Mutual Persuasion*

Giuseppe Dari-Mattiacci
University of Amsterdam

Davide Grossi
University of Groningen

September 12, 2018

Abstract

Two agents have to collectively select one of two options. They are endowed with a personal bias, each in favor of a different option, and they observe a private signal with known quality. They then need to reveal their private signal to the other agent, but may decide to withhold some of the evidence the private signal provides, in order to persuade the other agent in the direction of their own bias. We present a Bayesian model capturing this form of persuasion. The model applies to a variety of phenomena, including political discussions, settlement negotiations and trade.

Keywords: persuasion, opinion formation, information manipulation, litigation, settlement, trade.

JEL codes: K41, D03, D83.

1 Introduction

People routinely try to persuade each other about their respective political views, future market trends, the quality of goods for sale, the outcome of adjudication, the abilities of a baseball coach, and many other serious or trivial topics. Yet, while trying to persuade another, one also tries to grasp useful information from the other's statements. While attempting to convince the seller that a good is of low value (and hence that the price should be low), a buyer will also typically try to infer whether the seller's claim that the good is in fact of high value (and hence the price should be high) is accurate. Stubbornly sticking to one's prior and therefore foregoing a good deal is a naive strategy. If the good is indeed of high quality, the buyer is better off paying a high price and going through with the transaction rather than leaving the good to someone else.

*__PRELIMINARY DRAFT__. The authors wish to thank the participants of the workshop "Reasoning in Social Context" (Amsterdam, November 2015) and of LOFT16 (Maastricht, July 2016), the reviewers of LOFT'16, J. Reinganum, A. Daughety, M. Pivato, M. Raskin, M. Siniscalchi for useful comments and discussions. Davide Grossi acknowledges support for this research from the UK Engineering and Physical Sciences Council (EPSRC) under the research grant EP/M015815/1.

Similarly, a voter may want to convince her friend to choose a candidate that seems particularly close to her preferences but, while doing so, she may also want to revise her assessment of the candidate if her friend brings convincing arguments in favor of an alternative candidate.

In this paper, we refer to these and similar scenarios as *mutual persuasion*. We introduce a Bayesian model that allows us to predict if information exchange will result in persuasion and, if so, whether persuasion will be unidirectional (with a persuading b or vice versa) or bidirectional (with a persuading b *and* b persuading a). We show that bidirectional persuasion is indeed a possible outcome of information exchange, so that both parties might leave the conversation having reversed their opinions. In some contexts, this might be a desirable outcome. Think of a plaintiff and a defendant engaged in settlement negotiations: a reversal of opinions will lead them to settle the case for some amount and hence save the costs of going to trial.

In recent years a growing literature has focused on economic aspects of processes of argumentation, deliberation, opinion formation and group decision-making in general (e.g., Austen-Smith and Feddersen (2005), Gerardi and Yariv (2007), Visser and Swank (2007), List et al. (2013), Dickson, Hafer, and Landa (2015)). Our model builds on recent literature on Bayesian persuasion (stemming from Kamenica and Gentzkow, 2011), which has typically focused on the attempts of one agent (a sender) to persuade another (a receiver or decision-maker). The asymmetric setting where a sender tries to persuade a receiver is the predominant setup in the economic literature on persuasion (cf. the aforementioned Kamenica and Gentzkow, 2011 but also, for a different model, Glazer and Rubinstein, 2001, 2006, 2012, 2004). We are not aware of studies of the symmetric setting we explore here, where parties act both as senders and receivers.

In our model, two parties exchange information about a particular event. They cannot lie but can introduce noise in the information they reveal in an attempt to induce the counterpart to misinterpret such information in a way more favorable to them. In other words a party may shade information in order to make it less informative to the other party. Persuasion critically depends on the parties' ability to shade. Quite realistically, we find that if a party's ability to shade is too great, the other party becomes too skeptical and persuasion fails. To persuade, parties must be able to shade but should not shade too much.

Information shading is so relevant in reality that it is explicitly prohibited in certain circumstances. For instance, the Brady rule[1] in the United States obliges prosecutors to disclose materially exculpatory evidence to the defense. Yet, there is no such obligation in private discussions and in negotiations. Only in some cases do contractual parties incur liability for failure to disclose information, while giving false information is generally punished (Kronman, 1978).

The model applies to independent decisions. With independent decisions, the parties exchange arguments that are instrumental to making two different choices. A party might be interested in and derive utility from the decision

[1] Brady v. Maryland, 373 U.S. 83 (1963).

taken by the other party, but the parties' decisions are independent actions. Political discussions and sports talk belong to this category, which can be referred to as opinion formation. The model, however, could be adapted to apply to joint decisions, where the parties' opinions after arguments are exchanged to determine a joint decision to, for instance, trade a good or settle a lawsuit. [2]

The paper is organized as follows. Section 2 contains the model, which is analyzed and solved in Section 3. Section 4 presents the results of our analysis and concludes.

