

Staff Working Paper/Document de travail du personnel 2019-10

Limiting Sender's Information in Bayesian Persuasion



by Shota Ichihashi

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

Bank of Canada Staff Working Paper 2019-10

February 2019

Limiting Sender's Information in Bayesian Persuasion

by

Shota Ichihashi

Canadian Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
sichihashi@bankofcanada.ca

Acknowledgements

I am thankful to Gabriel Carroll, Matthew Gentzkow, Paul Milgrom, Kota Murayama, Ilya Segal, and Takuo Sugaya for valuable comments and suggestions. I am grateful for financial support from Yoshida Scholarship Foundation and Koret Fellow Stanford Graduate Fellowship. The opinions expressed in this article are the author's own and do not reflect the view of Bank of Canada.

Abstract

This paper studies how the outcome of Bayesian persuasion depends on a sender's information. I study a game in which, prior to the sender's information disclosure, the designer can restrict the most informative signal that the sender can generate. In the binary action case, I consider arbitrary preferences of the designer and characterize all equilibrium outcomes. As a corollary, I solve a problem of how to maximize a receiver's payoffs by restricting the sender's information: Whenever the designer can increase the receiver's payoffs by restricting the sender's information, the receiver-optimal way coincides with an equilibrium of the game in which the receiver persuades the sender.

Bank topic: Economic models
JEL codes: D82, D83

Résumé

Dans cette étude, j'examine comment le résultat de la persuasion dans un jeu bayésien dépend de l'information dont dispose un expéditeur. Pour ce faire, je considère un jeu dans lequel, avant que l'expéditeur ne transmette la moindre information, le concepteur est autorisé à limiter le contenu du signal le plus porteur d'information que l'expéditeur peut générer. Partant de l'hypothèse que chaque joueur est confronté à un choix binaire, je me penche sur les préférences arbitraires du concepteur et décris toutes les situations d'équilibre. Comme corollaire, je résous un problème pour déterminer comment maximaliser les bénéfices du destinataire en restreignant l'information du côté de l'expéditeur : chaque fois que le concepteur peut améliorer les bénéfices du destinataire en restreignant l'information que peut collecter l'expéditeur, l'issue optimale pour le destinataire correspond à une situation d'équilibre où le destinataire persuade l'expéditeur.

Sujet : Modèles économiques
Codes JEL : D82, D83

Non-technical Summary

Online retailers such as Amazon collect consumers' information like purchase histories and personal characteristics. This information can be used to offer consumers better services, for example, to design recommendation systems that maximize consumers' value and minimize their time spent searching; however, such information can also be used improperly to mislead consumers and maximize retailers' profits. Whether a regulator can increase consumers' welfare by restricting these types of services and information collection is a non-trivial question.

To address this question, this paper considers a game-theoretic model of information disclosure, in which the sender strategically discloses information to the receiver, who takes an action that affects the welfare of each player. The main focus is on how a regulator can improve the receiver's welfare by restricting the information available to the sender. In the example of online sellers, imagine that a seller (i.e., sender) discloses consumers' (i.e., receiver) information about product values, which depend on the product's characteristics and consumers' characteristics. The question is whether a regulator can improve consumer welfare by restricting the seller's information collection, which, in turn, limits the scope for strategic information disclosure. The paper first characterizes all possible outcomes when the regulator restricts the sender's information in arbitrary ways. Moreover, I derive a necessary and sufficient condition under which the regulator can strictly improve the receiver's welfare and social welfare by restricting the sender's information.

In the example of online sellers, the main results have the following implication. Regulation benefits consumers if it allows sellers to collect just enough information to provide accurate recommendations without strategically manipulating consumers' purchase behavior in their favor.

1 Introduction

This paper studies how the outcome of Bayesian persuasion depends on a sender's information. Suppose that the sender (he) discloses information to a receiver (she), who then takes an action that affects the welfare of both players. Now, can the receiver benefit if the sender has less information to potentially disclose? How does social welfare depend on information available to the sender? To answer these questions, I consider a model in which, prior to the sender's information disclosure, a "designer" (it) can restrict the most informative signal that the sender can generate. I assume that the receiver has a binary action choice but impose no special assumptions on the state space and the preferences.

The model would fit the context of delegation in an organization. Imagine that the principal (the designer or the receiver) wants to learn division-specific information to make an appropriate decision, and the principal asks an agent (the sender) in the division to design an experiment that reveals information. However, because the agent puts more priority on the division, the agent might strategically design an experiment that benefits the division instead of the principal. To avoid this problem, can the principal set a rule determining the types of information that the agent cannot acquire?

Another potential application is the effect of online privacy regulations. Suppose that an online seller (the sender) launches a new product. The product values are idiosyncratic and unknown to consumers (the receiver). However, the seller can learn about the values, say, by combining its knowledge of product characteristics and observable individual characteristics such as consumers' browsing histories. If the seller can obtain precise estimates of product values, it can strategically disclose this information to influence consumers' purchase decisions and increase sales.¹ Now, can a regulator (the designer) improve consumer welfare through online privacy regulations that prevent the seller from observing individual characteristics and obtaining fine-grained estimates of valuations?

¹For example, the seller may design a recommendation system that determines how and whether to recommend the new product as a function of the estimate of each consumer's product valuation. This is similar to the idea of [Rayo and Segal \(2010\)](#) where a website designs a rule to display ads to communicate the relevance of ads to users.

The main result considers arbitrary preferences of the designer and characterizes all equilibrium payoffs of the sender and the receiver. In other words, the result characterizes the set of all possible equilibrium outcomes when we consider the arbitrary information of the sender.

The main result has several notable implications. First, the set of all equilibrium payoff profiles may have an empty interior, in which case restricting the sender's information does not affect the receiver's payoffs and can only reduce the sender's payoffs. I show that this occurs if the sender prefers one action in all states of the world. Second, as a direct corollary, I obtain the set of all *efficient* outcomes that the designer can attain by restricting the sender's information.² Moreover, I show that if the designer's payoffs are increasing in the expected payoffs of the receiver and the sender, the designer implements an efficient outcome.³ Combining these observations, we can derive the equilibrium strategy of the designer, who puts arbitrary (nonnegative) weights on the payoffs of the sender and the receiver.

One interesting case is when the designer aims to maximize the receiver's payoff. I show that whenever restricting the sender's information strictly benefits the receiver (relative to the original Bayesian persuasion), the receiver-optimal way coincides with an equilibrium of the "flipped game" in which *the receiver persuades the sender*. Thus, we can obtain the designer's equilibrium strategy by solving the original Bayesian persuasion and the flipped game in which the preferences of the sender and the receiver are switched.

This work relates to several strands of existing literature. First, the paper refers to work on information design (e.g., [Kamenica and Gentzkow 2011](#); [Kolotilin 2015](#); [Taneva 2015](#); [Bergemann and Morris 2016](#)). We may view the current model as a new information design problem, where the underlying game is an information design problem as well. Second, the paper also relates to works on Bayesian persuasion with multiple senders, such as [Gentzkow and Kamenica \(2017\)](#) and [Li and Norman \(2018\)](#). However, we should note that the designer in this paper is not a sender because the receiver does not observe a realization drawn by the designer's signal. Third, the cheap-talk literature studies the idea of restricting the sender's information ([Fischer and Stocken 2001](#); [Ivanov 2010, 2015](#); [Frug 2016](#)) and sequential communication among multiple senders ([Ambrus et al.](#)

²Here, an outcome is said to be efficient when there is no state-contingent action plan of the receiver that gives both the sender and the receiver greater payoffs.

³This result is not a priori obvious, because some efficient outcomes may not arise as an equilibrium regardless of how the designer restricts the sender's information.

2013). For example, [Ivanov \(2010\)](#) studies the problem of restricting the sender’s information and provides a sufficient condition under which information restrictions increase the receiver’s payoff. The literature typically uses a version of [Crawford and Sobeli \(1982\)](#)’s model, and it either derives a sufficient condition for information restrictions to benefit the receiver or characterizes the receiver-optimal outcome in a particular class of restrictions. In contrast, I characterize all outcomes across all restrictions of the sender’s information in the context of Bayesian persuasion. As by-products, I also obtain information restrictions maximizing the receiver’s welfare or social welfare.

