Bayesian Persuasion in the Digital Age

Gabriel Martinez-Roa*

For the most recent version click here

November 14, 2019

Abstract

In digital platforms, agents have vast access to information, but the quality of it is unclear. I study this phenomenon as a Bayesian persuasion game with multiple senders that have partial control over the beliefs of a receiver. The receiver knows the signal chosen by all senders but randomly observes the realization of only one such signal. Senders can pool their signals, so the receiver is uncertain about the informativeness of the message received. This uncertainty may incentivize a sender to provide more or less information compared to the benchmark without uncertainty about the sender. The central insight is that each sender’s signal is chosen to affect the average correlation between messages and the state of the world, given the communication strategies of other senders. I derive policy recommendations to improve the quality of the information on platforms like social media.

JEL Classification: D82, D83, D91

Keywords: Bayesian Persuasion, Uncertainty, Social Media

1 Introduction

The question of how and when information can change people’s beliefs about the world, and consequently their actions, is of great importance in economics and other social sciences. In the current digital age, people have vast access to information on practically any topic they are interested in learning about. Intuitively, the great variety of information and diversity of sources should lead to well-informed actions. However, paradoxically, there is growing concern about misinformation, and people’s apparent disregard of factual evidence (See Lazer et. al, 2018). The internet has both allowed for an increase in the number of sources where people get information, by lowering the cost of entry, but this increase in the number of sources obscures the quality and reliability of each of them. Many of the new sources mimic reputable sources in form, but not in the editorial rigor and quality of information. According to the Pew Reseach Center (2018), 68% of Americans reported

*University of Wisconsin-Madison, Department of Economics; E-mail gamartinez@wisc.edu Website: https://gabriel-martinez-roa.github.io
getting news from social media\footnote{In their survey, Facebook, Twitter, Youtube or Reddit were the main pathways to news for 81% of the U.S. adults that get news from social media.}. Of them, however, 57% report that they expect social media news to be largely inaccurate.

In this paper I propose a simple model where the uncertainty about the quality of information is endogenous. While I allow information providers to freely choose how much information to produce - a Blackwell experiment - the receiver does not process all the information that is available. Instead, the receiver knows aggregate information about the choices of the senders, but observes the realization of a single experiment which is anonymous. I explore how these restrictions in communication affect three important aspects of communication: the incentives for senders to produce information, the type of information that senders use to maximize their expected influence on the receiver, and the evolution of receivers’ beliefs - how much they learn in expectation.

I use the framework to study important policies that are commonly used to promote the disclosure of better information. I find that a hands-off approach of allowing any type of information to circulate in a platform can have negative and disproportionate effects in undermining the credibility of all sources even when the proportion of “noisy senders” is relatively small. I also illustrate how increasing the variety of sources with heterogeneous interests can be detrimental to the quality of information, due to the anonymity assumption. In fact, I show that platforms can improve the quality of communication by allowing senders to self-verify their identity. In particular, senders and receivers can benefit from eliminating the uncertainty about the biases of the receiver. With such transparency, the sender gains control of the evolution of beliefs of the decision maker; hence the incentives to produce information increase. Further, the ability to separate from other senders eliminates the incentives to exaggerate one’s position in order to influence the average informativeness of messages; thus increasing the quality of the information produced.

This paper departs from previous literature by adopting two key assumptions. Senders choose Blackwell experiments in the vein of the growing Bayesian persuasion literature, but the receiver does not observe the realization of all such experiments. The first assumption allows me to focus on the incentives to produce information, rather than on the incentives to disclose information. As noted by Kamenica and Gentzkow (2011), assuming that senders can commit to a Blackwell experiment, characterizes an upper bound on the extent to which communication can occur. In this paper, the commitment assumption is somehow relaxed. While senders can still freely choose a Blackwell experiment and the choice is observed. The message space is shared across senders, so they can effectively pool with other senders to hide their incentives to persuade and affect the overall process of communication.

Previous research has studied how the uncertainty about the incentives (or bias) of a sender affects a sender’s incentives to disclose the information they have (see Morgan and Stocken, 2003, Dimitrakas and Sarafidis, 2005, Li and Madarasz, 2008 and the appendix in Chakraborty and Harbaug, 2010). That is, the sender knows the state of the world and uses cheap-talk to fully or partially disclose such information. Similarly, Shin (1994) and Wolinsky (2003) study similar setups in communications games with verifiable information, in the vein of Grossman (1981) and
This paper fills a gap by examining these questions within the Bayesian persuasion framework. Abstracting from the incentives to disclose allows me to uncover a new mechanism where opposing biases unambiguously lowers the incentive to provide information. Further, since Bayesian persuasion characterizes an upper bound in communication, the negative results uncovered in this paper are informative about any more complex model that also considers the incentives to disclose information.

The second assumption in the model departs from previous literature by endogenizing the uncertainty about the informativeness of messages received. While Le Treust and Tomala (2017) and Tsakas and Tsakas (2018) consider models with noisy bayesian persuasion, they characterize optimal signals for given exogenous and known garblings of the experiment chosen. Instead each sender in my model only knows that their chosen signal will be pooled with the signals of other senders, so the garbling process is endogenous. Further, In section 5.1 I discuss how the model can be seen as an extension of Gentzkow and Kamenica’s 2017 series of papers on competitive persuasion where the assumption that receivers process all the information available is dropped. Relaxing this assumption provides a more nuanced perspective about the role of increasing the number of senders.

To illustrate some of the main mechanisms present in the model, let’s consider a simple extension of Kamenica and Gentzkow’s (2011) Bayesian persuasion model (KG11 henceforth). While they find that even a sender with an extreme bias can benefit from persuasion whenever they can commit to a signal structure. In this paper, I show how the value for providing information depends on the beliefs of the receiver about the source of information. For example, respectable information providers like think-tanks, journalists, or analysts can develop standards to disclose all the information they have (even if they are biased), but if other actors can pool their messages with them – if they can appear to look like a respectable source – the receiver faces a double inference problem. From any message received she must infer both the true state of the world and the reliability of the source. This dimension of uncertainty benefits some senders, while harms others, and shapes their choice of an optimal signal.

Let’s consider a simple example. Suppose there is an election and there are two candidates in the ballot, A and B, a voter wants to vote for the best candidate and has no preferred party. That is, the voter casts a vote for whichever candidate is more likely to be the best. The common prior belief is that candidate A is more likely to be the best, say with 80% confidence. Before the election the voter will read about the candidates and then make up her mind. Suppose there are two kinds of journalist producing information about candidates: a good one (g) that wants to reveal as most information as possible (an unbiased journalist), and a bad one (b) that wants the voter to choose the wrong candidate. Each of them can choose a signal which is a function from the state space, \( \Theta = \{A, B\} \) into a distribution over recommendations, \( \Delta(M) \) where \( M = \{a, b\} \).

Let’s suppose the good journalist chooses a perfectly informative signal, denoted \( \pi_g \), while the bad journalist chooses a perfectly uninformative signal, denoted \( \pi_b \).

