Bayesian Persuasion with Private Types

Tommaso Coen

October 15, 2020

Abstract

I study an extension to the simple Judge/Prosecutor example in Kamenica & Gentzkow (2011), where the Judge has a private type. The distribution of types (G) and the one of states (F) are commonly known. I provide a simple method to determine the optimal persuasion mechanism, under the assumption that G is single peaked. I then provide some partial results for general G's, for the case where the state space is non binary and for situations where there are multiple receivers who care about each other's actions.

Contents

1 Introduction

Take the standard bayesian persuasion example from Kamenica & Gentzkow (2011) [1] of a prosecutor trying to convince a judge that a defendant is guilty.

> The judge (Receiver) must choose one of two actions: to acquit or convict a defendant. There are two states of the world: the defendant

is either guilty or innocent. The judge gets utility 1 for choosing the just action (convict when guilty and acquit when innocent) and utility 0 for choosing the unjust action (convict when innocent and acquit when guilty). The prosecutor (Sender) gets utility 1 if the judge convicts and utility 0 if the judge acquits, regardless of the state. The prosecutor and the judge share a prior belief Pr (guilty) = 0.3.

I will consider a simple extension to this basic example: the judge assigns different utilities to convicting a guilty defendant (say utility in this case is $r \in [0, 1]$) and acquitting an innocent one (associated with utility $1 - r$). Throughout the analysis r is private information of the judge, with $r \sim G$.

I will study the problem first considering a sender (she) who can choose from a relatively small class of persuasion mechanisms, which I will call simple. Under these mechanisms, she sends a binary signal that only depends on the state, without acquiring any information on Receiver's type. I'll later move on to study the problem for a sender who can choose from a wider class of persuasion mechanisms, where she can ask receivers to report their types and condition her action recommendation on these as well as on the state. In the first case, I will be able to fully characterize the optimal persuasion mechanism as a function of the distribution of types using a concavification result that closely resembles the one in Kemenica & Gentzkow (2011). In the second case, I will be able to provide similar results only for single peaked distributions of types.

When studying settings such as the interaction between a judge and a prosecutor, where the two players know each other well, the assumption that types are commonly known might be justified. Anyway, there are several situations which could be modeled as a similar persuasion problem, in which such assumption would be absolutely irrealistic. Think for example about one off interactions, where the sender is matched repeatedly at random with receivers from an heterogeneous population.

EXAMPLE 1
A firm (Sender) produces m different goods which are either high or low quality. Consumers (Receivers) are randomly exposed to one of these products that they never tried before, they want to buy it if it's high quality while if it's low quality they prefer to stick to some other known outside option. They are heterogeneous in how much they value their outside option. The firm has to choose its branding strategy, which consists of branding some products as high quality and others as low quality (an example of a sender who is constrained by the nature of the problem to stick to simple mechanisms). Consumers know from their previous experience how frequently the firm brands low quality goods as high quality and vice versa. We are interested in examining how the optimal branding policy varies as a function of the share of high quality goods produced by the firm and the distribution of outside option values.

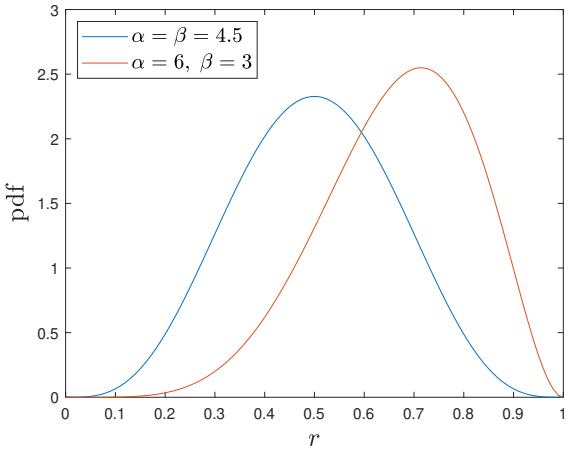


Figure 1: Beta densities for different values of the parameters

To make Example 1 more concrete, suppose that each new product of the firm is high quality with probability $p = 0.5$, that consumers receive a benefit of 1 or 0 depending on whether the product is high or low quality, and that benefits from outside options are distributed according to a beta with parameters $\alpha = \beta = 4.5$ (Figure 1). As I show below, the optimal branding strategy consists in never branding an high quality product as a low quality one, and to brand a low quality product as high quality about 11% of the times. This would increase the share of consumers who decide to try the product form 50% (Under no branding) to 63.5%.

It sounds intuitive that under an optimal strategy, the firm should never brand an high quality product as low quality. If α and β are greater than one (and more generally if the distribution of types is unimodal) this is indeed true for all values of p. Also, building our intuition on the Kamenica & Gentzkow example, we would expect no branding to be optimal if and only if p exceeds some threshold p^*. This also turns out to be true if the distribution of types is unimodal. In particular, for our example we have $p^* = 0.69$.

When $p < p^*$, optimal branding varies with p, but the posterior probabilities that consumers assign to the product being high quality conditional on its branding do not change. On the contrary, changes in the parameters of the distribution affect these posteriors as well as the optimal strategy. For example if in our example we change the parameters to $\alpha = 6$ and $\beta = 3$ (Figure 1), the firm should brand low quality products as high quality about 31% of the times. In this case, the share of consumers buying the product jumps from 14.5% to 53.4%! To give a more comprehensive look at how p and the shape of the distribution of types impact the value of presuasion (defined as the difference between

the share of consumers trying the new product under no branding and under optimal branding), refer to Figure 2.

For more general classes of distributions (beyond unimodal ones) the optimal strategy may change with p in a less straightforward and intuitive way. Also, for these distributions of types we may incur in counterintuitive results. For example an optimal branding strategy may prescribe to brand high quality products as low quality, or we might have that no branding is optimal for both high and low values of p.

The main purpose of the current paper is to show how we can derive these results using a concavification method which is substantially analogous to the one in Kamenica & Gentzkow (2011) [1] and is therefore closely related to their contribution. Under the assumption that Sender observes the utility of the receiver, [1] relates the optimal persuasion mechanism to the concave closure of the sender value function without restricting the utility functional form of Sender and Receiver as I do.

Papers focusing on private types for receivers are Rayo & Segal (2010) [5], Kolotilin & al (2017) [3], and Kolotilin (2018) [2]. Rayo & Segal assume through most of their paper that types are uniformly distributed on $[0, 1]$, and characterize optimal disclosure in a setting with two dimensional sender information. Also, they do not rely on any concavification procedure similar to the one used here and in Kamenica & Gentzkow. Kolotilin & al. look at a more general utility function for Sender than the one assumed here. Their main result is that persuasion mechanisms where Sender can condition the signal he selects on a Receiver's report about his type are equivalent (as far as Sender and Receiver expected utilities are concerned) to mechanisms where Sender can only condition his signal on the state. Kolotilin (2018) characterizes optimal information disclosure in a setting similar to the one in Kolotilin et al. (2017) using a linear programming approach. He shows that to check that a mechanism is optimal it suffices to show that no better mechanism can be obtained via some simple deviation from it.