2 The model

2.1 Setup

Consider two parties, a and b, who need to select one of two options $\gamma_i \in \{A, B\}$ (with $i \in \{a, b\}$) independently of each other. Either option may be the best-suited—i.e., may be the one that correctly tracks the state of the world—$\theta \in \{A, B\}$, but this is only imperfectly known to the parties. Parties have an objective preference for choosing the best option and an idiosyncratic preference for choosing the option they feel affinity with. All other things equal, a has an inclination towards A and b towards B. More precisely, a party receives a payoff equal to 1 if she chooses the best option, $\gamma_i = \theta$. She also receives an additional positive payoff v_i, if the best option happens to be her preferred option. The same additional payoff accrues if the other chooses that option. Note that the parties receive 0 from choosing the wrong option. In particular, choosing the wrong option yields 0 even if this option is the party's favorite choice.

To illustrate, consider two entrepreneurs (a and b) who are about to invest in an innovative project. The project can be pursued using either of two technologies (γ_i) but it is not known ex ante which of the two technologies leads to a success. If the entrepreneur chooses the right technology (θ), the project succeeds, yielding profits equal to 1 (profits are 0 if the entrepreneur chooses the wrong technology). In addition, each entrepreneur has a comparative advantage with a different technology and earns an additional payoff v_i—for instance, a fee for services provided to customers—every time that technology is adopted (including by the other entrepreneur) *and* is successful. The payoffs of this *biased truth-tracking game* are recapitulated in Table 1.

Therefore, although party a prefers option A, he or she will choose B if it is clear that B is the best option. However, since θ is unknown, party a's choice

[2] With respect to settlement negotiations, our approach contributes a novel reason why parties go to trial. Extant models focus either on divergent (irrational) priors (Shavell, 1982), where parties simply do not exchange information and, hence, litigate because their initial priors are different, or on asymmetric information (Bebchuk, 1984; Reinganum and Wilde, 1986; Spier, 1994), where parties would like to exchange information but cannot credibly do so before trial. In contrast, in our model, parties can credibly exchange information prior to trial but choose not to do so in order to obtain a more favorable settlement. More generally, in our model, ex post asymmetric information does not result from an ex ante commitment problem (as in Akerlof, 1970) but from the parties' strategic decisions to withhold some information in order to induce persuasion and obtain a better outcome.

			b	
		A	B
a	A	$1+2v_a, 1$	$1+v_a, 0$
	B	$v_a,\ 1$	$0, 0$

$\theta = A$

			b	
		A	B
a	A	$0, 0$	$0, 1+v_b$
	B	$1, v_b$	$1, 1+2v_b$

$\theta = B$

Table 1: Biased truth-tracking game. Nature chooses the left or the right matrix (i.e., the value of θ).

will be based on the probability that an option is the best, given the available information. Accordingly, party a's dominant strategy is to choose A if, and only if $\Pr(\theta = A \mid \cdot)(1 + v_a) > \Pr(\theta = B \mid \cdot)$ where the left-hand side of the inequality is the expected difference in payoffs from choosing A as compared to choosing B and the right-hand side is the expected payoff from choosing B. Noting that $\Pr(\theta = B \mid \cdot) = 1 - \Pr(\theta = A \mid \cdot)$, the inequality can be rewritten as

$$\Pr(\theta = A \mid \cdot) > \frac{1}{2 + v_a} \equiv t_a \tag{1}$$

where $0 < t_a < \frac{1}{2}$ can be thought of as a's idiosyncratic inclination for A (or a's ideological position), so that a chooses option A even when option B is more likely to be the (objectively) correct choice, that is, in cases where t_a is less than $\frac{1}{2}$.

Similarly, b's dominant strategy is to choose B if, and only if $\Pr(\theta = B \mid \cdot)(1 + v_b) > \Pr(\theta = A \mid \cdot)$, where the left-hand side is the expected payoff from choosing B and the right-hand side is the expected payoff from choosing A. As before, we can rewrite the inequality in a more convenient form as

$$\Pr(\theta = A \mid \cdot) < \frac{1 + v_b}{2 + v_b} \equiv t_b \tag{2}$$

where $\frac{1}{2} < t_b < 1$. Intuitively, parties can be thought of preferring false positives to false negatives with respect to their own biases, or to suffer from a confirmation bias.