I later justify the use of direct messages in this setting and discuss the role of the message spaces for each sender.
\[ \pi_g(a|A) = 1 \quad \pi_g(b|A) = 0 \quad \pi_b(a|A) = \frac{1}{2} \quad \pi_b(b|A) = \frac{1}{2} \]

\[ \pi_g(a|B) = 0 \quad \pi_g(b|B) = 1 \quad \pi_b(a|B) = \frac{1}{2} \quad \pi_b(b|B) = \frac{1}{2} \]

The voter is aware of these two types of journalists and knows how each of them produce recommendations. However, she cannot read all of the information produced (in this example two pieces of information), instead she consumes whichever information she receives first. Let’s assume that she is equally likely to receive the recommendation from either type of journalist before the election. Then, her posterior belief after receiving each recommendation is as follows.

\[
\mu_1(A|a) = \frac{\pi_g(a|A) \cdot 0.5 \cdot \mu_0(A) + \pi_b(a|A) \cdot 0.5 \cdot \mu_0(A)}{\sum_{\theta \in \{A,B\}} (\pi_g(a|\theta) \cdot 0.5 \cdot \mu_0(\theta) + \pi_b(a|\theta) \cdot 0.5 \cdot \mu_0(\theta))} \approx 0.92
\]

\[
\mu_1(A|b) = \frac{\pi_g(b|A) \cdot 0.5 \cdot \mu_0(A) + \pi_b(b|A) \cdot 0.5 \cdot \mu_0(A)}{\sum_{\theta \in \{A,B\}} (\pi_g(b|\theta) \cdot 0.5 \cdot \mu_0(\theta) + \pi_b(b|\theta) \cdot 0.5 \cdot \mu_0(\theta))} \approx 0.57
\]

The receiver’s best response after either message is to vote for candidate A. Since this is the same action she would have taken had she not received any information, it is clear that no relevant information was communicated to the receiver. Further, no sender has a profitable deviation in terms of the signals they can choose. The receiver is choosing her default action regardless of the message received, so the probability of choosing the wrong candidate is maximized given the prior information (20\% probability making a mistake) . The bad journalist is getting his first best. In contrast, even though the good journalist is obtaining his worst possible payoff, he is already providing as much information as possible, so it has no profitable deviation.

Notice that if the receiver faced no uncertainty, as in KG11, the bad journalist would be indifferent between any uninformative signal. This is no longer true in this model. For example, if the bad journalist were to choose instead a signal that always recommends the receiver her default action (to vote for the candidate A), then the receiver will vote for B whenever she receives information recommending b. The receiver’s posteriors after each message would be

\[
\mu_1(A|a) = \frac{0.8}{0.8 + 0.1} \approx 0.89 \quad \mu_1(A|b) = \frac{0}{0 + 0.1} = 0
\]

After receiving message b, the receiver can infer that it must be coming from the good journalist, so she can also infer that the best candidate is B. Clearly, relevant information is communicated to the receiver now, and the probability of mistakes is down to 10\%. Therefore, the bad journalist has strict preferences between uninformative signals, and the value of the perfectly informative signal chosen by the good journalist depends on the signal chosen by the bad journalist. Furthermore, this also illustrates that signal structures cannot be reduced to the distribution over posterior beliefs about the state of the world that they induce as in KG11.

The techniques used in this paper are closer to Bergemann and Morris (2017). In their setup with a single sender, his problem reduces to choosing a signal that satisfies an obedience constraint. Here,
each sender faces an extra feasibility constraint defined by the actions of other senders. Theorems 1 and 2 characterize geometrically and in terms of its informativeness the set of feasible signals given any profile of actions of the other senders. The remaining of the theorems characterize equilibrium for large classes of games – games with a sender that has incentives for the receiver to not obtain valuable information, games with senders with state independent preferences, games with senders that have private information and games were senders can verify their true motives. The structure of the remaining of the paper is as follows. In the next section I present the model. In section 3 I characterize equilibrium in general and the new constraint given by the set of feasible signals to each sender. Section 5 specializes to the aforementioned classes of games. Section 6 reviews previous work in the literature and 6.1 studies how the assumptions in this model compare to the assumptions made in previous research on Bayesian persuasion with multiple senders. Finally, section 7 concludes.

2 Model

There are $S$ senders and a single receiver. The receiver and each sender has preferences (possibly state-dependent) over the action chosen by the receiver, $a \in A$. Let $u_s(a, \theta)$ be the utility for sender $s \in S$ and $u_r(a, \theta)$ be the utility for the receiver, where $\theta \in \Theta$ is the state of the world. Both $\Theta$ and $A$ are assumed finite. Each sender can choose a signal - or Blackwell experiment - $\pi_t : \Theta \to \Delta(M)$, which specifies a probability distribution over the common message space, $M$, for each state of the world. $M$ is assumed finite, but rich enough: $M > \Theta$.

The timing is as follows. First, each sender chooses simultaneously their signal, $\pi_s$, without knowing the state of the world. The receiver observes the profile of strategies, $\{\pi_s\}_{s \in S}$, and the realization of the signal of one of the senders which is chosen at random. Then the receiver updates her beliefs and takes an action. Finally, all players’ payoffs, $\{u_r, \{u_s\}_{s \in S}\}$, are realized.

Per usual, players have a common prior regarding the state of the world denoted $\mu_0 \in \Delta(\Theta)$. However, players also have a common prior belief about the likelihood that the receiver will observe the realization of their signal. Denote this prior by $\nu_0 \in \Delta(S)$ since the identity of the sender identifies each signal.

Note that persuasion is costless. Senders’ payoffs do not depend on the signal they choose. This allows me to focus on the role that the uncertainty about which signal realization will be observed by the receiver has on the incentives for senders. I also focus attention on pure strategy equilibria. This is without loss, since a mixed strategy is a distribution over Blackwell experiments, so it is itself a Blackwell experiment. Thus, any mixed strategy is equivalent to a pure strategy.

An important assumption is that all senders have access to the same messages, $M$. So in general, after observing a message, the receiver remains uncertain about which signal produced

$^3$Senders have no restriction on how correlated their messages can be with the state of the world; they can finely choose how much information to produce.

$^4$As established by lemmas, it is without loss to consider $M$ to be finite.

$^5$The mechanisms explored by Li and Norman (2018) regarding mixed strategies are mute here since the receiver observes the realization of only one signal.
the message. Since the correlation between a message and the state of the world can be chosen differently by different senders, the receiver has a double inference problem. From the message observed, the receiver must infer both about the state of the world and the quality of the message. This assumption that all senders have access to the same message space captures the fact that in digital media it is relatively easy for sources to remain anonymous or for sources to misrepresent their credibility. In section ??, I relax this assumption by allowing senders to have access to exclusive messages and discuss its implications.

In regards to the beliefs, I assume \( \mu_0 \) and \( \nu_0 \) to be independent. This choice is for exposition purposes. The main results do not change by allowing these probabilities to be correlated. Note that the probability measure \( \nu_0 \) can be interpreted as the prevalence of each sender in the platform. It measures how likely it is for each sender to have their content seen by the receiver. Next I introduce some notation to keep track of the evolution of the receiver’s beliefs.

Let \( p_0 = \mu_0 \cdot \nu_0 \) denote the joint distribution. After receiving a message, \( m \), the receiver learns about the state of the world and about the source of the message. Denote the joint posterior by \( p_1(\theta, s|m) \in \Delta(\Theta \times S) \) with marginal distributions for the state of the world denoted by \( \mu_1(m) = \sum_s p_1(\cdot|m) \in \Delta(\Theta) \), and for the source of the message by \( \nu_1(m) = \sum_\Theta p_1(\cdot|m) \in \Delta(S) \). We have that, given a profile of signals \( \{\pi_s\}_{s \in S} \), each message in its support induces the following posterior belief:

\[
p_1(\theta, s|m) = \frac{\pi_s(m|\theta)p_0(\theta, s)}{\sum_\Theta \sum_S \pi_{s'}(m|\theta')p_0(\theta', s')} \quad \forall \; \theta \in \Theta, \; s \in S, \; \text{given } m \in M;
\]

Since the observed message is stochastic, let the distribution over posteriors \( \tau \in \Delta(\Delta(\Theta \times S)) \) that is induced by \( \{\pi_s\}_{s \in S} \) be written as \( \tau(\{\pi_s\}) \). This distribution is characterized by two properties:

1. Every point in its support, \( p_1(\cdot|m) \), is defined by the previous equation; i.e. using Bayes rule.

2. The probability mass on any posterior \( p \) in the support of \( \tau \), is given by the total probability of receiving a message that leads to such a posterior:

\[
\tau(p) = \sum_{m:p_1(\cdot|m)=p} \sum_\Theta \sum_S \pi_s(m|\theta)\nu_0(s)\mu_0(\theta).
\]

The solution concept is Perfect Bayesian Equilibrium (henceforth simply equilibrium). It consists of a receiver’s (possibly mixed) strategy for any belief, the profile of sender strategies, and beliefs that are bayesian after any message in the support of \( \{\pi_s\}_{s \in S} \).

