



**HAL**  
open science

# The Value of Mediated Communication

Andrés Salamanca

► **To cite this version:**

| Andrés Salamanca. The Value of Mediated Communication. 2016. hal-01289379v4

**HAL Id: hal-01289379**

**<https://hal.science/hal-01289379v4>**

Preprint submitted on 1 Sep 2017 (v4), last revised 22 Aug 2019 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Value of Mediated Communication<sup>☆</sup>

Andrés Salamanca Lugo<sup>1</sup>

*Toulouse School of Economics*

---

## Abstract

Kamenica and Gentzkow [Bayesian persuasion. (2001). *A.E.R.*, 101, p. 2590-2615] consider a model in which a sender chooses a public communication device for signaling his information to an uninformed receiver, who then takes an action that affects the welfare of both individuals. In their model, the sender is fully committed to truthfully communicate the signal to the receiver, so that they abstract from incentive compatibility issues. By considering mediated communication, we provide an analytical framework overcoming this overly restrictive assumption. Specifically, we are able to characterize incentive constraints by a set of linear inequalities, which allows us to formulate the sender's problem as a linear programming problem. As a result, we can use an alternative geometric approach based on Duality Theory to transform the sender's problem into a simplified problem without incentive constraints that can be solved using concavification arguments.

*Keywords:* Bayesian persuasion, mediated communication, incentive compatibility, virtual utility.

*JEL Classification:* D82, D83.

---

## 1. Introduction

This paper provides an analytical framework for studying *Bayesian persuasion* problems in which the sender cannot commit himself to truthfully communicate his information to the receiver, so that incentive compatibility becomes one of the major issues for communication to be meaningful. By allowing the two players to communicate with a neutral third party, we are able to solve two analytical problems that could possibly prevent a tractable analysis of incentive compatibility: first, truthful revelation of information when communication is direct may considerably limit the ability of the sender to credibly signal his information by himself.

---

<sup>☆</sup>This paper makes part of my Ph.D. dissertation written at Toulouse School of Economics. I am deeply grateful to Françoise Forges for her continuous guidance and advice. I also wish to thank Thomas Mariotti for his thoughtful comments. This paper also benefited from helpful comments of seminar audiences at Toulouse School of Economics, Paris School of Economics, Paris Game Theory Seminar, the 5th World Congress of the Game Theory Society and the 8th UECE Lisbon Meetings in Game Theory and Applications.

*Email address:* [asalamanca1.tse@gmail.com](mailto:asalamanca1.tse@gmail.com) (Andrés Salamanca Lugo)

<sup>1</sup>This version: September 1, 2017. First version: Mars 16, 2016.

In particular, Forges (1985) and Farrell (1993) propose some examples in which no substantive communication can occur between the players. However, it is well known that the set of implementable outcomes may be strictly larger when players use mediated rather than direct communication (see for instance, Forges (1985, 1990)). Second, revelation of influential information with direct communication requires the sender to be indifferent between all signals he sends with positive probability. This is a strong form of incentive compatibility that reflects the fact that the sender sends a random signal by himself. In contrast, when the players use a mediator to perform the randomization of the signals on behalf of the sender, incentive compatibility will only demand each type of the sender to prefer the expected allocation designated for him.

In a recent pioneering work, Kamenica and Gentzkow (2011) offer a general approach to Bayesian persuasion under full commitment on the part of the sender. They consider a sender-receiver game in which before learning his type (ex-ante stage), the sender publicly chooses a signaling strategy, i.e., a conditional distribution of signals for each of his types, that he will use for transmitting his information to an uninformed receiver<sup>2</sup>. The sender produces a signal according to his true type and the corresponding distribution of signals. He cannot distort the signal realization, nor can he misrepresent his information (full commitment assumption). The receiver observes the signal realization and then takes an action that affects the welfare of both individuals. Drawing on a geometric approach developed by Aumann and Maschler (1995), Kamenica and Gentzkow characterize the sender's optimal value of persuasion. They first construct a *non-revealing payoff function* over prior beliefs,  $\hat{a}(\cdot)$ , describing the (ex-ante) expected equilibrium payoffs the sender can achieve in the absence of communication. Then, they compute the *concavification* of  $\hat{a}(\cdot)$ , denoted  $cav \hat{a}(\cdot)$ , i.e., the smallest concave function that is larger or equal to  $\hat{a}(\cdot)$ . Their main result establishes that, for given prior beliefs  $p$ , the sender's optimal expected payoff (value of persuasion) is  $cav \hat{a}(p)$ .

Under full commitment, restricting attention to either direct or indirect communication systems does not change the value of persuasion. Yet, in the absence of commitment, the communication system determines the strategic opportunities the sender has for manipulating his private information, thus, it might have an impact on the sender's payoffs. We consider the Bayesian persuasion setup studied by Kamenica and Gentzkow (2011), but with limited commitment and mediated communication. Under mediation, the sender reports a type to a neutral trustworthy mediator who then recommends an action to the receiver. The sender's report is not verifiable either by the mediator or the receiver, which allows the sender to strategically manipulate his private information. The mediator's recommendation is not binding, that is, the receiver is free to choose any action different from the recommended one. The mediator can only create value by controlling the flow of information between both players. He introduces noise in the communication, which may relax the incentive constraints faced by the sender. In our setting, a very broad form of the *revelation principle* applies and, thus, without loss of generality, we can restrict attention to mediation protocols in which the sender reports his type truthfully and the receiver obeys the prescribed recommendation. A mediation protocol in which the sender

---

<sup>2</sup>Kamenica and Gentzkow (2011) refer to a signaling strategy simply as a signal. In order to distinguish the conditional distribution from its realizations, they call the latter *signal realization*. This paper follows the terminology developed in the literature of communication games.

always reports the truth and the receiver always follows the recommendation is called a *communication equilibrium* (see Myerson (1986) and Forges (1986)). The sender’s problem is thus to select a communication equilibrium maximizing his ex-ante expected payoff.

The Lagrange multipliers associated to the truth-telling incentive constraints yield “shadow prices”,  $\gamma$ , that can be used to define the sender’s *virtual utility* (see Myerson (1991, ch. 10)). These virtual utility scales incorporate into the sender’s utility function the signaling costs associated with incentive compatibility<sup>3</sup>. Considering virtual utilities rather than real utilities, we construct a *fictitious persuasion problem* in which there are no truth-telling incentive constraints. For this game, the non-revealing payoff function,  $\hat{a}(\cdot; p, \gamma)$ , depends on the prior probability  $p$  and the signaling costs  $\gamma$ . Our main result (Theorem 1) says that the sender’s optimal expected payoff at the prior belief  $p$ , denoted  $a^*(p)$ , equals the value of persuasion in the fictitious game with virtual scales defined by the optimal signaling costs, namely,

$$a^*(p) = \min_{\gamma} \text{cav } \hat{a}(p; p, \gamma).$$

Ex-post inefficiencies derived from the signaling costs are geometrically expressed in the form of non convexities in the subgraph (or hypograph) of  $a^*$ . That is, the optimal value of persuasion may not be a concave function of the prior beliefs, as it is in the full commitment model. In fact, it may contain convex segments lying strictly below  $\text{cav } \hat{a}(\cdot)$ . Moreover, it may also exhibit discontinuities.

We characterize the optimal mediation protocol through a constrained splitting of the prior belief into a distribution over posterior beliefs. Given the optimal signaling costs  $\gamma^*$ , the posterior beliefs induced by any optimal mediation protocol correspond to the points on the domain of  $\hat{a}(\cdot; p, \gamma^*)$  for which the convex combination of their images yields  $a^*(p)$ . The corresponding distribution of posteriors is constrained by the Bayes plausibility (martingale property) together with *complementary slackness* conditions from Duality Theory (see Proposition 1). Complementary slackness says that if a shadow price is positive, then the associated informational incentive constraint must be binding.

We also provide an upper-bound on the number of messages that the sender requires to transmit in order to attain the value of persuasion. Under full commitment, the greatest number of messages the sender needs to convey does not exceed his number of types. However, when misrepresentation is problematic, the sender might need to transmit one extra message with every binding incentive constraint. Hence the number of messages is bounded by the total number of types plus the total number of incentive constraints (see Proposition 2). To understand the idea behind this result, consider for instance a situation in which the sender could be a “good” type or a “bad” type. Assume that he would prefer to be perceived as the good type, so that the incentive constraint asserting that the bad type should not gain by imitating the good type is binding. The good type of the sender would like to communicate his true type to the receiver (which requires every type to send a different message). However, since the receiver will face difficulties preventing the bad type from claiming to be the good type, the latter will need to find

---

<sup>3</sup>The approach is similar to the one used in Auction theory to define the virtual surplus that takes into account the bidders’ information rents (see Myerson (1981)).

a way to separate himself from the bad type. To do this, he may commit to recommend some additional action that might be unpleasant for him but highly hurtful for the bad type.

We conclude the paper with some discussions about the cheap-talk implementation of the optimal mediation protocols and the extension of our approach to general information design problems.