The rest of the paper is organized as follows. [Section 2](#) describes the model. [Section 3](#) proves the main result and applies it to simple examples. [Section 4](#) characterizes the set of efficient outcomes that the designer can implement. This section also derives information restrictions maximizing the receiver’s payoffs and the sum of each player’s payoffs. [Section 5](#) discusses the case in which the receiver has more than two actions. [Section 6](#) concludes.

2 Model

There are three players: the sender, the receiver, and the designer. The receiver has a payoff function $u(a, \omega)$ that depends on her action $a \in \{0, 1\}$ and the state of the world $\omega \in \Omega$. For simplicity, I assume that Ω is finite.⁴ The sender has a payoff function $v(a, \omega)$ that depends on the receiver’s action and the state of the world. Both the sender and the receiver maximize expected payoffs. I do not specify the preferences of the designer, who plays an auxiliary role. The sender, the receiver, and the designer share a prior $b_0 \in \Delta(\Omega)$.⁵ Without loss of generality, normalize $u(0, \omega) = v(0, \omega) = 0$ and write $u(\omega) := u(1, \omega)$ and $v(\omega) := v(1, \omega)$ for any $\omega \in \Omega$.

A *signal* (S, μ) consists of a finite realization space S and a function $\mu : \Omega \rightarrow \Delta(S)$. Hereafter, I often write μ instead of (S, μ) . Given state ω , signal μ , and realization $s \in S$, let $\mu(s|\omega)$ denote the probability that $\mu(\omega)$ draws s .

The timing of the game is as follows. First, the designer chooses a signal, say μ_D . The sender then chooses from any signals weakly less informative than μ_D . The informativeness is in the

⁴The main result remains true if Ω is a compact metric space.

⁵ $\Delta(X)$ denotes the set of all probability distributions over set X .

sense of [Blackwell \(1953\)](#). Let μ denote the sender's choice. Nature draws the state of the world $\omega \sim b_0$ and a signal realization $s \sim \mu(\omega)$. The receiver observes the sender's choice μ and its realization $s \in S$, and then chooses her action $a \in \{0, 1\}$. The solution concept is a subgame perfect equilibrium in which the receiver breaks ties in favor of the sender.

If it is clear from the context, I often use “equilibrium” to mean a subgame perfect equilibrium of the *subgame* in which the designer has chosen some signal. For instance, if I write “the sender's equilibrium payoffs decrease as the designer chooses less informative signals,” the equilibrium payoffs are the ones in the corresponding subgames.

The receiver learns about the state of the world only from the sender's signal; however, the designer's signal caps the most informative signal available to the sender. We may view the designer's strategy as a restriction on what information the sender is allowed to collect or on testing procedures and experiment techniques. If the designer chooses the fully informative signal, we obtain a standard Bayesian persuasion.

3 Main Result

I begin with introducing several notions. First, call any signal $\mu : \Omega \rightarrow \Delta(S)$ with $S = \{0, 1\}$ a *straightforward* signal. The set of all straightforward signals is denoted by $\mathcal{S} := \Delta(\{0, 1\})^\Omega$. Given any $\mu \in \mathcal{S}$, let $\mathbb{E}_\mu[u]$ and $\mathbb{E}_\mu[v]$ denote the payoffs of the receiver and the sender when the receiver follows signal realizations of μ , i.e., the receiver chooses $a = 1$ with probability $\mu(1|\omega)$ at each ω .

Define the set \mathcal{F} of all feasible payoff vectors as

$$\mathcal{F} := \{(\mathbb{E}_\mu[u], \mathbb{E}_\mu[v])\}_{\mu \in \mathcal{S}}.$$

Some payoff vectors in \mathcal{F} may not arise as an equilibrium regardless of the designer's choice, because the corresponding straightforward signals may not respect the incentives of the sender and the receiver.

We say that the designer can *implement* $(u, v) \in \mathcal{F}$ (or (u, v) is *implementable*) if there exists a signal μ such that in an equilibrium of the subgame in which the designer has chosen μ , the

receiver and the sender obtain expected payoffs of u and v , respectively.

Finally, I define the following payoffs, which enable us to concisely describe the set of implementable outcomes. First, let

$$\underline{u} := \max_{a \in \{0,1\}} \left[a \cdot \int_{\Omega} u(\omega) db_0(\omega) \right] \quad (1)$$

denote the receiver's payoff from always choosing her ex ante optimal action. Second, let

$$v_{NO} := \max_{a \in A^*} \left[a \cdot \int_{\Omega} v(\omega) db_0(\omega) \right], \quad \text{where } A^* := \arg \max_{a \in \{0,1\}} \left[a \cdot \int_{\Omega} u(\omega) db_0(\omega) \right], \quad (2)$$

denote the sender's expected payoff when the receiver chooses her ex ante optimal action in all states of the world, breaking ties in favor of the sender (if A^* is not a singleton). Third, let

$$\underline{v} := \max_{a \in \{0,1\}} \left[a \cdot \int_{\Omega} v(\omega) db_0(\omega) \right] \quad (3)$$

denote the sender's payoff when he always chooses his ex ante optimal action.

Regardless of the sender's strategy, the receiver can secure \underline{u} by taking one action deterministically. Also, the sender can always attain v_{NO} by disclosing no information. In contrast, there is no guarantee that the sender can secure \underline{v} because he never chooses an action by himself.

The main result concerns the question of what payoff vectors are implementable. In other words, how does the outcome of Bayesian persuasion depend on the information available to the sender? If the receiver has a binary action choice, the following result gives a comprehensive answer to this question.

Theorem 1. *Define A_1 and A_2 as follows.*

$$\begin{aligned} A_1 &:= \{(u, v) \in \mathcal{F} : u > \underline{u} \text{ and } v \geq \underline{v}\}, \\ A_2 &:= \{(u, v) \in \mathcal{F} : u = \underline{u} \text{ and } v \geq v_{NO}\}, \end{aligned}$$

where \underline{u} , v_{NO} , and \underline{v} are defined by (1), (2), and (3). The designer can implement a feasible payoff vector $(u, v) \in \mathcal{F}$ if and only if $(u, v) \in A_1 \cup A_2$. If the designer can publicly randomize signals, (u, v) is implementable if and only if (u, v) is in the convex hull of $A_1 \cup A_2$.

A_2 is nonempty because the designer can implement (\underline{u}, v_{NO}) by giving the sender no information (i.e., the designer chooses μ such that $\mu(\omega)$ is independent of ω). In contrast, A_1 can be empty or nonempty depending on the preferences of the sender and the receiver, as the following examples depict.

Example 1. Suppose that there are three equally likely states: $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $b_0(\omega) = 1/3$ for all $\omega \in \Omega$. The payoffs from $a = 1$ are shown in Table 1: The sender and the receiver agree about preferred actions at states ω_1 and ω_3 , but they disagree at ω_2 .

	ω_1	ω_2	ω_3
the receiver	1	-1	-1
the sender	1	1	-1

Table 1: Payoffs from $a = 1$

In Figure 1, the rectangle delineated by the thick black line is the set \mathcal{F} of feasible payoff vectors. The blue vertical line and red horizontal line correspond to the payoffs from ex ante optimal actions for the receiver and the sender, respectively, i.e., $\underline{u} = 0$ and $\underline{v} = 1/3$. The sender's payoff v_{NO} from the receiver's ex ante optimal action is 0, as the receiver prefers $a = 0$ at the prior belief.