**Definition 1 (Equilibrium)** An equilibrium is a collection of signals, \( \{\pi_s^*\}_{s \in S} \), a response function \( a^* : \Delta(\Theta \times S) \rightarrow \Delta(A) \), and beliefs \( p_1 \in \Delta(\Theta \times T) \) such that:

\[\text{support}(\{\pi_s\}_{s \in S}) := \bigcup_{s \in S} \bigcup_{\theta \in \Theta} \text{support}(\pi_s(\theta))\]
1. \( a^*(p_1) \in \Delta(A) \) maximizes the receivers expected utility for each belief \( p_1 \in \Delta(\Theta \times S) \)

\[
\text{support } a^*(p_1) \subseteq \operatorname{argmax}_{a \in A} E_{p_1} [u(a, \theta)];
\]

2. The beliefs of the receiver are Bayesian after any message in the support of \( \{\pi_s\}_{s \in S} \). That is \( p_1(\theta, s|m) \) is given by equation (1) for all posterior beliefs with \( \tau(p_1) > 0 \);

3. \( \pi_s \) is a best response for sender \( s \) to the strategies of other senders \( \{\pi_s^*\}_{s' \neq s} \) and the receiver’s response \( a^*(p) \) to every induced belief:

\[
\pi_s^* \in \operatorname{argmax}_{\Pi} E_{\tau(\pi_s^*)} u_s(a^*(p_1(\cdot|m)), \theta) \quad \forall \ s \in S, \quad \text{given } \{\pi_{s'}^*\}_{s' \neq s}.
\]

Importantly, \( a^*(p) \) is a well defined function everywhere. At each belief where the receiver is indifferent, some mixed strategy is assumed, and this selection rule is commonly known to all senders. Common selections in the literature are sender-optimal strategies, and sender-pessimal strategies. In the case of a single sender, the game can be interpreted as one of an information designer. Therefore, these selections can be motivated in the spirit of partial implementation and robust implementation from the mechanism design literature. In this model, however, different senders may have different optimal and pessimal strategies for beliefs where the receiver is indifferent. In the spirit of partial implementation, I allow the selection to be part of the equilibrium concept.

Though equilibrium is not guaranteed to exists in general, it does exist for a large class of games, as illustrated in section 5. In the appendix I discuss the issue of existence. It stems from the fact that, as in most sender-receiver games, the correlation of messages with each state of the world is endogenous. Therefore, whenever there are at least two senders that can unilaterally change the sign of the average correlation between messages and the state of the world, equilibrium may fail to exist.

In the next section I develop a geometric interpretation that is useful in characterizing equilibria for specific applications. In general, the problem does not allow for traditional concavification techniques, because the posteriors that a sender can induce over the state of the world, \( \mu_1 \), not only have to be a mean preserving spread of the prior (also known as Bayesian-plausible beliefs), but also must be “feasible” given the strategies of other senders. In the next section I formalize the feasibility constraints for each sender.

3 Model Simplification

In this section I simplify the analysis of the model in three steps. First I find a unique signal, \( \bar{\pi} \), that summarizes the actions of the senders and is a sufficient statistic for the receiver’s beliefs, thus her optimal actions. Second, I characterize how the actions of the all other senders constraints the influence of a sender over \( \bar{\pi} \). This constraints depend on how all of the other senders’ correlate
their messages with the states of the world, and on the messages in the support of their signals. Finally, I justify restricting the cardinality of the message space to $|M| = |A|$. As in previous literature, messages can be restricted to be recommendations, but in contrast with previous literature, recommendations need not always be followed in equilibrium.

We can start by noting that the receiver’s problem is fairly simple. Given the profile of equilibrium signals and the message she observes, $\{\pi_s^*\}_{s \in S}$ and $m^*$ respectively, the receiver forms posterior beliefs $p_1(\theta, s|m^*)$ for every pair $(\theta, s) \in \Theta \times S$ and chooses an action that maximizes her expected utility $E_{p_1}[u(a, \theta)]$. Since neither the receiver’s payoff, nor any of the senders’ payoffs depend on $s$, the posterior beliefs about the identity of the sender, $\nu_1$ should be irrelevant in terms of characterizing the outcomes of the game. This is indeed the case. The following lemma establishes that without loss of generality we can restrict attention to functions $a^*(p(\theta, s|m))$ that are constant over $s$.

**Lemma 1 (The marginal posterior, $\mu_1$, as a sufficient statistic for $a^*$)** If $((\pi_s^*)_{s \in S}; a^*(p); p)$ is a perfect Bayesian equilibrium, there exists an outcome equivalent equilibrium $((\pi_t^*)_{t \in T}; a'; p)$ where $a': \Delta(\Theta \times S) \rightarrow \Delta(A)$, such that $a'(p(\theta, s)) = a'(p(\theta, s'))$ for all $s, s' \in S$.

**Corollary 1** It is without loss of generality to restrict the definition of equilibrium to receiver’s response functions that depend only on the belief about the state of the world. That is to functions $a^*: \Delta(\Theta) \rightarrow \Delta(A)$.

After restricting attention to equilibria where $a^*$ only depends on the marginal belief over the state of the world, we can simplify the set of equilibrium signals, $\{\pi_s\}$. If two sets of equilibrium signals induce the same marginal distribution, $\mu_1$, they will also induce the same distribution over actions by the receiver and so are outcome equivalent. This implies that, instead of keeping track of the whole profile of signals, we can focus on the posterior marginal distribution they induce. Furthermore, we can define a unique signal that is a sufficient statistic for $\mu_1$.

**Lemma 2 (The average signal, $\bar{\pi}$, as a sufficient statistic for $\mu_1(m)$)** For any profile of signals, $\{\pi_s\}_{s \in S}$, the marginal posterior $\mu_1$, is a function of the marginal prior, $\mu_0$, and the average signal, $\bar{\pi}$, alone; where

$$\bar{\pi}(m|\theta) = \sum_{s \in S} \pi_s(m|\theta) \nu_0(s) \quad \forall \ m \in \text{support}(\{\pi_s\}_{s \in S}), \quad \forall \ \theta \in \Theta.$$  

Denote the distribution over marginal posteriors induced by an expected signal by $\tau^\theta(\bar{\pi}) \in \Delta(\Delta(\Theta))$.

The previous lemma establishes that the signal chosen by each sender, affects the posterior beliefs of the receiver only through the average signal, $\bar{\pi}$. Whenever a message is used by multiple senders, each of them has only partial control over the posterior belief of the receiver. In fact, if the way the receiver responds to a message only depends on $\bar{\pi}$, the average correlation between messages and the state of the world. Therefore, senders choose their communication strategies in order to influence this average correlation.
Remark 1 We can relax the assumption that the receiver knows the profile of signals chosen by the senders, \( \{\pi_s\}_{s \in S} \), and assume instead that the receiver only knows the average communication strategy, \( \bar{\pi} \).