*Related literature.* Our analytical framework is the same as in Kamenica and Gentzkow (2011), except that we consider a more general interaction situation in which communication is mediated and the sender may strategically manipulate his private information. In that respect, our paper relates to the recent literature on information design known as Bayesian persuasion. To our best knowledge, this literature so far has been rather unsuccessful in developing a tractable approach for an explicit analysis of the sender's informational incentive compatibility. It is worth mentioning that Kolotilin, Li, Mylovanov and Zapechelnyuk (2016) study Bayesian persuasion with a privately informed receiver. In their framework, the sender designs a communication device that gathers information from the receiver and then sends a recommendation to the receiver conditional on her report and the sender's *true* type. In addition to the strategic incentive constraints ensuring that the receiver will follow the recommendation, the sender is also led to consider informational incentive constraints guaranteeing that the receiver finds it optimal to report truthfully her information. In their setup, the communication device is a mediation rule unable to verify the receiver's private information, but capable of identifying the sender's type.

This paper also relates to the literature on contracting with limited commitment. This literature considers a principal-agent setup in which the principal (receiver) is imperfectly informed about the agent's (sender's) type. The principal cannot contractually commit herself to chose any action, however, she may extract information from the agent by using a communication system. Bester and Strausz (2001) study direct communication in which the agent simply sends a single message to the principal. In contrast, Bester and Strausz (2007) allow the principal to use general communication devices which may enlarge the set of implementable contracts. Contrary to Bayesian persuasion, here the communication device (mechanism) is designed by the uninformed party, i.e., the principal. However, because the agent cannot commit to truthfully transmit his information, informational incentive compatibility is a relevant matter.

Mitusch and Strausz (2005) and, Golstman, Hörner, Pavlov and Squintani (2009) compare different communication protocols in the framework of Crawford and Sobel (1982). In this regard, our paper is also connected with this literature. They study the conditions under which mediation improves upon direct communication. As with contracting problems with adverse selection and limited commitment, it is also assumed that the communication procedures are designed to maximize the ex-ante welfare of the receiver. However, due to the particular structure of pay-offs, it turns out that this is also equivalent to maximize the ex-ante welfare of the sender.

Finally, it is worth mentioning that our model is mathematically analogous to a problem of mechanism design by an informed principal (as in Myerson (1983)) in which contracting takes place at the ex-ante stage<sup>4</sup>. Indeed, by using the concept of virtual utility, we borrow some

---

<sup>4</sup>Myerson (1983) considers contracting at the *interim* stage, whereby the principal chooses the mechanism after

analytical tools that were developed by Myerson (1983) in order to characterize his neutral optimum.

This paper is organized as follows. In the next section we present a motivating example. Section 3 is devoted to formally describing the basic interaction scenario. The concept of communication equilibrium is also defined. Section 4 introduces the mediated persuasion problem and the virtual utility approach. The main results are presented. It also contains some examples illustrating our findings. Finally, Section 5 presents some concluding discussions.

## 2. Motivating Example

In this section we study an example which motivates several aspects of our analytical framework. First, it illustrates how, in the absence of commitment, the necessity of players to trust each other may lead to inefficiencies derived from the signaling costs associated with incentive compatibility. Second, the example shows that mediated communication may help the sender to reach equilibrium outcomes that cannot be achieved under direct communication. Finally, the example provides an instance in which, by allowing for more signaling opportunities, mediated communication compels the sender to disclose more information than when he is fully committed.

Consider the following sender-receiver game. Player 1 (the sender) has a privately known type that may be  $H$  with probability  $p = 1/10$  or  $L$  with probability  $1 - p = 9/10$ , and player 2 (the receiver) must choose an action from the set  $J = \{j_1, j_2, j_3\}$ . Payoffs for both players depend on the sender's type and the receiver's action as follows:

	$j_1$	$j_2$	$j_3$
$H$	1,3	3,1	-5,-3
$L$	-1,-3	2,-1	0,0

We can set this example in an economic situation described as follows. The informed player is a financial analyst knowing whether the general state of the financial markets is more favorable for investments in portfolios  $j_1$  and  $j_2$  (type  $H$ ) rather than in  $j_3$  (type  $L$ ). The uninformed player is an investor who must select among these three different portfolios offered by the analyst<sup>5</sup>. Thus, each portfolio generates an expected return for the investor that depends on the state. On the other hand, the analyst's preferences are explained by fact that he gets profits with investments in the portfolio  $j_2$  but he wants also to give good advice to the investor.

The expected payoff of the receiver, as a function of her belief  $q \in [0, 1]$  about the type  $H$ , is depicted in Figure 1. Thick lines denote her best-reply payoff (optimal actions appear above the corresponding best-reply payoff). At  $q = 1/5$  (resp.  $q = 1/2$ ) any randomization between  $j_3$  and  $j_2$  (resp.  $j_2$  and  $j_1$ ) is a best reply for the receiver.

Given the receiver's best-replies, the sender's (ex-ante) expected payoffs, as a function of the belief  $q \in [0, 1]$ , are represented in Figure 2. In particular, according to the prior belief  $p =$

---

she has received her private information. This is a more involved problem, as the choice of the mechanism may signal information about the principal's type.

<sup>5</sup>To keep the example as simple as possible, we suppose that the investor has no outside option.

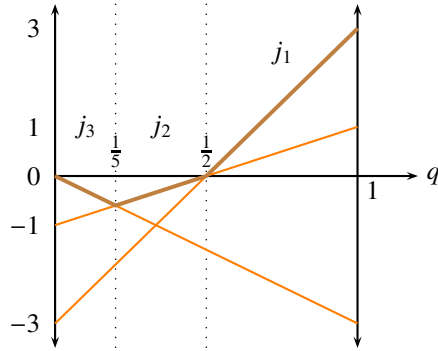


Figure 1: Receiver's expected payoffs (thin lines) and best-reply expected payoffs (thick lines)

$1/10$ , in the absence of communication, the receiver will choose action  $j_3$ . Thus, leaving the sender with an expected payoff equal to  $-1/2 (= -5 \times p + 0 \times (1 - p))$ .

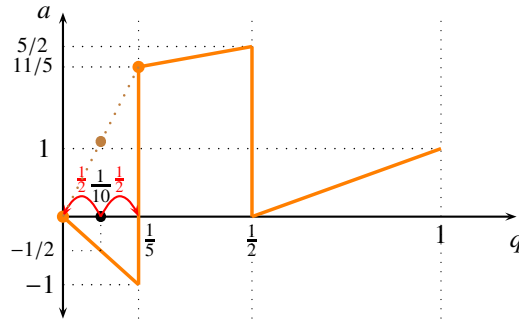
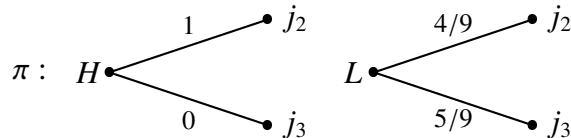


Figure 2: Non-revealing payoffs and optimal split under full commitment

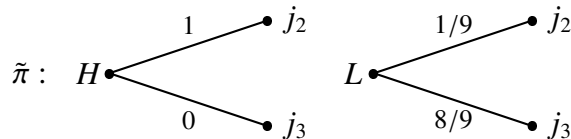
Assume now that, in order to persuade the receiver to change her action, the sender publicly *commits* to disclose his information according to a signaling strategy (i.e., a conditional distribution of signals for each of his types). This commitment assumption means that the sender cannot distort the signal realization, nor can he misrepresent his private information. The problem for the sender is then to induce posterior beliefs leading the receiver to choose actions maximizing his expected payoff. Without loss of generality, the sender may choose to send “recommendations” in  $J$ . Then, the unique optimal signaling strategy can be described as follows:



According to  $\pi$ , the sender recommends  $j_2$  with probability 1 when he is type  $H$  and randomizes between  $j_2$  and  $j_3$  with probabilities  $4/9$  and  $5/9$  respectively, when he is type  $L$ . After receiving the recommendation to play  $j_2$  (resp.  $j_3$ ), the receiver forms a posterior belief  $p_2 = \frac{1}{5}$  (resp.  $p_3 = 0$ ). Since  $j_2$  (resp.  $j_3$ ) is optimal for the receiver at  $p_2$  (resp.  $p_3$ ), she will follow the recommendation. Both actions are recommended with an expected probability equal to

$1/2(= 1 \times p + 4/9 \times (1-p))$ , and thus the sender's expected payoff is  $11/10(= 1/2 \times 0 + 1/2 \times 11/5)$ . Figure 2 illustrates this situation.

Let us now consider the situation in which commitment cannot be (legally) enforced, for instance because the sender's private information is not verifiable by any authority. The question is then: why would the sender like to maintain his commitment? Knowing that the receiver will follow the recommendation, type  $L$  of the sender will send message  $j_2$  with probability 1. But then a message that is sent with probability 1 by both types will convey no information to the receiver. Therefore, in anticipation of this behavior, the receiver will rationally choose her optimal action at  $p = 1/2$ . As a consequence, the information transmitted by the sender is not credible and thus  $\pi$  has no persuasive effect. In the absence of commitment, the signaling strategy must be part of a Nash equilibrium of the underlying cheap-talk game. Thus, truthful revelation of information requires the sender to be indifferent between the distinct outcomes that his messages lead the receiver to choose. This is a strong form of incentive compatibility that may considerably limit the ability of the sender for credibly signaling his information. In this game, the most the sender can get in the absence of commitment is  $1/6$ . To achieve this payoff, he can use the signaling strategy<sup>6</sup>



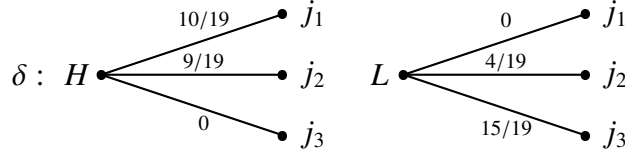
After receiving the recommendation to play  $j_2$  and  $j_3$ , the receiver forms posterior beliefs  $p_2 = 1/2$  and  $p_3 = 0$ , respectively. At  $p_3$  action  $j_3$  is her unique best-reply, so she follows the recommendation. At  $p_2$  action  $j_2$  is optimal for the receiver, yet following this recommendation (with probability 1) will lead type  $L$  of the sender to recommend  $j_2$  with probability 1, thus making communication not credible. Instead, the receiver may randomize between actions  $j_1$  and  $j_2$  with probabilities  $2/3$  and  $1/3$ , respectively, which is also optimal for her at  $p_2$ . In this way, the receiver makes type  $L$  indifferent between the recommendations  $j_1$  and  $j_2$  and, thereby communicating his information according to  $\tilde{\pi}$  remains optimal for the sender.