2.2 Signals

Parties base their decisions on the three following pieces of information:

1. A common prior equal to $\frac{1}{2}$.
2. A private signal $s_i \in \{A, B\}$ for each party i. The signal has quality $q = \Pr(s_i = \theta \mid \theta) \in \left(\frac{1}{2}, 1\right)$, that is, it is more likely than not to be accurate although not perfectly accurate (full accuracy would make information exchange irrelevant). Note that $\Pr(s) = \frac{1}{2}$ (the prior).

		θ	
		A	B
s_i	A	q	$1-q$
	B	$1-q$	q

Table 2: Quality of the private signals s_i

3. A revealed signal $r_i \in \{A, B\}$ conveyed to party i by the other party. Such signal is a selective revelation of the sender's private signal. The idea is that a party can decide to withhold some evidence when revealing the signal, thereby reducing its quality. Selective revelation can be intuitively thought of as a form of "jamming" of the original private signal. If evidence is withheld, there is a probability that the other party will interpret the evidence in a different way, which is possibly more favorable to the sender than the original signal. Hence, let

$$p_i^{s_i} \equiv \Pr\left(r_j = s_i \mid s_i\right) \in [p, 1)$$

describe the probability that the receiver j (with $j \neq i$) interprets the revealed signal r_j in the same way as sender i interpreted the original signal s_i. The boundary value $p \in \left(\frac{1}{2}, 1\right]$ captures the maximum level of "jamming" that the sender can choose. This limit can derive from some natural limitation on shading or from the law.

		s_a	
		A	B
r_b	A	p_a^A	$1-p_a^B$
	B	$1-p_a^A$	p_a^B

		s_b	
		A	B
r_a	A	p_b^A	$1-p_b^B$
	B	$1-p_b^A$	p_b^B

Table 3: Quality of the revealed signal r_j

It is worth stressing the following features of the above setup:

- The sender i chooses p_i^A and p_i^B but cannot decide r_j directly. Intuitively, each party can only choose whether and how much to "jam" the original signal but cannot impose a particular interpretation of the same signal on the other party.
- The sender's choice is a continuous choice between p and 1.
- The receiver j observes r_j but does not observe p_i^A and p_i^B directly; however, the receiver will anticipate the sender's strategy and infer p_i^A and p_i^B in the equilibrium.

2.3 The persuasion game

The model has four time steps:

1. At T_0 Nature chooses t_a, t_b, q, and p, which are common knowledge.
2. At T_1 each party learns his or her signal s_i privately; the signals are independent random draws from the same conditional distribution described in Table 2.
3. At T_2 the parties simultaneously choose their revelation strategies, exchange signals r_j and learn the signal revealed to them by the other party.
4. At T_3 the parties make their choices $\gamma_i \in \{A, B\}$, that is they play the biased truth-tracking game (Table 1) based on the beliefs about θ they acquired at T_2.

The action of the game takes place at T_2 when the parties strategically select the quality of the signal to transmit to the other party depending on the private signal they receive. The strategies are described in Table 3 above. A *revelation strategy* for party i is denoted $p_i = \left\{p_i^A, p_i^B\right\}$. We call this type of strategic interaction a *mutual persuasion game.*[3] It should be clear that such mutual persuasion game can be thought of as a Bayesian extensive form game with simultaneous and observable actions. We spell this out in detail in the Appendix. The analysis provided in the following sections studies a specific perfect Bayesian Nash equilibrium for this game, depending on the parameters t_a, t_b, q, and p.

3 Analysis

In this section, we first illustrate a benchmark case in which each party has direct access to the signal drawn by the other party. We then analyze parties' strategic interaction. We start we examining the parties' optimal choices at T_3, and then use them to analyze their rational revelation strategies at T_2. The key result we are after is the characterization of conditions under which a party's revelation results in successful persuasion.

3.1 Benchmark: sincere revelation

Let us first characterize the benchmark case where parties revel their signals sincerely (that is, we impose $p_a^A = p_a^B = p_b^A = p_b^B = 1$). Table 4 recapitulates the probabilities $\Pr(s_a, s_b \mid \theta)$, which are easily derived from Table 2.

Using Table 4, we now turn to the probabilities $\Pr(\theta \mid s_a, s_b)$ of the state of the world given the two individual signals. This is obtained by repeated application of Bayes' rule,[4] and is given in Table 5.

Note that, given our assumption $q < \frac{1}{2}$, we have $\frac{q^2}{q^2+(1-q)^2} > \frac{1}{2}$ and $\frac{(1-q)^2}{q^2+(1-q)^2} < \frac{1}{2}$. Although the game is perfectly symmetric, given this information, party a

[3] The term "persuasion game" is used also in Glazer and Rubinstein (2006) (cf. also the related work in Glazer and Rubinstein (2001)and Glazer and Rubinstein (2012)) to refer to a different game, consisting of an extensive form strategic interaction where a speaker tries to convince a listener to take a specific action.