To study the best response for each sender, we can characterize the posteriors that a sender can induce given the strategies of other senders. I call these, the feasible posteriors for sender \( s \) given the strategies of other senders. Notice that for every sender, we can think of each pure strategy signal, \( \pi_s \), as a rectangular (row) stochastic matrix of dimension \( \Theta \times M \). That is, a matrix with non-negative elements and rows that sum up to 1. Each signal, \( \pi_s \), is an element of the convex set \( \Pi = \prod_{\theta \in \Theta} \Delta(M) \). Since the expected signal, \( \bar{\pi} \), is the weighted average of elements in \( \Pi \), it is itself an element in \( \Pi \). That is:

\[
\bar{\pi} = \sum_{s \in S} \nu_0(s)\pi_s \in \Pi. \tag{2}
\]

Given this structure, we can summarize the strategies of every other sender by a single element in \( \Pi \), the weighted average of the other senders’ signals. Let such signal be called the \( -s \) average signal, and be denoted by \( \bar{\pi}_{-s} \). It is computed as follows:

\[
\bar{\pi}_{-s} := \sum_{s' \neq s} \frac{\nu_0(s')}{1 - \nu_0(s)} \pi_{s'} \in \Pi \tag{3}
\]

It is clear that any two profiles of signals chosen by the other senders that define the same \( -s \) average signal are equivalent, for strategic purposes, to sender \( s \). This is because each expected signal, \( \bar{\pi} \), can always be expressed as the weighted average of \( \pi_s \) and \( \bar{\pi}_{-s} \) as follows:

\[
\bar{\pi} = \nu_0(s)\pi_s + (1 - \nu_0(s))\bar{\pi}_{-s} \tag{4}
\]

Equation (4) defines a one-to-one relationship between any strategy for sender \( s \), and the expected signal, \( \bar{\pi} \), given the \( -s \) average signal and the prior prevalence of sender \( s \), \( \nu_0(s) \in (0,1) \). Let’s denote such function by \( \bar{\pi}(\pi_s|\bar{\pi}_{-s},\nu_0(s)) : \Pi \to \Pi \), and its image by \( \Pi_\theta(\bar{\pi}_{-s},\nu_0(s)) \). We can then write a sender’s expected payoff as a function of his own signal and the signals of other players, as follows:

\[
v_s(\pi_s, \bar{\pi}_{-s}) := E_{\tau(\bar{\pi}(\pi_s|\cdot))} E_{\mu} u_s(a^*(\mu), \theta) \tag{5}
\]

\[= E_{\tau(\bar{\pi}(\pi_s|\cdot))} \hat{v}_s(\mu). \tag{6}
\]

Each sender’s preferences now depend directly on their signal, \( \pi_s \), and the strategies of other senders, \( \bar{\pi}_{-s} \). To characterize their best response, it suffices to look at the set of average signals that a sender can induce; i.e. at the image of function \( \bar{\pi}(\pi_s|\bar{\pi}_{-s},\nu_0(s)) \). To derive some intuition about the shape of this set of feasible average signals, figure illustrates its geometric properties for a binary \( \Theta \), binary \( A \) game with \( M = A \). Lemma formalizes these properties for the general case.
Lemma 3 (Properties of the set of feasible $\bar{\pi}$’s for sender $s$) Let $\bar{\pi}(\Pi|\pi_{-s},\nu_0(s))$ be the set of feasible average signals to sender $s$ given $\pi_{-s}$ and $\nu_0(s)$, and let $\succsim$ denote Blackwell’s (1953) partial order over $\Pi$:

1. $\bar{\pi}(\Pi|\pi_{-s},\nu_0(s))$ is a contraction of $\Pi$ with all distances shrunk by $\nu_0(s)$ and $\pi_{-s} \in \bar{\pi}(\Pi|\pi_{-s},\nu_0(s))$.

2. The correspondence $\bar{\pi}(\Pi|\pi_{-s},\nu_0(s))$ is continuous with respect to $\nu_0(s)$ and $\pi_{-s}$, and is strictly increasing in $\nu_0(s)$. That is, $\nu_0(s) < \nu_0(s)'$ implies $\bar{\pi}(\Pi|\pi_{-s},\nu_0(s)) \subset \bar{\pi}(\Pi|\pi_{-s},\nu_0(s)')$.

3. For any $\pi_{-s} \in \Pi$, there exists $\bar{\pi}_h, \bar{\pi}_l \in \bar{\pi}(\Pi|\pi_{-s},\nu_0(s))$ such that $\bar{\pi}_h \succsim \pi_{-s} \succsim \bar{\pi}_l$ with at least one strict inequality.

Property 3 in lemma 3 establishes that senders can provide more or less information relative to $\pi_{-s}$, but not arbitrarily so. Property 1 and $\nu_0(s) < 1$ imply that there are some signals that are not feasible to $s$; in particular signals that are far from $\pi_{-s}$. Lastly, property 2 establishes that the choice set depends smoothly on its parameters and that increasing the prevalence of a sender increases the set of feasible average signals available to this sender.

Notice that senders are not restricted to use signals with full support. Indeed, whenever a sender, $s$, is the only one using some message, say message $m^*$, not only will the receiver identify the sender after receiving message $m^*$, i.e. $\nu_1(s|m^*) = 1$, but the sender has full control over the correlation between this message and each state of the world:

$$\bar{\pi}(m^*|\theta) \propto \pi_s(m|\theta) \quad \forall \theta \in \Theta.$$  

We could suspect that the number of messages available is affects the set of equilibrium outcomes. I show, next how this is not the case. In actuality, if the message space is as large as the action
space, increasing the cardinality of the message space does not affect the distribution of outcomes that can occur in equilibrium.

**Lemma 4 (Message space simplification)** If $|A| \leq |M|$, there exists an equilibrium of the game where the message space is $M$ if and only if there is an outcome equivalent equilibrium in the game where the message space is $A$.

This lemma establishes that it is without loss to restrict the cardinality of the message spaces, which greatly simplifies the search for equilibria. Though similar in spirit to proposition 1 in KG11, there is a key distinction between their model and this one. Messages can be assumed straightforward, i.e. each message recommends an action. However, the recommendations are not always followed in equilibrium. Theorem 2 provides a sufficient condition for the receiver to disregard any recommendation that does not agree with her default action, $a^*(\mu_0)$.

Even though we can assume there is exactly one message available to recommend each action, it is with loss to assume that $M = A$ and that recommendations are followed, i.e. $a^*(\mu_1(a)) = a$. Different senders may choose different messages to induce the same action, or senders might differ on which action should a particular message induce. Clearly, given an equilibrium, we can label the messages according to the action they induce, but there is no guarantee that each message will induce a different action.

**Corollary 2 (Recommendations may not be followed)** It is without loss of generality to assume that signals are recommendations, $M = A$; so $\Pi = \prod_{\theta \in \Theta} \Delta(A)$. However, recommendations need not be followed in equilibrium, i.e. $a^*(\mu_1(a)) = a$ does no necessarily hold.

## 4 Equilibrium Characterization

From the last section we have that each sender’s objective is

$$
\max_{\pi_s} v_s(\pi_s, \pi_{-s}) = E_{\tau^{\theta}(\pi_s)} [\hat{v}_s(\mu)] \\
\text{s.t. } \pi_s \in \Pi_s(\pi_{-s}, \nu_0(s))
$$

(7)  
(8)

It is clear that the sender cannot induce some Bayes-plausible distributions of posteriors since $\Pi_s(\pi_{-s}, \nu_0(s))$ is a strict subset of $\Pi$. Moreover, the value of proving information not only depends on the concave closure of $\hat{v}_s(\mu)$.

**Definition 2** Say that a sender is not providing information if $\pi_s(a|\theta) = \pi_s(a|\theta')$ for all $a \in A$ and for all $\theta \in \Theta$. Otherwise we say that the sender is providing information.