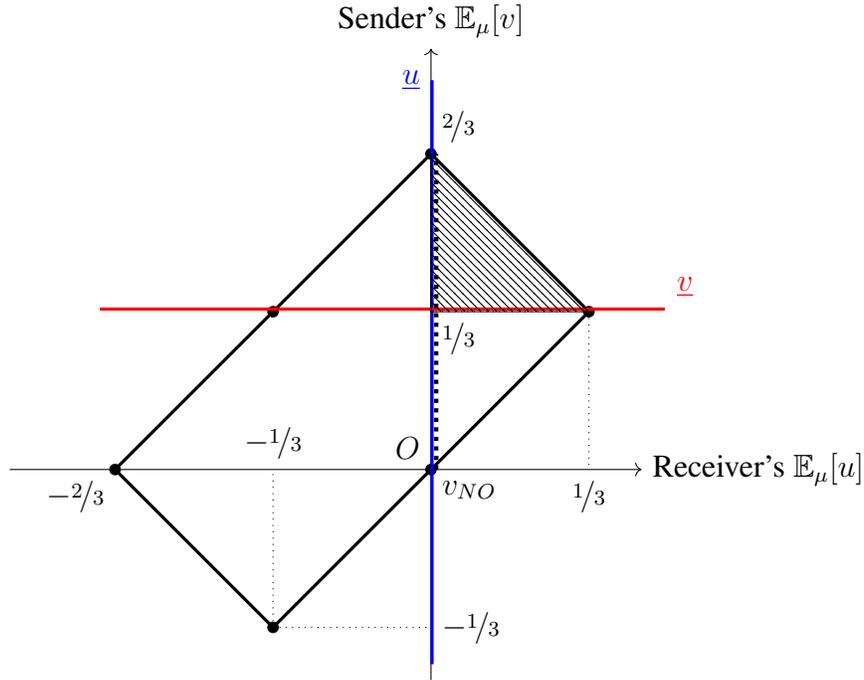


Figure 1: Implementable Payoff Vectors when $A_1 \neq \emptyset$

A_1 is the dashed triangular area including all edges except the one connecting $(0, 1/3)$ and $(0, 2/3)$. A_2 is the thick dashed line connecting $(\underline{u}, v_{NO}) = (0, 0)$ and $(0, 2/3)$. Note that $A_1 \cup A_2$ is not convex. However, if the designer can publicly randomize signals, it can achieve any expected payoff vectors in the triangle connecting $(0, 0)$, $(1/3, 1/3)$, and $(0, 2/3)$. Thus, in this example, the designer (with public randomization) can implement any payoff vectors that give the receiver weakly greater payoffs than the payoff under no information.

Figure 1 tells us two more things. First, if the sender can choose any signal as in a standard Bayesian persuasion, payoff vector $(0, 2/3)$ is realized, because it maximizes the sender's payoff among all implementable payoff vectors. Second, the designer can move the equilibrium outcome from $(0, 2/3)$ to the receiver's best outcome $(1/3, 1/3)$ by restricting the sender's information. Specifically, the designer chooses a signal that only discloses whether the state is ω_1 . In this case, the designer can strictly increase the receiver's payoff by limiting the sender's information. The next example shows that this is not necessarily the case.

Example 2. I modify the payoffs of the sender as in Table 2: The sender strictly prefers $a = 1$ in all states of the world.

	ω_1	ω_2	ω_3
the receiver	1	-1	-1
the sender	1	1	1

Table 2: Payoffs from $a = 1$

In Figure 2, the rectangle delineated by the thick black line is the set \mathcal{F} of feasible payoff vectors. The blue vertical line and red horizontal line correspond to the payoffs from ex ante optimal actions for the receiver and the sender, respectively, i.e., $\underline{u} = 0$ and $\underline{v} = 1$.

Note that $A_1 = \emptyset$ because for any $(u, v) \in \mathcal{F}$ with $u > \underline{u}$, $v < \underline{v}$ holds. Thus, the designer can only achieve payoff vectors in A_2 , which is the dashed line connecting $(\underline{u}, v_{NO}) = (0, 0)$ and $(0, 2/3)$. Thus, in this example, limiting the sender's information has no impact on the receiver and may reduce the sender's payoff.

The last observation in Example 2 is general: Whenever the sender prefers one action in all states of the world, restricting the sender's information has no impact on the receiver's payoffs.

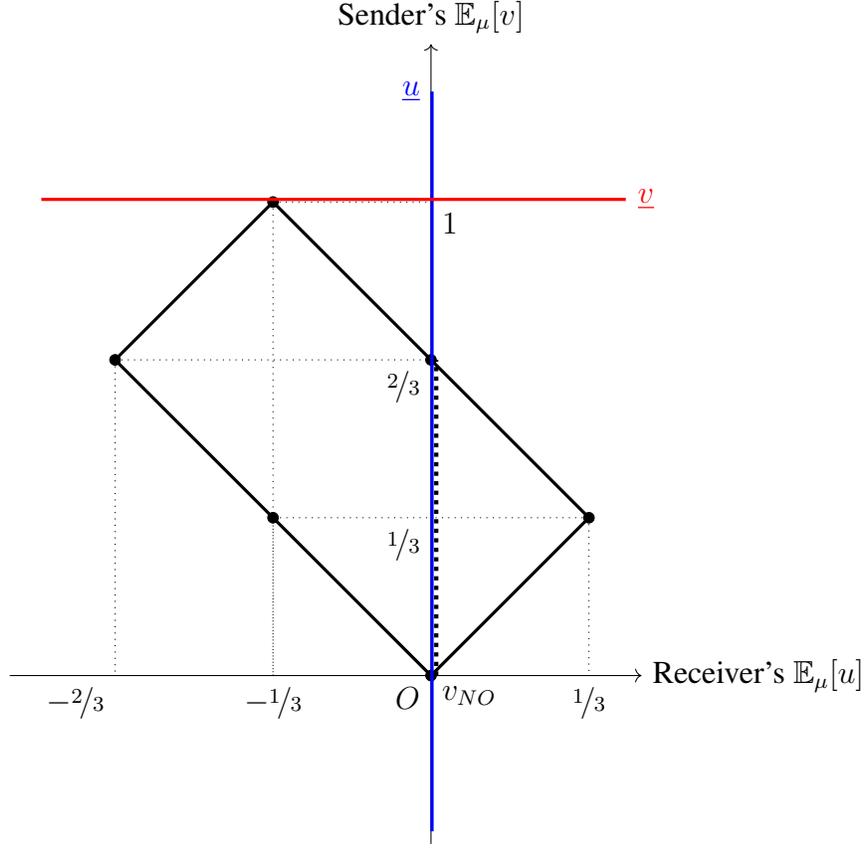


Figure 2: Implementable Payoff Vectors when $A_1 = \emptyset$

Corollary 1. *Suppose that the sender prefers one action in all states of the world, i.e., either $v(\omega) > 0$ for all $\omega \in \Omega$ or $v(\omega) < 0$ for all $\omega \in \Omega$ holds. Then, restricting the sender's information does not affect the receiver's payoff. Namely, $A_1 = \emptyset$, and thus the receiver's equilibrium payoff is \underline{u} regardless of the designer's strategy.*

Proof. Without loss of generality, suppose $v(\omega) > 0$ for all $\omega \in \Omega$. Take any $(u, v) \in \mathcal{F}$ and let μ be $(\mathbb{E}_\mu[u], \mathbb{E}_\mu[v]) = (u, v)$. $u > \underline{u}$ implies that the receiver takes different actions at different states if she follows μ . This implies $v < \underline{v}$, because the sender can achieve \underline{v} only when the receiver chooses $a = 1$ in all states. Thus, $A_1 = \emptyset$. \square

3.1 Proof of Theorem 1

The proof of Theorem 1 consists of several steps. First, I define the incentive compatibility of straightforward signals for the receiver and the sender.

Definition 1. A straightforward signal μ is incentive-compatible for the receiver (sender) if the receiver (sender) weakly prefers action a after observing each possible realization $a \in \{0, 1\}$ drawn by μ .

The following lemma reduces the incentive compatibility to a single inequality: A straightforward signal μ is incentive-compatible for the receiver (sender) if and only if her (his) expected payoff from following μ is weakly greater than the payoff under no information.⁶ The result relies on the assumption of binary action choice. The proof is in Appendix A.

