# Bayesian Persuasion under Partial Commitment*

Daehong Min[†a]

[a]*Division of Social Science, New York University Abu Dhabi*[‡]

March 14, 2020

**Abstract**

The "full commitment" assumption in the Bayesian persuasion literature might not always hold: the sender might be tempted to deviate from an information structure to which he commits. To incorporate this possibility, I study a model in which the sender's commitment to an information structure binds with an *exogenously* given probability less than one. This gives rise to a novel mode of communication which is between cheap talk and Bayesian persuasion. I focus on the welfare implication of partial commitment. Not surprisingly, the sender's *ex ante* payoff *weakly* increases as the commitment probability increases. Then I show that, in the uniform-quadratic setting, both the sender and the receiver are *strictly* better off as the mode of communication changes from cheap talk through communication under partial commitment to Bayesian persuasion. This Pareto-ranking holds for any level of the conflict of interests between two players. The sender's opportunistic behavior under partial commitment undermines Pareto-optimal welfare achievable under Bayesian persuasion. The sender under partial commitment can improve upon the highest welfare under cheap talk via two contrasting classes of information structures: (1) one which involves truth-telling or (2) one which involves a pure noise-generating process.

*JEL Classification*: D60, D82, D83

*Keywords*: Communication, Bayesian Persuasion, Cheap Talk, Imperfect Commitment

*This paper is mainly based on a part of my Ph.D. dissertation written at the University of Arizona. I am grateful for comments by Andreas Blume, Martin Dufwenberg, Asaf Plan, Stanley Reynolds, Maria Goltsman, Yong-Gwan Kim, Lucas Siga, John Wooders, and seminar participants at the 27th International Conference on Game Theory at Stony Brook, Games 2016 (the 5th World Congress of Game Theory Society), and 2019 Koreas Allied Economic Associations Annual Meeting.

†Corresponding Author: daehong.min@nyu.edu

‡Present Address: New York University Abu Dhabi, A5/1129, Abu Dhabi, Saadiyat Island, United Arab Emirates

# 1 Introduction

The Bayesian persuasion literature assumes that the sender's commitment to an information structure is always binding. That is, once the sender chooses an information structure, the sender should truthfully report the messages realized from that chosen information structure to the receiver. For example, consider a pharmaceutical company (the sender) seeking the approval of a newly developed drug from the FDA (the receiver). The drug approval procedure requires the company to conduct a drug experiment and report data from it. The pharmaceutical company can choose a design for the drug experiment (an information structure). However, by law, the company needs to "truthfully" report data to the FDA: it should report all data (messages) from the experiment as they are; hiding or falsifying data are legally prohibited. If law enforcement is perfect, the pharmaceutical company in this example can be thought of as the sender whose commitment is fully binding.

However, it is not hard to find situations in which the sender's commitment is not binding and compromised. The story of Dr. Robert Fiddes[1] is a good example. Dr. Fiddes had been an outstanding figure in the field of drug experiments: he had conducted more than 200 experiments that lead to the approval decisions by the FDA. However, it turned out that he manipulated data from his study by falsifying data that are against the approval and even creating fake data in favor of the study. A surprising fact is that the FDA could not detect his misconduct, and Dr. Fiddes knew this beforehand. Dr. Fiddes said that he would not be caught unless his former employee blew the whistle. He knew that the FDA's monitoring system is not good enough to detect his misconduct. Simply, he knew that his commitment to truthfully disclosing data is not binding: he knew that he could report data at his discretion without being caught.

This short anecdote above demonstrates the possibility of imperfect commitment. I incorporate this possibility by relaxing the full-commitment assumption. Namely, I study a model of partial commitment in which the sender's commitment to an information structure is binding with a probability *less than one*. In the model, there are two players, the sender and the receiver. The sender moves first and chooses an information structure that governs what messages are realized contingent on the unknown true state. Then, with an *exogenously* given probability, the sender learns that his commitment is *not* binding (as Dr. Fiddes did). If this event occurs, the sender can choose the message he wants to send to the receiver. With the other probability, the sender learns that his commitment is binding (as the other "honest"

[1] See an article in the New York times, 'A Doctor's Drug Trials Turn Into Fraud' by By Kurt Eichenwald and Gina Kolata (https://www.nytimes.com/1999/05/17/business/a-doctor-s-drug-trials-turn-into-fraud.html) or Beach (2001)

drug experimenters believed the FDA's monitoring is austere enough to detect their misconduct). If this event occurs, the sender truthfully reports messages from the previously chosen information structure to the receiver. The receiver knows the probability that the sender's commitment is binding *but* cannot tell whether the message she observes comes from the sender whose commitment is binding or not. Finally, the receiver makes an inference on the true state based on the message she observes and takes an action that affects both players' payoffs.[2]

In this paper, the focus is mainly on the welfare implication of partial commitment: Are the sender and the receiver better off or worse off as the commitment probability increases? The sender under partial commitment can send messages at his discretion with a probability greater than zero not as in the full-commitment environment (Bayesian persuasion). Note that as the commitment probability decreases, the sender is more likely to be free from his commitment and use that "freedom" to favor himself as Dr. Fiddes did. Hence, it might be tempting to say that the sender is better off as the commitment probability decreases. However, in fact, the sender is *weakly better off* as the commitment probability *increases*. Simply, having the chance to commit cannot "hurt" the sender. Note that the sender with a higher commitment probability can always reproduce an outcome he could obtain with a lower commitment probability. For example, under no commitment, the sender and the receiver communicate as in the cheap-talk models; the sender under full commitment can simply reproduce any outcome under cheap talk by committing himself to a cheap-talk strategy[3] which induces an outcome under cheap talk. Thus the sender cannot be worse off by having higher commitment probabilities.

Then, I study this model of partial commitment with the well-known uniform-quadratic setting in Crawford and Sobel (1982). That is, I assume that the state space and the common prior over the state space are the unit interval, $[0, 1]$, and the uniform distribution, respectively. In addition, both the sender's and receiver's preferences are represented by quadratic loss functions; the sender's preference differs from the receiver's by a positive bias. In this uniform-quadratic setting, I show that the *ex ante* payoffs of both players are *strictly* improved as they move from the no-commitment through the partial-commitment to the full-commitment environments.

I first compare welfare under full commitment with that under partial commitment. I note that there exists the unique equilibrium outcome under full commitment: the sender

[2]One may think of the partial commitment model as a model lying between the cheap-talk and the Bayesian persuasion models: in one extreme case that the commitment probability is zero, the model collapses to the cheap-talk model; in the other extreme case, the model collapses to the Bayesian persuasion model.

[3]Note that, as we will see later, any cheap-talk strategy is mathematically equivalent to an information structure.

commits to a full information structure that tells the exact true state, and the receiver learns the exact true state, which is a Pareto-optimal outcome. Then, I show that this outcome is not attainable under partial commitment. Once the sender has a chance to send messages at his discretion (even with an arbitrarily small probability), the sender is not "free" from the Incentive Compatibility constraints (henceforth IC constraints) that must be satisfied at an equilibrium as in the cheap-talk models. Then it is immediate that the unique equilibrium outcome under full commitment cannot be achieved under partial commitment since that unique outcome is not incentive compatible for the sender: when the sender's commitment is not binding, the sender would be better off by sending a message which exaggerates the true state by the amount of his bias rather than telling the truth.

Then I show that the sender under partial commitment can improve upon the highest welfare under no commitment (or cheap talk) via two classes of information structures which have contrasting features: (1) for the high biases, the information structures which involves truth-telling and (2) for the low biases, the information structures which involves the pure noise-generating process in Blume, Board, and Kawamura (2007). Basically, I study each subgame after the sender chooses one of these information structures at the start of the game. Then, I show that, in each subgame, it is possible to construct an equilibrium with a strictly higher welfare level than the highest welfare under no commitment. These two classes of information structures are sufficient to establish this welfare improvement result for almost all exogenously given parameters, i.e., for any commitment probability and almost all biases which can be arbitrarily large.

The first class of information structures generates the message which is equal to the true state at low states and a single message at all the other high states. I call an information structure falling into this class a *semi-full information structure* as it reveals the true state at the low states but does not at the high states. When the sender's bias is high, there is no information transmission under no commitment. However, under partial commitment, if the sender chooses this type of information structure, there must be information transmission with a probability greater than zero as this information structure will be "on" with the commitment probability and conveys information according to its structure. Thus the receiver can take better actions which are also better for the sender compared to one single action induced under no commitment. Hence both players are strictly better off compared to the no-commitment environment.

The information structures which belong to the second class are convex combinations of two information structures, (1) the pure noise-generating process and (2) an information structure which is equivalent to the *front-loading strategy* in Blume et al. (2007). Note that I can obtain a set of such information structures by varying the weight for the convex combination from zero to one. I call an information structure in this class a *convexified information structure*. I focus on the subgames after the sender chooses one of these information structures. I first note that, in these subgames, it is possible to replicate the equilibrium in Blume et al. (2007) which Pareto-dominates all the equilibria under cheap talk. Then I show that, for almost all low biases and any commitment probability, this way of improving welfare is possible by adjusting the weight on the pure noise-generating process. Furthermore, if the commitment probability is higher than a certain level, there exists a convexified information structure which achieves the efficiency bound established in the mediation model by Goltsman, Hörner, Pavlov, and Squintani (2009) and that in Blume et al. (2007).

## 1.1 Related Literature

This paper is related to two strands of communication literature, cheap talk and Bayesian persuasion. Following the work by Kamenica and Gentzkow (2011), there have been many papers that study variations of the Bayesian persuasion model as summarized well in Kamenica (2019). For example, Gentzkow and Kamenica (2017) study the Bayesian persuasion with multiple senders; Wang (2013) and Alonso and Câmara (2016b) study the persuasion problems with multiple receivers.[4] While most of the studies assume that the sender's commitment is always binding, this paper relaxes this full-commitment assumption. Recently, Fréchette, Lizzeri, and Perego (2019), Lipnowski, Ravid, and Shishkin (2019), and Nguyen and Tan (2019a) study a similar setting in which the sender's commitment is limited. Fréchette et al. (2019) experimentally study how limited commitment interacts with the verifiability of messages. Their way to model partial commitment is identical to this paper. However, they consider a more straightforward setting to obtain clear predictions for the experimental study.[5] Lipnowski et al. (2019) also model partial commitment in a similar way as in this paper; the sender can replace the message from an information structure with an exogenously given probability which can depend on the true state; in this paper, however, that probability is independent of the true state. Furthermore, assuming that the sender has a state-independent preference, they provide the characterization of the equilibrium and show that the receiver can be better off under the sender's limited commitment power. While I

[4]Besides these works, there are many other extensions such as Alonso and Câmara (2016a), Jain (2018), Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), and Kolotilin (2018). Furthermore, there is growing literature on various applications of Bayesian persuasion. For these, refer to Arieli and Babichenko (2019), Bizzotto, Rüdiger, and Vigier (2020), Luo and Rozenas (2018), Martinez-Roa (2020), Min (2017), Nguyen and Tan (2019b), and Yoder (2019).

[5]While they consider a model with the sender whose preferences is state-independent and binary states, the model in this paper is not limited to these restrictions.

allow the sender's preference to be state-dependent, I provide the welfare implication of partial commitment without the characterization of the equilibrium. Lastly, Nguyen and Tan (2019a) incorporate partial commitment in their model by assuming that the sender bears some manipulation costs.[6] In this paper, the sender obtains the manipulation chance at no cost with an exogenously given probability.

In the cheap talk literature, there have been papers studying various ways to improve communication and welfare. Goltsman et al. (2009) show that a mediation scheme can improve welfare and also establish the efficiency bound for it. Blume et al. (2007) show that adding a small amount of noise in the communication process can improve welfare; furthermore, they show how the efficiency bound in Golsman et al. (2009) can be attained with a simple noise. The main finding in this paper adds an observation to this strand of papers: the sender's partial commitment can improve welfare even when the probability that commitment is binding is arbitrarily small and the sender's bias is arbitrarily large. This paper closely relates to Blume et al. (2007) and a result in this paper resembles that in Kim and Pogach (2014). On the one hand, when the sender commits to the degenerate convexified information structure, the subsequent game can be thought of as the noise model in Blume et al. (2007). I show that the welfare improvement is possible by the sender's commitment to the convexified information structures based on results in Blume et al. (2007). On the other hand, the equilibrium structure constructed in the subgame after the semi-full information structure resembles that of so-called type 1 equilibria in the honest model by Kim and Pogach (2014).

Lastly, this paper is also related to the literature on imperfect commitment in games. For example, Bagwell (1995) shows that a strategic benefit from commitment immediately disappears if that commitment is not observable with a very small probability; Fershtman and Kalai (1997) demonstrates that commitment via delegation is beneficial even if the delegation is not observable. In this paper, the sender's commitment is not perfect in the sense that he can take other "actions" than those to which he committed if he realizes that his commitment is not binding.

The remainder of this paper is structured as follows. In section 2, I formally introduce the model and provide a (weak) comparative statics of the model. In section 3, I focus on the uniform-quadratic case and present the welfare results. Section 4 concludes this paper.

[6] Hence, their model collapses to either Bayesian persuasion or cheap talk depending on whether the cost of manipulation is 0 or $\infty$.

## 2 The Model

There are two players: a sender (S) and a receiver (R). Two players have a common prior distribution, $F(\omega)$, on the state space, $\Omega$. Assume that $\Omega$ is compact and $F(\omega)$ has a density which is positive everywhere.

The game proceeds as follows. The game starts with S's commitment to an information structure. An information structure, $\{\pi(\cdot|\omega)\}_{\omega\in\Omega}$, is a set of conditional distributions over a message realization space, $M$, where $M$ is a set containing $\Omega$ (i.e., $\Omega \subset M$). I denote an information structure by $\pi \in \Pi$, where $\Pi$ is the set of all information structures. Importantly, S commits to a $\pi$ *before* he learns the realization of the true state. Once S chooses a $\pi$, he publicly announces his commitment to that $\pi$.