Assume now that, instead of communicating his information directly to the receiver, the sender sends a confidential report about his type to an impartial mediator commissioned to produce a recommendation for the receiver. Because information is nonverifiable, even with the help of a mediator, the sender may strategically manipulate his information. For instance, consider a mediator that recommends actions according to  $\pi$ . This mediation rule is not incentive compatible for the sender. Indeed, type  $L$  would have incentives to report that he is type  $H$ . As a consequence, the sender cannot get  $11/10$ , yet he can do better than  $1/6$ . Consider a mediator that recommends actions as follows:

---

<sup>6</sup>The signaling strategy  $\tilde{\pi}$  is part of the unique cheap-talk equilibrium of this game that maximizes the sender's (ex-ante) expected payoff.





According to this mediation protocol, the receiver will infer posterior beliefs  $p_1 = 1$ ,  $p_2 = 1/5$  and  $p_3 = 0$ , where as before  $p_j$  denotes the probability of type  $H$  conditional on the recommendation to play  $j$ . For any  $j \in J$ , action  $j$  is a best-reply at  $p_j$ , then following the recommendation is optimal for the receiver. On the other hand, given that the receiver is obedient, no type of the sender has incentives to misrepresent his information. Indeed, by reporting to the mediator that he is type  $H$ , type  $L$  gets an expected payoff equal to  $8/19 (= -1 \times 10/19 + 2 \times 9/19 + 0 \times 0)$ , while by telling the truth he gets  $8/19 (= -1 \times 0 + 2 \times 4/19 + 0 \times 15/19)$ . That is, type  $L$  is indifferent, and thus he has no incentives to lie. A similar analysis reveals that type  $H$  has no incentives to misrepresent his information either.

The mediation protocol  $\delta$  gives the sender an (ex-ante) expected payoff equal to  $109/190 (\approx 0.57 > 1/6)$ . No other mediation rule can guarantee a higher expected payoff to the sender. In this particular game, mediation facilitates incentive compatibility, thus allowing the sender to achieve an outcome that cannot be attained under direct communication. Mediation alleviates the conflict between the incentives of both players. However, its potential benefits are reduced by the degree of such a conflict. In particular, the sender's expected payoff in the absence of commitment is larger under mediation than under direct communication, but lower than what he gets under full commitment.

Finally, notice that  $\delta$  requires the sender to transmit 3 ( $> 2$ ) different messages with positive probability. That is, the sender communicates more information than just his true type (which has two possible values). Unlike the full commitment case, in which the number of messages need not exceed the number of types, when misrepresentation is problematic, the sender might need to disclose more information. The idea is that the sender requires to signal as much information as when he is fully committed (i.e., he needs to induce posterior beliefs  $p_3 = 0$ ,  $p_2 = \frac{1}{5}$ ), but also he needs to make such revelation credible to the receiver. Because type  $L$  has incentives to imitate type  $H$ , the latter type will need to find a way to separate himself from type  $L$ . To do this, type  $H$  can commit himself to recommend action  $j_1$  (with probability  $10/19$ ), something that is unpleasant for type  $L$ . In this manner, the receiver can discriminate between both sender's types, so that whenever  $j_1$  is recommended, she deduces that this message can only come from type  $H$ , i.e.,  $p_1 = 1$ , and thus, she follows the recommendation.

### 3. Basic Game

Our basic framework is a two-person finite Bayesian game in which player 1 has no decision to make, but is the only player to have private (nonverifiable) information. Let  $K$  be the (finite) set of types of player 1. A type  $k \in K$  is chosen according to<sup>7</sup>  $p \in \Delta(K)$ , and only player 1 is informed about  $k$ . We assume that  $p^k > 0$  for every  $k \in K$ . Player 2 chooses an action in a

<sup>7</sup>For any finite set  $A$ ,  $|A|$  denotes its cardinality and  $\Delta(A)$  denotes the set of probability distributions over  $A$ .

(finite) set  $J$ . When action  $j$  is chosen by player 2 and player 1 is of type  $k$ , then player 1 and player 2 get respective payoffs  $a_j^k$  and  $b_j^k$ . We refer to this *basic game* as  $\Gamma(p)$ .

A (*mediated*) *communication device*  $\delta$  is a mapping  $\delta : K \rightarrow \Delta(J)$ , namely a vector of probability distributions  $(\delta^k)_{k \in K}$  over  $J$  for every  $k \in K$ . By adding a communication device  $\delta$  to the game  $\Gamma(p)$ , one generates an extended game  $\Gamma_\delta(p)$ , which is played as follows:

1. A type  $k \in K$  is randomly chosen according to  $p$ .
2. Player 1 learns his type  $k \in K$ .
3. Player 1 sends a confidential report  $k' \in K$  to a mediator.
4. The mediator chooses an action  $j \in J$  with probability  $\delta_j^{k'}$ .
5. The mediator recommends the action  $j$  to player 2.
6. Player 2 chooses an action and both players receive payoffs as in  $\Gamma(p)$ .

For obvious reasons, we refer to player 1 in  $\Gamma_\delta(p)$  as the sender, and player 2 as the receiver. In this game, a strategy for the sender is a transition probability  $\tau : K \rightarrow \Delta(K)$  where  $\tau(k' | k)$  is the probability to report  $k'$  if his type is  $k$ . A strategy  $\tau$  is called *sincere* if  $\tau(k | k) = 1$  for every  $k \in K$ , namely, if the sender always reveals honestly his type to the mediator. A strategy for the receiver in  $\Gamma_\delta(p)$  is a transition probability  $\varsigma : J \rightarrow \Delta(J)$  where  $\varsigma(i | j)$  is the probability to choose  $i$  when  $j$  is recommended by the mediator. A strategy  $\varsigma$  is called *obedient* if  $\varsigma(j | j) = 1$  for every  $j \in J$ , i.e., if the receiver always follows the recommendation made by the mediator. When both players are sincere and obedient, respectively, in  $\Gamma_\delta(p)$ , the (ex-ante) expected payoff of the sender is

$$a(\delta; p) := \sum_{k \in K} p^k \sum_{j \in J} \delta_j^k a_j^k. \quad (3.1)$$

The communication device  $\delta$  is *incentive-compatible for the sender* if and only if the sincere strategy is a best response for the sender in  $\Gamma_\delta(p)$  whenever the receiver is obedient, that is,

$$\sum_{j \in J} \delta_j^k a_j^k \geq \sum_{j \in J} \delta_j^{k'} a_j^k, \quad \forall k, k' \in K. \quad (3.2)$$

The *informational incentive constraints* in (3.2) reflect the fact that neither the receiver nor the mediator can verify the sender's private information (adverse selection problem).

Suppose action  $j$  is recommended to the receiver according to the communication device  $\delta$ , provided that the sender is sincere in  $\Gamma_\delta(p)$ . Then, the receiver computes posterior probabilities  $p_j(\delta) = (p_j^k(\delta))_{k \in K}$  given by

$$p_j^k(\delta) = \frac{\delta_j^k p^k}{\sum_{k' \in K} \delta_j^{k'} p^{k'}}. \quad (3.3)$$

The communication device  $\delta$  is *incentive-compatible for the receiver* if and only if the obedient strategy is a best response for the receiver in  $\Gamma_\delta(p)$  whenever the sender is sincere, namely,

$$\sum_{k \in K} p_j^k(\delta) b_j^k \geq \sum_{k \in K} p_j^k(\delta) b_i^k, \quad \forall i, j \in J. \quad (3.4)$$

The *strategic incentive constraints* in (3.4) characterize the receiver's inalienable right to control her action in  $J$  (moral hazard problem). By definition of the posterior probabilities in (3.3), both

sides of (3.4) are divided by the total probability of receiving the recommendation to play  $j$ . Then, the strategic incentive constraints can be equivalently written as

$$\sum_{k \in K} \delta_j^k p^k b_j^k \geq \sum_{k \in K} \delta_j^k p^k b_i^k, \quad \forall i, j \in J. \quad (3.5)$$

We define  $Y(q)$  as the set of receiver's optimal actions at belief  $q \in \Delta(K)$ , i.e.,

$$Y(q) = \left\{ y \in \Delta(J) \mid \sum_{k \in K} q^k \sum_{j \in J} y_j b_j^k = \max_{j \in J} \sum_{k \in K} q^k b_j^k \right\}.$$

Let  $\pi_j(\delta) := \sum_{k \in K} p^k \delta_j^k$  be the probability of sending the recommendation  $j$  when  $\delta$  is implemented. Then,  $\delta$  is incentive compatible for the receiver if and only if for each  $j \in J$ ,  $\pi_j(\delta) > 0$  implies that  $j$  is optimal for the receiver given the posterior probabilities  $p_j(\delta)$ , i.e.,  $j \in Y(p_j(\delta))$ .