Lemma 1. *A straightforward signal μ is incentive-compatible for the receiver if and only if $\mathbb{E}_\mu[u] \geq \underline{u}$. μ is incentive-compatible for the sender if and only if $\mathbb{E}_\mu[v] \geq \underline{v}$.*

The “only if” part is straightforward. To see why the “if” part holds, suppose that the receiver’s ex ante optimal action is $a = 0$. As $\mathbb{E}_\mu[u] \geq \underline{u}$, it must be the case that the receiver weakly prefers $a = 1$ at realization 1. Indeed, if the receiver strictly prefers $a = 0$, then following μ gives the receiver a strictly lower payoff than \underline{u} . Moreover, whenever the receiver prefers $a = 1$ at realization 1, she also prefers $a = 0$ at realization 0, because her ex ante optimal action is $a = 0$. This implies that μ is incentive-compatible for the receiver.

The next lemma ensures that the designer can without loss of generality focus on incentive-compatible straightforward signals that the sender has no incentive to garble. As the proof suggests, the result does not rely on the assumption of binary action.

Lemma 2. *A payoff vector (u, v) is implementable if and only if the designer can implement it using a straightforward signal μ with the following properties:*

1. μ is incentive-compatible for the receiver.
2. If the designer chooses μ , the sender prefers to choose μ given the receiver’s optimal behavior.

Proof. I show the “only if” part. Let μ be the designer’s choice of a signal to implement (u, v) . Consider an equilibrium of the subgame following μ , and let $\mu^*(a|\omega)$ denote the probability with which the receiver takes action a in each state ω . Note that we can view $\mu^* : \omega \mapsto (\mu^*(0|\omega), \mu^*(1|\omega))$

⁶A similar calculation appears in [Alonso and Câmara \(2016\)](#).

as an incentive-compatible straightforward signal, which draws realization a with probability $\mu^*(a|\omega)$ at each state ω .⁷ Now, the set of signals available to the sender is smaller when the designer chooses μ^* , which is less informative than μ ; however, the sender can achieve the same payoff as before (i.e., the payoff under μ) by choosing μ^* . Thus, the sender has no incentive to garble μ^* . \square

The next lemma connects the incentive compatibility for the receiver with that for the sender.

Lemma 3. *If (u, v) is implementable and $u > \underline{u}$, then $v \geq \underline{v}$.*

Proof. Suppose (u, v) is implementable and let $\mu \in \mathcal{S}$ be a straightforward signal satisfying $(\mathbb{E}_\mu[u], \mathbb{E}_\mu[v]) = (u, v)$ and Points 1 and 2 of Lemma 2. Suppose to the contrary that $u > \underline{u}$ but $v < \underline{v}$. First, $v < \underline{v}$ implies that there is a realization $s \in \{0, 1\}$, say $s = 0$, at which the receiver strictly prefers $a = 0$ but the sender strictly prefers $a = 1$. Second, $u > \underline{u}$ implies that the receiver strictly prefers $a = 1$ at realization 1. These observations imply that the sender can garble μ to strictly increase his payoff: He can choose signal μ' that sends realization 1 with probability $\varepsilon > 0$ when μ is supposed to send 0. For a small ε , the receiver prefers to follow realizations, which strictly increases the sender's payoff. This contradicts Point 2 of Lemma 2. \square

Proof of Theorem 1. I prove that any $(u, v) \in A_1 \cup A_2$ is implementable. Take any $(u, v) \in A_1 \cup A_2$ and let $\mu \in \mathcal{S}$ be an incentive-compatible straightforward signal satisfying $(\mathbb{E}_\mu[u], \mathbb{E}_\mu[v]) = (u, v)$. Such μ exists because of $u \geq \underline{u}$ and Lemma 1.

First, suppose $(u, v) \in A_1$. Because $u > \underline{u}$ and $v \geq \underline{v}$, by Lemma 1, both the receiver and the sender prefer action s after observing realization $s \in \{0, 1\}$. That is, if the designer chooses μ , the sender prefers to choose μ and the receiver prefers to follow realizations. Thus, (u, v) is implementable.

Second, suppose $(u, v) \in A_2$. If $(u, v) = (\underline{u}, v_{NO})$, the designer can implement (u, v) by choosing μ' that discloses no information (i.e., $\mu'(\omega)$ is independent of $\omega \in \Omega$).

Suppose $(u, v) \in A_2$, where $u = \underline{u}$ and $v > v_{NO}$. I consider two cases. One is when the receiver is indifferent between two actions at both signal realizations drawn by μ . In this case, the designer can implement (u, v) by choosing μ : the sender has no incentive to garble μ because the receiver chooses the best action for the sender at each realized posterior, as a result of her tie-breaking rule.

⁷Proposition 1 of Kamenica and Gentzkow (2011) implies that μ^* is incentive-compatible for the receiver.

The other case is when the receiver strictly prefers to follow one signal realization and is indifferent between two actions at the other realization. Without loss of generality, suppose that the receiver strictly prefers $a = 1$ at realization 1 and she takes $a = 0$ at realization 0 being indifferent between $a = 0$ and $a = 1$. Note that the only case we have to consider is when the sender strictly prefers $a = 0$ at realization 1. Consider whether the sender has an incentive to garble μ . Take any μ' that is a straightforward signal and a garbling of μ . If μ' induces the receiver to take $a = 1$ when she is supposed to take $a = 0$ at μ , then μ' does not increase the sender's payoff. If μ' induces the receiver to take $a = 0$ when she is supposed to take $a = 1$ at μ , then the receiver's payoff from μ' is strictly less than \underline{u} . This implies that μ' cannot be incentive-compatible for the receiver, who can always ignore information and obtain \underline{u} . To sum up, the sender never benefits from garbling μ . Thus, the designer can implement (u, v) by choosing μ . Finally, we do not need to consider the case in which the receiver's incentive is strict at both signal realizations, because it contradicts $\mathbb{E}_\mu[u] = \underline{u}$.

Next, I prove that if (u, v) is implementable, then $(u, v) \in A_1 \cup A_2$. Take any implementable $(u, v) \in \mathcal{F}$ and let μ be $(\mathbb{E}_\mu[u], \mathbb{E}_\mu[v]) = (u, v)$. By [Lemma 2](#), assume that μ is an incentive-compatible straightforward signal for the receiver and that the sender has no incentive to garble μ .

First, $u \geq \underline{u}$ holds because the receiver can always ignore information. Second, $v \geq v_{NO}$ must hold, because the sender can always choose to disclose no information regardless of the designer's choice. If $u = \underline{u}$, then $(u, v) \in A_2$ holds; if $u > \underline{u}$, then [Lemma 3](#) implies $v \geq \underline{v}$ and thus $(u, v) \in A_2$.

Finally, if the designer can publicly randomize signals and the receiver observes which signal is realized, the designer can implement any point in $\text{conv}(A_1 \cup A_2)$ by the definition of convex hull. \square

4 Implementable Outcomes on the Pareto Frontier

In this section, I characterize the designer's optimal strategy more explicitly, assuming that the designer's payoff is increasing in the payoffs of the sender and the receiver. For instance, the designer might care about the receiver's payoff alone, or the sum of the sender's and the receiver's

payoffs. Given such objectives, how should the designer restrict the sender’s information?

The analysis here consists of three steps. First, I characterize the set of implementable outcomes on the “Pareto frontier” \mathcal{P} of feasible outcomes \mathcal{F} . Second, I show that whenever the designer’s payoff is nondecreasing in each player’s payoffs, the designer implements an outcome in the set $\mathcal{P} \cap (A_1 \cup A_2)$ derived in the first step. As corollaries, I obtain the designer’s strategy when it maximizes the receiver’s welfare or social welfare.