After his commitment to a $\pi$ is announced, S *privately* learns the true state and also if his commitment is binding or not. With probability $\alpha > 0$, S learns that his commitment is binding. I denote S whose commitment is binding by $S_b$. Since his commitment is binding, $S_b$ plays a simple role as a "truthful mediator" between $\pi$ and R: if the true state is $\omega'$, he observes a message realization, $m \in M$, from the $\pi(m|\omega')$ and truthfully reports that message to R.

With the other probability, $1-\alpha > 0$, S learns that his commitment is not binding. I denote S whose commitment is not binding by $S_c$. Since his commitment is not binding, $S_c$ does not need to be a truthful mediator as $S_b$ does. Instead, $S_c$ can send any $m \in M$ at his discretion: he can discard a message realization, $m \in M$, from a $\pi$ and report any $m \in M$ that he wants.[7] I denote a strategy of $S_c$ by $\sigma(\cdot) : \Pi \times \Omega \to \Delta M$. Especially, in a specific subgame after S's commitment to a $\pi'$, $S_c$'s strategy in that subgame is reduced to a mapping from $\Omega$ to $\Delta M$. That is, in that subgame, if the true state is $\omega'$, $S_c$ who employs $\sigma'(\cdot)$ reports $m \in M$ according to his discretion, $\sigma'(m|\omega';\pi')$, where $\sigma'(m|\omega';\pi')$ is a distribution over $M$ given that the true state is $\omega'$.

To distinguish the messages sent by $S_c$ from those sent by $S_b$, I refer the messages sent by $S_c$ to the *cheap-talk messages*; I refer the message sent by $S_b$ to merely messages unless there is possible confusion. I also assume that the value of $\alpha$ (the probability that S's commitment is binding) is common knowledge for both S and R.

Finally, R observes a message, $m \in M$, from S (either $S_c$ or $S_b$). R is aware of the possibility that the message is sent by $S_c$ and knows the probability of it, $1-\alpha$. However, R cannot tell if the message is sent by $S_c$ or $S_b$ (or putting it differently, she cannot distinguish

[7] Note that this is a restriction. I restrict the set of messages that $S_c$ can use to be $M$ on which an information structure is constructed. However, one can easily imagine a situation where $S_c$ can send a message that is not in $M$. In that case, that message will immediately tell R that she is hearing from $S_c$.

the cheap-talk messages from the messages generated from a $\pi$.) Given a message from S and the value of $\alpha > 0$, R makes an inference on the true state, $\omega \in \Omega$. Then, R takes an action, $a \in A$, based on her updated beliefs, where $A$ is any set that contains $\Omega$. Once R takes an action, $a \in A$, both players' payoffs are realized, and the game ends.

Both players' payoffs depend on both the action taken by R, $a \in A$, and the true state, $\omega \in \Omega$. I denote S's and R's payoff functions by $U^S(a, \omega)$ and $U^R(a, \omega)$, respectively.

## 2.1 Equilibrium

I use the Perfect Bayesian Equilibrium as the solution concept. (Henceforth, an equilibrium means a Perfect Bayesian Equilibrium). In the model, S chooses a $\pi \in \Pi$ and publicly announces his choice. Hence, a $\pi$ chosen by S induces a subgame. I denote a subgame after S's commitment to a $\pi'$ by $\Gamma(\pi')$. I start by defining an equilibrium in a subgame and then define an equilibrium of the entire game.

Consider a subgame after a $\pi'$, $\Gamma(\pi')$. An equilibrium in $\Gamma(\pi')$ consists of $S_c$'s strategy, R's strategy, and the set of posteriors of R. Note that I do not need to consider $S_b$'s strategy in $\Gamma(\pi')$ since $S_b$ behaves according to his commitment to $\pi'$. In $\Gamma(\pi')$, I denote $S_c$'s strategy by $\sigma(\pi') := \{\sigma(m|\omega, \pi')\}_{\omega \in \Omega}$, R's strategy by $a(m; \pi')$, and the set of R's posterior by $\mathcal{M}(\sigma(\pi'), \pi')$. Formally, an equilibrium in $\Gamma(\pi')$ is a triple, $(\sigma^*(\pi'), a^*(m; \pi'), \mathcal{M}(\pi', \sigma^*(\pi')))$ such that

(1) for any posterior $\mu(\omega|m; \pi') \in \mathcal{M}(\pi', \sigma^*(\pi'))$,

$$\mu(\omega|m; \pi') = \frac{(\alpha\pi'(m|\omega) + (1-\alpha)\sigma^*(m|\omega, \pi'))f(\omega)}{\int_\Omega (\alpha\pi'(m|\omega) + (1-\alpha)\sigma^*(m|\omega, \pi'))f(\omega)d\omega},$$

(2) given any posterior, $\mu(\omega|m; \pi') \in \mathcal{M}(\cdot)$, $a^*(m; \pi') \in \arg\max_{a \in A} \int_\Omega U^R(a, \omega)d\mu(\omega|m; \pi')$,

(3) for all $\omega \in \Omega$, if $\sigma^*(m'|\omega, \pi') > 0$ for a $m'$, $m' \in \arg\max_{m \in M} U^S(a^*(m; \pi'), \omega)$.

In words, at an equilibrium in $\Gamma(\pi')$, R needs to form a posterior according to the Bayes' rule, R's strategy should maximize R's expected payoff given any posteriors, and, finally, $S_c$'s strategy should be a best response to R's strategy. Now I define an equilibrium of the entire game as follows:

An equilibrium of the model is a quadruple, $(\pi^*, \sigma^*(\pi), a^*(m; \pi), \mathcal{M}(\pi, \sigma^*(\pi)))$, such that

(1) for any $\pi \in \Pi$, $(\sigma^*(\pi), a^*(m; \pi), \mathcal{M}(\pi, \sigma^*(\pi))$ constitutes an equilibrium in $\Gamma(\pi)$,

(2) $\pi^*$ is a maximizer of S's ex ante payoff: $\pi^* \in \underset{\pi\in\Pi}{\arg\max}\ EU^S(\pi)$, where

$$\begin{aligned} EU^S(\pi) &= \alpha \left[ \int_\Omega \left( \int_M U^S(a^*(m,\pi),\omega) d\pi(m|\omega) \right) dF(\omega) \right] \\ &\quad +(1-\alpha) \left[ \int_\Omega \left( \int_M U^S(a^*(m,\pi),\omega) d\sigma^*(m|\omega,\pi) \right) dF(\omega) \right]. \end{aligned}$$

The *ex ante* payoff that S can get by choosing a $\pi$ is simply a weighted average of the expected payoffs of $S_c$ and $S_b$ at an equilibrium of $\Gamma(\pi)$: the term after $\alpha$ is $S_b$'s expected payoff and the term after $1-\alpha$ is $S_c$'s expected payoff. At an equilibrium of the model, by choosing $\pi^*$, S decides to be in the subgame, $\Gamma(\pi^*)$, that gives him the highest payoff.

## 2.2 (Weak) Comparative Statics

Now I state the first proposition which is the weak comparative statics analysis on S's welfare.

**Proposition 1.** *The Sender is **weakly** better off as the commitment probability, $\alpha$, increases: if there is an equilibrium of the model with $\alpha_0$, the sender can always reproduce that equilibrium outcome in the model with $\alpha > \alpha_0$.*

*Proof.* Consider the model with a commitment probability, $\alpha_0$. Suppose that there is an equilibrium and denote the equilibrium information structure by $\pi^*$. Now consider the equilibrium path. After the sender commits to $\pi^*$, the sender is in $\Gamma(\pi^*)$, the subgame after $\pi^*$. Denote $S_c$'s equilibrium strategy in $\Gamma(\pi^*)$ by $\sigma^*(\pi^*) := \{\sigma^*(m|\omega;\pi^*)\}_{\omega\in\Omega}$. In addition, denote the receiver's equilibrium strategy in $\Gamma(\pi^*)$ by $a^*(m;\pi^*)$. The triple, $(\pi^*, \sigma^*(\pi^*), a^*(m;\pi^*))$, determines the equilibrium outcome of the model with $\alpha_0$. At this equilibrium, any $m \in M$ such that $\pi^*(m|\omega) > 0$ or $\sigma^*(m|\omega) > 0$ induces the receiver's posterior belief according to Bayes' Rule:

$$\mu_0(\omega|m) = \frac{(\alpha_0\pi^*(m|\omega) + (1-\alpha_0)\sigma^*(m|\omega))\, dF(\omega)}{\int_\Omega (\alpha_0\pi^*(m|\omega) + (1-\alpha_0)\sigma^*(m|\omega))\, dF(\omega)}.$$

Then, $a^*(m;\pi^*)$ is a maximizer of $\int_\Omega U^R(a,\omega) d\mu_0(\omega|m)$. Finally, $\sigma^*(\pi^*)$ is incentive compatible for $S_c$. That is, for all $\omega \in \Omega$, if $\sigma^*(m'|\omega) > 0$,

$$m' \in \underset{m\in M}{\arg\max}\ U^S(a^*(m;\pi^*),\omega).$$

Now consider the model with a higher commitment probability, $\alpha > \alpha_0$. Consider the following information structure, $\pi^c = \{\pi^c(m|\omega)\}_{\omega\in\Omega}$: for all $\omega \in \Omega$,

$$\pi^c(m|\omega) := \frac{\alpha_0}{\alpha}\pi^*(m|\omega) + (1-\frac{\alpha_0}{\alpha})\sigma^*(m|\omega;\pi^*),$$

which is a convex combination of the equilibrium $\pi^*$ and $\sigma^*(\pi^*)$ in the model with $\alpha_0$.

Suppose that the sender commits to $\pi^c$. The sender's commitment to $\pi^c$ induces $\Gamma(\pi^c)$, the subgame after $\pi^c$. In $\Gamma(\pi^c)$, let $S_c$ employs $\sigma^*(\pi^*)$. Then, the triple, $(\pi^c, \sigma^*(\pi^*), a^*(m; \pi^*))$, constitutes an equilibrium in $\Gamma(\pi^c)$. To see this, consider the receiver's incentive compatibility first. Given $(\pi^c, \sigma^*(\pi^*))$, the receiver's posterior belief about $\omega$ conditional on observing $m \in M$ is

$$\begin{aligned}
\mu(\omega|m) &= \frac{(\alpha\pi^c(m|\omega) + (1-\alpha)\sigma^*(m|\omega))\, dF(\omega)}{\int_\Omega (\alpha\pi^c(m|\omega) + (1-\alpha)\sigma^*(m|\omega))\, dF(\omega)} \\
&= \frac{\big(\alpha(\frac{\alpha_0}{\alpha}\pi^*(m|\omega) + (1-\frac{\alpha_0}{\alpha})\sigma^*(m|\omega)) + (1-\alpha)\sigma^*(m|\omega)\big)\, dF(\omega)}{\int_\Omega \big(\alpha(\frac{\alpha_0}{\alpha}\pi^*(m|\omega) + (1-\frac{\alpha_0}{\alpha})\sigma^*(m|\omega)) + (1-\alpha)\sigma^*(m|\omega)\big)\, dF(\omega)} \\
&= \mu_0(\omega|m).
\end{aligned}$$

Thus, the receiver's inference about $\omega$ after observing $m \in M$ in $\Gamma(\pi^c)$ is exactly same as her inference at the equilibrium of the model with $\alpha_0$. Accordingly, the receiver's strategy, $a^*(m; \pi^*)$, is optimal given $(\pi^c, \sigma^*(\pi^*))$.

Now only thing left to check is $S_c$'s incentive compatibility. Since the receiver employs $a^*(m; \pi^*)$ which is the same strategy in the equilibrium of the model with $\alpha_0$, $\sigma^*(\pi^*)$ is also a best response of $S_c$ in $\Gamma(\pi^c)$ as it is at the equilibrium of the model with $\alpha_0$.

Hence, the triple, $(\pi^c, \sigma^*(\pi^*), a^*(m; \pi^*))$, indeed constitutes an equilibrium in $\Gamma(\pi^c)$ in the model with the higher commitment probability, $\alpha > \alpha_0$. The sender can achieve this outcome by committing to $\pi^c$ and employing $\sigma^*(\pi^*)$ when the sender learns his commitment is not binding.

Note that $(\pi^*, \sigma^*(\pi^*), a^*(m; \pi^*))$ and $(\pi^c, \sigma^*(\pi^*), a^*(m; \pi^*))$ yield the same outcome: Both strategy profiles yield the same set of posteriors, the same distribution over the set of posteriors, and the same distribution over equilibrium actions. Accordingly, both strategy profiles yield the same *ex ante* payoffs for the sender.

The sender may be better off but cannot be worse off with a higher commitment probability: given a commitment probability, the sender can always obtain an *ex ante* payoffs that he could have obtained at any equilibrium of the model with any lower commitment probability. $\square$

Proposition 1 simply states that the *ex ante* payoff of S *cannot* decrease as the commitment probability, $\alpha$, increases. In the model with $\alpha$, S can achieve any equilibrium outcome in models with any $\alpha_0 < \alpha$ if he is willing to do so.

Note that this weak comparative statics result does not necessarily hold for R. Consider two models with two different commitment probabilities, $\alpha$ and $\alpha_0$, where $\alpha > \alpha_0$. If S can achieve a strictly better equilibrium outcome in the model with $\alpha$ than with $\alpha_0$, S will choose that outcome in the model with $\alpha$. However, it is unclear if R is also better off at that outcome chosen by S in the model with $\alpha$ compared to the outcome that S would choose in the model with $\alpha_0$.

However, there are cases that the weak comparative statics result also holds for R.