2 Model

Let's go back to the judge/prosecutor example from the introduction. Let $u(\omega, r, a)$ denote the utility that a judge of type r gets by choosing action a in state ω. By choosing to acquit, the judge gets utility r if the defendant is innocent ($\omega = 0$) and zero otherwise, therefore $u(\omega, r, 0) = r \cdot (1 - \omega)$. Similarly, if he convicts, he gets utility $(1 - r)$ if $\omega = 1$ and zero otherwise, so that $u(\omega, r, 1) = (1 - r) \cdot \omega$.

We can obtain a convenient expression for a utility $\tilde{u}$ representing the preferences of the judge if we take $\tilde{u}(\omega, r, a) = u(\omega, r, a) - u(\omega, r, 0)$. In this case we have

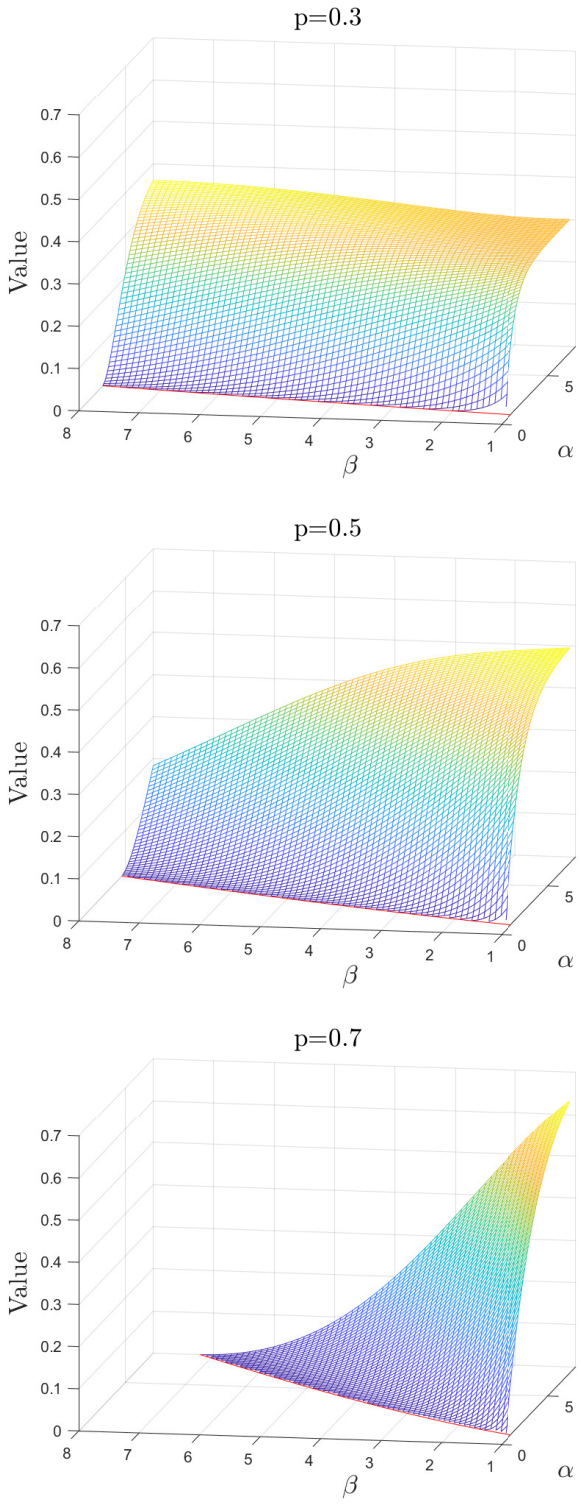


Figure 2: Value of persuasion for different values of p, when G is a $Beta(\alpha, \beta)$

$\tilde{u}(\omega, r, a) = a(\omega - r)$. This justifies the following setup.

2.1 Setup

There are two players: Sender and Receiver. Receiver makes a choice $a \in A = \{0, 1\}$. There are two payoff relevant random variables: the state of the world $\omega \in \Omega = \{0, 1\}$ and Receiver's private type $r \in R = [0, 1]$. Random variables ω and r are independent and have distributions F and G, where $p = \Pr(\omega = 1)$ characterizes F.

Utilities for Receiver and Sender are:

$$u(\omega, r, a) = a \cdot (\omega - r)$$
$$v(\omega, r, a) = a$$

Receiver's utility from inaction ($a = 0$) is normalized to zero, whereas his utility from action ($a = 1$) equals the benefit ω less the private cost r.

2.2 Persuasion mechanisms

In order to influence the action taken by Receiver, Sender could in principle design a mechanism that asks Receiver to report his private information and sends him a message conditional on his report and the realized state. A *persuasion mechanism* π asks Receiver to report $\hat{r} \in R$ and then recommends him to take one of two actions: for every $\omega \in \Omega$, it recommends action ($\hat{a} = 1$) with probability $\pi(\omega, \hat{r})$ and not to act ($\hat{a} = 0$) with complementary probability $1 - \pi(\omega, \hat{r})$, where $\pi : \Omega \times R \to [0, 1]$ is a measurable function. If $\pi(\omega, \hat{r})$ doesn't depend on $\hat{r}$, then π is a *simple mechanism*.

Receiver can manipulate a persuasion mechanism lying about his type or disobeying the recommendation. We can represent any manipulative strategy by a pair of functions (ρ, α) where $\rho : R \to R$ and $\alpha : R \times A \to A$, with the interpretation that if Receiver is of type r, he will report type $\rho(r)$ and take action $\alpha(r, \hat{a})$ whenever faced with recommendation $\hat{a}$. His interim utility if mechanism π is selected, $U_\pi(r)$, is the expected utility he gets under π if he chooses an optimal manipulative strategy

$$U_\pi(r) = \max_{\alpha, \rho} \int_\Omega \pi\big(\omega, \rho(r)\big) u\big(\omega, r, \alpha(1)\big) + \big(1 - \pi\big(\omega, \rho(r)\big)\big) u\big(\omega, r, \alpha(0)\big) dF(\omega).$$

Persuasion mechanism π is incentive compatible if it's optimal for Receiver to truthfully report his type and obediently follow the recommendation he receives. In this case we can write

$$U_\pi(r) = \int_\Omega \pi(\omega, r)(\omega - r) dF(\omega).$$

Two special cases: interim utility under no disclosure ($\underline{U}$) and under full disclosure ($\overline{U}$) will be used often below and deserve special notation

$$\underline{U}(r) = \max\{p - r, 0\} \quad \text{and} \quad \overline{U}(r) = \int_r^1 (\omega - r) dF(\omega) = p - r/p. \tag{1}$$

Let α_π and ρ_π attain the max above, the ex ante utility Sender gets from choosing persuasion mechanism π is

$$V(\pi) = \int_R \int_\Omega \alpha_\pi(1)\pi\big(\omega, \rho_\pi(r)\big) + \alpha_\pi(0)\big(1 - \pi\big(\omega, \rho_\pi(r)\big)\big) dF(\omega) dG(r)$$