It is clear that if a sender is not providing information, $\pi_s \preceq \pi$ for all $\pi \in \Pi$. If neither sender provides information, the receiver clearly cannot learn anything. However, the converse is not true. The average signal can be uninformative even if all senders are providing information. In sharp contrast, the receiver fully learns the state of the world if and only if $\pi_s = \pi_s'$ for all $s, s' \in S$ and
π_s is fully informative: π_s ⪰ π for all π ∈ Π. The lack of transparency in the model breaks the perfect correlation between the information a sender provides and how much the receiver learns. As a consequence we have the following result.

**Proposition 1 (The value of providing information)**

- A sender may choose to provide information even if ˆv_s(µ) is concave at µ_0
- A sender may not benefit from providing information even if the concave closure of ˆv_s(µ) at µ_0 is strictly larger than ˆv_s(µ_0).

**Example 1 (Informative {π_s}_s∈S, but no valuable learning)** Consider the following game. Θ = A = M = {0, 1} are binary sets. Suppose the receiver wants to match the state of the world with her action: u_r = -1{a ≠ θ} and has a prior µ_0(θ = 1) = 1/4. Consider three senders, S = {s_1, s_2, s_3} with equal prevalence ν_0(s) = 1/3 for all s ∈ S and preferences as follows u_{s_1} = u_{s_2} = u_r and u_{s_3} = -u_r. That is ˆv_{s_1}(µ) and ˆv_{s_2}(µ) are convex at every µ, while ˆv_{s_3}(µ) is concave everywhere. Then, the signals π_{s_1} = π_{s_2} such that π_{s_1}(θ|θ) = 1 for all θ ∈ Θ and π_{s_3}(θ|θ) = 0 for all θ ∈ Θ are part of an equilibrium of this game.(see the appendix).

Notice that all senders are providing information, including sender s_3 whose expected payoff is concave everywhere. Indeed π_s ⪰ π for all π ∈ Π and for all s ∈ S. However, the correlation between messages and states is opposite between the first two senders and the third one. Therefore, neither sender s_1 nor s_2 benefit from providing information. The induced posteriors are 1/7 and 2/5 both below 0.5, so the receiver learns nothing valuable and these senders have perfectly aligned preferences with the receiver. ♦

Studying the concave closure of each sender’s payoff is not particularly tractable, to identify equilibrium signals. We can instead translate problem (7) into a problem of choosing a decision rule (or distribution over actions) subject to constrains. Clearly, the rationality of the receiver requires the decision rules to be obedient (as shown by Bergemann and Morris, 2019), but also feasible given the actions of all the other senders and the sender’s own prevalence.

**Definition 3 (Obedience (Bergemann and Morris, 2019))** A decision rule, σ : Θ → Δ(A) is obedient for the receiver given the prior µ_0 ∈ Δ(Θ) if for each a ∈ A we have

\[
\sum_{θ∈Θ} u_r(a, θ)σ(a|θ)µ_0(θ) \geq \sum_{θ∈Θ} u_r(a', θ)σ(a|θ)µ_0(θ)
\]

for all a' ∈ A.

It is clear that each signal in Π induces a distribution over actions that is equivalent to an obedient decision rule, however, not all signals in Π are feasible to sender s, so a sender cannot induce all of the obedient decision rules.

---

5Along with the receiver’s response function, a^*(µ) = 1{µ ≥ 0.5}, and appropriate Bayesian beliefs
Definition 4 (Feasible for sender $s$) A decision rule $\sigma : \Theta \to \Delta(A)$ is feasible to sender $s$ given $\bar{\pi}_{-s}$ and $\nu_0(s)$, if there exists a function $\rho : M \to A$ and an average signal $\bar{\pi} \in \Pi_s(\pi_{-s}, \nu_0(s))$ such that

$$\sigma(a|\theta) = \sum_{m: \rho(m) = a} \bar{\pi}(m|\theta)$$

for all $\theta \in \Theta$ and all $a \in A$.

Note that the function $\rho$ need not be bijective and that not all decision rules are feasible. Nonetheless, the intersection of obedience and feasible decision rules is not empty. To see this note that the decision rule that assigns probability one to the default action, i.e. to $a^*(\mu_0)$, in all states of the world is always obedient. Further, by choosing the function $\rho(m) = a^*(\mu_0)$ for all $m \in M$ and any signal in $\Pi_s(\pi_{-s}, \nu_0(s))$, the decision rule is also feasible for sender $s$.

Theorem 1 (Constrained information Design) The following are equivalent

- There exists a $\pi_s \in \Pi$ with value $v^*_s$ to sender $s$ given $\pi_{-s}$ and $\nu_0(s)$.
- There exists a decision rule $\sigma : \Theta \to \Delta(A)$ that is obedient and feasible to $s$ given $\pi_{-s}$ and $\nu_0(s)$ such that $v^* = E_{\sigma}u_s(a, \theta)$.

Each sender, thus, chooses their favorite $\sigma$ among the obedient and feasible distributions given the actions of all other senders and their prevalence. Since the equilibrium signals must be a fixed point, we get the following corollary.

Corollary 3 In equilibrium all senders must choose the same decision rule.

Hence, when the interests of two or more senders with non-negligible prevalence are different, equilibrium may fail to exists. In the limit, when $|S| \to \infty$ and $\nu_0(s) \to 0$ for all $s \in S$, the existence of equilibrium can be guaranteed provided $\hat{v}_s$ is upper semi-continuous for all $s \in S$. In the next section I study important cases of games where equilibrium exists even away from this limit.

ADD SUFFICIENCY FOR EXISTENCE

5 Applications

(Streamline examples and simplify notation. See conference slides)

In this section I study the set of equilibria in three important classes of games to show some general implications of having uncertainty about the sender’s motives. First, the equilibrium signals chosen by senders may increase with the frequency with which the receiver observes a message from a uninformative signal, but beyond a threshold, no relevant information can be transmitted in equilibrium. This threshold depends on the prior belief and preferences of the receiver and can be arbitrarily small. Second, if there are multiple senders with diverse goals, every sender may choose

\footnote{The empty sum is assumed to be equal to zero.}
in equilibrium a signal that is less informative than the signal the signal they would choose if they
Lastly, I show that if each sender could choose to disclose their type, private information unravels
in equilibrium. I illustrate each of these results with examples in a binary state setting. Next, I
define two orders on the sets of equilibrium signals with respect to how informative they are for
the senders and the receiver, respectively. \( \Theta \)

**Definition 5 (Sender and Receiver orders on equilibrium signals)** Let \( E(\mathcal{G}) \) define the set
of equilibrium signals for game \( \mathcal{G} \) and the expected signal they induce. For any two games with the
same underlying state space, \( \mathcal{G} = \{ \Theta, \cdot \} \) and \( \mathcal{G}' = \{ \Theta, \cdot \} \), and non-empty equilibria we say that:

- More information is produced in \( \mathcal{G} \) than in \( \mathcal{G}' \) if for every \( (\{\pi_t\}_{t \in T}, \bar{\pi}) \in E(\mathcal{G}) \) there exists
  \( (\{\pi'_t\}_{t \in T}, \bar{\pi}') \in E(\mathcal{G}') \) such that \( \pi_t \lesssim \pi'_t \) for every \( t \in T' \). We write \( E(\mathcal{G}) \sqsubseteq_s E(\mathcal{G}') \).

- The receiver learns more in \( \mathcal{G}' \) than in \( \mathcal{G} \) if for every \( (\{\pi_t\}_{t \in T}, \bar{\pi}) \in E(\mathcal{G}) \) there exists
  \( (\{\pi'_t\}_{t \in T}, \bar{\pi}') \in E(\mathcal{G}') \) such that \( \bar{\pi} \lesssim \bar{\pi}' \). We write \( E(\mathcal{G}) \sqsubseteq_r E(\mathcal{G}') \).