[4] For instance: $\Pr(\theta = A \mid s_a = A, s_b = A) = \frac{\Pr(\theta = s_a = s_b = A)}{\Pr(s_a = s_b = A)}$.

		s_b: A	s_b: B
s_a: A		q^2	$q(1-q)$
s_a: B		$q(1-q)$	$(1-q)^2$

$\theta = A$

		s_b: A	s_b: B
s_a: A		$(1-q)^2$	$q(1-q)$
s_a: B		$q(1-q)$	q^2

$\theta = B$

Table 4: Joint probabilities of two independent individual signals

		s_b: A	s_b: B
s_a: A		$\frac{q^2}{q^2+(1-q)^2}$	$\frac{1}{2}$
s_a: B		$\frac{1}{2}$	$\frac{(1-q)^2}{q^2+(1-q)^2}$

$\Pr(\theta = A \mid s_a, s_b)$

		s_b: A	s_b: B
s_a: A		$\frac{(1-q)^2}{q^2+(1-q)^2}$	$\frac{1}{2}$
s_a: B		$\frac{1}{2}$	$\frac{q^2}{q^2+(1-q)^2}$

$\Pr(\theta = B \mid s_a, s_b)$

Table 5: Probability of θ given the individual signals

will choose A in three out of the four cases, that is, whenever at least one of the signals indicates A because his or her persuasion threshold t_a is less than $\frac{1}{2}$. Party a may choose B only if both signals indicate B and if $\frac{(1-q)^2}{q^2+(1-q)^2} < t_a$; she will still choose A if the latter condition is not satisfied, even though both signals indicate B. Similarly, party b will choose B more often than A. Here we see the parties' idiosyncratic preferences at work, absent any strategic interaction between them. Let us now turn to such strategic interaction.

3.2 Parties' optimal choices at T_3

Let us start with party a and characterize party a's optimal choice γ_a at T_3 by identifying the conditions (as a function of a's private signal and of b's revealed signal) under which a particular choice maximizes a's expected utility.

- If $s_a = A$, then party a chooses for A irrespective of b's revealed signal. This is the case because, as observed in the above benchmark analysis, $\Pr(\theta = A \mid s_a = A, s_b) \geq \frac{1}{2} > t_a$ irrespective of s_b.
- If $s_a = B$, then party a's choose depends on what a thinks b's signal is and this information can be inferred from the signal that b reveals to a. In particular, a will choose B, that is, will be convinced by b's revealed signal if $\Pr(\theta = A \mid s_a = B, r_b) \leq t_a$. The latter can be written as follows:

$$\begin{array}{l} \Pr(\theta = A \mid s_a = B, s_b = A) \Pr(s_b = A \mid s_a = B, r_b) \\ + \Pr(\theta = A \mid s_a = B, s_b = B) \Pr(s_b = B \mid s_a = B, r_b) \quad \leq \quad t_a \end{array} \tag{3}$$

  where the first term is the probability that $\theta = A$ when $s_b = A$ times the probability that $s_b = A$ given the argument r_b; similarly, the second term

is the probability that $\theta = A$ when $s_b = B$ times the probability that $s_b = B$ given the revealed signal r_b. Equation (3) is party a's *persuasion constraint*, that is, the constraint that needs to be satisfied for a to be persuaded by b's revealed signal.

It is worth observing that, given the payoffs in Table 1 party a is strictly better of whenever party b chooses A.The same analysis applies *mutatis mutandis* to b.

3.3 Parties' revelation strategies at T_2

The choices made at time T_3 depend on the expectation a party has about the state of the world. These expectations depend in turn on the party's private signal and on the revealed signal she receives. So parties can influence each other's choices through their revealed signals at T_2. We start from party b trying to convince party a to choose B. Party b wants to maximize the probability that party a chooses for B. To do so, party b needs to set up her revelation strategy $p_b = \left\{p_b^A, p_b^B\right\}$ in such a way that party a will receive the signal $r_b = B$ as often as possible while still "believing" it (in a Bayesian sense). In other words, party b maximizes $\Pr(r_b = B \mid s_b)$ subject to the persuasion constraint in (3) (solved below in (4)).

First of all, observe that, from Table 3:

$$\begin{aligned} \Pr(r_b = B \mid s_b = A) &= 1 - p_b^A \\ \Pr(r_b = B \mid s_b = B) &= p_b^B \end{aligned}$$

It follows that party b's optimal strategy to maximize these probabilities is:

- If $s_b = A$, party b will set p_b^A as low as possible, that is, at $p_b^A = p$, which maximizes the probability that party a will misinterpret B's revealed signal;
- If $s_b = B$, party b will set p_b^B as high as possible, that is, at $p_b^B = 1$ and provide all evidence to party a.