**Corollary 1.** *If $U^S(a,\omega) = -(\omega + b - a)^2$ and $U^R(a,\omega) = -(\omega - a)^2$, both the sender and the receiver are **weakly** better off as the commitment probability, $\alpha$, increases.*

*Proof.* One can easily show that $EU^S = EU^R - b^2$ should hold at any equilibrium when $U^S(a,\omega) = -(\omega + b - a)^2$ and $U^R(a,\omega) = -(\omega - a)^2$, where $EU^i$ is $i$'s *ex ante* payoff for $i = S, R$. By Proposition 1, $EU^S$ is weakly increasing in $\alpha$. Hence, $EU^R = EU^s + b^2$ is also weakly increasing in $\alpha$. □

It is a well-known fact that the *ex ante* interests of S and R are perfectly aligned if both players' preferences are represented by a quadratic loss function.[8] Corollary 1 is an immediate consequence of this well-known fact and Proposition 1.

# 3 Welfare under Partial Commitment in the Uniform-Quadratic Case

In this section, I study the partial-commitment model in a specific setup, the uniform-quadratic framework, which is widely used in the cheap-talk literature. From now on, I assume that the state space, $\Omega$, is the unit interval, $[0,1]$, and the common prior on $\Omega = [0,1]$ is the uniform distribution. Furthermore, I assume that both players' preferences are represented by a quadratic loss function: $U^S(a,\omega;b) = -(\omega + b - a)^2$ and $U^R(\omega,a) = -(\omega - a)^2$, where $b > 0$ is the term for S's bias. Thus, while R wants to match her action to the true state, S wants R to take the action that is equal to the true state plus his bias.

I focus on the welfare comparison of three different modes of communication, cheap talk, communication under partial commitment, and Bayesian persuasion. Here welfare is merely the sum of the ex ante payoffs of S and R. Then, I show that both S and R are *strictly* better off as they move from cheap talk to communication under partial commitment to Bayesian persuasion, which is true for almost all biases of S (even when $b$ is arbitrarily high).

Finally, note that $EU^S = EU^R - b^2$ holds at any equilibrium[9] in this uniform-quadratic case. Thus, when discussing welfare, it is enough to look at S's *ex ante* payoff. Furthermore, I

[8]It is worth mentioning that there is more general condition (called condition (M) in Crawford and Sobel (1982)) which ensures the perfect alignment of both players' *ex ante* preferences. Under the condition (M), S *ex ante* prefers an equilibrium to the other if and only if R *ex ante* prefers the former to the latter. Hence Corollary 1 also holds under the condition (M).

[9]Putting it differently, the equation above holds as far as the receiver maximizes her expected payoff given a posterior.

will interchangeably use 'more (less) information transmission' and 'the welfare improvement (loss)' since $EU^R$ increases in the amount of information transmitted and $EU^S$ is simply $EU^R - b^2$.

## 3.1 Benchmarks

Note that if $\alpha = 0$, the model is equivalent to the leading example in Crawford and Sobel (1982), and if $\alpha = 1$, the model corresponds to a special case of the Bayesian persuasion model by Kamenica and Gentzkow (2011). In this subsection, I summarize results in these two cases and set the benchmarks used to compare welfare in later sections.

### 3.1.1 Cheap Talk: Case when $\alpha = 0$

First, suppose $\alpha = 0$. Then, all messages that R observes are sent by $S_c$. Hence, S's commitment to a $\pi$ does not play any role; what matters is $S_c$'s strategy. Furthermore, R is aware of this. Thus, we return to the uniform-quadratic case in Crawford and Sobel (1982) (henceforth, CS model). I summarize the results from the CS model and define the "best" equilibrium of it below.

I define the Best Equilibrium in the CS model (henceforth, BECS) as the equilibrium with the highest *ex ante* payoff for S (thus the highest ex ante payoff for R). Given $b > 0$, there are multiple equilibria in the CS model. Every equilibrium is characterized by a partition of the state space. An equilibrium partition reflects the amount of information transmitted to R: the finer the partition is, the more information R obtains, the higher the R's ex ante payoff is. Thus the equilibrium with the finest partition gives the highest payoff to R, so does to S. Hence, given $b > 0$, the BECS is the equilibrium with the finest partition of the state space, $[0, 1]$.

Note that S's bias, $b > 0$, determines the finest partition attainable as an equilibrium outcome. For any $b \geq 1/4$, the finest equilibrium partition has only one element in it: the pooling equilibrium outcome is the unique one. Accordingly, the pooling equilibrium is the BECS in this case. For $b < 1/4$, the maximum number of elements in an equilibrium partition is equal to $[-\frac{1}{2} + \frac{1}{2}\left(1 + \frac{2}{b}\right)^{1/2}]$, where $[I]$ is the smallest integer greater than or equal to $I$. Thus, the BECS is the equilibrium with the partition which has $[-\frac{1}{2} + \frac{1}{2}\left(1 + \frac{2}{b}\right)^{1/2}]$ numbers of elements in it. For example, if $b = 1/8$, we have $[-\frac{1}{2} + \frac{1}{2}(17)^{1/2}] = [1.56...] = 2$. Thus given $b = 1/8$, the BECS is the equilibrium which partitions $[0, 1]$ into two intervals. In this manner, I can obtain the BECS for each $b > 0$. Then I can collect the BECS for every possible value of $b > 0$. This collection of the BECS will be the benchmarks used in later sections.

### 3.1.2 Bayesian Persuasion: Case when $\alpha = 1$

Now suppose $\alpha = 1$. Then, R knows that any $m \in M$ she observes is sent by $S_b$. Thus, the model collapses to the Bayesian Persuasion model by Kamenica and Gentzkow (2011). In this case, an information structure, $\pi$, to which S commits fully determines an outcome since there is no "interruption" by $S_c$. Thus, at an equilibrium, S commits to an information structure that maximizes his *ex ante* payoff among all other information structures. In this case, there exists the unique equilibrium outcome which is independent of $b > 0$. Furthermore, the unique equilibrium outcome is also a first-best one in the uniform-quadratic case. Remark 1 below formally states this.

**Remark 1.** *Suppose $\alpha = 1$. Then, for any $b > 0$, there exists the unique equilibrium outcome under which the sender fully reveals the true states. Thus, at any equilibrium, the sender's expected payoff is $-b^2$ and the receiver's expected payoff is 0. Furthermore, this outcome is Pareto efficient.*

The detailed proof of Remark 1 can be found in the Appendix. In the proof I simply use the fact that $EU^S = EU^R - b^2$ should hold at any equilibrium. As mentioned before, $EU^R$ increases as R obtains more information: R can take better action as she knows more about the true state. Then S can achieve the maximum *ex ante* payoff by giving R full information since $EU^S = EU^R - b^2$.[10] Thus S chooses an information structure, $\pi$, that reveals the exact true state to R. An example of such an information structure is the one that directly tells the exact true state as follows: for all $\omega \in [0, 1]$,

$$\begin{aligned} \pi^f(m|\omega) &= 1 \quad \text{if } \; m = \omega, \\ &= 0 \quad \text{if } \; m \neq \omega. \end{aligned}$$

Given this information structure, R learns the exact true state and takes the action that is equal to the true state, $a^*(m = \omega) = \omega$. Finally, R gets the payoff of 0, and S gets the payoff of $-b^2$. In addition, this outcome is a first-best one since the sum of both players' payoffs cannot be higher than $-b^2$.

It is not hard to imagine other equilibria in which S chooses an information structure that reveals the truth in a different way than the example above. However, any equilibrium information structure should be the one that fully reveals the true state. Thus, all equilibrium outcomes should be the same as that in the example above. Hence Remark 1 establishes the unique benchmark that will be used in later sections.

[10] Here $b$ is exogenously given. Thus S cannot alter the value of $b$. Then the only way S can change his ex ante payoff is to change the amount of information transmitted to R.

### 3.2 The Welfare Loss under Partial Commitment

In this subsection, I compare welfare under partial commitment with that under full commitment. In short, both S and R are *strictly* worse off as the communication environment changes from the full-commitment one to the partial-commitment one.

Remark 1 says that there exists the unique equilibrium outcome under full commitment. Remark 2 below tells us that the unique outcome in Remark 1 cannot be an equilibrium outcome if $\alpha < 1$ (i.e. if S has an opportunity to engage in cheap talk with a probability greater than 0).

**Remark 2.** *If $\alpha < 1$, there is no equilibrium in which the sender reveals full information to the receiver.*

*Proof.* Consider the model with $\alpha \in (0, 1)$. Now suppose that there is an equilibrium which achieves the unique outcome under full commitment. At that equilibrium (if it does exist), S fully reveals the true state. Thus $S_b$ must commit to an information structure that tells the true state. In addition to this, $S_c$ must tell the truth as well. Thus R optimally takes the action equal to the true state. Consider two equilibrium actions, $a' = \omega'$ and $a'' = \omega' + b$, at $\omega'$ and $\omega' + b$, respectively. In this hypothetical equilibrium, $S_c$ at state $\omega'$ sends the message which induces $a' = \omega'$, and he obtains the payoff of $-b^2$. However, $S_c$ can obtain the payoff of $0 > -b^2$ by deviating to the message which induces his most preferred action, $a'' = \omega' + b$. Thus, this cannot be an equilibrium. □

Remark 2 says that R should be uncertain about some states at any equilibrium in the model with $\alpha < 1$ as the full information outcome is not feasible. Recall that the more uncertain R is about the state, the lower $EU^R$ is. Then Remark 2 implies that, at any equilibrium under partial commitment, $EU^R$ should be strictly less than zero which is the payoff R would obtain under full commitment. Because $EU^S = EU^R - b^2$ should hold at any equilibrium, the loss in S's ex ante payoff immediately follows. This establishes the welfare loss under partial commitment: both S and R are *strictly* worse off as S has an opportunity to engage in cheap talk with any positive probability (even when that is arbitrarily small).

In the model with $\alpha < 1$, S can send any $m \in M$ at his discretion with a positive probability, which can be thought of as a data manipulation behavior.[11] Hence, it is intuitively clear that R is worse off when S can manipulate data with a positive probability. However, why does S also suffer from having a data manipulation chance though it seems S can benefit from it?

[11] Recall that an information structure, $\pi$, can be thought of as an experiment; the messages generated by $\pi$ can be considered as data from an experiment.

The mathematical formulation of two different cases clarifies the answer to the question above. The main difference between the full-commitment and partial-commitment cases is whether $S_c$ is "dormant" or not. If $\alpha = 1$, $S_c$ is dormant. In other words, S has only one future self, $S_b$, who will exactly behave as S wants at the start of the game. Hence, S solves a simple maximization problem given that R chooses an action that maximizes her expected payoff. That is, S's problem is as follows:

$$\max_{\pi \in \Pi} \quad EU^S = EU^R - b^2.$$

Now suppose that $\alpha < 1$. Since S assumes R's incentive compatibility as in $\alpha = 1$ case, his objective function is the same as before, $EU^S = EU^R - b^2$. However, since $S_c$ is "active" in this case, S should take into account how the other future self, $S_c$, will behave. As $S_c$ learns that his commitment is not binding and also the true state, $S_c$ will maximize the *ex post* payoff rather than the *ex ante* payoff that S wants to maximize. At an equilibrium, S rationally expects this. Thus S solves the same maximization problem above *but* with $S_c$'s IC behavior:

$$\max_{\pi \in \Pi} \; EU^S = EU^R - b^2 \;\; \text{s.t. } S_c\text{'s IC behavior.}$$

Hence, any small decrease in $\alpha$ from 1 immediately brings some non-trivial constraints into S's maximization problem. At the first glance, a data manipulation chance seems to make S more "free," *but* it actually imposes more restrictions on him.

By looking at two different maximization problems, it is obvious that S cannot be better off as $\alpha$ decreases, which we already saw in Proposition 1. However, Proposition 1 does not tell us whether S is *strictly* worse off or not; if the constraint in the second problem is slack, S can sustain the same payoff under full commitment. Remark 2 shows that the constraint is not slack in the quadratic-uniform case. Thus, S is *strictly* worse off, and so is R as $\alpha$ is drifting away from 1.

### 3.3 The Welfare Improvement under Partial Commitment

In this subsection, I show that both S and R obtain *strictly* higher *ex ante* payoffs under partial commitment than under no commitment. This welfare improvement result holds for any pair of exogenous parameters, $(\alpha, b) \in (0, 1) \times (0, \infty)$.

To establish this result, I hinge on two classes of information structures: one consisting of the *convexified information structures* and the other consisting of the *semi-full information structures*. I use the former for the low bias cases, $b \in (0, 1/2)$, and the latter for the high bias cases, $b \in (1/2, \infty)$. More precisely, I consider subgames after S announces his commitment to the information structures which belong to one of two classes. Then I construct a Preto-dominant equilibrium (relative to the BECS) in these subgames. In the model, S moves first and chooses which subgame he wants to be in later by choosing an information structure. Thus it suffices to show that there is a Pareto-dominant equilibrium in a subgame induced by an information structure.[12]

### 3.3.1 The Welfare Improvement via Convexified Information Structures

A convexified information structure is a convex combination of two different information structures, a *noise generating process* and a specific strategy called the *front-loading strategy* in Blume, Board, and Kawamura (2007). I start with the degenerate convexified information structure which puts weight 1 on the noise generating process defined below.

**Definition 1.** *The noise generating process is defined as* $\pi^n := \{\pi(m|\omega)\}_{\omega\in[0,1]}$, *where for each* $\omega \in [0,1]$,

$$\pi(m|\omega) \textit{ is the uniform distribution on } [0,1].$$

Note that any message generated by $\pi^n$ is a simple "noise." Thus $\pi^n$ does not convey any information about $\omega \in [0,1]$.

Now suppose that S publicly announces his commitment to $\pi^n$ at the start of the whole game. S's announcement induces a subgame in which $S_b$ sends the noise according to $\pi^n$ with probability $\alpha > 0$. Then $S_c$ in $\Gamma(\pi^n)$ faces R who believes that any $m \in M$ she observes is the noise sent by $S_b$ with probability $\alpha > 0$. Thus this subgame, $\Gamma(\pi^n)$, coincides with the noise model by Blume et al. (2007); they consider the model in which the sender sends a message to the receiver, but that message is replaced by a meaningless noise with an exogenously given probability, $\epsilon > 0$. Note that the commitment probability, $\alpha > 0$, in $\Gamma(\pi^n)$ is equivalent to the noise level, $\epsilon > 0$, in Blume et al. (2007).