Define the Pareto frontier \mathcal{P} of \mathcal{F} as follows.⁸

$$\mathcal{P} := \{(u, v) \in \mathcal{F} : \nexists (u', v') \in \mathcal{F} \text{ s.t. } u' \geq u, v' \geq v, \text{ and } (u, v) \neq (u', v')\}.$$

Next, define the set U of the receiver’s feasible payoffs by

$$U := \{u \in \mathbb{R} : \exists v, (u, v) \in \mathcal{F}\}$$

and function $f : U \rightarrow \mathbb{R}$ by

$$f(u) := \max \{v \in \mathbb{R} : (u, v) \in \mathcal{F}\}. \quad (4)$$

$f(u)$ is the maximum payoff of the sender consistent with the receiver’s payoff u . Note that $(u, v) \in \mathcal{P}$ implies $v = f(u)$. $f(\cdot)$ is concave because \mathcal{F} is a convex set.

The following result characterizes $\mathcal{P} \cap (A_1 \cup A_2)$ —the set of implementable payoff vectors on the Pareto frontier. To state the result concisely, I define two types of Bayesian persuasion: the *original game* and the *flipped game*. In the original game, the sender can choose any signal and the receiver chooses an action. In the flipped game, the receiver of the original game, who has payoff function $u(\cdot)$, chooses a signal. Then, the sender of the original game, who has payoff function $v(\cdot)$, chooses an action to maximize his own payoff. In order to pin down equilibrium payoffs in these games, I assume that, in the original game, the receiver breaks ties in favor of the sender, who breaks ties in favor of the receiver given her best responses. I impose the analogous tie-breaking rule for the flipped game with “the receiver” and “the sender” flipped.

Theorem 2. *Let u_O and u_F be the receiver’s equilibrium payoffs in the original and flipped games,*

⁸ \mathcal{P} can be different from the Pareto frontier of implementable outcomes.

respectively. Then, the set of implementable outcomes on the Pareto frontier is as follows.

$$\mathcal{P} \cap (A_1 \cup A_2) = \begin{cases} \{(u, f(u)) : u_O \leq u \leq u_F\} & \text{if } u_O < u_F \\ \{(u_O, f(u_O))\} & \text{if } u_O \geq u_F, \end{cases}$$

where $A_1 \cup A_2$ is defined in [Theorem 1](#), and $f(u)$ is the maximum payoff of the sender consistent with u as defined by (4).

Proof. First, suppose that $u_O \geq u_F$ and that the designer can implement $(u^*, v^*) \in \mathcal{P}$. $u^* > u_O$ cannot hold because it implies $u^* > \underline{u}$ and thus $v^* \geq \underline{v}$ for (u^*, v^*) to be in A_1 . However, this means that the receiver could attain $u^* > u_O \geq u_F$ in the flipped game, which is a contradiction.⁹ Also, $u^* < u_O$ cannot hold because it implies that $v^* > f(u_O)$ as $(u^*, v^*) \in \mathcal{P}$, but $f(u_O)$, which is the sender's equilibrium payoff in the original game, gives the sender the maximum payoff among implementable payoff vectors. Thus, $u^* = u_O$ holds; moreover, because $f(u_O)$ gives the sender the maximum payoff among all $(u, v) \in \mathcal{F}$ with $u = u_O$, $(u^*, v^*) = (u_O, f(u_O))$. Finally, the designer can implement $(u_O, f(u_O))$ by not limiting the sender's information, because $(u_O, f(u_O))$ is the equilibrium payoff profile of the original game. This shows $\mathcal{P} \cap (A_1 \cup A_2) = \{(u_O, f(u_O))\}$.

Second, suppose that $u_O < u_F$ and that the designer can implement $(u^*, v^*) \in \mathcal{P}$. $u^* < u_O$ cannot hold for the same reason as above. Also, $u^* > u_F$ cannot hold: It implies $u^* > u_O \geq \underline{u}$ and thus $v^* \geq \underline{v}$ by [Lemma 3](#). This implies $(u^*, v^*) \in A_1$. However, u_F maximizes u among all payoff vectors (u, v) in A_1 , because u_F is a solution of $\max_{(u,v) \in \mathcal{F}} u$ subject to the sender's incentive constraint $v \geq \underline{v}$ (recall [Lemma 1](#)). This is a contradiction. Thus, $u^* \in [u_O, u_F]$. $v^* = f(u^*)$ holds from $(u^*, v^*) \in \mathcal{P}$.

Next, suppose $(u^*, v^*) = (u^*, f(u^*))$ and $u^* \in [u_O, u_F]$. I prove $(u^*, v^*) \in \mathcal{P}$. Note that $f(\cdot)$ is strictly decreasing on $u \geq u_O$: If there is some $(\delta_1, \delta_2) \subset [u_O, \max U]$ on which f is nondecreasing, then $f(u)$ must be nondecreasing on $[u_O, \delta_2]$ because f is concave. However, this implies that there is some $(u, v) \in \mathcal{F}$ such that $u > u_O$ and $v \geq v_O$. This contradicts that (u_O, v_O) is a solution of the original game with tie-breaking. Now, if $f(\cdot)$ is strictly decreasing on $u \geq u_O$, then for any $(u, v) \in \mathcal{F}$, $u > u^*$ implies $v \leq f(u) < f(u^*) = v^*$. Thus, $(u^*, v^*) \in \mathcal{P}$.

⁹Namely, if the receiver were to persuade the sender, the receiver could attain u^* with a straightforward signal that is incentive-compatible for the sender. Such a signal exists because $v^* \geq \underline{v}$.

It remains to show that the designer can implement $(u^*, v^*) = (u^*, f(u^*))$ and $u^* \in [u_O, u_F]$, i.e., $(u^*, v^*) \in A_1 \cup A_2$. If $u^* = u_O$, then $v^* = f(u_O)$, which is implementable by not limiting the sender's information. If $u^* \in (u_O, u_F]$, then $u^* > \underline{u}$ holds. Also, $v^* \geq f(u_F)$ holds because $f(\cdot)$ is decreasing on $u \geq u_O$ as shown above. Because $f(u_F)$ is the sender's equilibrium payoff in the flipped game, $f(u_F) \geq \underline{v}$ holds. By [Theorem 1](#), $(u^*, v^*) \in A_1$. \square

Figures 1 and 2 correspond to the first and second cases of [Theorem 2](#), respectively. First, consider [Figure 1](#). If the sender persuades the receiver, the sender chooses a signal to maximize his payoff $\mathbb{E}_\mu[v]$ subject to the constraint $\mathbb{E}_\mu[u] \geq \underline{u}$. From the figure, we can derive $(u_O, v_O) = (0, 2/3)$. In the flipped game where the receiver persuades the sender, the receiver chooses a signal to maximize her payoff $\mathbb{E}_\mu[u]$ subject to the constraint $\mathbb{E}_\mu[v] \geq \underline{v}$. This implies $(u_F, f(u_F)) = (1/3, 1/3)$, which establishes $u_F > u_O$. Thus, the set of implementable outcomes on the Pareto frontier is the nondegenerate segment connecting these two points. In contrast, in [Figure 2](#), we can conclude $(u_O, v_O) = (0, 2/3)$ and $(u_F, f(u_F)) = (-1/3, 1)$ by the same argument as above. As $u_O \geq u_F$, the only implementable outcome on the Pareto frontier is (u_O, v_O) , which arises in the absence of the designer.

Note that [Theorem 2](#) alone may be insufficient to characterize the designer's equilibrium strategy. For instance, if the designer aims to maximize $u+v$ but the first-best outcome ($\arg \max_{(u,v) \in \mathcal{F}} u+v$) is outside of $\mathcal{P} \cap (A_1 \cup A_2)$, the designer might instead implement an inefficient outcome. The next lemma shows that this never happens.

Lemma 4. *Suppose that the designer's payoff is nondecreasing in the expected payoffs of the sender and the receiver. Then, there is an equilibrium in which the designer implements an outcome in $\mathcal{P} \cap (A_1 \cup A_2)$.*

Proof. I derive a contradiction by assuming that no points in $\mathcal{P} \cap (A_1 \cup A_2)$ maximize the designer's payoff. Suppose that the designer implements $(u^*, v^*) \notin \mathcal{P}$, which is Pareto-dominated by some $(u_D, v_D) \in \mathcal{P}$. If $(u^*, v^*) \in A_1$, then $(u_D, v_D) \in A_1$ as $u_D \geq u^* > \underline{u}$ and $v_D \geq v^* \geq \underline{v}$. By [Theorem 1](#), $(u_D, v_D) \in A_1$ is implementable and gives the designer a weakly greater payoff, which is a contradiction.