If π is incentive compatible, we can write

$$V(\pi) = \int_R \int_\Omega \pi(\omega, r) dF(\omega) dG(r). \tag{2}$$

Finally, denoting by $\underline{V}$ the ex ante utility for Sender under the no-disclosure mechanism, we define the value of information design to the sender as the difference between the highest expected utility she can get by choosing optimally π and the expected utility she gets under the no disclosure mechanism

$$Value = \sup_\pi V(\pi) - \underline{V}.$$

3 Simple Mechanisms

Let's first study the case of a Sender whose choice is limited to simple mechanisms. I'll refer to this case as a *constrained Sender*, while I'll later call *unconstrained* a Sender who is allowed to choose any persuasion mechanism.
Simple mechanisms represent an interesting class of persuasion mechanisms (e.g. the branding choice in Example 1) and are especially simple to study. The following proposition shows how we can extend the concavification procedure from Kamenica & Gentzkow to our setting, and use it to find the optimal simple persuasion mechanism.

PROPOSITION 1. *Let $\overline{G}$ be the concavification [1] of G. Also, let p^0 be the greatest $q \leqslant p$ and p^1 be the smallest $q \geqslant p$ such that $\overline{G}(q) = G(q)$.*

1. *The value of information design for a constrained Sender is $\overline{G}(p) - G(p)$.*
2. *The simple mechanism π such that $p^0 = \mathbb{E}[\omega | m = 0]$ and $p^1 = \mathbb{E}[\omega | m = 1]$ is optimal among simple mechanisms.*

Proof.
Under no information disclosure types below p will choose $a = 1$ while types above p will choose $a = 0$. Expected utility for Sender under no disclosure

[1] $\overline{G} = \min\{H \mid H \text{ concave, and } H(r) \geqslant G(r) \text{ for all } r\}$

(denote it $\underline{V}$) coincides therefore with $G(p)$.

All that matters for a simple mechanism are the posteriors it induces with its recommendations. These are $p^0 = \mathbb{E}[\omega|m = 0]$ and $p^1 = \mathbb{E}[\omega|m = 1]$. Again, conditional on observing message m, types below p^m will choose $a = 1$ and types above p^m will choose $a = 0$. Expected action under this mechanism is therefore $V(\pi) = \mathbb{E}[G(p^m)]$, where π determines the measure defining the expectation. As a consequence

$$\begin{aligned} \Delta V(\pi) &= V(\pi) - \underline{V} \\ &= \mathbb{E}[G(p^m)] - G(p) \end{aligned} \tag{3}$$

and

$$Value = \sup_{\pi} \mathbb{E}[G(p^m)] - G(p).$$

The Sender problem simply becomes

$$\begin{aligned} \max_{p^0, p^1} \quad & \alpha G(p^0) + (1-\alpha) G(p^1) - G(p) \\ \text{s.t.} \quad & p = \alpha p^0 + (1-\alpha) p^1 \end{aligned} \tag{4}$$

where the unique constraint is given by the Bayes consistency requirement. From Figure 3 it is evident how both the statements of the proposition are simple consequences of this fact.

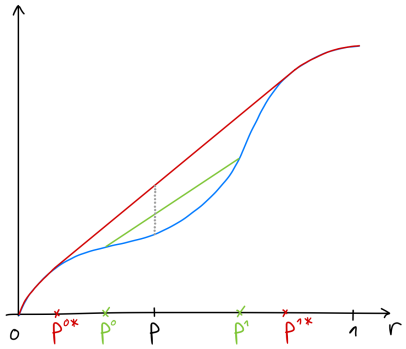


Figure 3: Choosing $(p^0, p^1) = (p^{0*}, p^{1*})$ clearly solves the Sender problem

□

Proposition 1 underpins the results listed in the analysis of Example 1 in the introduction. In that discussion I stated that if G is single peaked, persuasion is profitable only if p is below a certain threshold p^*. I also said that in that case $p^0 = 0$ (Sender never "lies" when recommending her least favorite action) and $p^1 = p^*$, independently from p. These facts become intuitive if one looks at Figure 4, which shows the application of the proposition to several single peaked distributions. Similarly, this graphical device allows us to spot settings in which counterintuitive results arise, as described in the following corollaries.

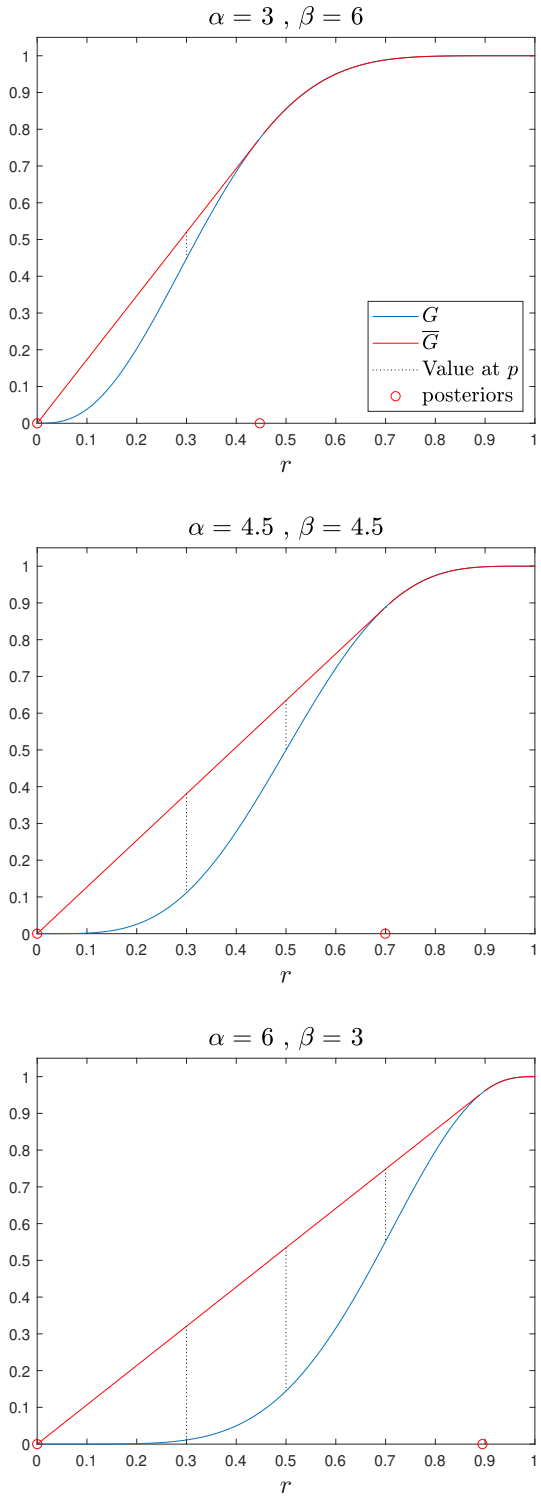


Figure 4: three different G's (in this case a $Beta(\alpha, \beta)$), their concavifications $\overline{G}$'s, the posteriors p^0 and p^1 associated to the optimal simple mechanism and the value of persuasion at three possible values of p.