Blume et al. (2007) construct an equilibrium in which the sender uses the so-called *front-loading strategy.* I call such an equilibrium a front-loading equilibrium. As $\Gamma(\pi^n)$ is equivalent to the noise model by Blume et al. (2007), I can replicate the front-loading equilibrium in $\Gamma(\pi^n)$ by having $S_c$ employ the front-loading strategy; the front-loading strategy partitions the state space, $[0,1]$, into $N$-interval as follows: Given a $b \in (\frac{1}{2N^2}, \frac{1}{2(N-1)2})$ for $N = 2, 3, ..,$

$$S_c \text{ randomizes uniformly on } M\backslash\{m_2, m_3, ..., m_N\} \text{ if } \omega \in [0, \omega_1],$$

[12] There are two possible cases: the subgame in which the Pareto-dominating equilibrium is constructed is on the equilibrium path or off the equilibrium path of the equilibrium of the model. In the first case, I can directly argue the welfare improvement. In the second case, I can still argue the welfare improvement: if that subgame is off the equilibrium path, it implies that there exists another subgame which is on the equilibrium path *and* gives the same or a higher *ex ante* payoff to the sender (and the receiver) as the subgame in which I constructed an equilibrium.

$$S_c \text{ sends } m_i \text{ for } i = 2, ..., N \text{ if } \omega \in (\omega_{i-1}, \omega_i], \text{ where } \omega_N = 1 \text{ and } M = [0,1].$$

Note that the number of elements in the partition, $N$, depends on $b > 0$. For example, $b = 1/6 \in (\frac{1}{2 \cdot 2^2}, \frac{1}{2 \cdot (2-1)^2})$. Thus $N = 2$, and the front-loading strategy partitions $[0,1]$ into two intervals. Given $b > 0$, I denote the *front-loading* strategy by $\sigma_{fl}(b)$. Lastly, one can clearly see why it is called the front-loading strategy; almost all messages, $[0,1] \backslash \{m_2, m_3, ...m_N\}$, are loaded at the "front" interval, $[0, \omega_1]$.

At a front-loading equilibrium, R's inference is as follows. First, when R observes $m \in [0,1] \backslash \{m_1, m_2, ...m_N\}$, R believes that, with probability $\alpha$, $m$ is the noise sent by $S_b$, and, with probability $1-\alpha$, $m$ is sent by $S_c$'s types who are in $[0, \omega_1]$. Thus, R's optimal action is a weighted average of two midpoints, $\frac{1}{2}$ of $[0,1]$ and $\frac{\omega_1}{2}$ of $[0, \omega_1]$.[13] Secondly, when R observes $m_i \in \{m_2, m_3, ..., m_N\}$, R believes that $m_i$ is sent by $S_c$'s types who are in $(\omega_{i-1}, \omega_i]$ with probability 1 since the event that $m_i$ is generated by $\pi^n$ (or sent by $S_b$) has measure zero. Thus, R's optimal action is merely the midpoint of the interval, $(\omega_{i-1}, \omega_i]$.

It is worth noting that a "meaningful" front-loading equilibrium exists only if $0 < b < 1/2$; if $b \geq 1/2$, the only incentive compatible $\sigma_{fl}(b)$ for $S_c$ is the one with $N = 1$ which is equivalent to $\pi^n$; it does not partition the state space (i.e., at every state, $S_c$ uniformly randomizes on $M = [0,1]$). Consequently, a front-loading equilibrium when $b \geq 1/2$ is the same as the 'babbling' equilibrium. Hence, it is immediate that, if $b \geq 1/2$, the front-loading equilibrium yields the same outcome as a BECS in the no-commitment case.

However, if $b < 1/2$, a front-loading equilibrium Pareto-dominates BECS under no commitment, as shown in Blume et al. (2007). More importantly, this welfare improvement is only possible if the noise level, $\epsilon \in (0,1)$, is less than an upper bound, $\bar{\epsilon}(b)$, where the upper bound, $\bar{\epsilon}(b)$, is a function of $b$, and its definition is relegated to the Appendix for expositional convenience. This result can be interpreted in the framework of this paper (by treating the noise level, $\epsilon$, as the commitment probability, $\alpha$). The following remark states the interpretation of their result in our context.

**Remark 3.** *If $\alpha \in (0, \bar{\epsilon}(b))$ and $b \in (0, 1/2) \backslash \{b = \frac{1}{2N^2}$ for all integers $N \geq 2\}$, the Sender's partial commitment to $\pi^n$ **strictly** improves welfare upon BECS under no commitment. Otherwise, the Sender's partial commitment to $\pi^n$ does not.*[14]

[13] The R's optimal action, $a^*(m)$, with the appropriate weights is as follows:

$$a^*(m) = \frac{\alpha(1-0)\frac{(1+0)}{2} + (1-\alpha)(\omega_1 - 0)\frac{\omega+0}{2}}{\alpha(1-0) + (1-\alpha)(\omega_1 - 0)} = \frac{\alpha\frac{1}{2} + (1-\alpha)\omega_1\frac{\omega_1}{2}}{\alpha + (1-\alpha)\omega_1}.$$

[14] For the detailed proof, refer to the proof for Proposition 9 in Blume et al. (2007).

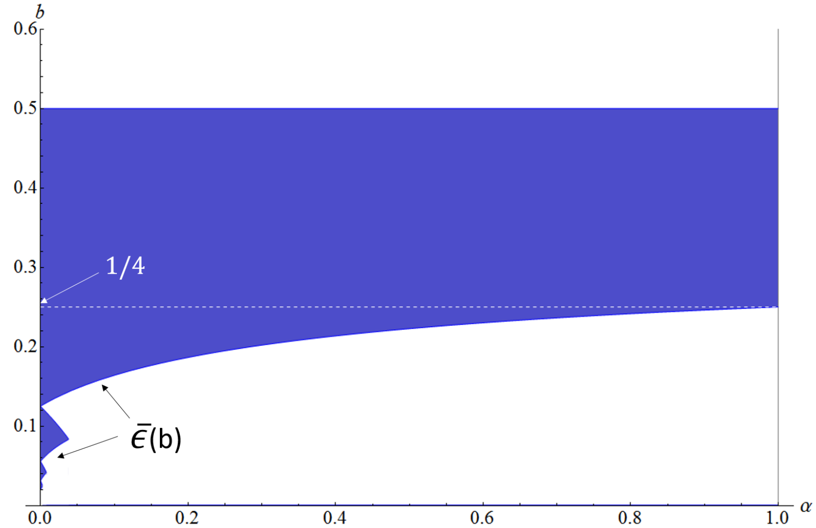


Figure 1: Summary of Remark 3

Figure 1 summarizes Remark 3. The shaded area in Figure 1 indicates the set of parameter pairs, $(\alpha, b)$, in which S's partial commitment to $\pi^n$ results in a welfare improvement. The unshaded area indicates the set of parameter pairs in which S's partial commitment to $\pi^n$ cannot improve welfare by constructing a front-loading equilibrium.

S's commitment to $\pi^n$ leads both S and R to the subgame in which the noise is generated with a positive probability as in the noise model by Blume et al. (2007). As discussed in their paper, the presence of noise has two effects on welfare. The presence of noise "loosens" $S_c$'s incentive compatibility constraints. Due to this loosening effect[15], we can construct a front-loading equilibrium that has either *finer* or *evener* partition than a BECS, which implies *potentially* more information transmission (thus a welfare improvement) at a front-loading equilibrium than at a BECS. However, the noise by itself hinders the information transmission, which implies there are some negative effects that may offset the positive effect above. The condition, $\alpha < \bar{\epsilon}(b)$, in Remark 3 ensures that the positive effect of the noise dominates the negative one. However, if $\alpha \geq \bar{\epsilon}(b)$, the negative effect dominates the positive effect. Thus we cannot have the welfare improvement result in the unshaded area in Figure 1. Lastly, recall that when $b \geq 1/2$, the front-loading equilibrium is equivalent to the BECS in the no-commitment case. Thus we have the unshaded area above $b = 1/2$ in Figure 1.

However, I can use a non-degenerated convexified information structure to show the welfare improvement is possible in the unshaded area under $b = 1/2$ in Figure 1. I start by

[15]In their paper, they refer to this effect as the *strategic* effect.

constructing a non-degenerated convexified information structure. Given a parameter pair, $(\alpha, b) \in (0,1) \times (0,\infty)$, choose a $p \in (0,1)$. Then, consider a "hypothetical" subgame in which R believes that $S_b$ sends the message according to $\pi^n$ with probability $\alpha \cdot p$ (instead of $\alpha$); denote this subgame by $\mathcal{H}(\pi^n, \alpha p)$; construct the front-loading equilibrium in $\mathcal{H}(\pi^n, \alpha p)$. Denote the front-loading strategy in that equilibrium by $\sigma_{fl}(b, \alpha p)$. Lastly, to get a non-degenerate convexified information structure, I simply make a convex combination of $\pi^n$ and $\sigma_{fl}(b, \alpha p)$ with weight $p \in (0,1)$ on $\pi^n$. I formally provide the definition of the non-degenerate convexified information structure below.

**Definition 2.** *A non-degenerate convexified information structure is denoted by $\pi^c(p)$, where, for some $p \in (0,1)$,*

$$\pi^c(p) := p\{\pi^n(m|\omega)\}_{\omega\in[0,1]} + (1-p)\{\sigma_{fl}(m|\omega; b, \alpha p)\}_{\omega\in[0,1]},$$

*where $\sigma_{fl}(b, \alpha p)$ is the $S_c$'s strategy at the front-loading equilibrium in the subgame in which the Receiver believes that $S_b$ sends $m$ according to $\pi^n$ with probability $\alpha p$.*

Note that $\pi^c(p)$ is a valid information structure since, given any $\omega \in [0,1]$,

$$\begin{aligned}\int_M \pi^c(m|\omega; \alpha p)dm &= p\int_0^1 \pi^n(m|\omega)dm + (1-p)\int_0^1 \sigma_{fl}(m|\omega; \alpha p)dm \\ &= p \cdot 1 + (1-p)\cdot 1 = 1\end{aligned}$$

A simple "interpretation" of $\pi^c(p)$ is as follows: with probability $p$, $\pi^c(p)$ generates $m$ according to $\pi^n$, and with probability $1-p$, it generates $m$ according to $\sigma_{fl}(m|\omega; \alpha p)$.

Note that $\pi^c(p)$ depends on the value of $p \in (0,1)$: for each $p \in (0,1)$, we have a different $\pi^c(p)$. It is also worth mentioning that, if $b \geq 1/2$, the construction above gives us $\pi^n$. To see this, recall that the only incentive compatible $\sigma_{fl}(b)$ for $b \geq 1/2$ is the one having no partition. Thus, given $b \geq 1/2$, $\sigma_{fl}(m|\omega; b, \alpha p) = \pi^n$ for any $p \in (0,1)$; then $\pi^c(p) = p\pi^n + (1-p)\pi^n = \pi^n$.

Now suppose that $b < 1/2$, and S announces his commitment to $\pi^c(p)$ with a certain value of $p \in (0,1)$. In $\Gamma(\pi^c(p))$, I can easily construct an equilibrium by having $S_c$ employ $\sigma_{fl}(b, \alpha p)$ which is built in $\pi^c(p)$. If $S_c$ plays $\sigma_{fl}(b, \alpha p)$ in $\Gamma(\pi^c(p))$, R believes that $S_b$ sends $m$ according to $\pi^c(p)$ with probability $\alpha$ and $S_c$ does so according to $\sigma_{fl}(b, \alpha p)$ with probability $1-\alpha$. That is, R's posterior belief is as follows:

$$\mu(\omega|m) = \frac{\alpha\pi^c(m|\omega; p) + (1-\alpha)\sigma_{fl}(m|\omega; b, \alpha p)}{\int_0^1 \left(\alpha\pi^c(m|\omega; p) + (1-\alpha)\sigma_{fl}(m|\omega; b, \alpha p)\right) d\omega}$$

$$= \frac{\alpha(p\pi^n(m|\omega) + (1-p)\sigma_{fl}(m|\omega; b, \alpha p)) + (1-\alpha)\sigma_{fl}(m|\omega; b, \alpha p)}{\int_0^1 (\alpha\pi^c(m|\omega; p) + (1-\alpha)\sigma_{fl}(m|\omega; b, \alpha p))\, d\omega}$$
$$= \frac{\alpha p\pi^n(m|\omega; p) + (1-\alpha p)\sigma_{fl}(m|\omega; b, \alpha p)}{\int_0^1 (\alpha\pi^c(m|\omega; p) + (1-\alpha)\sigma_{fl}(m|\omega; b, \alpha p))\, d\omega}.$$

Note that $\mu(\omega|m)$ is exactly the same as the posterior belief at the front-loading equilibrium in $\mathcal{H}(\pi^n, \alpha p)$: R believes that $S_b$ sends $m$ according to $\pi^n$ with probability $\alpha p$ and $S_c$ does so according to $\sigma_{fl}(b, \alpha p)$ with probability $1-\alpha p$. Thus, given the R's posterior belief above, it is immediate that $S_c$'s strategy, $\sigma_{fl}(b, \alpha p)$, is incentive compatible in $\Gamma(\pi^c(p))$ because $\sigma_{fl}(b, \alpha p)$ is incentive compatible in $\mathcal{H}(\pi^n, \alpha p)$.