**Definition 1 (Communication equilibrium).**

A communication device  $\delta$  is a communication equilibrium of  $\Gamma(p)$  if and only if the sincere and obedient strategies form a Nash equilibrium of  $\Gamma_\delta(p)$ , that is,  $\delta$  satisfies the incentive constraints in (3.2) and (3.5). We let  $\mathcal{D}^*(p)$  denote the set of communication equilibria of  $\Gamma(p)$ .

REMARK 1. Communication equilibria are defined by a set of linear inequalities, hence the set  $\mathcal{D}^*(p)$  is a convex polyhedron. Furthermore, this set is closed, bounded and non-empty.

A communication equilibrium  $\delta$  is *fully revealing* (FR) if it recommends different actions for every type of the sender, so that the receiver can infer the true state by looking at the prescribed recommendation. It is *non-revealing* (NR) if  $\delta^k = \delta^{k'}$  for every  $k, k' \in K$ , so that no useful information is revealed to the receiver. It is *partially revealing* (PR) if it is neither NR nor FR.

Thanks to a very general form of the *revelation principle* for Bayesian games (see Forges (1985, 1986)), there is no loss of generality in restricting attention to communication equilibria, in the following sense: assume that the game  $\Gamma(p)$  is extended by allowing the players to communicate for a possibly infinite number of stages through a *general* communication device, sending signals to every player at every stage but also receiving messages from them. Such devices may involve preplay communication, before player 1 learns his type, but also interplay communication, after player 1 has learnt his type but before player 2 chooses his action. The set of all Nash equilibrium payoffs of all extensions of  $\Gamma(p)$  by general communication devices coincides with the set of all communication equilibrium payoffs.

**4. Mediated Persuasion**

In the basic game  $\Gamma(p)$ , player 1 has the option to remain silent and let player 2 to choose an action given her prior belief  $p$ . He can also design a communication system to signal his private information, and try to persuade player 2 to change her action. We assume that player 1 publicly chooses a mediated communication device  $\delta$  (i.e., a mediator) *before* learning his type. Then both players interact as in  $\Gamma_\delta(p)$ . Because the selection of the communication device is done at the ex-ante stage, this choice is by itself uninformative. The problem of player 1 is

then to choose a communication equilibrium maximizing his ex-ante payoff, namely, to select a communication device solving

$$\max_{\delta \in \mathcal{D}^*(p)} a(\delta; p). \quad (4.1)$$

We shall refer to this optimization problem as the *primal problem for p*.

REMARK 2. The optimization problem in (4.1) is a linear programming problem: the objective function is linear in  $\delta$  (see (3.1)) and the feasible set is defined by a system of linear inequalities in  $\delta$  (see Remark 1).

**Definition 2 (Value of persuasion).**

The optimal value of the primal problem for  $p$  will be called the *value of persuasion at p* and is denoted  $a^*(p)$ .

REMARK 3. Viewed as a correspondence defined on  $\Delta(K)$ , the set of communication equilibria is upper-hemicontinuous. Then,  $a^*$  is an upper-semicontinuous function. It may however fail to be continuous.

*4.1. Mediated Persuasion Under Verifiable Information*

Before proceeding with the analysis of the primal problem, let us consider the more simplified persuasion game in which the type of the sender is verifiable by the mediator but not by the receiver. In such a situation, the informational incentive constraints are not relevant, so that our framework reduces to Kamenica and Gentzkow's (2011).

**Definition 3.**

For a given prior  $p$ , we denote as  $\mathcal{D}(p)$  the set of communication devices satisfying the strategic incentive constraints in (4.5).

Under verifiable information, the problem for the sender can thus be expressed as

$$\max_{\delta \in \mathcal{D}(p)} a(\delta; p). \quad (4.2)$$

In this case, the sender has nothing to communicate to the mediator. The only thing he has to do is to choose a communication device that will recommend an action to the receiver depending on his *true* type. The verifiability assumption is thus equivalent to the full commitment assumption.

Given any prior belief  $q$ , the maximal ex-ante utility that the sender can expect in the absence of communication is

$$\hat{a}(q) := \max_{y \in Y(q)} \sum_{k \in K} q^k \sum_{j \in J} y_j a_j^k.$$

We refer to the function  $\hat{a}$  as the *non-revealing payoff function*. Let  $\text{cav } \hat{a}$  be the *concavification* of  $\hat{a}$ , i.e., the smallest concave function that is larger or equal to  $\hat{a}$ . As observed by Aumann and Maschler (1995) and Kamenica and Gentzkow (2011), the optimal value of the relaxed primal problem (4.2) is  $\text{cav } \hat{a}(p)$ . Also there exists a subset  $I \subseteq J$  of actions with  $|I| \leq |K|$  and posterior probabilities  $\{p_i\}_{i \in I}$  with  $p_i \in \Delta(K)$  for every  $i \in I$ , such that there exists a unique probability vector  $\rho \in \Delta(I)$  satisfying

$$\sum_{i \in I} \rho_i p_i = p \quad (4.3a)$$

and

$$\sum_{i \in I} \rho_i \hat{a}(p_i) = \text{cav } \hat{a}(p). \quad (4.3b)$$

Then, it is possible to “split” the total prior probability  $p$  into a set of conditional distributions  $\{p_i\}_{i \in I}$ , such that (i), for every  $i \in I$ , the posterior probabilities that the receiver computes after receiving the recommendation to play  $i$  are  $p_i \in \Delta(K)$ ; and (ii) the sender guarantees an expected payoff equal to  $\text{cav } \hat{a}(p)$ . Condition (4.3a) is called *Bayes plausibility*, while condition (4.3b) is an optimality requirement. Since the distribution  $\rho$  is unique, the receiver will update her prior beliefs from  $p$  to  $p_i$  with probability  $\rho_i$ .

Notice that in the previous result the number of signals required for achieving the optimal value of (4.2) is bounded by the number of types of the sender<sup>8</sup>. Namely, the most the sender needs to transmit to the receiver is just  $k$ , which has  $|K|$  possible values.

Given the set  $\{p_i\}_{i \in I}$  and the corresponding distribution of posteriors  $\rho$ , an optimal communication device can be computed from the following formula:

$$\delta_j^k = \begin{cases} \frac{p_j^k \rho_j}{p^k}, & \text{if } j \in I \\ 0, & \text{otherwise} \end{cases}, \quad \forall k \in K. \quad (4.4)$$

Finally, we observe that for any  $p \in \Delta(K)$ , we have that

$$\hat{a}(p) \leq a^*(p) \leq \text{cav } \hat{a}(p). \quad (4.5)$$

The first inequality follows from the fact that  $\hat{a}(p)$  can always be achieved by a NR communication equilibrium. The second inequality is due to the fact that  $\mathcal{D}^*(p) \subseteq \mathcal{D}(p)$ .

We start the analysis of the primal problem in Section 4.2 by dealing only with its optimal value. Then, in Section 4.3 we shall characterize its optimal solutions.

#### 4.2. The Virtual Persuasion Game

As we have seen, when there are no informational incentive constraints, the solution to the primal problem can be easily characterized. Informational incentive constraints complicate matters by interconnecting the signals in different states. However, we can integrate the welfare effects of incentive compatibility into the objective function using duality theory. The idea is that the set of communication equilibria is defined by a system of linear inequalities (see Remark 1) for which the dual variables can be used to define the sender’s *virtual utility*. These virtual utility scales incorporate into the utility function the signalling costs associated with the incentive compatibility. Using the concept of virtual utility we can transform the original primal problem into a simplified problem without informational incentive constraints but with a different objective function.

Let  $\gamma(k' | k) \geq 0$  be the dual variable (or Lagrange multiplier) for the constraint that type  $k$  of the sender should not gain by reporting  $k'$  in the primal problem for  $p$ . Following Myerson

---

<sup>8</sup>This result follows from Carathéodory’s theorem.

(1991, sec. 10.5), we define the *virtual utility* of the sender from the action  $j$ , when his type is  $k$ , w.r.t. the prior  $p$  and the duals  $\gamma$  to be

$$\alpha_j^k(p, \gamma) = \frac{1}{p^k} \left[ \left( p^k + \sum_{k' \in K} \gamma(k' | k) \right) a_j^k - \sum_{k' \in K} \gamma(k | k') a_j^{k'} \right]. \quad (4.6)$$

In order to understand formula (4.6), we disentangle its components. The terms of the form  $\gamma(k' | k)$  measures the *information rent* that type  $k$  can extract by pretending to be type  $k'$ . On the other hand, the terms of the form  $\gamma(k | k')$  measure the *signaling cost* that type  $k$  must incur in order to reduce the misrepresentation of type  $k'$ . Virtual utility is thus defined as the actual utility plus the total information rents minus the total signaling costs. Notice that multiplying type  $k$ 's utility  $a_j^k$  by the positive constant  $\frac{1}{p^k} (p^k + \sum_{k' \in K} \gamma(k' | k))$  is decision-theoretically inessential. That is, the unique decision-theoretic difference between the real utility scale and the virtual utility scale is given by the signaling costs. Hence, the virtual utility of the sender is a distorted utility that magnifies the difference between his actual utility and the utility of the types that would be tempted to imitate him.