- If \( E(\mathcal{G}) \sqsubseteq_s E(\mathcal{G}') \) and \( E(\mathcal{G}) \sqsubseteq_r E(\mathcal{G}') \) we write \( E(\mathcal{G}) \sqsubseteq E(\mathcal{G}') \).

The symbol \( \lesssim \) denotes the Blackwell order over the space of signals, \( \Pi \).

The orders \( \sqsubseteq_s \) and \( \sqsubseteq_r \) capture separately the effect of uncertainty on the strategic decisions of
senders, and on how much the receiver can learn in equilibrium. Surprisingly, these two orders not
necessarily agree. In the next section, for example, I show how if the prevalence of a noisy sender
increases, every other sender chooses more informative signals, but the average signal is still more
noisy.

### 5.1 Games with a noisy sender

In many situations, a sender has no incentives to provide information to the receiver. KG11 showed
that whenever a sender’s payoff, \( v_t(a^*(\mu^\Theta), \theta) \), as a function of a receiver’s beliefs is concave at the
prior, there is no value to providing information. This result relies on the fact that any distribution
over posterior beliefs induced on a rational receiver must be Bayesian-Plausible, and that any Bayes
plausible distribution over posterior beliefs can be implemented by some signal. Therefore, when \( \hat{c}(\cdot) \) is not concave, there exists some Bayes-plausible distribution over beliefs that provide the
sender with a larger expected payoff. However, in this model where the receiver faces uncertainty
about the sender, the set of Bayes-plausible distributions over beliefs that would benefit a sender
might only be induced by expected signals that are not feasible for such a sender. We see that the
non-concavity of \( v_t(a^*(\mu^\Theta), \theta) \) at the prior is necessary, but no longer a sufficient condition for a
sender to benefit from persuasion, given the strategies of other senders. In this section I illustrate
a case where the uncertainty about the sender can prevent any relevant information from being
communicated in equilibrium.

---

9I adopt the equivalence relationship that two equilibria are the same if they are outcome equivalent.
10Since Blackwell’s order does not define a lattice for general state spaces, the strong set order cannot be used for
the sets of equilibria.
Let’s consider a game with symmetric message spaces, $M_t = M$ for all $t \in \mathcal{T}$, where there is a noisy sender, say $\hat{t} \in \mathcal{T}$, for which $v_i(a^*(\mu^\Theta), \theta)$ is concave at $\mu^\Theta_0$. Since this sender has no incentives to provide information to the receiver, he prefers expected signals, $\pi$ that are less informative. Its choice of $\pi_i$, thus minimizes the informativeness of every recommendation. Clearly, in light of equation\[23]\ the larger the matrix $A_{\hat{t}}$ is, the closer that $\pi$ will be to $\pi_i$. Since $A_{\hat{t}}$ is a primitive given by the conditional probabilities $\{\mu_0(\hat{t}|\theta)\}_{\theta \in \Theta}$ it is no surprise that sender $\hat{t}$ has a larger control over the beliefs of the receiver, whenever it is more likely that the receiver will see the outcome of $\pi_i$. As stated in the next theorem, an increase in the probabilities $A_{\hat{t}}$ gives incentives to all other senders to choose more informative signals, even if the noisy sender chooses the same strategy in equilibrium. However, the expected signal weakly decreases in informativeness. Furthermore, after a threshold on the elements of $A_{\hat{t}}$, all other senders find unprofitable to provide information, even if they have incentives to do so in the absence of uncertainty (KGI). Their feasible $\pi$ provide the same expected payoff as the prior. Surprisingly this threshold can be relatively low.

**Theorem 2 (Equilibrium signals with a noisy sender)** Suppose there is a sender $\hat{t} \in \mathcal{T}$, that prefers the receiver to be un-informed\[24]\ Consider two priors $\mu_0$ and $\mu'_0$ such that $\mu_i^\Theta = \mu_0^\Theta$ and $\mu_0(t|\theta, t \neq \hat{t}) = \mu'_0(t|\theta, t \neq \hat{t})$ for all $\theta \in \Theta$, but $\mu_0(\hat{t}|\theta) \leq \mu'_0(\hat{t}|\theta)$ for all $\theta$ with strict inequality $\exists \theta$. Then, there exists thresholds $\gamma(\mu_i^\Theta, u_r(\cdot), \theta)$ for each state such that:

- $E(\mathcal{G}) \subseteq E(\mathcal{G}')$ but $E(\mathcal{G}) \supsetneq E(\mathcal{G}')$ if $\mu'_0(\hat{t}|\theta) \leq \gamma(\mu_0^\Theta, u_r(\cdot), \theta)$ for some $\theta \in \Theta$,

- $a^*(\pi^*_n) = a^*(\mu_i^\Theta)$ for every equilibrium $\pi^*$ if $\mu'_0(\hat{t}|\theta) > \gamma(\mu_i^\Theta, u_r(\cdot), \theta)$ for all $\theta \in \Theta$ or $q(\theta_i) = \mu_0(\theta_i|\hat{t})$;

where the threshold $\gamma(\cdot) = \left[\frac{1}{\Theta} \left(\frac{q(\theta) - \mu_0(\theta|\hat{t})}{1-q(\theta)}\right) + 1\right]^{-1}$ and $q(\theta_i)$ is the belief closest to the prior that persuades the receiver to choose an action different than the default:

$$q(\theta) = \arg\inf_{\mu^\Theta(\theta) \in (\mu_0^\Theta(\theta), 1]} \{|\mu^\Theta(\theta) - \mu_0^\Theta(\theta)| ; a^*(\mu^\Theta) \neq a^*(\mu_0^\Theta)\}.$$  

In the last result I am allowing for the prevalence of the noisy type to vary across states along the sequence. the threshold not only depends on how prevalent is the noise type in each state, but also how easy it is to convince the receiver of choosing something different than the default action in each state, measured by the difference between $q(\theta)$ and $\mu_0(\theta)$, which is state-specific. If the prevalence of the sender is independent of the state, we can define a uniform threshold as in the next corollary.

**Corollary 4 (Equilibrium signals with a noisy sender and $\mu_0$ a product distribution)** Consider a sequence of games as in theorem\[23]\ with the extra assumption that $\mu_0 \in \Delta(\mathcal{T}) \times \Delta(\Theta)$. That is $\mu_0(\hat{t}|\theta)_n = \mu_0^\Theta(\hat{t})_n$ for all $\theta \in \Theta$ and for all $n \in \mathbb{N}$, and $\mu_0^\Theta_n$ is constant along the sequence. Then:

- $E(\mathcal{G}_n) \subseteq E(\mathcal{G}_{n+1})$ but $E(\mathcal{G}_n) \supsetneq E(\mathcal{G}_{n+1})$ if $\mu_0^\Theta(\hat{t})_{n+1} \leq \max_{\theta \in \Theta} \gamma(\mu_0^\Theta, u_r(\cdot), \theta)$,

\[11\] i.e. $v_i(a^*(\mu^\Theta), \theta)$ is concave at the prior belief
\[ a^\ast(\bar{\pi}^\ast_n) = a^\ast(\mu_0^\Theta) \] for every equilibrium \( \bar{\pi}^\ast \) if \( \mu_0^T(t)_{n+1} > \max_{\theta \in \Theta} \gamma(\mu_0^\Theta, u(\cdot), \theta); \)

Note that if we consider priors \( \mu_0 \in \Delta(\Theta) \times \Delta(T) \), i.e. where nature independently chooses a sender and the state of the world, the right hand side of the inequalities that define the threshold are constants along the sequence, and the left hand side does not depend on the state, so they can be reduced to a single inequality. Furthermore, note that the threshold does not depend on the number of senders. Rather, it has a negative relationship with the distance between the prior and the closest belief that induces the receiver to choose a different action. This means that if the receiver needs a lot of information to choose an action different than the default, even moderate probabilities of receiving noisy signals might prevent any relevant information to be communicated in equilibrium. We see that whenever the prevalence of the noisy sender is larger than the threshold for all states, the receiver’s behavior is observationally equivalent to one where she disregards any information and chooses the default action regardless. The following example builds on the example presented in the introduction to illustrate this result in a binary state space with two actions.