Corollary 1.1. *When the judge type is commonly known, the prosecutor always recommends conviction of guilty defendants. This is not necessarily the case if the type of the judge is private.*

Indeed, in Figure 5a we see a bimodal G such that under the optimal simple mechanism $p^0 > 0$. This simple mechanism implies that sometimes Sender recommends action zero (*acquit*) in state one (*guilty*).

Corollary 1.2. *When the judge type is commonly known, the prosecutor benefits from persuasion if and only if p is sufficiently low. If the type of the judge is private, the set of p's such that persuasion is beneficial to the prosecutor doesn't even need to be convex.*

As an example, just look at Figure 5b, which shows a case where Sender benefits from persuasion if $p < p^*$ or if $p^{**} < p < p^{***}$ and doesn't otherwise. Also, note that in this very same example, different values of p are associated with different posteriors under the optimal mechanism. Indeed, if $p \in [0, p^*]$, the optimal mechanism implies that $p^0 = 0$ and $p^1 = p^*$, while if $p \in [p^{**}, p^{***}]$, under the optimal mechanism $p^0 = p^{**}$ and $p^1 = p^{***}$.

4 Single Peaked Distributions of Types

Proposition 1 characterized optimal simple mechanisms, i.e. optimal strategies for senders who try to influence the action of receivers simply announcing (possibly lying) what the binary state ω is. In many situations, Sender has more sophisticated persuasion mechanisms available to influence the action of Receiver, and it's natural to ask whether these sophisticated mechanisms can improve upon simple ones. The following result shows that, if G is single peaked, [2] then the mechanism provided by Proposition 1 is optimal among all persuasion mechanisms, not just simple ones.

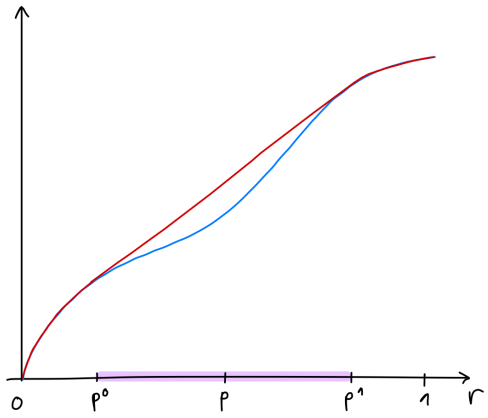
PROPOSITION 2. *If G is single peaked then an unconstrained Sender can choose optimally from simple mechanisms.*

The first part of the proof will follow an argument used in Kolotilin et al. to show that $V(\pi)$ can be described in terms of U_π for any persuasion mechanism π. The same result will be used to impose restrictions on the shape of U_π for an optimal π. With this in place, it will be enough to show that any U_π satisfying those restrictions can be induced by a simple mechanism.

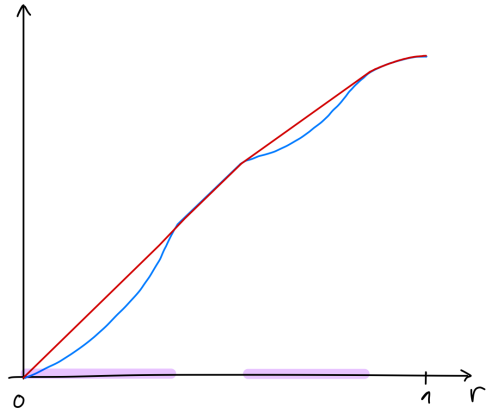
Proof. Kolotilin et al. exploit a parallel with the standard problem of mechanism design with transfers to show (among other things) that

- $U_\pi(r)$ is convex
- $U'_\pi(r_+) = -\int_\Omega \pi(\omega, r) dF(\omega)$

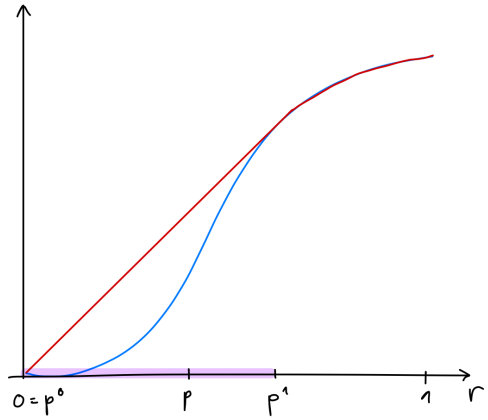
[2] G is single peaked if it is convex on $(0, k)$ and concave on $(k, 1)$, for some $k \in [0, 1]$.



(a) $p^0 \neq 0$



(b) non convex set of p



(c) single peaked G

Figure 5: Examples: G (in blue), $\overline{G}$ (in red) p's under which persuasion is beneficial to Sender (in purple)

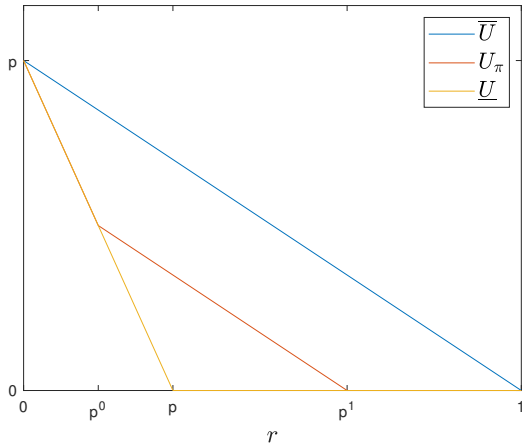


Figure 6: $\underline{U}$, $\overline{U}$ and U_π for the simple mechanism π associated with posteriors p^0 and p^1.

- $U_\pi(0) = \mathbb{E}[\omega]$

Using these results they show (substitute U'_π in equation 2 and integrate by parts), that

$$V(\pi) = g(0)\mathbb{E}[\omega] + \int_0^1 U_\pi(r)g'(r)dr$$

Note that $\underline{U}(0) = \overline{U}(0) = 1$ and that $\underline{U}(1) = \overline{U}(1) = 0$. Since $\underline{U}$ and $\overline{U}$ bound any U_π from below and from above, respectively, we must have $U_\pi(0) = 1$ and $U_\pi(1) = 0$. Let k be the mode of G. Since $g'(r)$ is positive if $r \leqslant k$, maximizing $V(\pi)$ requires maximizing $U_\pi(r)$ pointwise for $r \in [0, k]$. This, combined with the fact that U_π is convex, implies that U_π is linear in $[0, k]$. Similarly, as $g'(r)$ is negative if $r \geqslant k$, maximizing $V(\pi)$ requires minimizing $U_\pi(r)$ pointwise for $r \in [k, 1]$. Again, convexity implies that $U_\pi(r)$ falls linearly after k, with the same slope s that characterizes it on $[0, k]$,[3] until it becomes zero. In brief, $U_\pi = \max\{p - s \cdot r, 0\}$ for some $s \in [p, 1]$ (see Figure 6 for a graphical representation).
Since ω is binary, any simple mechanism is determined by the two posteriors $p^0 = \mathbb{E}[\omega|m = 0]$ and $p^1 = \mathbb{E}[\omega|m = 1]$, whose only constraint is that $p^0 \leqslant p \leqslant p^1$. Choose simple mechanism π' such that $p^0 = 0$ and $p^1 = p/s$. Under this mechanism Types greater than p^1 will always pick $a = 0$ and get zero expected utility, while types smaller than p^1 will choose $a = m$ and get an expected utility which is linear in their type. We have that $U_{\pi'} = U_\pi$, which ends the proof. □