In a situation where information is not verifiable, so that misrepresentation is possible, some types of the sender may get some information rents from having private information. Also, some types may be compelled to incur in signaling costs in an effort to distinguish themselves from the types that try to mimic them. This new compromise in the payoff maximization goals of the different types of the sender is described by the virtual utility and mathematically measured by the dual variables.

Let us assume that, as a consequence of the pressure that a type might feel in getting the receiver to trust him, the sender begins to act as if he were maximizing his virtual utility (Myerson (1991, sec. 10.8) refers to this idea as the *virtual utility hypothesis*). Thus, for some fixed prior beliefs  $p$  and signaling costs  $\gamma$ , consider the  $(p, \gamma)$ -virtual persuasion problem, a fictitious game that differs from the original persuasion game in the following. First, the sender's types are verifiable by the mediator (but not by the receiver), so that there are no informational incentive constraints. Second, the sender's payoffs are in the virtual utility scales  $(\alpha_j^k(p, \gamma))_{j \in J, k \in K}$  instead of  $(a_j^k)_{j \in J, k \in K}$ .

Let  $\hat{\alpha}(\cdot; p, \gamma)$  denote the non-revealing (virtual) payoff function of the  $(p, \gamma)$ -virtual persuasion problem. As already observed in Section 4.1, the value of persuasion in the  $(p, \gamma)$ -virtual game is given by  $\text{cav } \hat{\alpha}(p; p, \gamma)$ .

Although the  $(p, \gamma)$ -virtual game gives us some insights on how to simplify the sender's problem by removing the informational incentive constraints, it does not say anything about the "dual" relationship between the optimal value of the primal problem for  $p$ ,  $a^*(p)$ , and the value of persuasion in the virtual game. Furthermore, it leaves open the question of determining the optimal signaling costs incurred by the sender in order to distinguish himself from the types that are tempted to imitate him. However, there exists an answer to the first question that will make the second question redundant. The following result is a consequence of strong duality.

**Theorem 1.**

For any prior  $p \in \Delta(K)$  we have that

$$a^*(p) = \min_{\gamma \geq 0} \text{cav } \hat{\alpha}(p; p, \gamma). \quad (4.7)$$

We refer to the minimization problem in (4.7) as the dual problem for  $p$ .<sup>9</sup>

*Proof.* The Lagrangian of the primal problem for  $p$  is

$$\begin{aligned}\mathcal{L}(\delta, p, \gamma) &= \sum_{k \in K} p^k \sum_{j \in J} \delta_j^k a_j^k + \sum_{k \in K} \sum_{k' \in K} \gamma(k' | k) \left[ \sum_{j \in J} \delta_j^k a_j^k - \sum_{j \in J} \delta_j^{k'} a_j^{k'} \right] \\ &= \sum_{k \in K} p^k \alpha^k(\delta; p, \gamma)\end{aligned}$$

where  $\gamma \geq 0$  and  $\delta \in \mathcal{D}(p)$ .

Then, the dual problem for  $p$ , associated to the primal problem for  $p$ , is given by

$$\begin{aligned}\min_{\gamma \geq 0} \max_{\delta \in \mathcal{D}(p)} \mathcal{L}(\delta, p, \gamma) &= \min_{\gamma \geq 0} \max_{\delta \in \mathcal{D}(p)} \sum_{k \in K} p^k \alpha^k(\delta; p, \gamma) \\ &= \min_{\gamma \geq 0} \text{cav } \hat{\alpha}(p; p, \gamma)\end{aligned}$$

By strong duality, the value of the primal problem equals the value of its dual<sup>10</sup>. Thus  $a^*(p) = \min_{\alpha \geq 0} \text{cav } \hat{\alpha}(p; p, \gamma)$ .  $\square$

Fix a prior  $p \in \Delta(K)$  and let  $\gamma^*(p)$  be an optimal solution of the dual problem for  $p$ . Then, the key implication of Theorem 1 is that the value of persuasion in the original game coincides with the value of persuasion in the  $(p, \gamma^*(p))$ -virtual game, namely,

$$a^*(p) = \text{cav } \hat{\alpha}(p; p, \gamma^*(p)).$$

Thus, instead of saying that incentive compatibility restricts the sender's ability to signal his information, we may say that he is compelled to modify his actual preferences from the real to the virtual scales  $(p, \gamma^*(p))$ .

**Definition 4 (Value of information).**

The value of information for the sender at the prior  $p$  is the difference between the value of persuasion at  $p$  and the non-revealing value at  $p$ , i.e.,  $a^*(p) - \hat{\alpha}(p)$ .

We say that the sender benefits from his private information at  $p$  if the value of information at  $p$  is positive. Theorem 1 provides a necessary and sufficient condition for the sender to benefit from his information.

**Corollary 1.**

The sender benefits from his information at  $p$  if and only if for all  $\gamma \geq 0$ ,

$$\text{cav } \hat{\alpha}(p; p, \gamma) > \hat{\alpha}(p).$$

<sup>9</sup>We notice that the right-hand side of (4.7) is not, strictly speaking, the dual problem associated to (4.1). However, as shown in the proof of Theorem 1, its optimal value equals the optimal value of the dual.

<sup>10</sup>The following characterization also results from strong duality. Let  $\alpha(\delta; p, \gamma)$  be the ex-ante expected payoff the sender gets in the  $(p, \gamma)$ -virtual game when he uses the communication device  $\delta \in \mathcal{D}(p)$ . Let us define, for every  $p \in \Delta(K)$ , the auxiliary zero-sum game  $G(p)$ , in which the maximizing player chooses  $\delta \in \mathcal{D}(p)$ , the minimizing player chooses  $\gamma \geq 0$  and the payoff is  $\alpha(\delta; p, \gamma)$ . Then the value of  $G(p)$  exists and equals  $a^*(p)$ .

The motivating example of Section 2 will help us to understand the meaning and significance of the virtual utility. Also, it will provide some interesting conclusions about the effects of informational incentive compatibility.

*Example 1.* Consider again the motivating example studied in Section 2. The non-revealing value function  $\hat{a}$  joint with its concavification  $\text{cav } \hat{a}$  are depicted in Figure 3.

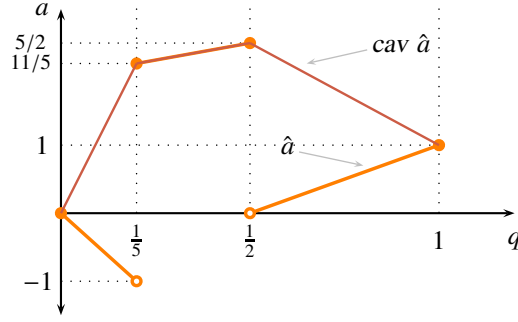


Figure 3: Function  $\hat{a}$  and its concavification

We fix the prior probability of type  $H$  to be  $p < \frac{1}{5}$  and we denote  $p_j$  the posterior belief about type  $H$  that the receiver will infer after receiving the recommendation to choose action  $j$ . Under full commitment, (which is equivalent to assume that information is verifiable by the mediator), an optimal communication device must split the total probability  $p$  into the posteriors  $p_3 = 0$  and  $p_2 = \frac{1}{5}$  with probabilities  $\rho_3 = 1 - 5p$  and  $\rho_2 = 5p$ , respectively (see Figure 3). The value of persuasion under the verifiability assumption is

$$\text{cav } \hat{a}(p) = 0\rho_3 + \frac{11}{5}\rho_2 = 11p$$

Thus, the unique optimal communication device is given by

$$\delta_2^H = 1, \quad \delta_2^L = 1 - \delta_3^L = \frac{4p}{1-p}$$

This communication device is, however, not incentive compatible for the sender. This is so because type  $L$  would have incentives to report that he is type  $H$ . As a consequence, the sender cannot achieve the expected payoff  $\text{cav } \hat{a}(p)$  when information is not verifiable, yet he can do better than  $\hat{a}(p)$  as we will see in the sequel.

By solving the dual problem for  $p < \frac{1}{5}$ , we have that the optimal value of the dual variables (Lagrange multipliers) is

$$\gamma^*(H | L) = \frac{10p(1-p)}{3-11p} := \gamma^*(p), \quad \gamma^*(L | H) = 0$$

Because type  $L$  has incentives to lie, it is natural that  $\gamma^*(H | L) > 0$ . Since type  $H$  cannot take any advantage from his private information (lying is not profitable),  $\gamma^*(L | H) = 0$ . The  $(p, \gamma^*(p))$ -virtual utility game can thus be described by the following payoffs matrix:



$\alpha, b$	$j_1$	$j_2$	$j_3$
$H$	$1 + \frac{\gamma^*(p)}{p}, 3$	$3 - \frac{2\gamma^*(p)}{p}, 1$	$-5, -3$
$L$	$-(1 + \frac{\gamma^*(p)}{1-p}), -3$	$2(1 + \frac{\gamma^*(p)}{1-p}), -1$	$0, 0$

We notice that type  $L$ 's virtual utility is just a positive multiple of his actual utility. Therefore, both the virtual game and the actual persuasion problem are decision-theoretically equivalent in state  $L$ . On the other hand, in state  $H$ , the sender's virtual utility magnifies the difference between his true type and the type that would be tempted to imitate him.