In other words: if b receives a signal $s_b = B$, she will reveal to a all the evidence associated with that signal so hat a interprets the signal in the same way as b did. If instead b receives a signal $s_b = A$, she will give a only part of the evidence in the hope that a will interpret it differently, that is, as a signal for B. The problem is that if p_b^A is set too low, evidence becomes unreliable and party a will not be swayed.

Therefore the quality p_b of party b's revealed signal r_b becomes:

The following is worth stressing. Party a does not observe p_b directly. If this were the case, party a could infer that $s_b = A$ by simply observing $p_b < 1$. This is not possible since the quality of the signal is not observable. However, a will expect b to act strategically and hence will anticipate the value of p_b and use it in calculating her expectations. A similar analysis applies to party a.

		s_b	
		A	B
r_b	A	p_b^A	0
	B	$1-p_b^A$	1

Table 6: Quality p_b of the revealed signal r_b

3.4 Solving the persuasion constraint

To specify party b's strategy when $s_b = A$ recall that if party a has a private signal $s_a = A$, he or she cannot be convinced, hence party b bases his or her strategy on the assumption that party a has a signal $s_a = B$, which is the only case when r_b may matter. This is important because, in order to set b's optimal strategy, we need to calculate $\Pr(s_b \mid r_b)$ in (3), that is, party a's expectation of b's signal given b's revelation, which in turn depends on the conditional probability of s_b given that a has observed $s_a = B$.

Now recall equation (3). Our analysis of the benchmark case gave us exact values for $Pr(\theta = A \mid s_a = B, s_b = A) = \frac{1}{2}$ and for $Pr(\theta = A \mid s_a = B, s_b = B) = \frac{(1-q)^2}{q^2+(1-q)^2}$. We need to obtain exact values for $\Pr(s_b = A \mid s_a = B, r_b)$ and for $\Pr(s_b = B \mid s_a = B, r_b)$, where $r_b \in \{A, B\}$.

From Table 6 it is easy to see that, if party a observes $r_b = A$, she can be certain that b's individual signal was A, irrespectively of s_a. That is, we have $\Pr(s_b = A \mid r_b = A) = 1$. The persuasion constraint in (3) is therefore not satisfied and, as expected, a will not be persuaded in this case.

If, instead, party a observes $r_b = B$, then we need now to compute $\Pr(s_b = A \mid s_a = B, r_b = B)$:

$$
\begin{aligned}
& \Pr(s_b = A \mid s_a = B, r_b = B) \\
= \; & \frac{Pr(s_b = A, s_a = B, r_b = B)}{Pr(s_a = B, r_b = B)} \\
= \; & \frac{Pr(s_b = A, r_b = B \mid s_a = B) Pr(s_a = B)}{Pr(s_a = B, r_b = B)} \\
= \; & \frac{Pr(r_b = B \mid s_b = A) Pr(s_b = A \mid s_a = B)}{Pr(r_b = B \mid s_a = B)} \\
= \; & \frac{Pr(r_b = B \mid s_b = A) Pr(s_b = A \mid s_a = B)}{Pr(r_b = B \mid s_b = A) Pr(s_b = A \mid s_a = B) + Pr(r_b = B \mid s_b = B) Pr(s_b = B \mid s_a = B)} \\
= \; & \frac{(1-p_b^A) Pr(s_b = A \mid s_a = B)}{(1-p_b^A) Pr(s_b = A \mid s_a = B) + 1 Pr(s_b = B \mid s_a = B)} \\
= \; & \frac{(1-p_b^A) 2q(1-q)}{(1-p_b^A) 2q(1-q) + (q^2 + (1-q)^2)}
\end{aligned}
$$

The first equality applies the definition of conditional probability. The second equality holds by multiplication of probabilities. The third equality assumes that a learns signal $s_a = B$ (i.e., it assumes that $\Pr(s_a = B) = 1$ in the numerator) and applies multiplication of probabilities in the denominator with the

assumption that $\Pr(s_a = B) = 1$. The fourth equality expands the denominator summing the conditional probabilities conditioned on the two events $s_b = A$ and $s_b = B$ (under the condition $s_a = B$). The fifth equality substitutes the values of table 6 in the expression. Finally, the sixth equality substitutes the values obtained via the two following series of equalities, which make use of the values of Table 2:

$$\begin{aligned}
&Pr(s_b = A \mid s_a = B) \\
&= Pr(s_b = A \mid \theta = A)Pr(\theta = A \mid s_a = B) + Pr(s_b = A \mid \theta = B)Pr(\theta = B \mid s_a = B) \\
&= 2q(1-q)
\end{aligned}$$

and

$$\begin{aligned}
&Pr(s_b = B \mid s_a = B) \\
&= Pr(s_b = B \mid \theta = A)Pr(\theta = A \mid s_a = B) + Pr(s_b = B \mid \theta = B)Pr(\theta = B \mid s_a = B) \\
&= (1-q)^2 + q^2
\end{aligned}$$

More concisely, to obtain the persuasion value we first apply Bayes' rule to calculate the probability of s_b conditional on $s_a = B$ (this is done from Table 2 and using $1 - q$ as the new prior for $\theta = A$ given the observation of $s_a = B$). Then, this probability is used as distribution of s_b in a second application of Bayes' rule.