Next, suppose $(u^*, v^*) \in A_2$, which implies $u^* = \underline{u}$. There are three cases to consider. First, suppose $v^* = \max_{u \in U} f(u)$. In other words, v^* gives the sender the maximum payoff among all

feasible payoffs. This implies $v_D = v^* \geq \underline{v}$. As (u_D, v_D) Pareto-dominates (u^*, v^*) , $u_D > u^* = \underline{u}$ must hold. This implies $(u_D, v_D) \in A_1$, which is a contradiction.

Second, suppose that $v^* < \max_{u \in U} f(u)$ and that all elements of convex set $\arg \max_{u \in U} f(u)$ are strictly smaller than u^* . Note that for any $(u, v) \in \mathcal{F}$ such that $u > u^*$, $v < f(u^*)$ holds because $f(\cdot)$ is decreasing for $u \geq \max(\arg \max_{u \in U} f(u))$. Thus, $(u^*, f(u^*)) \in \mathcal{P}$. Also, because $u^* = \underline{u}$, $(u^*, f(u^*))$ corresponds to the equilibrium of the original game. Thus, the designer can implement $(u^*, f(u^*)) \in \mathcal{P}$ by not limiting the sender's information. However, $(u^*, f(u^*))$ gives the sender a weakly greater payoff than (u^*, v^*) . This is a contradiction.

Finally, suppose that $v^* < \max_{u \in U} f(u)$ and that all elements of convex set $\arg \max_{u \in U} f(u)$ are strictly greater than u^* . Define $u^{**} := \max(\arg \max_{u \in U} f(u))$ and $v^{**} = f(u^{**})$. Then, $u^{**} > u^* \geq \underline{u}$ and $v^{**} \geq \max(\underline{v}, v^*)$ hold. This implies that $(u^{**}, v^{**}) \in \mathcal{P} \cap A_1$ is implementable and gives the sender a weakly greater payoff, which is a contradiction. \square

I consider two objectives of the designer: maximizing the receiver's payoff or maximizing the sum of the sender's and the receiver's payoffs. Recall that u_O and u_F are the receiver's equilibrium payoffs in the original and flipped games, respectively.

Corollary 2. *Suppose that the designer's objective is to maximize the receiver's payoff. Then, an equilibrium outcome is on the Pareto frontier and the receiver obtains a payoff of $\max(u_O, u_F)$. If $u_O \geq u_F$, then the designer can maximize the receiver's payoff by not limiting the sender's information. If $u_O < u_F$, then the designer's choice is equal to an equilibrium strategy of the receiver in the flipped game.*

Proof. The only part that does not directly follow from [Theorem 2](#) is the last sentence. Let μ_F denote a straightforward signal that is an equilibrium strategy of the receiver in the flipped game. $v_F := \mathbb{E}_{\mu_F}[v] \geq \underline{v}$ holds because the sender chooses an action in the flipped game. Also, $u_F \geq \underline{u}$ holds because $u_F > u_O \geq \underline{u}$. By [Lemma 1](#), if the designer chooses μ_F , then the sender prefers to choose μ_F and the receiver prefers to follow its signal realizations. Thus, the designer can implement (u_F, v_F) by μ_F . \square

[Corollary 2](#) states that limiting the sender's information to maximize the receiver's payoff causes no efficiency loss, and thus the receiver-optimal information restriction moves an equilibrium outcome along the Pareto frontier. Also, the result shows that we can solve a two-stage

disclosure game among the designer, the sender, and the receiver using a pair of Bayesian persuasion games between two players. In particular, whenever restricting the sender's information can benefit the receiver, the receiver-optimal way is as if the receiver discloses information and delegates the sender to choose an action.

To state the next result, let $A_{SW} := \arg \max_{(u,v) \in \mathcal{F}} u + v$ denote the set of all payoff vectors that maximize “social welfare” defined by the sum of payoffs of the sender and the receiver. Define $\bar{u}_{SW} := \max \{u : \exists v, (u, v) \in A_{SW}\}$ and $\underline{u}_{SW} := \min \{u : \exists v, (u, v) \in A_{SW}\}$.

Corollary 3. *Suppose that the designer's objective is to maximize social welfare. If $u_O \geq u_F$, then it is optimal for the designer not to limit the sender's information. If $u_O < u_F$, then there are three cases:*

1. *If $u_F < \underline{u}_{SW}$, then it is optimal for the designer to choose the receiver's equilibrium signal of the flipped game, which yields social welfare $u_F + f(u_F)$.*
2. *If $\bar{u}_{SW} < u_O$, then it is optimal for the designer not to limit the sender's information, which yields social welfare $u_O + f(u_O)$.*
3. *Otherwise, the designer attains social welfare $\max_{(u,v) \in \mathcal{F}} u + v$.*

Proof. The case of $u_O \geq u_F$ and Point 3 directly follow from [Theorem 2](#). As Points 1 and 2 are symmetric, I prove Point 1. If $u_F < \underline{u}_{SW}$, then $u_F + f(u_F) < \underline{u}_{SW} + f(\underline{u}_{SW})$, or equivalently, $f(u_F) - f(\underline{u}_{SW}) < \underline{u}_{SW} - u_F$. As $f(\cdot)$ is concave, for any $(u, f(u)) \in \mathcal{P}$ such that $u \in [u_O, u_F]$, we get $f(u) - f(u_F) \leq u_F - u$, which implies $u + f(u) \leq u_F + f(u_F)$. Thus, $(u_F, f(u_F))$ maximizes social welfare among all implementable payoff profiles. \square

Corollaries 2 and 3 are useful only if the original and flipped games are easy to solve. I show that it is the case so long as there is a unique welfare-maximizing outcome. Let $\mu_\alpha \in \mathcal{S}$ denote a straightforward signal that maximizes $\mathbb{E}_\mu[\alpha u + (1 - \alpha)v]$ among all signals. Assume that such μ_α is unique. Note that $\mu_\alpha(1|\omega)$, the probability that μ_α recommends action 1 at state ω , is 1 whenever $\alpha u(\omega) + (1 - \alpha)v(\omega) > 0$ and 0 whenever $\alpha u(\omega) + (1 - \alpha)v(\omega) < 0$. The receiver's payoff $\mathbb{E}_{\mu_\alpha}[u]$ and the sender's payoff $\mathbb{E}_{\mu_\alpha}[v]$ are increasing and decreasing in α , respectively. Define α_O and α_F

as follows.

$$\alpha_O := \min \{ \alpha \in [0, 1] : \mathbb{E}_{\mu_\alpha}[u] \geq \underline{u} \},$$

$$\alpha_F := \max \{ \alpha \in [0, 1] : \mathbb{E}_{\mu_\alpha}[v] \geq \underline{v} \}.$$

μ_{α_O} and μ_{α_F} are equilibrium signals in the original and flipped games, respectively. The reason is as follows. For example, the sender in the original game maximizes $\mathbb{E}_\mu[v]$ subject to the receiver's incentive compatibility $\mathbb{E}_\mu[u] \geq \underline{u}$. This enables the sender to focus on outcomes on the Pareto frontier. Given the receiver's incentive, μ_{α_O} maximizes the sender's payoff among the signals whose outcomes are on the Pareto frontier. Thus, μ_{α_O} is the sender's equilibrium signal in the original game.

The next example applies this observation to a simple linear environment.