Figure 4 illustrates the non-revealing value function  $\hat{a}$  of the virtual game and its concavification. According to Theorem 1, the optimal value of the primal problem for  $p < \frac{1}{5}$  is

$$a^*(p) = \text{cav } \alpha(p; p, \gamma^*(p)) = p + \gamma^*(p) < 11p = \text{cav } \hat{a}(p)$$

We observe that ex-post inefficiencies are incurred in an optimal solution of the primal problem. The optimal value is, however, ex-post efficient in terms of the virtual utility scales. Thus, instead of saying that incentive compatibility forces the sender to incur in (ex-post) inefficiencies, we may say that incentive compatibility compels the sender to behave according to his virtual utilities (Myerson (1991, ch. 10) refers to this idea as the *virtual utility hypothesis*).

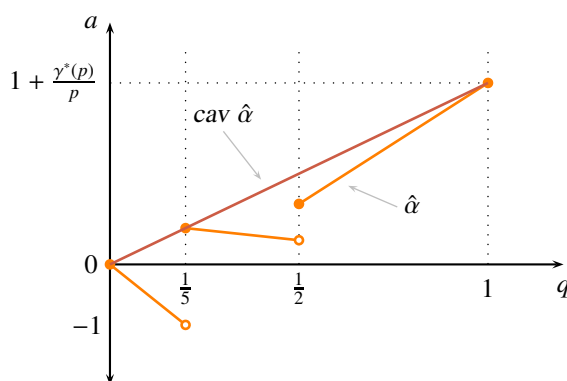


Figure 4: Function  $\hat{a}$  and its concavification

A similar analysis for the case  $p > \frac{1}{2}$  shows that optimal value of the primal problem is

$$a^*(p) = \text{cav } \alpha(p; p, \gamma^{**}(p)) = p + 6\gamma^{**}(p) < 4 - 3p = \text{cav } \hat{a}(p)$$

where,  $\gamma^{**}(p) := \gamma^{**}(L | H) = \frac{2p(p-1)}{5p-1}$  and  $\gamma^{**}(H | L) = 0$  are the optimal solutions of the dual problem for  $p$ .

Whenever  $p \in [1/5, 1/2]$ , the receiver's optimal action is to choose  $j_2$ , which is the preferred action for both types of the sender. Thus,  $a^*(p) = \hat{a}(p) = \text{cav } \hat{a}(p)$ . The value function  $a^*$  looks like in Figure 5.

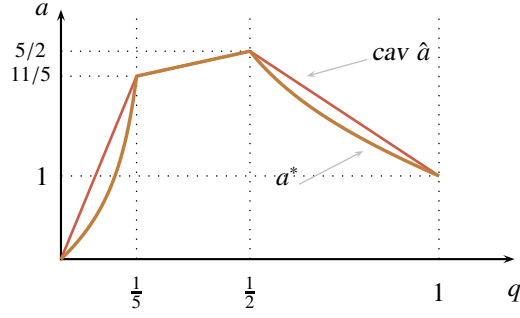


Figure 5: Functions  $a^*$  and  $cav \hat{a}$  for Example 1

*Example 2.* As a further illustration of our methodology, we study the following game, proposed by Forges (1990). It has been extensively analyzed in the literature of strategic information transmission. Payoffs for both players depend on the sender's type and the receiver's action as follows:

$a, b$	$j_1$	$j_2$	$j_0$	$j_3$	$j_4$
$H$	3,0	4,4	0,7	10,9	6,10
$L$	6,10	10,9	0,7	4,4	3,0

This example has a natural interpretation in terms of a job assignment scenario. An employer must decide whether to hire a candidate and, if so, to assign the employee to one of four possible jobs. The candidate may be one of two types. Type  $L$  performs better in job 1 but prefers job 2; he is bad at job 3, and even worse at job 4. Type  $H$  is similar but with jobs reversed. Sender's prior probability of type  $H$  is  $p \in (0, 1)$ .

By performing a similar analysis as in Example 1, it can be shown that the value of persuasion in this game (depicted in Figure 6) is given by

$$a^*(p) = \begin{cases} 6 + 14p + \frac{28p(5p-1)}{4-15p}, & \text{if } p < \frac{1}{5} \\ \frac{44}{5}, & \text{if } \frac{1}{5} \leq p \leq \frac{4}{5} \\ 20 - 14p - \frac{28(1-p)(5p-4)}{15p-11}, & \text{if } p > \frac{4}{5} \end{cases}$$

In particular, we have that for any  $p < 1/5$  (resp.  $p > 4/5$ ) only type  $L$  (resp.  $H$ ) has incentives to lie, so that  $\gamma(H | L) > 0$  (resp.  $\gamma(L | H) > 0$ ).

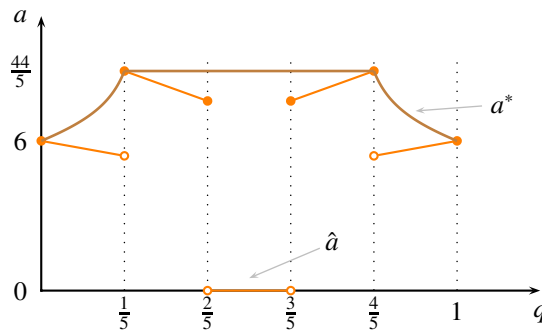


Figure 6: Functions  $a^*$  and  $\hat{a}$  for example 2

The nature of this game is similar to that of Example 1, except that here there is an outside option: not to hire the candidate, i.e., action  $j_0$ .

### 4.3. Optimal Mediators

So far we have focused on the optimal value of the primal problem. Our aim now is to characterize its optimal solutions. For that, let us start considering the case in which the optimal value of the dual variables is zero at some prior  $p$ , so that incentive constraints are not essential. In such a situation virtual utilities coincide with real utilities and the value of persuasion at  $p$  is  $cav \hat{a}(p)$ . Then, according to (4.3b), the posterior beliefs  $\{p_i\}_{i \in I}$  induced by any optimal communication device correspond to the points on the domain of  $\hat{a}$  for which the convex combination of their images yields  $cav \hat{a}(p)$ . The corresponding distribution of posteriors  $\rho = (\rho_i)_{i \in I}$  is the (unique) solution of a well determined system of linear equations given by (4.3a). Hence, given  $\{p_i\}_{i \in I}$  and  $\rho$ , an optimal communication device can be easily computed using formula (4.4).

Now consider the situation in which there are binding informational incentive constraints. On one hand, according to the virtual utility hypothesis, the sender might distort his preferences from the actual to the virtual scales, exaggerating the difference from the types that try to mimic him. On the other hand, as we have shown in the motivating example (Example 1), the sender may require to transmit more messages than his number of types. Then, the number of unknowns in (4.3a) (messages) increases while the number of equations (types) remains the same. The system in (4.3a) may become underdetermined and therefore infinitely many distributions of posteriors may be consistent with the same prior probability. Thus, additional conditions are required to characterize any optimal communication device.

To understand this issue, let us return to Example 1. Let  $p < \frac{1}{5}$  and consider the  $(p, \gamma^*(p))$ -virtual game. According to Figure 4 and condition (4.3b), the optimal value of the primal problem can be achieved by splitting the total probability  $p$  in either of the following collection of posteriors:

- (i)  $p_3 = 0, p_2 = \frac{1}{5}$ , or
- (ii)  $p_3 = 0, p_1 = 1$ , or
- (iii)  $p_3 = 0, p_2 = \frac{1}{5}, p_1 = 1$ .

In case (i), Bayes plausibility implies that  $(\rho_2, \rho_3) = (5p, 1 - 5p)$  and thus formula (4.4) yields  $\delta_2^H = 1 - \delta_3^H = 1$  and  $\delta_2^L = 1 - \delta_3^L = \frac{4p}{1-p}$ . But this communication device is not incentive compatible for the sender. In case (ii) we have that Bayes plausibility implies that  $(\rho_1, \rho_3) = (p, 1 - p)$  and therefore formula (4.4) yields  $\delta_1^H = \delta_3^L = 1$ . This communication device is a FR communication equilibrium giving an expected payoff to the sender equal to  $p$  which is strictly lower than  $a^*(p)$ . Finally, in case (iii), Bayes plausibility does not uniquely identify a distribution of posteriors. In particular, any probability vector  $(\rho_1, \rho_2, \rho_3)$  satisfying  $\rho_1 + \frac{\rho_2}{5} = p$  is a feasible distribution of posteriors. An additional condition is thus required in order to identify the correct distribution of posteriors.

