-1\Big]$$

$$= \frac{v_R(1)w(1)(v_R(1)+v_S(\bar{c}))^3}{\big[v_R(\mu_1)(1-\bar{c})(v_R(1)+v_S(\bar{c}))\big]^2}\cdot\frac{w(\bar{c})}{v_S'(1-\bar{c})} \quad (<0).$$

Thus, $\underline{U}_S^*(\mu_1)$ is indeed increasing and concave in Case (ii).

Based on the above results, the following lemma gives a complete description of the shape of $\underline{U}_S^*(\cdot)$, as well as the smallest concave function everywhere above it, named $\overline{U}_S^{RS}(\cdot)$:

Lemma 6. $\overline{U}_S^{RS}(\mu) = w(1)\mu$ for $\mu \in [0, \underline{\mu})$, and $\overline{U}_S^{RS}(\mu) = \underline{U}_S^*(\mu)$ for $\mu \in [\underline{\mu}, 1]$.

Proof of the lemma. We observes that $\underline{U}_S^*(\mu)$ is discontinuous at $\mu = \underline{\mu}$. Specifically, as μ decreases to $\underline{\mu}$, $\bar{c}(\mu)$ will decrease to 0, which means $\underline{U}_S^*(\mu)$ will converge to $v_S(\underline{\mu}) = v_S(0) + v_S'\underline{\mu} = w(1)\underline{\mu}$; while $\underline{U}_S^*(\mu) = 0$ for $\mu \in [0, \underline{\mu})$. As μ decreases to $\underline{\mu}$, the right-hand limit of $\frac{d\underline{U}_S^*(\mu)}{d\mu}$ is

$$\lim_{\mu \searrow \underline{\mu}} \frac{d\underline{U}_S^*(\mu)}{d\mu} = \frac{v_R(1)w(1)w(0)}{v_R(\underline{\mu})(v_R(1)+v_S(0))} = w(1).$$

Because for $\mu \in [0, \underline{\mu})$, $\underline{U}_S^*(\mu)$ is always below the line segment connecting $(0,0)$ and $(\underline{\mu}, \underline{U}_S^*(\underline{\mu}))$, whose slope is also equal to $w(1)$, we conclude that this line segment and $\underline{U}_S^*(\mu)$ in Case (ii) constitute a concave function.

On the other hand, because $\underline{U}_S^*(\mu)$ is continuous at $\mu = \hat{\mu}(\bar{c}^*)$, to prove that $\underline{U}_S^*(\mu)$ is already concave over $[\underline{\mu}, 1]$, we only need to show that the left-hand limit of $\frac{d\underline{U}_S^*(\mu)}{d\mu}$ is (weakly) larger than its right-hand limit at $\mu = \hat{\mu}(\bar{c}^*)$. Suppose not, then in the neighborhood of $\hat{\mu}(\bar{c}^*)$ we can find some $\mu' < \hat{\mu}(\bar{c}^*)$ such that $\underline{U}_S^*(\mu')$ is strictly above the line we get in Case (i). However, this is impossible because there are more binding constraints (i.e., Receiver's "individual rationality" constraint) in Case (ii) than in Case (i). □

□

A.4 Proof of Theorem 5

Proof. The ranking of Sender's private information (as well as the public information) follows directly from Section 3.2 and 4.1. As for the total information, we only need to compare the BP solution and the RS solution, since the RW solution eventually fully reveals the state. Recall that s_1 denote the public signal and s_2 denote the private signal for Sender.

When $\mu_0 < \underline{\mu}$, (s_1, s_2) fully reveals the state in the RS solution, while the BP solution only partially discloses the state.

When $\underline{\mu} \leq \mu_0 < \hat{\mu}_1(\bar{c}^*)$, (s_1, s_2) in the RS solution splits μ_0 to $[0, \bar{c}] \cup \{1\}$ according to λ_2^{RS} with parameter $\bar{c}$ satisfying $\mu_0 = \hat{\mu}_1(\bar{c})$; while the BP solution splits μ_0 to $\{0, \frac{-v_R(0)}{v_R'}\}$ with $\lambda_1^{BP}(0) = \frac{-v_R(\mu_0)}{-v_R(0)}$, where $\frac{-v_R(0)}{v_R'} > \bar{c}$. Thus, to prove that λ_2^{RS} is a mean-preserving spread of λ_1^{BP} (and thus λ_2^{RS} is more Blackwell-informative than λ_1^{BP}), it suffices to show that $\lambda_2^{RS}(0) > \lambda_1^{BP}(0)$:

$$\begin{aligned}
\lambda_2^{RS}(0) =& \frac{v_R'(1-\mu_0)}{-w(0) + (v_R(1) + v_S(\bar{c}))(-w(0))^{\frac{v_S'}{v_S'+v_R'}}(-w(\bar{c}))^{-\frac{v_S'}{v_S'+v_R'}}} \\
=& \frac{v_R'(1-\mu_0)}{-w(0) + (v_R(1) + v_S(\bar{c}))\frac{-w(0)v_R(1)}{-v_R(\mu_0)(v_R(1)+v_S(\bar{c}))}} \\
=& \frac{-v_R(\mu_0)}{-w(0)} > \frac{-v_R(\mu_0)}{-v_R(0)} = \lambda_1^{BP}(0).
\end{aligned}$$

When $\hat{\mu}_1(\bar{c}^*) \leq \mu_0 < \frac{-v_R(0)}{v_R'}$, (s_1, s_2) in the RS solution splits μ_0 to $[0, \bar{c}^*] \cup \{1\}$

according to λ_2^{RS} with parameter $\bar{c}^*$, thus, we have

$$\lambda_2^{RS}(0) = \frac{v'_R(1-\mu_0)}{-w(0) + (v_R(1) + v_S(\bar{c}^*))(-w(0))^{\frac{v'_S}{v'_S+v'_R}}(-w(\bar{c}^*))^{-\frac{v'_S}{v'_S+v'_R}}},$$

which is linear in μ_0, and strictly positive. On the other hand, the BP solution splits μ_0 to $\{0, \frac{-v_R(0)}{v'_R}\}$ with $\lambda_1^{BP}(0) = \frac{-v_R(\mu_0)}{-v_R(0)}$, which is also linear in μ_0. Because we have $\lambda_2^{RS}(0) > \lambda_1^{BP}(0)$ for $\mu_0 = \hat{\mu}_1(\bar{c}^*)$ (which is already proved in the previous case) and for $\mu_0 = \frac{-v_R(0)}{v'_R}$ (where $\lambda_1^{BP}(0) = 0$), we conclude that $\lambda_2^{RS}(0) > \lambda_1^{BP}(0)$ for all $\mu_0 \in [\hat{\mu}_1(\bar{c}^*), \frac{-v_R(0)}{v'_R})$. Since $\frac{-v_R(0)}{v'_R} > \bar{c}^*$, λ_2^{RS} is more Blackwell-informative than λ_1^{BP}.

When $\mu_0 \geq \frac{-v_R(0)}{v'_R}$, (s_1, s_2) in the RS solution splits μ_0 to $[0, \bar{c}^*] \cup \{1\}$ according to λ_2^{RS} with parameter $\bar{c}^*$; while the BP solution reveals no information. □

A.5 Proof of Theorem 6

Proof. For each regime $r \in \{BP, RW, RS\}$, let $\overline{U}_R^r(\mu_0)$, $\overline{U}_S^r(\mu_0)$ and $\overline{TW}^r(\mu_0)$ denote Receiver's expected payoff, Sender's expected payoff, and the total welfare at the ex ante stage, respectively.

(1) BP solution: When $\mu_0 \in [0, \frac{-v_R(0)}{v'_R})$, we have $\overline{U}_R^{BP}(\mu_0) = 0$, $\overline{U}_S^{BP}(\mu_0) = \overline{TW}^{BP}(\mu_0) = \mu_0\big(\frac{v_S(0)}{-v_R(0)}v_R(1) + v_S(1)\big)$.

When $\mu_0 \in [\frac{-v_R(0)}{v'_R}, 1]$, we have $\overline{U}_R^{BP}(\mu_0) = v_R(0) + \mu_0 v'_R$, $\overline{U}_S^{BP}(\mu_0) = v_S(0) + \mu_0 v'_S$, and $\overline{TW}^{BP}(\mu_0) = w(0) + \mu_0(v'_R + v'_S)$.

(2) RW solution: When $\mu_0 \in [0, \underline{\mu})$, we have $\overline{U}_R^{RW}(\mu_0) = 0$, $\overline{U}_S^{RW}(\mu_0) = \overline{TW}^{RW}(\mu_0) = \mu_0 w(1) > \overline{U}_S^{BP}(\mu_0) = \overline{TW}^{BP}(\mu_0)$.