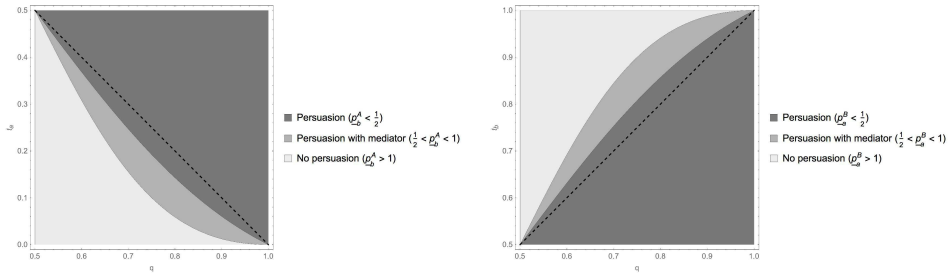
Now we can substitute the values obtained above and the values from Table 5 into a's persuasion constraint in (3), which becomes:

$$\begin{aligned}
&\frac{1}{2}\frac{\left(1-p_b^A\right)2q\left(1-q\right)}{\left(1-p_b^A\right)2q\left(1-q\right)+1\left(q^2+\left(1-q\right)^2\right)} \\
&+\frac{\left(1-q\right)^2}{q^2+\left(1-q\right)^2}\left(1-\frac{(1-p_b^A)2q(1-q)}{(1-p_b^A)2q(1-q)+(q^2+(1-q)^2)}\right) \quad \leq \quad t_a \qquad (4)
\end{aligned}$$

We now solve (4) for p_b^A to obtain the smallest value that satisfies (3):

$$\underline{p}_b^A = \frac{(1-q-t_a)}{(1-q)q(1-2t_a)}$$

We call $\underline{p}_b^A$ b's *persuasion value*, and it denotes the minimum value of p_b^A that party a can accept. Since revelation strategies are not observable and we assume that neither party is able to commit to a particular revelation strategy, party a will expect b to shade up to the limit p. Therefore, party a will be persuaded if $\underline{p}_b^A < p$ and will not be persuaded otherwise. Therefore, we can distinguish between two cases: if $\underline{p}_b^A > 1$, there there will be no persuasion irrespective of p; if instead $\underline{p}_b^A \leq \frac{1}{2}$ persuasion will occur irrespective of p. In the intermediate region, persuasion occurs for an appropriate value of p such that $p \geq \frac{(1-q-t_a)}{(1-q)q(1-2t_a)}$. The intuition is that, in this region, institutions that prevent parties from shading too much (we may call them "mediators") have the effect of expanding the scope of persuasive interactions, allowing party a to rely more convincingly on representations made by b and vice versa.



(a) Persuading party a (b) Persuading party b

Figure 1: Persuasion

4 Results

We discuss here how the model predicts the two parties may choose at time T_3 given the individual signals obtained at time T_1. Table 7 recapitulates the possible outcomes we discuss below.

		s_b	
		A	B
s_a	A	AB or AA	AB
	B	AA, BB, AB, BA	AB or BB

Table 7: Possible outcomes of the choice at time T_3 given the individual signals at time T_1

4.1 No persuasion

If both agents obtain individual signals which are in line with their private bias (i.e., $s_a = A$ and $s_b = B$) then no persuasion is possible because, as observed in Section 3.1, $\Pr(\theta = A \mid A, s_b) \geq \frac{1}{2} > t_a$ irrespective of s_b and $\Pr(\theta = B \mid B, s_a) \geq \frac{1}{2} > 1 - t_b$ irrespective of s_a.

4.2 Unidirectional persuasion

If one party receives a private signal, which is in line with her private bias, but the other party does not (e.g., without loss of generality, $s_a = B$ and $s_b = B$) then two outcomes are possible:

1. No persuasion happens and both parties choose according to their bias. This is the case if $P_b^A > p$.

2. The party who received the signal contradicting his bias is persuaded and agreement is reached. This is the case if $P_b^A \leq p$.

4.3 Bidirectional persuasion—reversal of opinions

This is the interesting case in which both parties receive private signals that contradict their bias. The following outcomes result from the interaction:

1. No persuasion happens. This is the case if no party is swayed by the other party's revealed signal, that is: $P_b^A > p$ and $P_a^B > p$.