Example 3. Suppose that ω is uniformly distributed on $[-3, 2]$, $u(\omega) = \omega$, and $v(\omega) = 2 + \omega$.¹⁰ Note that μ maximizes $\mathbb{E}_\mu[\alpha u + (1 - \alpha)v] = \mathbb{E}_\mu[\omega + 2(1 - \alpha)]$ for some $\alpha \in [0, 1]$ if and only if μ has the following cutoff structure: There is a unique $c \in [-2, 0]$ such that the receiver takes $a = 1$ and $a = 0$ if $\omega > c$ and $\omega < c$, respectively. Let $\mu(c)$ denote such a straightforward signal given cutoff c . Each player's payoff from $\mu(c)$ is as follows.

$$\mathbb{E}_{\mu(c)}[u] = \frac{2 - c}{5} \cdot \frac{2 + c}{2} = \frac{4 - c^2}{10},$$

$$\mathbb{E}_{\mu(c)}[v] = 2 \cdot \frac{2 - c}{5} + \frac{4 - c^2}{10} = \frac{12 - 4c - c^2}{10}.$$

Thus, in (u, v) -space, the Pareto frontier \mathcal{P} is the arc connecting $(\mathbb{E}_{\mu(0)}[u], \mathbb{E}_{\mu(0)}[v]) = (2/5, 6/5)$ and $(\mathbb{E}_{\mu(-2)}[u], \mathbb{E}_{\mu(-2)}[v]) = (0, 8/5)$. In the original game, the sender chooses the lowest $\alpha \in [0, 1]$, or equivalently, the lowest cutoff $c = -2(1 - \alpha) \in [-2, 0]$ such that $\mathbb{E}_{\mu(c)}[u] \geq \underline{u} = 0$. This implies $c = -2$. Thus, $(\mathbb{E}_{\mu(-2)}[u], \mathbb{E}_{\mu(-2)}[v]) = (0, 8/5)$ is the equilibrium outcome of the original game. In the flipped game, the receiver (who discloses information) chooses as high $c \in [-2, 0]$ as possible subject to the constraint that $\mu(c)$ is incentive-compatible for the sender. This leads to $c = -1$,

¹⁰To motivate the example, imagine that ω is the product value for a consumer relative to the consumer's outside option. The seller's payoff from purchase, $2 + \omega$, partly reflects the consumer's welfare (say, because the seller has a reputational concern), but the seller obtains the payment of 2 from the purchase.

which is the highest cutoff such that the sender prefers $a = 0$ after knowing $\omega < c$. Thus, the equilibrium payoff vector of the flipped game is $(\mathbb{E}_{\mu(-1)}[u], \mathbb{E}_{\mu(-1)}[v]) = (3/10, 3/2)$.

Figure 3 presents the Pareto frontier, the implementable efficient outcomes, and the relevant payoff vectors. By Corollary 2, if the payoff of the designer is equal to that of the receiver, the designer chooses $\mu(-1)$ to achieve a payoff of $3/10$. Thus, information restriction strictly benefits the receiver though it does not achieve the receiver-optimal outcome $(2/5, 6/5)$. Finally, signal $\mu(-1)$ also maximizes the social welfare $\mathbb{E}_{\mu}[2 + 2\omega]$. Thus, in this example, the designer can globally maximize social welfare by restricting the sender's information.

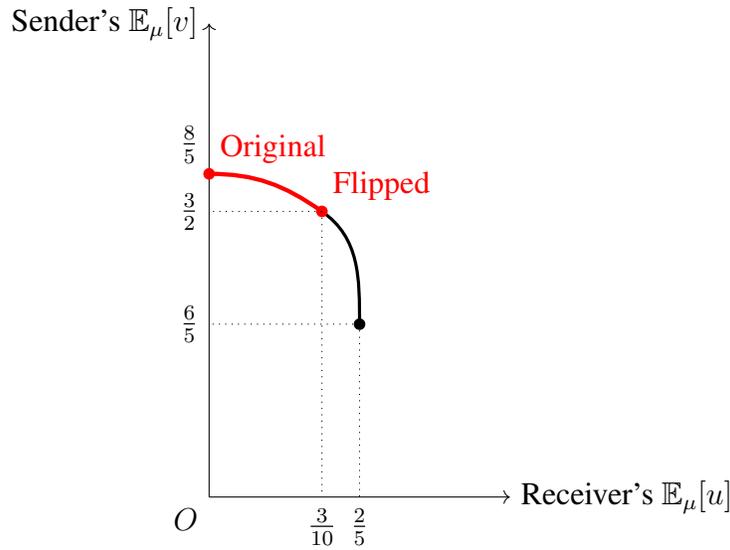


Figure 3: \mathcal{P} (union of black and red lines) and $\mathcal{P} \cap (A_1 \cup A_2)$ (red line)

5 Discussion: General Action Space

Theorem 1 does not extend to the case in which the receiver has more than two actions, and it is beyond the scope of this paper to consider such a case. (Appendix B provides a three-action example in which the results fail.) To the best of my knowledge, a tractable solution technique for general Bayesian persuasion problems has yet to be developed, which makes it challenging to analyze how the payoffs of each player depend on the sender's information.

One exception is when the state is binary. In this case, regardless of the action space, the designer cannot increase the receiver's payoff by limiting the sender's information. Thus, the

designer, who puts a nonnegative weight on each player's payoff, chooses not to limit the sender's information.

Proposition 1. *Take any Bayesian persuasion such that the state is binary and the sender has a unique equilibrium strategy. The designer can maximize the receiver's payoff by not restricting the sender's information.*

See [Appendix C](#) for the proof. A key step is to show that if the sender has no incentive to garble signal μ , then μ must be less informative than the equilibrium signal of the original game. Thus, restricting the sender's information simply reduces the amount of information revealed to the receiver.

6 Conclusion

In this paper, I study the problem of restricting the sender's information in Bayesian persuasion. Assuming that the receiver has a binary choice, I consider arbitrary restrictions and characterize the set of all outcomes that can arise in equilibrium. Thus, the paper gives a comprehensive answer to the question of how the sender's information affects the outcome of Bayesian persuasion with binary action. In particular, the result enables us to characterize an information restriction that maximizes the receiver's welfare or social welfare. I show that we can determine whether limiting the sender's information can improve these welfare criteria by solving the original Bayesian persuasion and the flipped version of it, in which the receiver discloses information and the sender chooses an action.

References

- Alonso, Ricardo and Odilon Câmara (2016), "Persuading voters." *American Economic Review*, 106(11), 3590–3605.
- Ambrus, Attila, Eduardo M Azevedo, and Yuichiro Kamada (2013), "Hierarchical cheap talk." *Theoretical Economics*, 8(1), 233–261.

- Bergemann, Dirk and Stephen Morris (2016), “Information design, Bayesian persuasion, and Bayes correlated equilibrium.” *American Economic Review*, 106(5), 586–91.
- Blackwell, David (1953), “Equivalent comparisons of experiments.” *The Annals of Mathematical Statistics*, 24(2), 265–272.
- Crawford, Vincent P and Joel Sobeli (1982), “Strategic information transmission.” *Econometrica*, 50(6), 1431–1451.
- Fischer, Paul E and Phillip C Stocken (2001), “Imperfect information and credible communication.” *Journal of Accounting Research*, 39(1), 119–134.
- Frug, Alexander (2016), “Dynamic cheap talk with static informational control.” *Economics Letters*, 143(C), 118–120.
- Gentzkow, Matthew and Emir Kamenica (2017), “Competition in persuasion.” *The Review of Economic Studies*, 84(1), 300–322.
- Ivanov, Maxim (2010), “Communication via a strategic mediator.” *Journal of Economic Theory*, 145(2), 869–884.
- Ivanov, Maxim (2015), “Dynamic information revelation in cheap talk.” *The BE Journal of Theoretical Economics*, 15(2), 251–275.
- Kamenica, Emir and Matthew Gentzkow (2011), “Bayesian persuasion.” *American Economic Review*, 101(6), 2590–2615.
- Kolotilin, Anton (2015), “Experimental design to persuade.” *Games and Economic Behavior*, 90(C), 215–226.
- Li, Fei and Peter Norman (2018), “On Bayesian persuasion with multiple senders.” *Economics Letters*, 170(C), 66–70.
- Rayo, Luis and Ilya Segal (2010), “Optimal information disclosure.” *Journal of Political Economy*, 118(5), 949–987.
- Taneva, Ina A (2015), “Information design.” *ESE Discussion Papers*.