Single-peaked distributions confirm being an interesting case since, also when Sender can choose from very general persuasion mechanisms, basic properties

[3] any steeper would fail convexity, any flatter would fail pointwise minimization.

of the optimal mechanism carry over without major modifications from the case where the receiver type is commonly known. This is expressed in the following corollary.

Corollary 2.1. *If the distribution of types is single peaked, optimal behavior of a prosecutor is qualitatively similar if judge types are private or commonly-known. In either case she always recommends conviction of guilty defendants and benefits from persuasion if and only if p is sufficiently low (see Figure 5c).*

5 Multimodal Distributions of Types

When G is multimodal we can still use Proposition 1 to find the optimal simple mechanism and the value associated with it. Anyway Proposition 2 doesn't hold anymore, and there could be persuasion mechanisms that outperform any simple mechanism, so that we are only able to obtain a lower bound to the value of persuasion:

$$Value \geqslant \overline{G}(p) - G(p).$$

A different strategy for finding bounds (both upper and lower) on the value of persuasion for general G's uses direclty equation 3. Consider the random variable $X = p^m$ and rewrite equation 3 as

$$\Delta V(\pi) = \mathbb{E}[G(X)] - G(\mathbb{E}[X])$$

The RHS of this equation is called the Jensen gap, as it's the difference between the two sides of Jensen's inequality, and there are several bounds on this quantity in the mathematical literature. These could be applied directly to our problem.
Here I provide an example of this second strategy to bound the value of persuasion, using bounds on the Jensen gap from Liao & Berg (2019) [4].

$$\sigma^2 \inf h(x) \leqslant \mathbb{E}(G(X)) - G(\mathbb{E}(X)) \leqslant \sigma^2 \sup h(x), \tag{5}$$

where $\mu = \mathbb{E}[X]$, $\sigma^2 = Var(X)$ and

$$h(x) \doteq \frac{G(x) - G(\mu)}{(x-\mu)^2} - \frac{G'(\mu)}{x-\mu}.$$

Bayes plausibility requires that $\mu = p$, and we can bound the variance of X from above realizing that this is maximized in the case of full disclosure. Hence, replacing $\mu = p$ and $\sigma^2 \leqslant p \cdot (1-p)$ in the expression above gives us upper and lower bounds on the value of persuasion for a general G.

6 Extensions

6.1 Continuous State Space

A natural extension, would be to relax the assumption that the distribution of states F is Bernoulli, and work with types $\omega \in [0,1]$ and a general F. In Propo-

sition 1, points p^0 and p^1 turned out to be posterior means under the optimal persuasion mechanism. What prevents us from extending the proposition to this more general setting is that, while any pair of posterior means satisfying $p^0 \leqslant p \leqslant p^1$ can be induced by some persuasion mechanism if the state is binary, this is no longer true for larger state spaces. This becomes evident if we plot $\overline{U}$ when F is a beta distribution and compare it to the Bernoulli case. We know from the proof of Proposition 2 that if p^0 and p^1 are induced by a simple mechanism π, then

$$U_\pi(r) = \begin{cases} \underline{U}(r) & \text{for } r \in [0, p^0] \cup [p^1, 1] \\ \underline{U}(p^0) - s(r - p^0) & \text{for } r \in [p^0, p^1] \end{cases}$$

Also, U_π must be bounded above by $\overline{U}$. When the state space is non binary, $\overline{U}$ stops being a linear function (see equation 1), which excludes a lot of potential (p^0, p^1) pairs, as the interim utility they would induce is no longer bounded by $\overline{U}$. This is illustrated in Figure 7. Still, we can use posteriors means associated

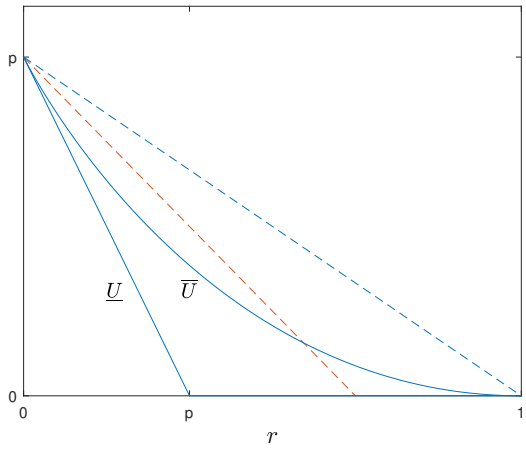


Figure 7: $\underline{U}$ and $\overline{U}$ when F is a Beta$(.5, 1)$. The dashed blue line is $\overline{U}$ when $\omega \in \{0, 1\}$ (i.e. Bernoulli F) and the red dashed line is an example of U_π which would be implementable if the state were binary but which is not when $\omega \in [0, 1]$

with the optimal simple mechanism in the binary-state case to get an upper bound on $Value$, and any pair posterior means which is feasible under F to obtain a lower bound. This would provide rather sharp bounds if the variance of F is large (fixed p, it is maximal if F is Bernoulli), and progressively worse ones as it decreases, as fewer and fewer U_π's can be implemented by simple mechanisms. This can be seen if we take F to be a $Beta(t\alpha, t\beta)$ and let t vary. In this way $p = \mathbb{E}[\omega] = \alpha/(\alpha + \beta)$ is fixed while variance decreases as t increases. In particular F converges in distribution to a Bernoulli as $t \to 0$. We see in Figure

8 that increasing t reduces the scope for persuasion, and widens the bounds.

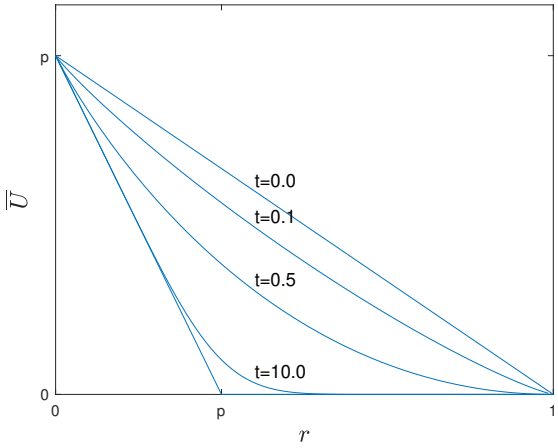


Figure 8: $\overline{U}$ when F is a Beta$(t, 2t)$ for different values of t.