2. Only one of the parties succeeds in persuading the other party, and agreement is reached. This is the case if $P_b^A > p$ but $P_a^B \leq p$, and hence both parties choose A, or $P_a^B > p$ but $P_b^A \leq p$, and hence both parties choose B.[5]

3. Both parties succeed in persuading the other party and a reversal of opinions occurs. This is the case if both $P_b^A \leq p$ and $P_a^B \leq p$.

5 Conclusions

We have presented a model of Bayesian persuasion that describes the revelation strategies of two parties who try to persuade each other while also attempting to gauge useful information from the discussion. We show under what conditions a party will persuade the other and vice versa. At the core of our result is the simple observation that a party wants to persuade the other only when he or she anticipates that the information held by the other party can be safely disregarded. Since a party's revealed information is pivotal for the other party's decision only when the other party has concurrent information, parties take a relatively aggressive stand in persuasion and it is possible that *both* parties succeed in persuading the other. This framework applies both to independent decisions (which option to choose) and to joint decisions (whether to settle a case). We offer a novel explanation of why negotiations fail: parties might leave the negotiation table with divergent opinions because each of them has strategic incentives not to reveal all the information at his or her disposal to the other party. This might happen even if information is perfectly verifiable.

References

Akerlof, George A. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics* 84 (3):488.

Austen-Smith, D. and T. Feddersen. 2005. "Deliberation and Voting Rules." In *Social Choice and Strategic Decisions*, Studies in Social Choice and Welfare. Springer.

Bebchuk, Lucian Arye. 1984. "Litigation and Settlement under Imperfect Information." *The RAND Journal of Economics* 15 (3):404–415.

[5] Obviously this cannot happen if $t_a = 1 - t_b$ as in such case both parties would have the same level of bias towards their preferred option.

Dickson, Eric, Catherine Hafer, and Dimitri Landa. 2015. "Learning from Debate: Institutions and Information." *Political Science Research and Methods* 3 (3):449–472.

Gerardi, D. and L. Yariv. 2007. "Deliberative Voting." *Journal of Economic Theory* 134:317–338.

Glazer, J. and A. Rubinstein. 2001. "Debates and Decisions: On a Rationale of Argumentation Rules." *Games and Economic Behavior* 36 (2):158–173.

———. 2004. "On Optimal Rules of Persuasion." *Econometrica* (72):1715–1736.

———. 2006. "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach." *Theoretical Economics* 1:395–410.

———. 2012. "A Model of Persuasion with a Boundedly Rational Agent." *Journal of Political Economy* 120:1057–1082.

Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6):2590–2615.

Kronman, Anthony. 1978. "Mistake, Disclosure, Information, and the Law of Contracts." *Journal of Legal Studies* 7 (1):1–34.

List, C., R. Luskin, J. Fishkin, and McLean I. 2013. "Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls." *The Journal of Politics* 75 (1):80–95.

Osborne, M. J. and A. Rubinstein. 1994. *A Course in Game Theory*. MIT Press.

Reinganum, Jennifer F and Louis L Wilde. 1986. "Settlement, Litigation, and the Allocation of Litigation Costs." *The RAND Journal of Economics* 17 (4):557–566.

Shavell, Steven. 1982. "Suit, Settlement, and Trial: A Theoretical Analysis Under Alternative Methods for the Allocation of Legal Costs." *Journal of Legal Studies* 11 (1):55–82.

Spier, Kathryn E. 1994. "Pretrial Bargaining and the Design of Fee-Shifting Rules." *The RAND Journal of Economics* 25 (2):197–214.

Visser, B. and O. Swank. 2007. "On Committees of Experts." *The Quarterly Journal of Economics* 122 (1):337–372.

A Appendix

The persuasion game is a Bayesian extensive game with simultaneous and observable actions (cf. Osborne and Rubinstein (1994)). Here we present it explicitly in such general form, that is, as a tuple$(G, (T_i, P_i, \mu_i)_{i \in N})$ where:

- $G = (N, H)$ is an extensive game form with $N = \{a, b\}$ and $H = \{\emptyset\} \cup \{A, B\}^2 \cup \{A, B\}^2; \{A, B\}^2$,that is, the set of all histories of the game (the empty history, the histories consisting of the simultaneous revelation fo two signals, and the histories consisting of the simultaneous revelation of two signals followed by the simultaneous revelation of two choices). The set of terminal histories is denoted $Z = \{A, B\}^2; \{A, B\}^2$.