Duality theory implies a relationship between the primal and dual problems that is known as *complementary slackness*. Specifically, it says that if a dual variable is positive, then the associated informational incentive constraint must be binding. Conversely, if a constraint fails to bind, then the associated dual variable must be zero. Complementary slackness provides us the additional equations we needed. Consider again Example 1 with  $p < \frac{1}{5}$ . As we have already

shown, the optimal value of  $\gamma^*(H | L)$  is strictly positive. Then, according to the complementary slackness, the constraint asserting that the type  $L$  should not gain by reporting  $H$  is binding, i.e.,

$$2\delta_2^L - \delta_1^L = 2\delta_2^H - \delta_1^H$$

The previous equality joint with formula (4.4) yield the additional restriction

$$5\rho_1(1 - p) = 2\rho_2(1 - 5p)$$

This equality together with Bayes plausibility implies that the optimal distribution of posteriors is

$$\rho_1 = \frac{2p(1-5p)}{3-11p}, \quad \rho_2 = \frac{5p(1-p)}{3-11p}, \quad \text{and} \quad \rho_3 = 1 - \rho_1 - \rho_2$$

Given these posteriors, formula (4.4) gives the optimal communication device solving the primal problem for  $p$ . We conclude that, in order to achieve the optimal value  $a^*(p)$ , the sender requires to induce a split of the total prior probability  $p$  into posterior beliefs  $p_3 = 0$ ,  $p_2 = \frac{1}{5}$  and  $p_1 = 1$ . Hence, an optimal communication device transmits 3 ( $> |K| = 2$ ) different messages (recommendations) with positive probability. The sender transmits as much information as when misrepresentation is not problematic (i.e., he induces posteriors  $p_3 = 0$  and  $p_2 = \frac{1}{5}$ ), however he also requires to send the message  $j_1$  to make such revelation credible to the receiver.

Drawing on the virtual utility hypothesis, the increased number of messages can be justified by the fact that incentive compatibility obliges type  $H$  to use an additional costly signal, from which he gains positive virtual utility. The optimal probability of sending message  $j_1$  (i.e.,  $\rho_1$ ) is determined by the minimization of the signaling costs incurred in recommending  $j_1$ .

Optimality conditions from strong duality theory imply the following result:

**Proposition 1.**

*Let  $\delta$  be a communication device satisfying the informational incentive constraints for the sender. Then,  $\delta$  is an optimal solution of the primal problem for  $p$  if and only if there exists a vector  $\gamma \geq 0$  such that*

$$\gamma(k' | k) \left[ \sum_{j \in J} (\delta_j^k - \delta_j^{k'}) a_j^k \right] = 0, \quad \forall k, k' \in K \quad (4.8)$$

and

$$\sum_{k \in K} p^k \alpha^k(\delta; p, \gamma) = \text{cav } \hat{a}(p; p, \gamma) \quad (4.9)$$

Condition (4.9) is the counterpart of conditions (4.3a) and (4.3b). It says that the optimal communication device induces a distribution of posterior beliefs giving the sender an ex-ante expected virtual payoff equal to the concavification of the non-revealing virtual payoff function  $\alpha(\cdot; p, \gamma)$  evaluated at the prior distribution  $p$ . Condition (4.8) is the complementary slackness.

Following Myerson (1991, sec. 10.5), we say that a type  $k$  *jeopardizes* another type  $k'$  at the prior  $p$  if the optimal value of the dual variable  $\gamma(k' | k)$  at  $p$  is positive.

#### 4.4. Extreme Communication Equilibria and the Number of Signals

Our aim in this section is to provide an upper-bound on the number of recommended actions required to achieve the value of persuasion. For that, we exploit the geometric properties of the set of communication equilibria.

Recall that, for any fixed  $p \in \Delta(K)$ , the feasible set of the primal problem for  $p$  is a convex polytope (bounded polyhedron). Then, the sender's expected payoff achieves its maximum at an extreme point of  $\mathcal{D}^*(p)$  (or a convex combination of them).

##### **Definition 5 (Extreme communication equilibrium).**

*The communication device  $\delta$  is an extreme communication equilibrium of  $\Gamma(p)$  if it is an extreme point of  $\mathcal{D}^*(p)$ .*

Using a basic result from the theory of linear programming, it is possible to characterize the number of messages sent with positive probability in any extreme communication equilibrium. A solution of a system of linear inequalities is an extreme point of the corresponding feasible set if and only if it can be obtained as the unique solution to a system of equations derived from equality constraints by setting a subset of variables equal to zero (see for instance Schrijver (1998)). These are called *basic feasible solutions*. As a consequence, the number of non-zero components in any extreme point is no greater than the number of binding constraints. Thus, a way to identify an upper-bound on the number of actions with positive probability in an extreme communication equilibrium is to determine how many incentive constraints can be binding.

The previous insight was applied by Forges (1994) to show that whenever the sender has only two types (i.e.,  $|K| = 2$ ), the number of recommended actions in an extreme communication equilibrium cannot exceed 4. This bound corresponds to the number of types ( $|K| = 2$ ) plus the number of informational incentive constraints. Unfortunately, the reasoning in the proof of this result relies strongly on the fact that  $|K| = 2$ . However, a similar statement can be proved for the general case  $|K| \geq 2$  by modifying the sender's problem<sup>11</sup>. The idea is as follows: let  $\bar{\delta}$  denote a solution of the primal problem for  $p$ . Now, replace each  $\delta_j^k$  in the definition of the primal for  $p$  by the variable  $\theta_j \bar{\delta}_j^k$ , with  $\theta_j \geq 0$  and add  $|K|$  constraints of the form  $\sum_j \theta_j \bar{\delta}_j^k = 1$  for all  $k \in K$ . By keeping fixed  $\bar{\delta}$  and  $p$ , we obtain a linear programming problem on  $\theta$ . For this problem, the strategic incentive constraints are redundant, thus we end up with  $|K|^2$  ( $= |K| + |K|(|K| - 1)$ ) constraints. Then, applying the previous insights, there exist a solution of the modified problem, denoted  $\tilde{\theta}$ , with at most  $|K|^2$  positive components. By construction, the communication device  $\tilde{\delta}$  defined by  $\tilde{\theta}_j \bar{\delta}_j^k$  is also an optimal solution of the primal problem for  $p$ . Since, all actions  $j$  for which  $\theta_j = 0$  have zero probability in  $\tilde{\delta}$ , we are able to find an upper bound on the number of signals.

##### **Proposition 2.**

*For any  $p \in \Delta(K)$ , there exists a solution of the primal problem for  $p$  for which the number of actions with positive probability does not exceed  $|K|^2$ .<sup>12</sup>*

<sup>11</sup>The same method is also applied by Bester and Strausz (2007).

<sup>12</sup>As it is inferred from the proof, we can also establish a result somewhat stronger than Proposition 2. Suppose

*Proof.* Let  $p \in \Delta(K)$  and assume that only  $m (\leq |K|(|K| - 1))$  informational incentive constraints are linearly independent in the primal problem for  $p$ . Then, there exist a set  $M \subseteq K \times K$  such that  $|M| = m$  and any solution of the programming problem

$$\begin{aligned} \max_{\bar{\delta} \geq 0} \quad & \sum_{k \in K} p^k \sum_{j \in J} \bar{\delta}_j^k a_j^k \\ \text{s.t.} \quad & \sum_{j \in J} \bar{\delta}_j^k a_j^k \geq \sum_{j \in J} \bar{\delta}_j^{k'} a_j^{k'}, \quad \forall (k, k') \in M \\ & \sum_{k \in K} \bar{\delta}_j^k p^k b_j^k \geq \sum_{k \in K} \bar{\delta}_j^k p^k b_i^k, \quad \forall i, j \in J \\ & \sum_{j \in J} \bar{\delta}_j^k = 1, \quad \forall k \in K \end{aligned} \tag{4.10}$$

is a solution of the primal problem for  $p$ , and viceversa. We need to prove that there exists a solution of the primal problem for  $p$  for which no more than  $|K| + m$  actions are recommended with positive probability. Let  $\bar{\delta}$  be a solution of (4.10). Let  $N = \{j \in J \mid \pi_j(\bar{\delta}) > 0\}$  and  $n = |N|$ . If  $n \leq |K| + m$ , there is nothing to prove. Then, assume that  $n > |K| + m$ . Consider the linear programming problem

$$\begin{aligned} \max_{\theta \geq 0} \quad & \sum_{k \in K} p^k \sum_{j \in J} \theta_j \bar{\delta}_j^k a_j^k \\ \text{s.t.} \quad & \sum_j \theta_j \bar{\delta}_j^k a_j^k \geq \sum_j \theta_j \bar{\delta}_j^{k'} a_j^{k'}, \quad \forall (k, k') \in M \\ & \sum_{j \in J} \theta_j \bar{\delta}_j^k = 1, \quad \forall k \in K \end{aligned} \tag{4.11}$$

Because  $\bar{\delta}$  is a solution of (4.10), the vector  $\bar{\theta} \geq 0$  defined by  $\bar{\theta}_j = 1$  for all  $j \in N$  and  $\bar{\theta}_j = 0$  for all  $j \in J \setminus N$  solves the linear program (4.11). By a fundamental result of linear programming (see Schrijver (1998)), we can always find a solution of (4.11) among the extreme points of its feasible set. Therefore, since (4.11) has  $|K| + m$  constraints, it has a basic feasible solution  $\tilde{\theta} \geq 0$  with no more than  $|K| + m$  strictly positive components. For every  $j \in J$  and  $k \in K$ , we define  $\tilde{\delta}_j^k = \tilde{\theta}_j \bar{\delta}_j^k \geq 0$ . Let  $\tilde{N} = \{j \in J \mid \pi_j(\tilde{\delta}) > 0\}$ . Then,  $|\tilde{N}| \leq |K| + m$ ,

$$\begin{aligned} \sum_{j \in J} \tilde{\delta}_j^k &= \sum_{j \in J} \tilde{\theta}_j \bar{\delta}_j^k = 1, \quad \forall k \in K \\ a(\tilde{\delta}; p) &= \sum_{k \in K} p^k \sum_{j \in J} \tilde{\theta}_j \bar{\delta}_j^k a_j^k = \sum_{k \in K} p^k \sum_{j \in J} \tilde{\theta}_j \bar{\delta}_j^k a_j^k = a(\bar{\delta}; p) \end{aligned}$$

and

$$\sum_{k \in K} \tilde{\delta}_j^k p^k b_j^k = \tilde{\theta}_j \sum_{k \in K} \bar{\delta}_j^k p^k b_j^k \geq \tilde{\theta}_j \sum_{k \in K} \bar{\delta}_j^k p^k b_i^k = \sum_{k \in K} \tilde{\delta}_j^k p^k b_i^k, \quad \forall i, j \in J$$

---

that it is possible to establish that at most  $m$  informational incentive constraints are binding, for instance by showing that some of them can be written as linear combinations of the others. Then, there is a solution of the primal problem for which the number of actions with positive probability does not exceed  $|K| + m$ .