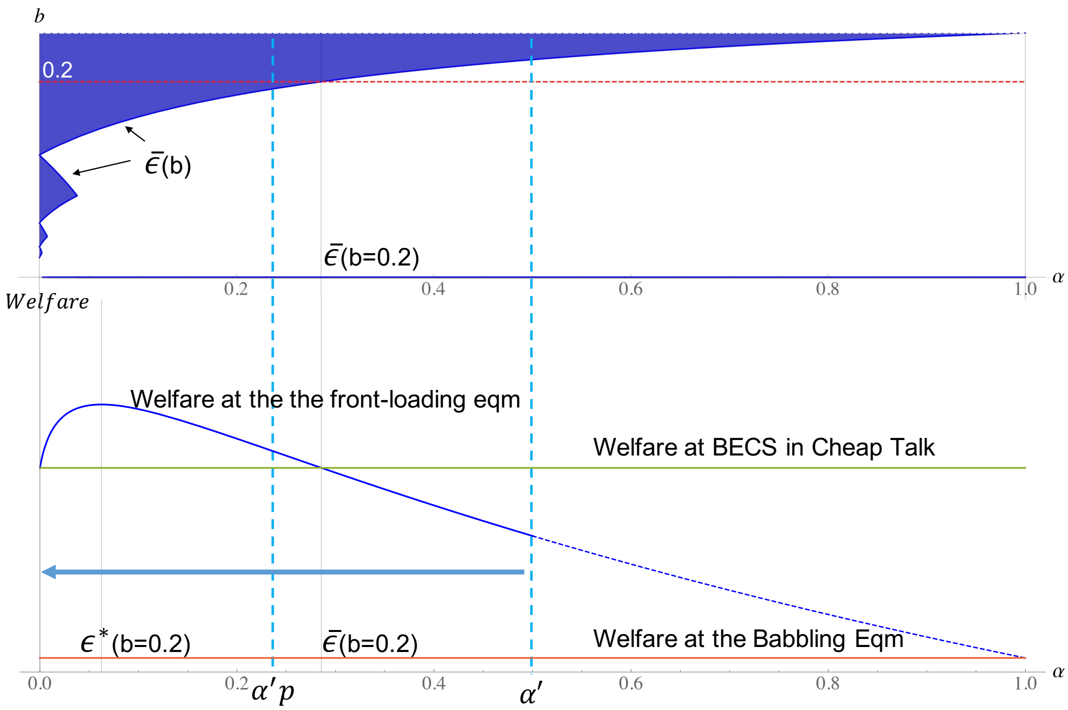


Figure 2: An Welfare Improvement via $\pi^c(p)$

With an appropriate choice of $p$, the equilibrium constructed above Pareto-dominates BECS under no commitment. Figure 2 demonstrates how to choose an appropriate $p$ for the welfare improvement via $\pi^c(p)$. At the bottom of Figure 2, I depict welfare at the front-loading equilibrium in $\Gamma(\pi^n)$ as a function of $\alpha$, given $b = 0.2$. Now suppose that $\alpha = \alpha' > \bar{\epsilon}(b = 0.2)$. Then, as mentioned above, welfare at the front-loading equilibrium in $\Gamma(\pi^n)$ is below that at BECS. Given $\alpha'$, now choose $p$ so that $\alpha' p < \bar{\epsilon}(b = 0.2)$ as shown in Figure 2. With that chosen $p$, construct $\pi^c(p)$ and the equilibrium in $\Gamma(\pi^c(p))$ as demonstrated above. Then, the outcome at the constructed equilibrium in $\Gamma(\pi^c(p))$ must be equivalent to that at the front-loading equilibrium in $\Gamma(\pi^n)$ in which the frequency of the noise generating process is equal to $\alpha' p$. Hence, the constructed equilibrium in $\Gamma(\pi^c(p))$ with that chosen $p$ must Pareto-dominate BECS as shown in Figure 2.

The non-degenerate convexified information structure gives us a way to control the frequency of the noise from $\pi^n$. Given any $\alpha > 0$, the frequency of the noise from $\pi^n$ is "anchored" to $\alpha$ in $\Gamma(\pi^n)$. However, with the non-degenerate convexified information structure, I can construct an equilibrium that has the same outcome as a front-loading equilibrium in $\Gamma(\pi^n)$ with *reduced frequencies of the noise*, any $\alpha p \in (0, \alpha)$. For example, given $\alpha = \alpha'$ in Figure 2, I can choose any welfare level which is on the solid line of welfare at the front-loading equilibrium. Lastly, it is worth mentioning that it is impossible to increase the frequency of the noise with $\pi^c(p)$; given any $\alpha$, the maximum frequency of the noise is $\alpha$.

Now note that welfare at the front-loading equilibrium is maximized at a certain level of the noise given $b > 0$. For example, in Figure 2, welfare is maximized at $\epsilon^*(b = 0.2)$. Following Blume et al. (2007), I refer to this noise level that maximizes welfare as the optimal noise level, which is denoted by $\epsilon^*(b)$. The optimal noise level is also a function of $b > 0$. For expositional convenience, the definition of $\epsilon^*(b)$ is also relegated to the Appendix as $\bar{\epsilon}(b)$. Lastly, it is worth mentioning that, given any $b > 0$, $\bar{\epsilon}(b) \geq \epsilon^*(b)$. Then now I have Lemma 1, which immediately follows with the notion of $\epsilon^*(b)$ and the fact that I can choose any welfare level among the welfare levels at the front-loading equilibria in $\Gamma(\pi^n)$ with different noise levels within $(0, \alpha)$, given $\alpha \in (0, 1)$. I refer to the equilibrium outcome with the optimal noise level, $\epsilon^*(b)$, as the optimal noise outcome.

**Lemma 1.** *If* $(\alpha, b) \in [\epsilon^*(b), 1) \times (0, 1/2) \backslash \{b = \frac{1}{2N^2}$ for all integers $N \geq 2\}$, $\pi^c(p^*)$ *can implement the optimal noise outcome, where* $p^* = \frac{\epsilon^*(b)}{\alpha}$.

*Proof.* Suppose that $\alpha \geq \epsilon^*(b)$ and $b \in (0, 1/2) \backslash \{b = \frac{1}{2N^2}$ for all integers $N \geq 2\}$. Then, there are two cases: (1) $\alpha = \epsilon^*(b)$ or (2) $\alpha > \epsilon^*(b)$. In case (1), $p^* = 1$ and the optimal noise outcome can be achieved by constructing the front-loading equilibrium in $\Gamma(\pi^n = \pi^c(p^* = 1))$. In case (2), I can choose $\pi^c(p^*)$ such that $\alpha p^* = \epsilon^*(b)$. Then, I can construct an equilibrium which has the optimal noise outcome by having $S_c$ employ $\sigma_{fl}(m|\omega; \alpha p^*)$ in $\Gamma(\pi^c(p^*))$. $\square$

Lemma 1 says that if S's commitment power is "high enough" (i.e., $\alpha \geq \epsilon^*(b)$), it is possible to achieve the optimal noise outcome. It is worth mentioning that welfare at the optimal noise outcome coincides with the highest welfare level in the mediation model by Goltsman, Hörner, Pavlov, and Squintani (2009).[16] Finally, Proposition 2 below concludes the welfare

[16] They consider a model of mediation in which the sender provides information to the mediator, who then recommends an action to the receiver. The optimal mediation scheme introduces a "noise" between the sender and receiver.

improvement via $\pi^c(p)$.

**Proposition 2.** *For any $\alpha \in (0,1)$ and $b \in (0,1/2)\backslash\{b = \frac{1}{2N^2}$ for all integers $N \geq 2\}$, the Sender's partial commitment to $\pi^c(p)$ **strictly** improves welfare upon BECS under no commitment:*

*(i) if $\alpha < \epsilon^*(b)$, the Sender's partial commitment to $\pi^c(p=1) = \pi^n$ strictly improves welfare upon BECS by achieving the front-loading equilibrium outcome,*

*(ii) if $\alpha \geq \epsilon^*(b)$, the Sender's partial commitment to $\pi^c(p^*)$ strictly improves welfare upon BECS by achieving the optimal noise outcome, where $p^* = \frac{\epsilon^*(b)}{\alpha}$.*

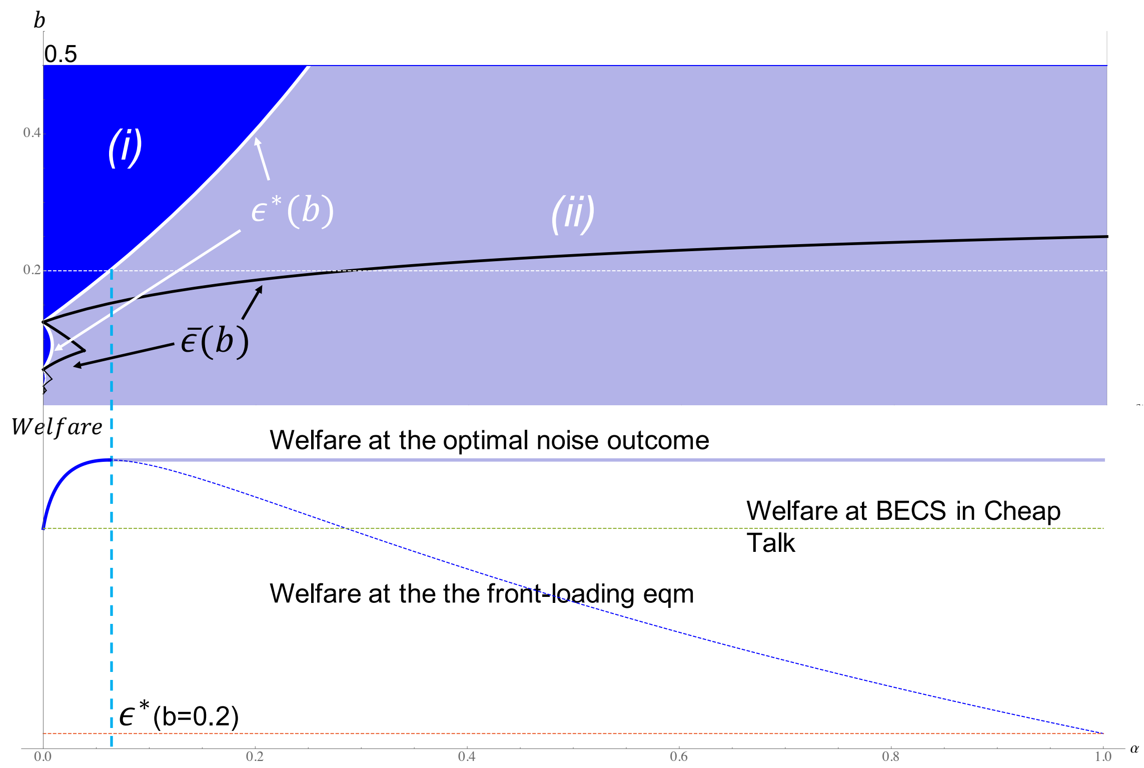


Figure 3: Summary of Proposition 2

Figure 3 summarizes Proposition 2.[17] At the bottom of Figure 3, I depict the welfare levels at different equilibrium outcomes given $b = 0.2$. If $\alpha < \epsilon^*(b = 0.2)$, there is no $\pi^c(p)$ which can implement the optimal noise outcome. However, as $\alpha < \epsilon^*(b = 0.2) < \bar{\epsilon}(b = 0.2)$, I can construct the front-loading equilibrium which has higher welfare than BECS. Furthermore, for $\alpha < \epsilon^*(b = 0.2)$, there is no benefit of reducing the noise level. Thus, given any $\alpha \in (0, \epsilon^*(b))$,

[17] I depict $\bar{\epsilon}(b)$ and $\epsilon^*(b)$ in Figure 3. As mentioned before, $\bar{\epsilon}(b) \geq \epsilon^*(b)$ for any given $b \in (0,1/2)$.

$\pi^c(p=1) = \pi^n$ is the information structure that improves welfare the most among all $\pi^c(p)$ for $p \in (0,1]$. If $\alpha \geq \epsilon^*(b=0.2)$, I can always find $\pi^c(p^*)$ which implements the optimal noise outcome given any $\alpha$. Hence, for any $\alpha \geq \epsilon^*(b=0.2)$, the highest welfare level is equal to that at the optimal noise outcome, which is depicted as a horizontal line. Thus, the solid line is the "envelope" of welfare that can be achieved with the set of $\pi^c(p)$. The top of Figure 3 simply visualizes the set of the parameter pairs, $(\alpha, b)$, with which this type of welfare improvement is possible as the shaded area in the figure. Lastly, note that a welfare improvement via $\pi^c(p)$ is not possible when $b = \frac{1}{2N^2}$ for all positive integers $N \in \mathbb{N}$. However, these cases have measure zero in the two-dimensional parameter space.

### 3.3.2 The Welfare Improvement via Semi-full Information Structures

So far, I have shown that, for the low-bias area, $(\alpha, b) \in (0,1) \times (0, 1/2)$, S's partial commitment to $\pi^c(p)$ can strictly improve welfare compared to the no-commitment case. In this section, I provide the result that a welfare improvement is also possible in the high-bias area, $(\alpha, b) \in (0,1) \times [1/2, \infty)$. Recall that (1) if $b \in [1/2, \infty)$, the only equilibrium under no commitment is the babbling equilibrium, and (2) welfare increases in the amount of information transmitted. These two facts make it easier to establish a welfare improvement result: given $(\alpha, b) \in (0,1) \times [1/2, \infty)$, I only need to find an information structure that induces a subgame having an equilibrium with a *certain degree of information transmission.* There indeed exists such an information structure which I define below:

**Definition 3.** *A semi-full information structure is defined as* $\pi^s(t) := \{\pi(m|\omega)\}$, *where given some* $t \in (0,1)$,

*(i) for* $\omega \in [0,t]$,
$$\pi(m|\omega) = 1 \text{ if } m = \omega, \quad = 0 \text{ if } m \neq \omega,$$

*(ii) for* $\omega \in (t,1]$, $\pi(m|\omega)$ *is the uniform distribution over* $(t,1] \subset M$.

In words, for each $\omega \in [0,t]$, $\pi^s(t)$ generates $m = \omega$ which reveals the true state; for every $\omega' \in (t,1]$, $\pi^s(t)$ randomly generates $m \in (t,1]$.