Finally, one result which is robust to this extension is that no disclosure is optimal whenever G is concave. Indeed equation 3 holds independently of the form of F, and if G is concave the value of persuasion must be zero as a consequence of Jensen's inequality. I formulate this in form of a remark for later reference.

Remark 6.0.1. If G is concave Sender the no disclosure mechanism is optimal independently of F.

6.2 Many Receivers

As shown in Example 1, a model where Receiver has private information may arise from the fact that Sender actually interacts with several receivers from an heterogeneous population. In this setting, the application of the results derived so far relies on the implicit assumption that receivers' payoffs do not depend on the actions of other individuals in the population. In this section I will find conditions on G under which the no disclosure mechanism is optimal in a many-receivers setting, under milder restrictions on how receivers can influence each other's utility with their actions.

Let $I = \{1, ..., N\}$ be the set of agents (receivers) and let $i \in I$ be associated with utility $u_i : R_i \times \Omega \times A \to \mathbb{R}_+$, mapping player's type $r_i \in R_i$, state $\omega \in \Omega = [0, 1]$ and profile of actions $a \in A = \times_{i \in I} A_i$ into the positive reals.
These elements define a game $\mathcal{G} = \langle I, \Omega, (A_i, R_i, u_i)_{i \in I} \rangle$, to which we can associate a directed graph Γ as follows:

- the set of players I coincides with the set of nodes of Γ

- the directed edge $(i, j) \in I^2$ belongs to the set of edges of Γ if and only if the action chosen by j can affect the utility of i

When there are $n > 1$ receivers, the definition of persuasion mechanism has to be generalized slightly. A *persuasion mechanism* π^n asks all receivers to report their types and then privately recommends them which action to take. For each profile of reported types $\hat{r} = (\hat{r}_i)_{i=1}^n$ and state ω, the mechanism selects a distribution over profiles of recommendations $p_{\omega,\hat{r}} \in \Delta(A)$, which is used to draw recommendations.[4]

I will now focus on games where for any $i \in I$ there exist functions $\varphi_i : \Omega \times A \to \{0, 1\}$ and $\phi_i : \Omega \times A \to [0, 1]$ such that we can write

$$u_i(\omega, r_i, a) = a_i \cdot \varphi_i(a) \cdot (\phi_i(\omega, a) - r_i). \tag{$\circledast$}$$

In these games, roughly speaking, we can say that each player's utility is affected by the actions of other players only through some aggregate state, that these actions concur to determine.

I'll extend the utility of Sender to the many receivers case as follows

$$v(\omega, r, a) = \sum_{i=1}^{N} a_i \cdot \varphi_i(a)$$

To make more concrete the kind of interactions on which this section focuses, let's consider two schematic examples.

EXAMPLE 2 (Advertisement on an Online Platform):
An home-goods retailer starts an e-commerce activity. There are N_C potential customers who might be interested in using the online platform and N_A firms who might be interested in buying online advertising space on it. The benefit that the platform would bring to a potential customer i is $\omega_i \in [0, 1]$. The customer and the retailer have a common prior on ω_i, and consumer i is willing to register on the platform if he thinks its quality is above $c_j \in [0, 1]$. The quality of the Advertiser j is willing to buy advertising space on the platform if the fraction of customers that use the platform is greater than $a_i \in [0, 1]$. The retailer values an advertiser twice as much as a customer and wants to maximize expected profit. [5]

EXAMPLE 3 (School choice):
N students try to get into one of two schools, each of which can admit only one

[4] $\pi : \Omega \times R^n \to \Delta(A)$ is a measurable function, and $\pi(\omega, r) = p_{\omega,r}$.

[5] in this case we have $\varphi_i(a^{i-1}) = 1$ for all receivers, while

$$\phi(\omega, a^n) = \begin{cases} \omega_i & \text{if } i \text{ is a customer} \\ |N_C|^{-1} \sum_{i \in N_C} a_i & \text{if } i \text{ is an advertiser} \end{cases}$$

student.

The mechanism assigning students to schools goes as follows:

- *the private school and the public school have a common priority over students. Student i is the i'th student of this ranking.*
- *student i chooses whether to apply to the private school ($a_i = 0$) or the public one ($a_i = 1$). Students' choices are simultaneous.*
- *student i is assigned to the school to which he applied if no higher-ranked student applied to that same school, otherwise he is left unmatched.*

Student i gets utility s_i from being assigned to the private school and utility $1 - s_i$ from being assigned to the public one. The information designer wants all students to apply to the public school. [6]

The following result states that for games of this kind, Sender cannot increase her expected utility via information design if G is concave.

PROPOSITION 3. *If $\mathcal{G}$ satisfies condition ✲, Γ is acyclic, and G is concave, then the no disclosure mechanism is optimal.*

The intuition is straightforward: the aciclicity of Γ allow us to order receivers in such a way that each of them only cares about the (distribution of) actions of receivers with a lower index than his. We can think of this as ranking them according to their influence in the game. I use this fact, together with the particular form of Sender utility, to set up an inductive argument that shows existence of an optimal mechanism that provides no information to Receiver whose action nobody cares about. Then I use this mechanism to come up with another optimal one which provides no information to the two receivers with the highest index, and so on. At every step, I need to show that non-disclosure is optimal in a different one-receiver problem, which is where I make use of the assumption on the concavity of G, through remark 6.0.1.

(Proof of PROPOSITION 3).
For the rest of the proof receivers are labeled such that Receiver j cares about the action of Receiver i only if $i < j$ (this can be done if and only if Γ is a directed acyclic graph). Also, let $\pi : \Omega \times R^n \to \Delta(A^n)$ be a generic persuasion mechanism

[6] Let $a^n = (a_1, \ldots, a_n)$. It's easy to check that, with a little abuse of notation,

$$u_i(r_i, \omega, a) = a_i \cdot \varphi(a^{i-1}) \cdot [\phi(\omega, a^{i-1}) - r_i]$$

with

$$\varphi(a^n) = \begin{cases} 1 & \text{if } a_i = a_j \text{ for all } i, j \leqslant n \\ 0 & \text{else} \end{cases}$$

and

$$\phi(\omega, a^n) = \begin{cases} 1 & \text{if } \omega = a_i = 1 \text{ for all } i \leqslant n \\ 0 & \text{else} \end{cases}$$

in the n-receivers case. Without loss of generality, we can assume that π is incentive compatible, so that it maps actual types (not just reported types) and the state into a distribution of receivers' actions (and not just messages). I'll denote by $\pi(\cdot|\omega, r^n)$ the distribution associated to state ω and types r^n.
Let $\pi_i(a_n|\omega, r^n, a^{i-1})$ be the marginal implied by mechanism π over the actions of Receiver i conditional on the actions of receivers $j < i$, and let π^j be the marginal over the actions of the first j receivers. By the chain rule of probability we can factor any π as follows

$$\begin{aligned}\pi(a^n|\omega, r^n) &= \prod_{i=1}^{n} \pi_i(a_i|\omega, r^n, a^{i-1}) \\ &= \pi^j(a^j|\omega, r^n) \cdot \prod_{i=j+1}^{n} \pi_i(a_i|\omega, r^n, a^{i-1})).\end{aligned}$$

If π is uninformative for Receiver i, I denote $\pi_i = \not{\pi}_i$. The proposition is proved by induction. The base step of the proof consists in showing that an optimal incentive compatible persuasion mechanism exists (which holds by the extreme value theorem [7]), while the inductive step consists of the following Claim.