- $T_i = \{A, B\}$ is the type set consisting of the two types of signals agent i may receive from Nature before playing. A profile of types is denoted $\mathbf{s} \in T_a \times T_b = \{A, B\}^2$ and s_i denotes the i^{th}-projection of $\mathbf{s}$.

- $P_i = \Delta(T_i)$ is the probability distribution over the type set of i, which in our case is $(0.5, 0.5)$.

- $u_i : \{A, B\}^2 \times Z \to \mathbb{R}$ is the utility function of player i associating a payoff to every terminal history of the game depending on a given type profile. In the persuasion game this function depends on the payoffs given by the biased truth-tracking game matrix. Given a terminal history z, let $z_c \in \{A, B\}^2$ (right projection of z) denote the profile of choices made in z, and η^A (resp. η^B) be the left (resp. right) payoff matrices of Table 1. Then the above utility function is defined as follows:

$$u_i(\mathbf{s}, z) = \eta_i^A(z_c) Pr(\theta = A \mid \mathbf{s}) + \eta_i^B(z_c) Pr(\theta = B \mid \mathbf{s})$$

  that is, i's (a posteriori) expected payoff of a voting profile given the profile of types $\mathbf{s}$. Notice that this function depends on the quality of signals q and the additional payoffs v_i (cf. Section 2).

In such a game a behavioral strategy $\sigma_i(t_i)$ is a function that, given a type (that is, the signal revealed by Nature), assigns to each non-terminal history h in the game tree a probability distribution in $\Delta(\{A, B\})$, that is, first a probability distribution over the possible signals to be revealed to the other party, and then a probability distribution over the possible choices to be made. Furthermore $\mu_i(h) \in \Delta(T_i)$ denotes the belief that the other player $-i$ has about the type of i at history h (possibly terminal). an option (for an equilibrium) is a pair (of pairs) $((\sigma_a, \sigma_b), (\mu_a, \mu_b))$ collecting, for each agent, her behavioral strategy and her belief about the type of the other player.

The strategies we are concerned with in this paper can be thought of as functions associating to each type a pair of (possibly probabilistic) revelatios and choices: $\sigma_i(t_i) = (r, c)$ where $r \in [0, 1]$ represents the probability that i reveals signal I (that is reveals the signal corresponding to her bias), and $c \in [0, 1]$ represents the probability that i chooses I according to her bias. If r and v are drawn from $\{0, 1\}$ then the strategy σ_i is said to be pure.

It is worth pointing to some natural classess of strategies. A fully biased strategy for agent i is the pure strategy $(1,1)$, accordint to which the agent reveals only signals, and chooses only, in accordance to her bias. An informative strategy is a strategy (r,c) where $r=1$ if $t_i = I$ and $r=0$ otherwise (with c arbitrary), that is a pure strategy in which the agent always reveals her signal sincerely. A truthful strategy is a strategy (r,c) where $c=1$ iff $Pr(\theta = I \mid t_i, r_{-i}) > t_i$ (with $r$arbitrary), that is i chooses for I iff she considers the likelihood of that state of the world larger than her threshold t_i, given her individual signal and the signal revealed by $-i$. In the paper we studied strategies that were pure in their choice component $c \in \{0,1\}$ and mixed in their revelation component r, but with a lower bound $p$on mixing, that is, $r \in (p,1]$. This class of strategies, to which we refer as strategies with pure choice and constrained mixed revelation, is the class upon which we focus in the paper. Within this class we signle out strategies of the following type: (r,c) where $r=1$ if $t_i = I$ and $r=0$ otherwise, and where $c=1$ iff $Pr(\theta = I \mid t_i, r_{-i}) > t_i$. Let us such strategies p-informative and truthful.

Theorem 1. *The strategy-belief pair* $((\sigma_a,\sigma_b),(\mu_a,\mu_b))$ *where* σ_i *is* p*-informative and truthful (for* $i \in \{a,b\}$*), and* μ_i *is such that, for any history* h *and* $\alpha \in \{A,B\}^2$*.*[6]

$$\mu_i(h\alpha)(s_i') = \frac{\sigma_i'(s_i')(h)(\alpha_i)\cdot \mu_i(h)(s_i')}{\sum_{s_i \in A,B} \sigma_i'(s_i')(h)(\alpha_i)\cdot \mu_i(h)(s_i')}$$

is a perfect Bayesian equilibrium in strategies with pure choice and constrained mixed revelation.

[6]Intuitively, the formula expresses a consistency criterion between the action performed by i in history h and $-i$'s belief about i's type at that history, given strategy σ_i. The formula states that $-i$'s belief about i's type at history $h\alpha$ is obtained through Bayesian update given $-i$'s belief about i's type at history h and the action performed by i at history h given strategy σ_i. In the persuasion game this reasoning is critical in the choice stage of the game (T_3) after the players have observed each other's revealed signal.