Now consider a subgame after S commits to $\pi^s(t)$, $\Gamma(\pi^s(t))$. Now suppose that $S_c$ employs the following strategy in $\Gamma(\pi^s(t))$:

For all $\omega \in [0,1]$, $\sigma(m|\omega)$ is the uniform distribution over $(t,1] \subset M = [0,1]$.

In words, $S_c$ randomly sends $m \in (t,1]$ at every state, $\omega \in [0,1]$. Denote $S_c$'s strategy above by $\sigma_s(t)$.

Assuming $S_c$ employs $\sigma_s(t)$ in $\Gamma(\pi^s(t))$, R's inference on $\omega$ and the corresponding optimal action in $\Gamma(\pi^s(t))$ are as follows. First, note $S_c$ only uses a part of $M$, $(t,1]$; $S_c$ does not "contaminate" the meaning of $m \in [0,t]$ which are generated by $\pi^s(t)$. Knowing this, if R observes $m \in [0,t]$, R believes that $m \in [0,t]$ is generated by $\pi^s(t)$ (or sent by $S_b$) with probability 1. Thus, when $m \in [0,t]$, R learns the true states and her optimal action is $a^*(m) = m$. Secondly, suppose R observes $m \in (t,1]$. Not as in the case of $m \in [0,t]$, $S_c$ uses $m \in (t,1]$. Thus, taking $\sigma_s(t)$ into account, R believes that, (1) with probability $\alpha$, $m$ is generated by $\pi^s(t)$, and (2) with probability $1-\alpha$, $m$ is sent by $S_c$. Then, on the one hand, R believes that, with probability $\alpha$, the true state is between $(t,1]$ since this is the information contained in $\pi^s(t)$. On the other hand, R believes that, with probability $1-\alpha$, the true state is between $[0,1]$ since $S_c$ sends $m \in (t,1]$ at every state in $[0,1]$. Accordingly, when R observes $m \in (t,1]$, her optimal action is a simple weighted average of $\frac{1+t}{2}$ and $\frac{1}{2}$ with some appropriate weight, and it is denoted by $\bar{a}(t)$.[18] As a summary, R's optimal (or IC) strategy, denoted by $a_s(m)$, is as follows:

$$\begin{aligned} a_s(m) &= m \quad \text{if } m \in [0,t], \\ &= \bar{a}(t) \;\; \text{if } m \in (t,1]. \end{aligned}$$

Given $a_s(m)$, I only need to check whether $\sigma_s(t)$ is incentive compatible for $S_c$ to argue that $(\sigma_s(t), a_s(m))$ actually constitutes an equilibrium in $\Gamma(\pi^s(t))$. Lemma 2 says that there is a set of $\pi^s(t)$ that induces subgames in which $\sigma_s(t)$ is incentive compatible.

**Lemma 2.** *For any $(\alpha, b) \in (0,1) \times [1/2, \infty)$, $(\sigma_s(t), a_s(m))$ constitutes an equilibrium in $\Gamma(\pi^s(t))$ for $\pi^s(t) \in \{\pi^s(t) | \pi^s(t) \text{ for } t \in T(\alpha,b)\}$, where $\emptyset \neq T(\alpha,b) \subset (0,1)$.*

I relegate the proof of Lemma 2 and the definition of $T(\alpha,b)$ to the Appendix again for expositional convenience. A sketch of proof is as follows. Suppose that $\bar{a}(t)$ is the highest action among any other actions in $a_s(m)$ (i.e., $\bar{a}(t) \geq t = \max_{m\in[0,t]} a_s(m)$). For $\sigma_s(t)$ to be incentive compatible, $S_c$ must not have an incentive to send $m \in [0,t]$ at every state, $\omega \in [0,1]$. As $U^S(\omega,a)$ is supermodular, it suffices to check whether $S_c$ has an incentive to send $m \in [0,t]$ at $\omega = 0$. Intuitively, as the bias is high enough (i.e., $b \geq 1/2$), $S_c$ prefers the highest action, $\bar{a}(t)$, to any other action, $a_s(m \in [0,t])$, even at the lowest state, $\omega = 0$: even at $\omega = 0$, $S_c$ wants to "exaggerate" the true state as much as possible. Hence, as far as $\bar{a}(t) \geq t$, $(\sigma_s(t), a_s(m))$ indeed constitutes an equilibrium. The condition, $\bar{a}(t) \geq t$, boils

[18] The R's optimal action, $\bar{a}(t)$, with the appropriate weights is as follows:

$$\bar{a}(t) = a^*(m \in (t,1]) = \frac{\alpha(1-t)}{\alpha(1-t) + (1-\alpha) \times 1} \left(\frac{1+t}{2}\right) + \frac{1-\alpha}{\alpha(1-t) + (1-\alpha) \times 1} \left(\frac{1}{2}\right)$$

.

down to a restriction on $t$. Thus, $(\sigma_s(t), a_s(m))$ constitutes an equilibrium in $\Gamma(\pi^s(t))$ if the "threshold" value, $t$, in $\pi^s(t)$ satisfies the restriction. Furthermore, the high bias ensures that the set of $t$ which satisfies the restriction is always non-empty.[19]

It is also worth noting that Lemma 2 holds even when $b$ is arbitrarily large and $\alpha$ is arbitrarily small. For example, $T(\alpha, b) = (0, \frac{1-\sqrt{1-\alpha}}{\alpha}]$ for any $b \in (\frac{1-\sqrt{1-\alpha}}{\alpha}, \infty)$. Note that $(0, \frac{1-\sqrt{1-\alpha}}{\alpha}]$ is nonempty for any $\alpha \in (0,1)$: as $\alpha \to 0$ (o*r* 1), $\frac{1-\sqrt{1-\alpha}}{\alpha} \to \frac{1}{2}$ (or 1).

Now, note that, at the equilibrium constructed in $\Gamma(\pi^s(t))$, there is a certain degree of information transmission: R learns the exact true states if she observes $m \in [0, t]$; she also learns that $\omega \in (t, 1]$ with a positive probability, $\alpha$, if she observes $m \in (t, 1]$. Thus, welfare at this equilibrium in $\Gamma(\pi^s(t))$ must be strictly higher than that at BECS (the babbling equilibrium) under no commitment. Proposition 3 summarizes the discussion in this subsection.

**Proposition 3.** *For any* $\alpha \in (0,1)$ *and* $b \in [1/2, \infty)$*, the Sender's partial commitment to* $\pi^s(t) \in \{\pi^s(t)|\pi^s(t)$ for $t \in T(\alpha, b)\}$ ***strictly*** *improves welfare upon BECS under no commitment.*

*Proof.* By Lemma 2, S's commitment to $\pi^s(t) \in \{\pi^s(t)|\pi^s(t)$ for $t \in T(\alpha, b)\}$ induces $\Gamma(\pi^s(t))$ in which $(\sigma_s(t), a_s(m))$ constitutes an equilibrium. At such an equilibrium, there is a certain degree of information transmission; for $b \geq 1/2$, BECS is the babbling equilibrium. Therefore, $EU^R$ is strictly higher at the equilibrium in $\Gamma(\pi^s(t))$ than BECS, so is $EU^S = EU^R - b^2$. $\square$

Lastly, it is worth mentioning the relation between the equilibrium constructed above and one of the equilibria constructed in the honesty model by Kim and Pogach (2014). Kim and Pogach (2014) consider a model of honesty where S is "honest" with a positive probability. In their model, S tells the exact true state with a positive probability. Hence, their model coincides with a subgame after S commits to a full information structure in this paper (which I do not consider). One of the equilibria in their model exhibits the same outcome structure as the equilibrium outcome in this subsection.[20] However, their equilibrium construction is qualitatively different from that in this paper: their construction requires a lower bound for $\alpha$ and an *upper* bound for $b$ while a *lower* bound for $b$ is required here.

I have considered two classes of information structures which improve welfare upon BECS when S has an opportunity to partially commit to one of these. Proposition 4 summarizes these results.

[19] While the proof only focuses on the case in which $b \geq 1/2$, the lowest value of $b$ which guarantees that $T(\alpha, b)$ is non-empty is 1/4.

[20] The equilibrium outcome in this subsection coincides with the outcome of the "simplest" Type 1 equilibrium in Kim and Pogach (2014). Furthermore, the equilibrium structure here also resembles the LSHP (low-separation-high-pooling) equilibrium in Kartik (2009).

**Proposition 4.** *For* $(\alpha, b) \in (0,1)\times(0,\infty)\backslash\{\frac{1}{2N^2}$ for all integers $N \geq 2\}$*, the Sender's partial commitment to either* $\pi^c(p = 1$ or $\frac{\epsilon^*(b)}{\alpha})$ *or* $\pi^s(t \in T(\alpha, b))$ ***strictly*** *improves welfare upon BECS under no commitment.*

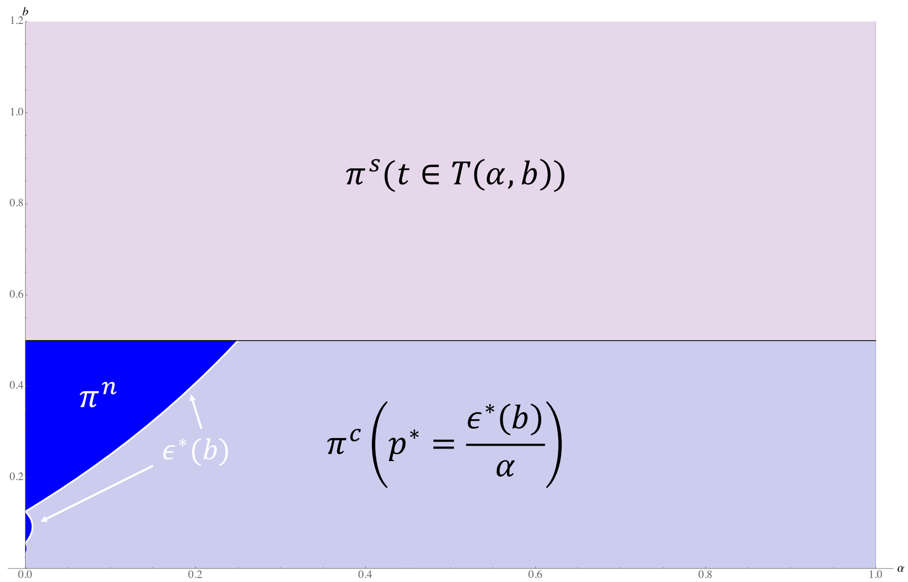


Figure 4: Proposition 4

The two classes of information structures, $\pi^c(p)$ and $\pi^s(t)$, have contrasting features: $\pi^c(p)$ involves a noise-generating process while $\pi^s(t)$ involves a full information structure. Note that these two contrasting information structures can improve welfare in two contrasting cases: $\pi^c(p)$ does for the *low* biases while $\pi^s(t)$ does for the *high* biases. How BECS changes according to the level of the bias gives us some intuition behind this result. On the one hand, when the bias is low, there is some information transmission at BECS. This information transmission requires some incentive compatibility constraints to be satisfied. To improve welfare upon such a BECS, we need to increase the amount of information transmitted. To do so, we need to "perturb" the incentive compatibility constraints so that providing more information is incentive compatible. Note that $\pi^c(p)$ plays that role by injecting the noise. Thus, $\pi^c(p)$ can achieve the welfare improvement when the bias is low. On the other hand, when the bias is high, there is no information transmission at BECS. To improve welfare, we just need some amount of information transmitted. Then, $\pi^s(t)$ which always transmits some information can achieve this along with the fact that S must communicate with R via $\pi^s(t)$ with a probability greater than 0.

Having S's commitment bind with a positive probability restricts S's ability to manipulate R's beliefs compared to the no-commitment case, which seems to only benefit R but not S.

In fact, S cannot be worse off by having his commitment bind with a positive probability as we see in Proposition1. Then, the question becomes "Can S be strictly better off compared to the no-commitment case?" and, if he can, "When and how S can be strictly better off?" In the uniform-quadratic setting, the answer to the first question is "Yes", and the answer to the second question is "Almost always via commitment to an information structure which belongs to one of two classes of information structures, $\pi^c(p)$ and $\pi^s(t)$.

The intuition from Proposition 1 is still valid in the uniform-quadratic case. That is, S obtains more "freedom" by "tying" his hands: By having the chance to commit to an information structure, S can have a strictly larger set of possible equilibria than the no-commitment case. On top of this, in that enlarged set of equilibria, there are equilibria that give strictly higher *ex ante* payoffs for him in the uniform-quadratic case. Lastly, as S is ex ante better off, so R is in this uniform-quadratic setting, which implies the welfare improvement.

# 4 Conclusion

Departing from the full-commitment assumption which is common in the Bayesian persuasion literature, this paper proposes a model of partial commitment in which the Sender's commitment is binding with a probability less than one. The model is in-between the cheap-talk (no commitment) and the Bayesian persuasion (full commitment) models, which are the benchmarks in this paper.

I first show that the Sender is *weakly* better off as the commitment probability increases: tightening a strap around his hands cannot make him worse off. When the preferences of the Sender and Receiver is *ex ante* well-aligned, the Receiver is also *weakly* better off as the commitment probability increases.

Then I study the model in a specific environment called the uniform-quadratic setting which is widely used in the cheap-talk literature. I show that welfare *strictly* increases as the mode of communication changes from cheap talk to the partial-commitment to Bayesian persuasion, which holds for any level of the Sender's bias. Interestingly, the Sender can achieve the best outcome in Blume et al. (2007) and Goltsman et al. (2009) by committing to a convexified signaling device with a high enough commitment power.