Claim 6.1. *If $\hat{\pi}$ is an optimal incentive compatible persuasion mechanism and $\hat{\pi}_j = \not{\pi}_j$ for all $j > i$, then replacing $\hat{\pi}_i$ with $\not{\pi}_i$ gives an optimal incentive compatible persuasion mechanism.*

Proof of Claim 6.1. In this proof I will denote $\mathcal{P}(\hat{\pi}^j) = \{\pi \mid \pi^j = \hat{\pi}^j\}$ the class of persuasion mechanisms π with marginal $\pi^j = \hat{\pi}^j$. Let $\hat{\pi}$ be an optimal incentive compatible persuasion mechanism with $\hat{\pi}_j = \not{\pi}_j$ for all $j > i$ and let $\tilde{\pi} \in \mathcal{P}(\hat{\pi}^{i-1})$ be such that $\tilde{\pi}_i = \not{\pi}_i$.

First, $\tilde{\pi}$ is incentive compatible. This clearly holds for all $j \geqslant i$, as a persuasion mechanism is trivially incentive compatible for those receivers to which it provides no information. It also holds for those $j < i$, as they only care about π^{i-1}, which coincides with the incentive compatible $\hat{\pi}^i$.

Second, note that we can focus on a constrained version of the Sender problem: since $\hat{\pi} \in \mathcal{P}(\hat{\pi}^{n-1})$ and $\hat{\pi}$ is optimal, we only need to show that $\tilde{\pi}$ maximizes Sender utility in $\mathcal{P}(\hat{\pi}^{n-1})$.

The next step is showing that this particular kind of constrained problem is equivalent to a one-receiver Sender problem where ω and p are replaced by $\omega_i = \phi_i(\omega, a)$ and $p_i = \mathbb{E}[\omega_i|\varphi_i(a) = 1]$.

[7] The objective function is continuous and the set of all persuasion mechanisms is relatively compact by Prokhorov's theorem as they are all supported on the same compact subset of $\mathbb{R}^{n+1}$.

Sender tries to maximize

$$\begin{aligned}\mathbb{E}[v(\omega, r^n, a^n)] &= \mathbb{E}\left[\sum_{i=1}^{n} a_i \cdot \varphi_i(a)\right] \\ &= \underbrace{\mathbb{E}[a_i \cdot \varphi_i(a)]}_{Pr[\varphi_i(a)=1]\cdot\mathbb{E}[a_i|\varphi_i(a)=1]} + \sum_{j\neq i} \mathbb{E}[a_j \cdot \varphi_j(a)].\end{aligned}$$

Since π^{i-1} is fixed by the constraints, $Pr[\varphi_i(a) = 1]$ and $\sum_{j=1}^{i-1} \mathbb{E}[a_j \cdot \varphi_j(a)]$ in the expression above are fixed as well. Having $\pi_j = \not{\pi}_j$ for $j > i$ also fixes $\sum_{j=i+1}^{n} \mathbb{E}[a_j \cdot \varphi_j(a)]$, so that the objective of Sender is to maximize $\mathbb{E}[a_i|\varphi_i(a) = 1]$, which is really close to her objective in the one-receiver problem.

We can look at Receiver i as facing a one-receiver version of the game, too. If $\varphi_i(a) = 0$, his utility is not affected by his action. Therefore, to determine which action is best, he only needs to consider scenarios in which $\varphi_i(a) = 1$. When this is the case, his utility is

$$u_i(\omega, r_i, a) = a_i \cdot (\omega_i - r_i), \tag{6}$$

Where $\omega_i = \phi_i(\omega, a)$. Under no disclosure, receiver i chooses $a_i = 1$ if and only if $r_i < p_i$, where $p_i = \mathbb{E}[\omega_n|\varphi_n(a) = 1]$.

Summing up: in the constrained problem described above, both Sender and Receiver i only consider those scenarios in which $\varphi_i(a) = 1$, and when this is the case their utilities coincide with those of the one-receiver setting. This problem is therefore equivalent to a Sender problem with one receiver where the binary state is $\omega_i = \phi_i(\omega, a)$ and $Pr(\omega_i = 1) = p_i$.

If G is concave, choosing $\pi_n = \tilde{\pi}_n$ is optimal by Remark 6.0.1 independently of p_n.

Δ

Since an optimal incentive compatible persuasion mechanism $\hat{\pi}$ exists, Claim 6.1 allows us to find a sequence of n optimal incentive compatible persuasion mechanisms, culminating in the uninformative one.

□

A Are Homogeneous Priorities Necessary in the School Choice Example?

In Example 3 we worked under the assumption that the public and the private school had the same priority over students. This was clearly enough to ensure applicability of proposition 3. One could wonder whether very similar, although not identical, priorities for the two schools would grant the same result.

Ergin acyclicity of school priorities represents a possible relaxation to the assumption of identical priorities. In our setting this is equivalent to requiring that, for any student ranked 3 and below in at least one of the two schools, the difference between his position in two rankings is at most one. The structure of priorities in Table 1 is clearly Ergin acyclic (all students ranked 3 and below in at least one school occupy the same position in the two rankings).

School 1	School 0
1	2
2	1
3	3
4	4
5	5
.	.
.	.
N	N

Table 1: school priorities

I will now show that, even when the distribution of types is concave, Ergin acyclicity of the structure of priorities doesn't imply that no-disclosure is optimal for Sender. To do this I'll consider the structure of priorities in Table 1 and a generic concave G supported on $[0,1]$ and I'll describe a persuasion mechanism that is strictly preferred by Sender to the no-disclosure mechanism (as it leads to both students choosing school 1 more often).

To ease the presentation I define $Stay_1 = Switch_2 = [0, 0.5]$ and $Stay_2 = Switch_1 = [0.5, 1]$, so that Receiver $i \in \{1, 2\}$ prefers to "stay" (i.e. he likes more the school where he ranks first) if $r_i \in Stay_i$ while he prefers to switch (i.e. he likes more the school where he ranks second) if $r_i \in Switch_i$.
Sender asks to players 1 and 2 to report their type and, based on this information, privately suggests them whether to "*stay*" or to "*switch*". I'll denote $\hat{r}_i$ the type reported by Receiver $i \in 0, 1$.