When $\mu_0 \in [\underline{\mu}, 1]$, we have $\overline{U}_R^{RW}(\mu_0) = -v_S(0) + \mu_0\big(v_S(0) + v_R(1)\big)$, $\overline{U}_S^{RW}(\mu_0) = v_S(0) + \mu_0 v'_S \geq \overline{U}_S^{BP}(\mu_0)$, and $\overline{TW}^{RW}(\mu_0) = \mu_0 w(1)$.

In fact, the RW solution implements the efficient outcome where (s_1, s_2) fully reveals the state and Receiver chooses $a = 0$ (or 1) in state 0 (or 1).

(3) RS solution: $\overline{U}_R^{RS}(\mu_0) = 0$ for $\mu_0 \in [0, \hat{\mu}_1(\bar{c}^*))$, and $\overline{U}_R^{RS}(\mu_0) = v_R(1) - (1-\mu_0)\frac{v_R(1)}{1-\hat{\mu}_1(\bar{c}^*)}$ for $\mu_0 \in [\hat{\mu}_1(\bar{c}^*), 1]$. Notice that $\underline{\mu} < \hat{\mu}_1(\bar{c}^*) < \frac{-v_R(0)}{v'_R}$, and immediately we have $\overline{U}_R^{RW}(\mu_0) \geq \overline{U}_R^{RS}(\mu_0) \geq \overline{U}_R^{BP}(\mu_0)$ for any $\mu_0 \in [0, 1]$.

Because the RW solution is always a candidate solution in the RS model, we have $\overline{U}_S^{RS}(\mu_0) \geq \overline{U}_S^{RW}(\mu_0)$. It follows that $\overline{U}_S^{RS}(\mu_0) \geq \overline{U}_S^{RW}(\mu_0) \geq \overline{U}_S^{BP}(\mu_0)$ for any $\mu_0 \in [0,1]$.

Since $\overline{U}_R^{RS}(\mu_0) \geq \overline{U}_R^{BP}(\mu_0)$ and $\overline{U}_S^{RS}(\mu_0) \geq \overline{U}_S^{BP}(\mu_0)$, we have $\overline{TW}^{RS}(\mu_0) \geq \overline{TW}^{BP}(\mu_0)$ for any $\mu_0 \in [0,1]$. On the other hand, because (s_1, s_2) in the RS solution only partially reveals the state (when $\mu_0 > \underline{\mu}$), there is some probability that the state is 1 but Receiver implements $a = 0$, which would induce efficiency loss; while the RW solution always induces the efficient outcome. Thus, we have $\overline{TW}^{RW}(\mu_0) \geq \overline{TW}^{RS}(\mu_0)$ for any $\mu_0 \in [0,1]$. □

References

Azarmsa, E., and L. W. Cong (2020): "Persuasion in relationship finance," *Journal of Financial Economics*, 138(3), 818–837.

Börgers, T. (2015): *An Introduction to the Theory of Mechanism Design.* Oxford University Press.

Brooks, B., and S. Du (2021): "On the Structure of Informationally Robust Optimal Auctions," *Working Paper.*

Condorelli, D., and B. Szentes (2020): "Information Design in the Holdup Problem," *Journal of Political Economy*, 128, 681 – 709.

Deb, R., and A.-K. Roesler (2021): "Multi-Dimensional Screening: Buyer-Optimal Learning and Informational Robustness," *Working Paper.*

Ding, W., and X. Zhu (2019): "Strategic Disclosure, Primary Market Uncertainty, and Informed Trading," *Working Paper.*

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

Laffont, J.-J., and D. Martimort (2002): *The Theory of Incentives: The Principal-Agent Model.* Princeton University Press.

Mensch, J. (2020): "Screening Inattentive Agents," *Available at SSRN: http://dx.doi.org/10.2139/ssrn.3405884.*

Pavan, A., and J. Tirole (2021): "Expectation Conformity in Strategic Cognition," *Working Paper.*

Roesler, A.-K., and B. Szentes (2017): "Buyer-Optimal Learning and Monopoly Pricing," *American Economic Review*, 107(7), 2072–80.

Szydlowski, M. (2021): "Optimal Financing and Disclosure," *Management Science*, 67(1), 436–454.