**Example 2 An election with a noisy sender (revisited)**

Consider a slightly modified version of the example in the introduction. The state space is \( \Theta = \{A, B\} \), the receiver wants to match the state of the world with his action, \( A = \{a, b\} \). The commonly known prior is \( \mu_0^\Theta = 0.8 \) the probability that the best candidate is \( A \). Before the election the receiver gets a recommendation, \( M_t = M = \{a, b\} \), from one out of two possible senders \( T = \{p, b\} \). There is a partisan journalist \( (p) \), and a bad journalist \( (b) \). The partisan journalist wants candidate \( B \) to be elected regardless of the state of the world, while the bad journalist wants the voter to elect the wrong candidate. The probabilities of receiving recommendations from the bad journalist is \( \mu_0^T = \alpha_b \), independent of \( \mu_0^\Theta \).

Then, the threshold \( \gamma(0.8, u(\cdot), B) = 0.4 \); i.e if \( \alpha_b > 0.4 \), then every equilibria is uninformative.\(^{12}\) Thus the sender with incentives to provide information must be significantly more prevalent than the noisy sender in order to be profitable to provide information. Further, if \( \alpha_b \leq 0.4 \), there is an equilibrium\(^{13}\) were the bad journalist always recommends either candidate \( B \) with equal probability, \( (\pi^*_b(a|A), \pi^*_b(a|B)) = (1/2, 1/2) \), and the optimal signal for the partisan journalist is \( (\pi^*_p(a|A), \pi^*_p(a|B)) = \left(\frac{3(2-\alpha_b)}{8(1-\alpha_b)}, 0\right) \), whose informativeness clearly increases in \( \alpha_b \). Finally the expected signal equals \( (\bar{\pi}(a|A), \bar{\pi}(a|B)) = \left(\frac{6+\alpha_b}{8}, \frac{\alpha_b}{2}\right) \), whose informativeness decreases in \( \alpha_b \). Figure 2 depicts an equilibrium for \( \alpha_b = 0.2 \) and \( \alpha_b = 0.4 \). The regions labeled “Obedience” and “Obedience*” correspond to the signals where the receiver chooses different actions after each of the two possible messages. In the first one, they obey the recommendation, while in the second one, they choose “a” whenever they receive the message “b”, and vice-versa, a form of obedience\(^{15}\). Outside

\(^{12}\)The fact that the sender is partisan instead of un-biased is irrelevant, it only matters for the comparative statics when equilibria is informative, since an unbiased sender always chooses a perfectly informative signal, so \( E(G_n) \) remains equally informative with respect to \( \bar{\pi} \).

\(^{13}\)Equilibrium is not unique, however, in every informative equilibrium the posterior \( \mu_0^\Theta = 0.5 \) is induced. In contrast with \( KG \) some equilibria do not fully disclose the true state when \( \Theta = A \).

\(^{14}\)Note that the perfectly informative signal is \( (\pi(b|A), \pi(b|B)) = (1, 0) \)

\(^{15}\)This two areas are equivalent in the sense that for every signal in one of them there exists an outcome equivalent
the obedience regions, the receiver always chooses A. Therefore, the biased journalist wants to induce an expected signal that is obeyed and maximizes the probability that the action B is chosen, as shown in the graphs. In contrast, the bad journalist is indifferent among all the expected signals that are not in the interior of the obedience sets, and they achieve his first best. The bad journalist has no incentives to deviate.

Theorem 2 and the previous example show that if the frequency with which a decision maker receives deceptive information is larger than a threshold that depends only on the receiver preferences and her prior belief, it becomes impossible to persuade her, and no relevant information is transmitted in equilibrium, even if all other sources design experiments that fully reveal the state of the world. Importantly, this threshold can be relatively low. In the example a 40% probability of receiving messages from the noisy sender was the threshold.

5.2 Senders with state-independent preferences

A common benchmark in the literature is to assume that sender’s preferences do not depend on the true state of the world. For example if they are sellers, they might want the buyer to buy their good regardless of its true quality. Politicians typically want the electorate to vote for their candidate regardless, etc. However, the literature has shown mixed results on the effect of uncertainty on the quality of information received. On the one hand, when communication occurs via “cheap talk”, Morgan and Stocken (2003) and Li and Madarasz (2008) have shown that uncertainty about the bias can improve communication of some states of the world. The uncertainty relaxes the incentive compatibility constraints and senders find it profitable to credibly disclose more information.

References

16 In the example this can be seen in the figure when $\alpha_b = 0.4$ where the partisan journalist is already fully revealing all states of the world.
for intermediate states. That is, for states where the receiver believes that the message may come from either sender, so she responds as if coming from the average sender (similar to this model). If biases have the opposite sign, for example, the expected bias can be relatively small, thus allowing for more communication. In contrast, when the state space is multidimensional Chakraborty and Hardaugh (2007) argue that communication is robust to uncertainty over expert’s preferences, only when the uncertainty is relatively small. In their construction of equilibria, the sender gain credibility by using comparative statements among which they are indifferent. If receivers are uncertain about the sender’s preferences, informative recommendations must keep all types of senders indifferent. Similarly Wolinsky (2003) shows that when communication occurs through verifiable messages, the uncertainty about the senders preferences leads to worse information than without uncertainty.

In this section, I show that when senders have state-independent preferences, the uncertainty about a sender’s preferences unambiguously decreases the incentives to produce information. Any gain in information quality must come from alleviating the incentive compatibility constraints to disclose the information produced. To simplify the exposition I assume that every sender has one preferred action, and is indifferent among all other actions.

**Theorem 3 (Single-Issue senders with opposed biases)** Let $\mathcal{T}$ be a set of senders with preferences such that for each $t \in \mathcal{T}$ there exists a unique action, $a_t \in A$, such that $v_t(a_t, \theta) = c > 0$ for all $\theta \in \Theta$ and $v_t(a', \theta) = 0$ for all $a' \neq a_t$. Assume $a^*(\mu_0^\Theta) \neq a_t$ for all $t \in \mathcal{T}$ (all senders have incentives to provide information), and for every $t \in \mathcal{T}$ $\exists \theta_t \in \Theta$ such that $a_t$ is the best response of the receiver when the state is $\theta_t$, and $t \neq s$ implies $\theta_t \neq \theta_s$ (their biases are opposed). Then:

- If $|\mathcal{T}| \leq |\mathcal{T}'|$, and $\mu_0^{\mathcal{T}'}(t|t \in \mathcal{T} \cap \mathcal{T}') = \mu_0^\Theta(t)$, then $E(G\{\cdot, \mathcal{T}\}) \subseteq E(G\{\cdot, \mathcal{T}'\})$. I.e. increasing the number of senders worsens the information in equilibrium.

- $E(G\{\cdot, \mathcal{T}\{M_t\}\}) \subseteq E(G\{\cdot, \mathcal{T}, \{M_t^s\}\})$ whenever $M_s = M$ for all $t \in \mathcal{T}$, and $M_t^s \cap M_t^s = \emptyset$ for all $t \neq s \in \mathcal{T}$. I.e. the information in equilibrium is worse when the receiver faces uncertainty about the sender’s motives.