I'll start by studying persuasion mechanisms of a particular form:

- No information is provided to Receiver $j = 3, 4, \ldots, N$.
- Receiver 1 gets the message "*stay*" if either him or Recerver 2 reported $\hat{r}_i \in Stay_i$.

- If both Receivers report $\hat{r}_i \in Switch_i$, they receive an identical (either "*stay*" for both, or "*switch*" for both) message, which allows them to coordinate. The message will be "*switch*" with probability $\alpha \in (0, 1]$ and "*stay*" with complementary probability.

- Receiver 2 receives the message *stay* whenever he reports $\hat{r}_2 \in Stay_2$, but he can receive the message *switch* (with probability $\beta \in (0, \alpha)$) if he reported a preference for switching and 1 reported a preference for staying (i.e. $\hat{r}_1 \in Stay_1$ and $\hat{r}_2 \in Switch_2$).

Intuitively, Receiver 1 (who is on top of the ranking for the Public school) always receives a message he can trust for coordinating with Receiver 2 and "switch" schools. Receiver 2 can be sometimes fooled by the message, and induced into switching when Receiver 1 doesn't. Table 1 shows the probabilities that Receivers 1 and 2 receive the message "*switch*" under all possible combinations $(\hat{r}_1, \hat{r}_2)$.[8]

$(\hat{r}_1, \hat{r}_2)$	$Switch_2$	$Stay_2$
$Stay_1$	0,β	0,0
$Switch_1$	α,α	0,0

Table 2: Signal structure. Columns refer to the interval in which falls $\hat{r}_2$, rows to the interval in which falls $\hat{r}_1$.

Note that any persuasion mechanism of this type, if incentive compatible (i.e. such that all types truthfully reveal their type and then obediently follow the recommendation of the message) strictly improves the expected utility of Sender. This happens because whenever the pair of messages sent to Receiver 1 and Receiver 2 allows them to coordinate, the number of Receivers choosing the public school is unchanged. On the contrary, whenever Receiver 2 is tricked into unilaterally switch to the public school, the number of Receivers applying to the public school increases by one.

Q: Does Receiver 1 (who is never "fooled" by the message) have an incentive to deviate unilaterally?

Suppose $r_1 \in Stay_1$. Being truthful and obedient gives him the maximum utility he could possibly get, and is therefore a weakly dominant strategy.

Now consider the case $r_1 \in Switch_1$. Suppose he reports $\hat{r}_1 \in Switch_1$. In this case he will trivially follow the message he receives, since that perfectly reveals the behavior of Receiver 2.

[8] remember that both players receive the same message if $(\hat{r}_1, \hat{r}_2) \in Switch_1 \times Switch_2$

We're left with studying when some $\hat{r}_1 \in Switch_1$ can gain from revealing he prefers to stay. This could only happen if, by falsely revealing $\hat{r}_1 \in Stay_1$, Receiver 1 could get Receiver 2 to switch more often than otherwise. This never happens provided $\alpha > \beta$, which we assumed.

A: no.

Q: Does Receiver 2 (who is sometimes "fooled" by the message) have an incentive to deviate unilaterally?

Suppose $r_2 \in Stay_2$. Being truthful and obedient gives him the maximum utility he could possibly get, and is therefore a weakly dominant strategy.

Now consider the case $r_2 \in Switch_2$. Suppose he reports $\hat{r}_2 \in Switch_2$. We can compute the probability he assigns to 1 switching, and use this to find a threshold $\underline{r}$ such that 2 will be obedient if $r_2 < \underline{r}$. Let $p_m = \Pr(a_1 = switch \mid m_2 = m)$. Then

$$p_{stay} = 0 \qquad\qquad p_{switch} = \frac{(1-\gamma)\alpha}{\gamma\beta + (1-\gamma)\alpha}$$

where $\gamma = G(1/2)$.

Suppose he observes $m_2 = switch$. He has no incentive to deviate and stay if expected utility from staying is lower than expected utility from switching, and therefore if $r_2 < p_{switch} \cdot (1 - r_2)$.

Suppose now he observes $m_2 = stay$ after having reported $\hat{r}_2 \in Switch_2$. He will never deviate by switching, since his expected utility from switching is zero (he knows for sure that Receiver 1 will stay).

Therefore Receiver 2 will never deviate if he prefers to stay, or if he prefers to switch and truthfully reveals so. We're left with studying when he can gain from revealing he prefers to stay when instead he prefers to switch.

If $r_2 \in Switch_2$ and $\hat{r}_2 \in Stay_2$, Receiver 1 will certainly stay, so that the best action player 2 can take after misreporting his type is to stay as well, obtaining r_2. This is equivalent to truthfully reporting type and then choosing to stay regardless the message. We have already shown that if can benefit only types $r_2 > \underline{r} = p_{switch}/(1 + p_{switch})$.

A: Receiver 2 has an incentive to deviate unilaterally only if $r_2 \in (\underline{r}, 1/2)$, and the deviation consists in choosing to stay regardless of the message he receives [9].

[9] if $\alpha \to 0$ and β is fixed and positive, $\underline{r} \uparrow 1/2$

To sum up: any persuasion mechanism of the form described in Table 2 is incentive compatible for everyone except for Receiver 2 when his type is in the interval $(\underline{r}, 0.5)$. When this happens, the optimal unilateral deviation for Receiver 2 consists in always applying to school 0 (the private one), ignoring the message he receives.

Persuasion mechanisms described in Table 3 include a minor modification that allows them to be incentive compatible for any Receiver. They work exactly as the ones described in Table 2, except for the fact that they always send message "*stay*" to Receiver 2 if $\hat{r}_2 \in (\underline{r}, 0.5)$. This leaves the incentive structure unchanged for all other types and just relabels the message sent to the deviating types to be consistent with their preferred action.

$(\hat{r}_1, \hat{r}_2)$	$Switch_2$		$Stay_2$
	[0,r]	(r,1/2)	
$Stay_1$	0,α	0,0	0,0
$Switch_1$	β,β	0,0	0,0

Table 3: Signal structure. Columns refer to the interval in which falls $\hat{r}_2$, rows to the interval in which falls $\hat{r}_1$.

For any choice of $\alpha \in (0, 1)$ and $\beta \in (0, \alpha)$, we obtain an incentive compatible persuasion mechanism that strictly [10] benefits Sender with respect to sending uninformative signals to all players.

References

[1] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, October 2011.

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

[3] Anton Kolotilin, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. Persuasion of a privately informed receiver. *Econometrica*, 85(6):1949–1964, 2017.

[4] J. G. Liao and Arthur Berg. Sharpening jensen's inequality. *The American Statistician*, 73(3):278–281, 2019.

[5] Luis Rayo and Ilya Segal. Optimal information disclosure. *Journal of Political Economy*, 118(5):949–987, 2010.

[10] this actually requires that G assigns a positive probability to both types in $[0, \underline{r}]$ and types in $[0.5, 1]$. Any concave G with full support satisfies this.