Appendix

A Proof of Lemma 1

Proof. Take any straightforward signal μ . If the receiver follows μ , her ex ante expected payoff is $\mathbb{E}_\mu[u] := \mathbf{P}_\mu(1) \cdot \int_\Omega u(\omega) d\mu(\omega|1)$. Here, given $s \in \{0, 1\}$, $\mathbf{P}_\mu(s)$ is the ex ante probability that μ sends signal realization s , and $\mu(\omega|s)$ is the conditional probability of state ω after observing signal realization s . The receiver's payoff from $a = 1$ conditional on signal realization 1 is

$$\int_\Omega u(\omega) d\mu(\omega|1) = \frac{\mathbb{E}_\mu[u]}{\mathbf{P}_\mu(1)}, \quad (5)$$

and the payoff from $a = 1$ conditional on observing 0 is

$$\int_\Omega u(\omega) d\mu(\omega|0) = \int_\Omega u(\omega) d\left(\frac{b_0(\omega) - \mathbf{P}_\mu(1)\mu(\omega|1)}{\mathbf{P}_\mu(0)}\right) = \frac{1}{\mathbf{P}_\mu(0)} \left(\int_\Omega u(\omega) db_0(\omega) - \mathbb{E}_\mu[u] \right).$$

The first equality follows from the law of total probability $\mathbf{P}_\mu(1)\mu(\cdot|1) + \mathbf{P}_\mu(0)\mu(\cdot|0) = b_0(\cdot)$, and the second equality is from (5).

Given these expressions, the receiver prefers to follow straightforward signal μ if and only if

$$\frac{\mathbb{E}_\mu[u]}{\mathbf{P}_\mu(1)} \geq 0 \geq \frac{1}{\mathbf{P}_\mu(0)} \left(\int_\Omega u(\omega) db_0(\omega) - \mathbb{E}_\mu[u] \right),$$

which reduces to

$$\mathbb{E}_\mu[u] \geq \underline{u} := \max \left(\int_\Omega u(\omega) db_0(\omega), 0 \right). \quad (6)$$

Note that (6) is the receiver's IC even if μ sends only one signal realization. In this case, as long as (6) holds, the receiver weakly prefers to follow the realization. \square

B Example for $|A| = 3$

Suppose that the receiver has three actions $A = \{a_1, a_2, a_3\}$, and there are two equally likely states $\Omega = \{\omega_1, \omega_2\}$. Table 3 summarizes the payoffs $(u(a, \omega), v(a, \omega))_{(a, \omega) \in A \times \Omega}$.

Note that both the receiver and the sender (weakly) prefer a_2 at the prior belief. Thus, each player obtains a payoff of 1 if he or she chooses an action under no information. Now, consider

	a_1	a_2	a_3
ω_1	2, 2	1, 1	0, -2
ω_2	-2, -2	1, 1	2, -2

Table 3: Payoffs (The first coordinate is the receiver’s payoff)

the state-contingent action plan (say μ^*) of taking a_1 and a_2 at ω_1 and ω_2 , respectively. This leads to the payoffs of $3/2$ to the receiver and the sender, which Pareto-dominates $(u, v) = (1, 1)$. I show that the designer cannot implement $(3/2, 3/2)$. Indeed, for the receiver to follow μ^* , she has to be fully informed of the state. However, if the receiver has full information, she would take a_1 and a_3 at ω_1 and ω_2 , respectively. This gives the sender a payoff of 0, which is lower than the payoff he can secure by disclosing no information. Thus, although $(3/2, 3/2)$ is feasible and satisfies $3/2 > 1 = \underline{u}$ and $3/2 > 1 = \underline{v}$, it is not implementable. This shows that [Theorem 1](#) does not extend to the case of $|A| > 2$. Also, other results such as [Corollary 2](#) do not extend either: The efficient payoff vector $(3/2, 3/2)$ arises in the equilibrium of the flipped game; however, this does not imply the designer can improve the receiver’s payoff by restricting the sender’s information.

C Proof of Proposition 1

Suppose $\Omega = \{0, 1\}$. We can identify $\Delta(\Omega)$ with the unit interval $[0, 1]$, where each $b \in [0, 1]$ represents the probability of $\omega = 1$. Let $b_0 \in [0, 1]$ denote the prior belief. Also, let $a(b) \in A$ denote the receiver’s best response given a belief b . Then, define $v(b) := bv(a(b), 1) + (1 - b)v(a(b), 0)$ as the sender’s expected payoff given b . As the state is binary, it is convenient to write a signal in terms of its distribution over posteriors $\tau \in \Delta(\Delta(\Omega)) = \Delta([0, 1])$ that is *Bayes’ plausible*, i.e., $\int_0^1 b d\tau(b) = b_0$. For any τ , let $Supp(\tau)$ denote the set of all posteriors that can arise with a positive probability under τ .

Proof of Proposition 1. Suppose to the contrary that the designer can restrict the sender’s information and give the receiver a strictly greater payoff than under the original game. Let $\tau_{BP} \in \Delta(\Delta(\Omega))$ denote the sender’s equilibrium strategy in the original game. Note that $|Supp(\tau_{BP})| \leq 2$.¹¹ $Supp(\tau_{BP}) = \{b_0\}$ cannot hold, because the sender would then disclose no information re-

¹¹This relies on the following general result: For any $A \subset \mathbb{R}^d$, if x is in the boundary of the convex hull of A , then x is a convex combination of at most d points of the boundary of A . Furthermore, even if the sender has multiple optimal signals, he can find one that generates two posteriors and maximizes the receiver’s payoff. Thus, we can

ardless of the designer's choice. Thus, suppose that $Supp(\tau_{BP}) = \{b_1, b_2\}$ with $b_1 < b_0 < b_2$. Suppose that there is some τ^* such that the sender chooses τ^* when the designer chooses τ^* , and the receiver's payoff is strictly greater under τ^* than τ_{BP} . Let $Supp(\tau^*) = \{b_1^*, b_2^*\}$ with $b_1^* < b_0 < b_2^*$. Because τ^* cannot be weakly less informative than τ_{BP} , either $b_1^* < b_1$ or $b_2^* > b_2$ holds; otherwise, each element of $Supp(\tau^*)$ can be expressed as a convex combination of b_1 and b_2 , and this implies that τ^* is less informative than τ_{BP} . Without loss of generality, suppose that $b_1^* < b_1$. In this case, b_1 can be expressed as a convex combination of b_1^* and b_2^* . Thus, under the information restriction τ^* , the sender can choose a signal τ' such that $Supp(\tau') = \{b_1, b_2^*\}$.

Now, I show that the sender strictly prefers τ' to τ^* . Since $v(b_1)$ is on the boundary of the convex hull of the graph of v , it satisfies

$$v(b_1) > \frac{b_2^* - b_1}{b_2^* - b_1^*} v(b_1^*) + \frac{b_1 - b_1^*}{b_2^* - b_1^*} v(b_2^*).$$

This inequality is strict since τ_{BP} is a unique equilibrium strategy. Then, we obtain

$$\begin{aligned} \frac{b_2^* - b_0}{b_2^* - b_1^*} v(b_1^*) + \frac{b_0 - b_1^*}{b_2^* - b_1^*} v(b_2^*) &= \frac{b_2^* - b_0}{b_2^* - b_1} \cdot \left(\frac{b_2^* - b_1}{b_2^* - b_1^*} v(b_1^*) + \frac{b_1 - b_1^*}{b_2^* - b_1^*} v(b_2^*) \right) + \frac{b_0 - b_1}{b_2^* - b_1} v(b_2^*) \\ &< \frac{b_2^* - b_0}{b_2^* - b_1} v(b_1) + \frac{b_0 - b_1}{b_2^* - b_1} v(b_2^*). \end{aligned}$$

Thus, the sender strictly prefers to garble τ^* to obtain τ' , which is a contradiction. \square

assume $|Supp(\tau_{BP})| \leq 2$ even with our tie-breaking rule.