# 5 Appendix

Proof of Remark 1

*Proof.* I start by specifying S's payoff from an arbitrary information structure, $\pi$. Suppose S chooses and commits to an information structure, $\pi$. Since R knows that S's commitment is always binging, R forms the following posterior when she observes $m \in M$,

$$\mu(\omega|m) = \frac{\pi(m|\omega)f(\omega)}{\int_0^1 \pi(m|\omega)f(\omega)d\omega}.$$

Given a $\mu(\omega|m')$, R chooses an optimal action $a^*(m') = \underset{a\in A}{argmax} \int_0^1 -(\omega - a)^2 d\mu(\omega|m')$. Note that $a^*(m') = -\int_0^1 \omega d\mu(\omega|m') = -E_{\mu_{m'}}[\omega]$. Then, S's expected payoff from inducing the posterior, $\mu(\omega|m')$, (which is, in turn, equal to inducing the action, $a^*(m')$) is equal to

$$\begin{aligned} E_{\mu_{m'}}[U^S(\cdot)] &= \int_0^1 -(\omega + b - a^*(m'))^2 d\mu(\omega|m') \\ &= -E_{\mu_{m'}}[\omega^2] + \left(E_{\mu_{m'}}[\omega]\right)^2 - b^2 = -\sigma^2_{\mu_{m'}}[\omega] - b^2, \end{aligned}$$

where $-\sigma^2_{\mu_{m'}}[\omega]$ is the variance of R's posterior, $\mu(\omega|m')$. Then, we can write S's *ex ante* payoff as follows:

$$E_\tau[E_{\mu_{m'}}[U^S(\cdot)]] = E_\tau[-\sigma^2_{\mu'_m}[\omega] - b^2] = -E_\tau[\sigma^2_{\mu'_m}[\omega]] - b^2,$$

where $E_\tau[\cdot]$ is the expectation with respect to the probability measure, $\tau$, which computes the probability that a posterior, $\mu(\omega|m')$, is induced (or probability that $m'$ is sent). Note that the term, $-E_\tau[\sigma^2_{\mu'_m}[\omega]]$, is the expected value of the variances of R's posteriors, $\{\mu(\omega|m)\}_{m\in M}$, and $-b^2$ is a constant. Thus, at an equilibrium, S chooses to commit to an information structure that minimizes $-E_\tau[\sigma^2_{\mu'_m}[\omega]]$. S can minimize $-E_\tau[\sigma^2_{\mu'_m}[\omega]]$ to zero by completely removing R's uncertainty about the states of the world. Thus, at any equilibrium, S chooses a $\pi$ that fully reveals the true states to R so that S obtains the maximized payoff of $-b^2$. Furthermore, a simple algebra shows that $-E_\tau[\sigma^2_{\mu'_m}[\omega]]$ is R's *ex ante* payoff. Thus, at any equilibrium, R's *ex ante* payoff is 0.

Lastly, given the uniform-quadratic setting, this payoff allocation is on the Pareto frontier: it is impossible to make S better off while making R not worse off and vice versa. Hence, the equilibrium outcome is a first-best outcome. □

Omitted definition of $\bar{\epsilon}(b)$ and $\epsilon^*(b)$

**Definition 4.** *$\bar{\epsilon}(b)$ is the upper bound of the noise levels which guarantee that welfare at the front-loading equilibrium is higher than that at BECS, and defined as follows:*

$$
\begin{aligned}
\bar{\epsilon}(b) &= \frac{2(1-2b(N-1)^2)(1+2b(N-1)N)}{N((2N-3)^2(1+2b(N-1))-8b^2(N-1)^3)} && \text{if} \quad \frac{1}{2N(N-1)} \leq b < \frac{1}{2(N-1)^2}, \\
&= \frac{2(1+2b(N-1)N)(2bN^2-1)}{(N-1)(1+2(1-b)N)(1+2N-4bN^2)} && \text{if} \quad \frac{1}{2N^2} < b < \frac{1}{2N(N-1)}, \\
&= 0 && \text{if} \quad b = \frac{1}{2N^2},
\end{aligned}
$$

*where $N = 2,3,4,5,...$ and $N$ is the number of elements in the partition at a front-loading equilibrium given $b \in (\frac{1}{2N^2}, \frac{1}{2(N-1)^2})$.*

For example, suppose that $b = \frac{1}{10}$. Then, $N = 3$ as $b = \frac{1}{10} \in [\frac{1}{12}, \frac{1}{8}) = [\frac{1}{2\cdot3\cdot(3-1)}, \frac{1}{2\cdot(3-1)^2})$. Thus, we can construct a front-loading equilibrium that partitions the state space into 3-interval. Following the definition above, $\bar{\epsilon}(b = \frac{1}{10}) = \frac{2(1-8b)(1+12b)}{3(9(1+4b)-64b^2)} = 0.0245$. Now suppose that $b = 1/15$. Then, $N = 3$ as $b = \frac{1}{15} \in (\frac{1}{18}, \frac{1}{12}) = (\frac{1}{2\cdot3^2}, \frac{1}{2\cdot3\cdot(3-1)})$. Thus, we can still construct a front-loading equilibrium with 3-interval partition. But, now, $\bar{\epsilon}(b = \frac{1}{15}) = \frac{2(1+12b)(18b-1)}{2(1+6(1-b)(7-36b)} = 0.08088$. Finally, if $b = \frac{1}{18} = \frac{1}{2\cdot3^2}$, $\bar{\epsilon}(b = 1/18) = 0$.

**Definition 5.** *The optimal noise level, $\epsilon^*(b)$, is the noise level that maximizes welfare at a front-loading equilibrium in $\Gamma(\pi^n)$ and is defined as follows:*

$$
\begin{aligned}
\epsilon^*(b) &= \frac{(1-2b(N-1)^2)(1-2bN^2)}{4(N-1)N(b+b^2(N-1)N-1)} && \text{if} \quad b \in (0,1/2) \backslash \{\frac{1}{2N^2} \text{ for } N = 2,3,...\}, \\
&= 0 && \text{if} \quad b = \frac{1}{2N^2} \quad \text{for} \quad N = 2,3,...,
\end{aligned}
$$

*where $N$ is the number of elements in the partition at a front-loading equilibrium given $b \in (\frac{1}{2N^2}, \frac{1}{2(N-1)^2})$.*

For example, suppose that $b \in (1/8, 1/2) = (\frac{1}{2\cdot2^2}, \frac{1}{2\cdot(2-1)^2})$. Then $N = 2$. Thus, we can construct a front-loading equilibrium that partitions the state space, $[0,1]$, into 2-interval. Then, the optimal noise level at this front-loading equilibrium is equal to $\epsilon^*(b) = \frac{(1-2b)(1-8b)}{8(b+2b^2+b-1)}$ for $1/8 < b < 1/2$.

Proof for Lemma 2

*Proof.* Consider $\Gamma(\pi^s(t))$ with some $t \in (0,1)$, and $(\sigma_s(t), a_s(m))$, where $\sigma_s(t)$ is the uniform distribution over $(t,1]$ for every $\omega \in [0,1]$, and

$$\begin{aligned} a_s(m) &= m \quad \text{if } m \in [0,t], \\ &= \bar{a}(t) \quad \text{if } m \in (t,1], \end{aligned}$$

where $\bar{a}(t) = \frac{\alpha(1-t)}{\alpha(1-t)+(1-\alpha)\times 1}\left(\frac{1+t}{2}\right) + \frac{1-\alpha}{\alpha(1-t)+(1-\alpha)\times 1}\left(\frac{1}{2}\right)$. Given $a_s(m)$, $\sigma_s(t)$ is incentive compatible for $S_c$ if condition (IC) below is satisfied:

(IC) $\qquad -(\omega + b - \bar{a}(t))^2 \geq -(\omega + b - a_s(m))^2$ for any $a_s(m)$, and for all $\omega \in [0,1]$.

There are two necessary conditions for condition (IC) to be satisfied.

(N1) $\quad t \leq b$

(N2) $\quad \max_{m\in[0,t]} a_s(m) = t \leq \bar{a}(t) \iff t \leq \frac{1-\sqrt{1-\alpha}}{\alpha}$

Condition (N1) simply says that $b$ must be higher than the "threshold value", $t \in (0,1)$. If $t > b$, $S_c$ at $\omega \in [0, t-b]$ can get his most preferred action by sending $m \in [b,t] \subset [0,t]$ instead of $m \in (t,1]$. For example, $S_c$ at $\omega = 0$ induces $\bar{a}(t)$ by sending $m \in (t,1]$ as specified in $\sigma_s(t)$. However, if $S_c$ at $\omega = 0$ sends $m = b \in [0,t]$, he could induce action $a_s(m = b \in [0,t]) = b$ which maximizes his payoff.

Condition (N2) simply states that $\bar{a}(t)$ should be the highest action that $S_c$ induces. If $t > \bar{a}(t)$, then there are $a_s(m) \in [\bar{a}(t), t]$. Then, it is immediate to see that $\sigma_s(t)$ is not incentive compatible. For example, $S_c$ at $\omega = 1$ prefers $a_s(m = t) = t$ to $\bar{a}(t)$ as $S_c$'s payoff, $(1 + b - a')^2$, increases in $a'$ up to $a' = 1 + b > 1 + t > t > \bar{a}(t)$.

Now assuming that (N1) and (N2) are satisfied, condition (IC) can be simplified as follows. First, note that $U^S(\omega, a; b)$ is supermodular (or, equivalently, $\frac{\partial^2 U^S}{\partial\omega\partial a} > 0$). Thus, if $a'' > a'$ and $\omega'$ prefers $a''$ to $a'$, any $\omega'' > \omega'$ also prefers $a''$ to $a'$. By (N2), $\bar{a}(t)$ is the highest action among all $a_s(m)$. Thus, if $-(b - \bar{a}(t))^2 \geq -(b - a_s(m))^2$ for any $a_s(m)$, condition (IC) must hold: if $S_c$ prefers $\bar{a}(t)$ to any other action (which is lower than $\bar{a}(t)$) at $\omega = 0$, does so $S_c$ at any other $\omega \in (0,1]$. Then, condition (IC) becomes

$$|b - \bar{a}(t)| \leq |b - a_s(m))| \text{ for all } a_s(m) \neq \bar{a}(t).$$

Now, note that $\bar{a}(t) \neq a_s(m) \in [0,t]$. Thus, $a_s(m) \leq \max_{m\in[0,t]} a_s(m) = t \leq b$ by (N1). Hence, $b - a_s(m) > 0$ for any $a_s(m) \neq \bar{a}(t)$. Thus, condition (IC) can be written as

$$|b - \bar{a}(t)| \leq b - a_s(m) \text{ for all } a_s(m) \neq \bar{a}(t).$$

Lastly, as any $a_s(m) \neq \bar{a}(t)$ is lower than or equal to $t$, $b-t \leq b-a_s(m)$ for any $a_s(m) \neq \bar{a}(t)$. Hence, condition (IC) finally becomes

$$|b - \bar{a}(t)| \leq b - t.$$

Now I need to show that, given any $(\alpha, b) \in (0,1) \times [1/2, \infty)$, there always exists $t \in (0,1)$ which satisfies $|b - \bar{a}(t)| \leq b - t$ assuming that (N1) and (N2) are true. Note that there are two cases when considering condition (IC): (1) $b \geq \bar{a}(t)$ and (2) $b \leq \bar{a}(t)$. With a simple algebra, we have

$$\begin{aligned} b - \bar{a}(t) &\geq 0 \text{ if and only if } k(t) := \alpha t^2 - 2\alpha b t + 2b - 1 \geq 0, \\ b - \bar{a}(t) &\leq 0 \text{ if and only if } k(t) := \alpha t^2 - 2\alpha b t + 2b - 1 \leq 0. \end{aligned}$$

Note that $k(t)$ is a quadratic function, and its discriminant is $4\alpha^2 b^2 - 4\alpha(2b-1)$. Denote the discriminant of $k(t)$ by $g(b)$. Note that $g(b) = 0$ always has two real roots, $\underline{b} = \frac{1-\sqrt{1-\alpha}}{\alpha}$ and $\bar{b} = \frac{1+\sqrt{1-\alpha}}{\alpha}$.[21] Thus, $g(b) \leq 0$ if $b \in [\underline{b}, \bar{b}]$; otherwise, $g(b) > 0$. Thus, depending on the value of $b$, we have three cases to consider: (1) $b \in [\underline{b}, \bar{b}]$, (2) $b \in [\frac{1}{2}, \underline{b})$, and (3) $b \in (\bar{b}, \infty)$.[22]

I denote by $T(\alpha, b)$ the set of the values of $t$ which satisfies (N1), (N2), and condition (IC). Then I show that $T(\alpha, b)$ is not an empty set for each case below.

*Case (1):* $b \in [\underline{b}, \bar{b}] = [\frac{1-\sqrt{1-\alpha}}{\alpha}, \frac{1+\sqrt{1-\alpha}}{\alpha}]$

In this case, $g(b) \leq 0$, which implies that $k(t)$ does not have any real root or has a unique real root. Thus, $k(t) \geq 0$ for any $t$, which implies that $b - \bar{a}(t) \geq 0$ for any $t \in (0,1)$. Hence, condition (IC) becomes $b - \bar{a}(t) \leq b - t$, which is trivially true under (N2). Thus, condition (IC) is always satisfied as far as (N1) and (N2) are true. For (N1) and (N2) to be true, $t \in (0, b) \cap (0, \frac{1-\sqrt{1-\alpha}}{\alpha})$. Note that $(0, b) \cap (0, \frac{1-\sqrt{1-\alpha}}{\alpha}) = (0, \frac{1-\sqrt{1-\alpha}}{\alpha})$ because $\frac{1-\sqrt{1-\alpha}}{\alpha} \leq b$ in this case. Hence, if $t \in (0, \frac{1-\sqrt{1-\alpha}}{\alpha})$, both (N1) and (N2) are true. Thus, $\sigma_s(t)$ is incentive compatible in $\Gamma(\pi^s(t))$ for $t \in (0, \frac{1-\sqrt{1-\alpha}}{\alpha})$ if $b \in [\underline{b}, \bar{b}]$. Lastly, in this case, $T(\alpha, b) = (0, \frac{1-\sqrt{1-\alpha}}{\alpha}) \subset (0,1)$; $T(\alpha, b)$ is not empty as $\frac{1-\sqrt{1-\alpha}}{\alpha} \in (\frac{1}{2}, 1)$ for any $\alpha \in (0,1)$.