Then,  $\tilde{\delta}$  is a solution of (4.10) for which no more than  $|K| + m$  actions are recommended with positive probability.  $\square$

We now provide an example that shows that the bound in Proposition 2 is actually tight. This means that without any further knowledge on the number of binding informational incentive constraints, the lowest possible upper bound on the number of recommended actions is  $|K|^2$ .

*Example 3.* Payoffs for both players depend on the sender's type and the receiver's action as indicated in the following matrix:

$a, b$	$j_1$	$j_2$	$j_3$	$j_4$	$j_5$
$H$	-2,0	2,4	1,7	$\frac{3}{2}, 9$	0,10
$L$	0,10	$\frac{3}{2}, 9$	1,7	2,4	-2,0

Let  $p = 1/2$  be the prior probability of type  $H$ . Then, the optimal solution of the dual problem for  $p$  is  $\gamma(H | L) = \gamma(L | H) = \frac{17}{42}$ . Complementary slackness implies that both informational incentive constraints are binding. Therefore, we expect an optimal solution of the primal problem to involve 4 messages. Indeed, the unique optimal solution is

$$\delta_2^H = \delta_4^L = \frac{8}{21}, \quad \delta_4^L = \delta_2^L = \frac{4}{7}, \quad \delta_5^H = \delta_1^L = \frac{1}{21},$$

which induces posterior probabilities  $p_1(\delta) = 0$ ,  $p_2(\delta) = \frac{2}{5}$ ,  $p_4(\delta) = \frac{3}{5}$ ,  $p_5(\delta) = 1$ . Thus, actions  $j_1$ ,  $j_2$ ,  $j_4$  and  $j_5$  are recommended with positive probability. Qualitatively similar results are obtained for any prior probability  $p \in (\frac{2}{5}, \frac{3}{5})$ .

## 5. Discussions

### 5.1. Cheap-Talk Implementation

In some environments, plain conversation between the players is more natural than mediated communication. Is it possible to achieve any communication equilibrium payoff by means of cheap-talk? Forges (1990) shows that there may exist communication equilibrium payoffs that cannot be implemented as *Nash* equilibrium payoffs of any long cheap-talk extension of  $\Gamma(p)$ . However, communication equilibrium payoffs can be implemented as *correlated* equilibrium payoffs (in the sense of Aumann (1974)). More precisely, a *feasibility theorem* holds: the set of all correlated equilibrium payoffs of all cheap-talk extensions of  $\Gamma(p)$  coincides with the set of all communication equilibrium payoffs. In fact, any communication equilibrium payoff can be achieved as a correlated equilibrium payoff of a cheap-talk extension of  $\Gamma(p)$  with only one stage of information transmission (see Forges (1985)).

### 5.2. Information Design Problems

The fundamental result we use to make the sender's problem more tractable is the revelation principle for general Bayesian games. This principle states that without loss of generality, the sender may restrict attention to communication equilibria, which are described by a set of linear inequalities. Therefore the sender's problem can be formulated as a linear programming problem. One advantage of this approach is that it does not depend on the number of receivers

or the fact that they are uninformed. Indeed, the revelation principle also applies for a persuasion problem with an arbitrary number of privately informed receivers. In this general setting, the mediator (communication device) first asks all informed players (sender and receivers) to simultaneously and confidentially reveal their individual types. Then he privately recommends an action to each receiver. In a communication equilibrium of this game, all informed players always report their types truthfully and the receivers always follow the prescribed recommendation. Here again, the set of communication equilibria is a convex polyhedron, thus our analytical framework readily extends to more general information design problems.

Unfortunately, our main results concerning the dual properties of the primal problem -Theorem 1 and Proposition 1- do not extend to general information design problems. The main difficulty comes from the fact that, because of the strategic externalities, each receiver's optimal action depends not only on the posterior beliefs she infers after receiving a recommendation, but also on the *non observed* recommendations made to the other receivers.

Our approach is reminiscent of a recent methodology developed by Taneva (2016) for the study of information design problems. She considers a basic game in which: (i) a set of multiple receivers have symmetric uncertainty about an unobserved payoff-relevant state with a commonly known prior distribution; and (ii), an information designer has preferences that depend on the state and the actions taken by the agents. The designer, before observing the realization of the state, commits to an information structure (i.e., a set of signals together with a signaling strategy). His problem then is to find an information structure which, for the given basic game, supports a Bayesian equilibrium that maximizes his expected payoff. Using the concept of *Bayes correlated equilibrium*, introduced by Bergemann and Morris (2016), Taneva characterizes the set of all Bayesian equilibria associated with all possible information structures for a given basic game. By doing so, she equivalently reformulates the designer's problem as a linear programming problem. The notion of Bayes correlated equilibrium considers only the strategic incentive constraints related to the "obedience" of the receivers. For that reason, Taneva's approach assumes full commitment on the part of the designer. In contrast, the concept of communication equilibrium captures the idea that players can strategically manipulate their information by imposing additional truth-telling incentive constraints. As a consequence, Taneva's formulation can be seen as a particular case of our analytical framework, one in which the receivers are uninformed (symmetric uncertainty) and the private information of the designer is verifiable by the mediator.

## 6. References

1. AUMANN, R. (1974). "Subjectivity and correlation in randomized strategies". *Journal of Mathematical Economics*, 1, pp. 67-96.
2. AUMANN, R. AND MASCHLER, M. (1995). *Repeated Games With Incomplete Information*, MIT Press.
3. BERGEMANN, D. AND MORRIS, S. (2016). "Bayes correlated equilibrium and the comparison of information structures in games". *Theoretical Economics*, 11, pp. 487-522.
4. BESTER, H. AND STRAUSS, R. (2001). "Contracting with imperfect commitment and the revelation principle: The single agent case". *Econometrica*, 69, pp. 1077-1098.



5. BESTER, H. AND STRAUZ, R. (2007). "Contracting with imperfect commitment and noisy communication". *Journal of Economic Theory*, 136, pp. 236-259.
6. CRAWFORD, V. AND SOBEL, J. (1982). "Strategic information transmission". *Econometrica*, 50, pp. 1431-1451.
7. FARRELL, J. (1993). "Meaning and credibility in cheap-talk games". *Games and Economic Behavior*, 5, pp. 514-531.
8. FORGES, F. (1985). "Correlated equilibria in a class of repeated games with incomplete information". *International Journal of Game Theory*, 14, pp. 129-150.
9. FORGES, F. (1986). "An approach to communication equilibria". *Econometrica*, 54, pp. 1375-1385.
10. FORGES, F. (1990). "Equilibria with communication in a job market example". *The Quarterly Journal of Economics*, 105, pp. 375-398.
11. FORGES, F. (1994). Non-zero sum repeated games and information transmission. In *Essays in Game Theory in Honor of Michael Maschler*, 65-95. (Ed.) Nimrod Megiddo, Springer-Verlag, New York.
12. GOLTSMAN, M., HÖRNER, J., PAVLOV, G. AND SQUINTANI, F. (2009). "Mediation, arbitration and negotiation". *Journal of Economic Theory*, 144, pp. 1397-1420.
13. KAMENICA, E. AND GENTZKOW, M. (2011). "Bayesian persuasion". *American Economic Review*, 101, pp. 2590-2615.
14. KOLOTILIN, A., LI, M., MYLOVANOV, T. AND ZAPECHELNYUK, A. (2015). Persuasion of a privately informed receiver. Working paper.
15. MITUSCH, K. AND STRAUZ, R. (2005). "Mediation in situations of conflict and limited commitment". *The Journal of Law, Economics and Organization*, 21, pp. 467-500.
16. MYERSON, R. (1981). "Optimal auction design". *Mathematics of Operations Research*, 6, pp. 58-73.
17. MYERSON, R. (1983). "Mechanism design by an informed principal". *Econometrica*, 51, pp. 1767-1797.
18. MYERSON, R. (1986). "Multi-stage games with communication". *Econometrica*, 54, pp. 323-358.
19. MYERSON, R. (1991). *Game Theory: Analysis of Conflict*, Harvard University Press.
20. SCHRIJVER, A. (1998). *Theory of Linear and Integer Programming*, John Wiley and sons.
21. TANEVA, I. (2016). "Information design". Working paper available at <https://sites.google.com/site/itaneva13/research>.