*Case (2):* $b \in [\frac{1}{2}, \underline{b}) = [\frac{1}{2}, \frac{1-\sqrt{1-\alpha}}{\alpha})$

Suppose $b \in [\frac{1}{2}, \underline{b})$. Then, the discriminant of $k(t)$ is strictly positive (i.e., $g(b) > 0$). Thus, $k(t) = 0$ has two real roots, $b \pm \frac{\sqrt{\alpha(\alpha b^2 - 2b + 1)}}{\alpha}$. Denote these two real roots by $k_1$ and $k_2$, where $k_1 < k_2$. Now choose $t \in [k_1, k_2]$.[23] Then, $k(t) \leq 0$, which implies $b \leq \bar{a}(t)$. Then,

[21] The discriminant of $g(b)$ is $64\alpha^2 - 64\alpha^3 = 64\alpha^2(1-\alpha)$, which is always strictly positive given $\alpha \in (0,1)$.

[22] It is easy to check that $\underline{b} = \frac{1-\sqrt{1-\alpha}}{\alpha} \in (\frac{1}{2}, 1)$ for any $\alpha \in (0,1)$, and $\bar{b} = \frac{1+\sqrt{1-\alpha}}{\alpha} \in (1, \infty)$ for any $\alpha \in (0,1)$.

[23] Note that $k_1 < b < k_2$. Thus, we cannot choose $t \geq k_2 > b$ as this immediately violates (N1). We can choose $t \leq k_1 < b$. However, in this case, $T(\alpha, b) = \emptyset$ when $b = 1/2$.

condition (IC) becomes $\bar{a}(t)-b \leq b-t$. Thus, $\sigma_s(t)$ is incentive compatible if $b \geq \frac{t+\bar{a}(t)}{2}$. Note that $b = \frac{1}{2}(\bar{a}(t)+t)$ can also be written as a function of $t$, $h(t) = 3\alpha t^2 - (4\alpha b+2)t + 4b - 1$. Then, given $b \in [\frac{1}{2}, \underline{b}]$, $h(t) = 0$ also has two real roots, $\frac{2\alpha b+1\pm\sqrt{4\alpha^2b^2-8\alpha b+3\alpha+1}}{3\alpha}$.[24] Denote these two real roots by $h_1$ and $h_2$, where $h_1 < h_2$. Thus, to have $h(t) \geq 0$ (or, equivalently, $b \geq \frac{1}{2}(\bar{a}(t)+t)$), either $t \leq h_1$ or $t \geq h_2$ must hold.

In summary, the restrictions imposed on $t$ are as follows. First, by (N1) and (N2), $t \in (0,b) \cap (0, \frac{1-\sqrt{1-\alpha}}{\alpha}) = (0,b)$, where the last equality is true as $b < \frac{1-\sqrt{1-\alpha}}{\alpha}$. Secondly, $t \in [k_1, k_2]$ and $t \in (0, h_1] \cup [h_2, 1)$ to have condition (IC) hold. Hence, $t \in T(\alpha, b)$ to have $\sigma_s(t)$ be incentive compatible in $\Gamma(\pi^s(t))$, where $T(\alpha,b) = (0,b) \cap [k_1,k_2] \cap \{(0,h_1] \cup [h_2,1)\}$.

Now to see $T(\alpha,b)$ is not an empty set, first note that $k_1 < b < k_2$ and $k_1 \geq 0$ if and only if $b \geq 1/2$. Thus, we have

$$\begin{aligned}(0,b) \cap [k_1,k_2] &= (0,b) \quad \text{for} \quad b = \frac{1}{2} \\ &= [k_1, b) \quad \text{for} \quad b \in \Big(\frac{1}{2}, \frac{1-\sqrt{1-\alpha}}{\alpha}\Big).\end{aligned}$$

Furthermore, $h_1 < b < h_2$ holds if $b < \frac{1-\sqrt{1-\alpha}}{\alpha}$. Lastly, $k_1 < h_1$. To see this, define $A := \alpha^2b^2 - 2\alpha b + \alpha$, and $B := 4\alpha^2b^2 - 8\alpha b + 3\alpha + 1$. Then, we have

$$\begin{aligned}h_1 - k_1 > 0 \quad &\iff \quad \frac{1+2\alpha b - \sqrt{B}}{3\alpha} - \frac{\alpha b - \sqrt{A}}{\alpha} > 0 \\ &\iff \quad 1 + 2\alpha b - \sqrt{B} - 3\alpha b + 3\sqrt{A} > 0 \\ &\iff \quad (1-\alpha b) + 3\sqrt{A} > \sqrt{B}.\end{aligned}$$

Note that $(1-\alpha b) > 0$, $A > 0$, and $B > 0$ if $b < \frac{1-\sqrt{1-\alpha}}{\alpha} < \frac{1}{\alpha}$. Thus,

$$\begin{aligned}(1-\alpha b) + 3\sqrt{A} > \sqrt{B} \quad &\iff \quad (1-\alpha b)^2 + 6(1-\alpha b)\sqrt{A} + 9A > B \\ &\iff \quad 6(1-\alpha b)\sqrt{A} + 9A + (1-\alpha b)^2 - B > 0 \\ &\iff \quad 6(1-\alpha b)\sqrt{A} + 6A > 0.\end{aligned}$$

Finally, we have

$$\begin{aligned}T(\alpha,b) &= (0,b) \cap \{(0,h_1] \cup [h_2,1]\} = (0,h_1] \quad \text{for} \quad b = \frac{1}{2} \\ &= [k_1,b) \cap \{(0,h_1] \cup [h_2,1]\} = [k_1,h_1] \quad \text{for} \quad b \in \Big(\frac{1}{2}, \frac{1-\sqrt{1-\alpha}}{\alpha}\Big).\end{aligned}$$

Thus, $T(\alpha,b) \neq \emptyset$ for $b \in [1/2, \frac{1-\sqrt{1-\alpha}}{\alpha})$.

[24] The discriminant of $h(t)$ is strictly positive (i.e., $4\alpha^2b^2 - 8\alpha b + 3\alpha + 1 > 0$) if $b > \frac{1+\frac{\sqrt{3}}{2}\sqrt{1-\alpha}}{\alpha}$ or $b < \frac{1-\frac{\sqrt{3}}{2}\sqrt{1-\alpha}}{\alpha}$. Note that, here, we assume $b \in [1/2, \frac{1-\sqrt{1-\alpha}}{\alpha})$. Thus, $b < \frac{1-\sqrt{1-\alpha}}{\alpha} < \frac{1-\frac{\sqrt{3}}{2}\sqrt{1-\alpha}}{\alpha}$.

*Case (3):* $b \in (\bar{b}, \infty) = (\frac{1+\sqrt{1+\alpha}}{\alpha}, \infty)$

Since $b > \bar{b}$, $k(t) = 0$ has two real roots, $k_1$ and $k_2$ as in case (2). Now choose $t \in (0, k_1)$. Then, $k(t) \geq 0$, which implies $b \geq \bar{a}(t)$. Thus, condition (IC) becomes $b - \bar{a}(t) \geq b - t$, which is trivially true under (N1) and (N2). Then, $T(\alpha, b)$ is the set of values of $t$ which satisfies (N1), (N2), and $t \in (0, k_1)$. Thus, $T(\alpha, b) = (0, b) \cap (0, \frac{1-\sqrt{1-\alpha}}{\alpha}) \cap (0, k_1)$. To see $T(\alpha, b) \neq \emptyset$, first note that $\frac{1-\sqrt{1-\alpha}}{\alpha} < \frac{1+\sqrt{1-\alpha}}{\alpha} < b$. Thus, $(0, b) \cap (0, \frac{1-\sqrt{1-\alpha}}{\alpha}) = (0, \frac{1-\sqrt{1-\alpha}}{\alpha})$. Furthermore, $0 < \frac{1-\sqrt{1-\alpha}}{\alpha} < k_1$ if $b > \frac{1+\sqrt{1-\alpha}}{\alpha}$. To see this, note that

$$\begin{aligned} k_1 - \frac{1-\sqrt{1-\alpha}}{\alpha} > 0 \quad &\iff \quad \alpha b - \sqrt{\alpha^2 b^2 - 2\alpha b + \alpha} - (1 - \sqrt{1-\alpha}) > 0 \\ &\iff \quad \alpha b - 1 + \sqrt{1-\alpha} > \sqrt{\alpha^2 b^2 - 2\alpha b + \alpha} \end{aligned}$$

Note that $\alpha b - 1 > 0$ and $A := \alpha^2 b^2 - 2\alpha b + \alpha > 0$ as $b > \frac{1+\sqrt{1-\alpha}}{\alpha} > \frac{1}{\alpha}$. Thus, we have

$$\begin{aligned} \alpha b - 1 + \sqrt{1-\alpha} > \sqrt{A} \quad &\iff \quad (\alpha b - 1)^2 + 2(\alpha b - 1)\sqrt{1-\alpha} + (1-\alpha) > A \\ &\iff \quad 2(1-\alpha) + 2(\alpha b - 1)\sqrt{1-\alpha} > 0. \end{aligned}$$

Thus, $T(\alpha, b) = (0, \frac{1-\sqrt{1-\alpha}}{\alpha}) \cap (0, k_1) = (0, \frac{1-\sqrt{1-\alpha}}{\alpha})$, which is not an empty set.

In summary, given any $\alpha \in (0, 1)$,

$$\begin{aligned} T(\alpha, b) &= (0, h_1) && \text{for } \; b = \frac{1}{2}, \\ &= [k_1, h_1] && \text{for } \; b \in \Big(\frac{1}{2}, \frac{1-\sqrt{1-\alpha}}{\alpha}\Big), \\ &= \Big(0, \frac{1-\sqrt{1-\alpha}}{\alpha}\Big) && \text{for } \; b \in [\frac{1-\sqrt{1-\alpha}}{\alpha}, \infty], \end{aligned}$$

where $k_1 = b - \frac{\sqrt{\alpha(\alpha b^2 - 2b + 1)}}{\alpha}$ and $h_1 = \frac{2\alpha b + 1 - \sqrt{4\alpha^2 b^2 - 8\alpha b + 3\alpha + 1}}{3\alpha}$. Thus, given any $(\alpha, b) \in (0, 1) \times [1/2, \infty)$, $T(\alpha, b) \neq \emptyset$, where $(\sigma_s(t), a_s(m))$ constitutes an equilibrium in $\Gamma(\pi^s(t))$ for $t \in T(\alpha, b)$.

$\square$

## References


Alonso, R., and O. Câmara (2016a): "Bayesian persuasion with heterogeneous priors," *Journal of Economic Theory*, 165, 672–706.

——— (2016b): "Persuading voters," *American Economic Review*, 106(11), 3590–3605.

Arieli, I., and Y. Babichenko (2019): "Private bayesian persuasion," *Journal of Economic Theory*, 182, 185–217.

Bagwell, K. (1995): "Commitment and observability in games," *Games and Economic Behavior*, 8(2), 271–280.

Beach, J. E. (2001): "Clinical trials integrity: A CRO perspective," *Accountability in Research*, 8(3), 245–260.

Bizzotto, J., J. Rüdiger, and A. Vigier (2020): "Testing, disclosure and approval," *Journal of Economic Theory*, p. 105002.

Blume, A., O. J. Board, and K. Kawamura (2007): "Noisy talk," *Theoretical Economics*, 2(4), 395–440.

Crawford, V. P., and J. Sobel (1982): "Strategic information transmission," *Econometrica: Journal of the Econometric Society*, pp. 1431–1451.

Fershtman, C., and E. Kalai (1997): "Unobserved delegation," *International Economic Review*, pp. 763–774.

Fréchette, G. R., A. Lizzeri, and J. Perego (2019): "Rules and commitment in communication," .

Gentzkow, M., and E. Kamenica (2017): "Bayesian persuasion with multiple senders and rich signal spaces," *Games and Economic Behavior*, 104, 411–429.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009): "Mediation, arbitration and negotiation," *Journal of Economic Theory*, 144(4), 1397–1420.

Jain, V. (2018): "Bayesian persuasion with cheap talk," *Economics Letters*, 170, 91–95.

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(6), 2590–2615.

Kartik, N. (2009): "Strategic communication with lying costs," *The Review of Economic Studies*, 76(4), 1359–1395.

Kim, K., and J. Pogach (2014): "Honesty vs. advocacy," *Journal of Economic Behavior & Organization*, 105, 51–74.

Kolotilin, A. (2018): "Optimal information disclosure: A linear programming approach," *Theoretical Economics*, 13(2), 607–635.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017): "Persuasion of a privately informed receiver," *Econometrica*, 85(6), 1949–1964.

Lipnowski, E., D. Ravid, and D. Shishkin (2019): "Persuasion via Weak Institutions," *Available at SSRN: https://ssrn.com/abstract=3168103 or http://dx.doi.org/10.2139/ssrn.3168103.*

Luo, Z., and A. Rozenas (2018): "Strategies of election rigging: Trade-offs, determinants, and consequences," *Quarterly Journal of Political Science*, 13(1), 1–28.

Martinez-Roa, G. (2020): "Bayesian Persuasion in the Digital Age," *Working Paper.*

Min, D. (2017): "Screening for Experiments," *Working Paper.*

Nguyen, A., and T. Y. Tan (2019a): "Bayesian Persuasion with Costly Messages," *Available at SSRN: https://ssrn.com/abstract=3298275 or http://dx.doi.org/10.2139/ssrn.3298275.*

——— (2019b): "Information control in the hold-up problem," *The RAND Journal of Economics*, 50(4), 768–786.

Wang, Y. (2013): "Bayesian persuasion with multiple receivers," *Available at SSRN: https://ssrn.com/abstract=2625399 or http://dx.doi.org/10.2139/ssrn.2625399.*

Yoder, N. (2019): "Designing incentives for heterogeneous researchers," *Available at SSRN: https://ssrn.com/abstract=3154143 or http://dx.doi.org/10.2139/ssrn.3154143.*