The intuition behind this result lies in the fact that if there was no uncertainty about the sender’s motives, each sender’s optimal signal, say $\pi^*_t$, is one that induces a posterior that partially obfuscates information. In particular, the $\pi^*_t$ induces a highly informative posteriors and a posterior that barely convinces the receiver to choose their preferred action. The receiver knows that when she receives a recommendation to not choose the sender’s preferred action, it is highly informative of the state of the world. When there is uncertainty about the sender’s motives, any recommendation can be highly informative if coming from a sender that is not biased towards that action, or less informative if coming from a sender with a bias towards that action. Each sender can recommend their preferred action more frequently leveraging the credibility of such recommendation when made by other senders. In equilibrium the expected signal is also less informative.

---

$^{17}$Lipnowski and Ravid (2018) have shown how the incentives to disclose information increase when the sender has strict preferences over multiple actions, the main mechanism established in the following theorem will still hold for more flexible preferences, see the notes in the proof of the theorem.
Example 3 (An election with 2 biased journalists) Consider a the same election setting discussed in the introduction and the previous example. The state space of which is the better candidate is $\Theta = \{A, B\}$, the receiver wants to vote for the best candidate, but in contrast with the previous example, can now also choose not to vote; i.e. the action space is $A = \{a, b, \emptyset\}$. Assume that the receiver votes only if she is at least $90\%$ sure that a candidate is the best, otherwise, she abstains from voting, $a^* = \emptyset$. The commonly known prior is $\mu^0_\Theta(A) = 0.8$, the belief that A is the better candidate. Before the election, the receiver gets a recommendation from one of two possible senders, $\mathcal{T} = \{\hat{a}, \hat{b}\}$. These are partisan journalists that want the receiver to vote for their favorite candidate regardless of the state of the world; i.e. $v_\hat{a}(a|\theta) = v_\hat{b}(b|\theta) = 1$ and $v_\hat{a}(b|\theta) = v_\hat{b}(a|\theta) = 0$. Let $\mu^T = \alpha$ denote the probability of receiving a recommendation from sender $\hat{a}$, independent of $\mu^0_\Theta$.

Consider the optimal signal that each sender will choose if their message spaces were disjoint, say $M_\hat{a} = \{a, b\}$ and $M_\hat{b} = \{a', b'\}$. From the discussion in section ?? we know that the strategy of one sender has no bearing on the strategy of the other sender. Upon receiving a message, the receiver learns what sender it came from, and thus the precision of the signal that generated the message. Each sender’s problem reduces to maximizing their expected payoff subject to inducing posteriors that are Bayesian plausible, as in KG11. Therefore the optimal signals are

\[
(\pi^*_\hat{a}(a|A), \pi^*_\hat{a}(a|B)) = \left( 1, \frac{4}{9} \right), \quad (\pi^*_\hat{b}(a'|A), \pi^*_\hat{b}(a'|B)) = \left( \frac{35}{36}, 0 \right),
\]

and the expected signal induces for each state a distribution over the four possible messages as follows:

\[
\bar{\pi}^* = \alpha \pi^*_\hat{a} + (1 - \alpha) \pi^*_\hat{b} : \Theta \to \Delta(\{a, b, a', b'\}).
\]

In contrast, if their message spaces are the same, $M_\hat{a} = M_\hat{b} = \{a, b\}$, the receiver faces uncertainty about the sender’s motives, in general, after any message received. and the equilibrium signals are

\[
(\pi^*_\hat{a}(a|A), \pi^*_\hat{a}(a|B)) = \begin{cases} 
(1, 1); & \alpha \leq \frac{7}{16} \\
(1, \frac{7}{16}); & \frac{7}{16} < \alpha \leq \frac{63}{64} \\
(1, \frac{4}{9}); & \frac{63}{64} < \alpha
\end{cases}, \quad (\pi^*_\hat{b}(a|A), \pi^*_\hat{b}(a|B)) = \begin{cases} 
\left( \frac{35}{36}, 0 \right); & \alpha \leq \frac{7}{16} \\
\left( \frac{63-64\alpha}{(1-\alpha)64}, 0 \right); & \frac{7}{16} < \alpha \leq \frac{63}{64} \\
(0, 0); & \frac{63}{64} < \alpha
\end{cases}
\]

with expected signal

\[
(\bar{\pi}^*(a|A), \bar{\pi}^*(a|B)) = \begin{cases} 
\left( \frac{35+\alpha}{36}, \alpha \right); & \alpha \leq \frac{7}{16} \\
\left( \frac{63}{64}, \frac{7}{16} \right); & \frac{7}{16} < \alpha \leq \frac{63}{64} \\
\left( \alpha, \frac{49}{9} \right); & \frac{63}{64} < \alpha
\end{cases}.
\]

Note that the optimal signal of each sender is less informative in the second case than in the first case, regardless of $\alpha$. Likewise the expected signal in the first case induces four posteriors that are a mean preserving spread of the two posteriors that the expected signal induces in the second case, so the latter is less informative. Note that for intermediate alphas, both senders obfuscate more information than in the benchmark without uncertainty about the signal. The equilibrium expected
signal $\bar{\pi}^*$ is in the interior of the signal space for any $\alpha \in (0, 1)$ when there is uncertainty between two senders, while with one sender the signal from which the receiver update beliefs is always in the boundaries of the signal space and is more informative. Figure 3 illustrates the equilibria for $\alpha = 0.25$ and for $\alpha = 0.75$.

5.3 Senders that can choose to disclose their motives

From the discussions in sections 5.1 and ?? we know that when a sender has access to exclusive messages, the receiver fully learns that the message came from such a sender after receiving one of these exclusive messages. The uncertainty arises when two or more senders use the same messages. In this section I explore the implications of giving senders the choice to self-report their type. In many platforms online it is common that senders can post information using their own profile name or anonymously. In many situations even non-anonymous posting can lead to uncertainty. Lazer et. al (2018) document how many accounts mimic in appearance reputable sources to gain credibility. If it is costless to choose a profile name that sounds like a respectable news paper, the receiver can correctly anticipate this, and interpret the information accounting for the uncertainty of the signal received. If every sender can mimic the messages of other senders, the situation is well approximated by assuming that senders have access to the same message space. However, in the late 2000’s and early 2010’s social platforms provided content producers with a way to verify their identity\footnote{Twitter adopted this policy in 2009, followed by other platforms like Google+, and Facebook in 2011 and 2012.}. This provides senders with exclusive badges that no other sender in the platform can use. Except for hackings, if a post has the verified account badge, people can have certainty about the identity of the source. In recent years, this feature has been extended not only to famous
sources, like public figures or well-known institutions, but to every individual that chooses to verify their account.

The purpose of this section is to study the equilibrium in a game where senders can choose to disclose their identities or to post anonymously, by using the messages that all senders have access to. I conclude the section discussing the implications of the results as well as the way it can be implemented. I argue that even though verifying the identity can solve the problem of impersonation, if a receiver is still uncertain about the motives of a source, knowing its identity might not be enough. I propose a way in which current policies can be enhanced in order to allow for sources to self-disclose their motives.

Consider the message space for each sender, $M_t$, to include a common part, $M$, and an exclusive part, $M^*_t$. The former consists of messages that do not necessarily disclose their type, since every sender can use them, while the second set of messages can only be used by sender $t$. That is $M_t = M \cup M^*_t$ for all $t \in \mathcal{T}$, where $\{M, \{M^*_t\}_{t \in \mathcal{T}}\}$ are $|\mathcal{T}| + 1$ pair-wise disjoint sets. For simplicity lets assume that there are no restrictions on the cardinality of these sets. From lemma ?? we can assume that $|M| = |\mathcal{A}|$ and $|M^*_t| = |\mathcal{A}|$.

**Theorem 4 (Exclusive messages and unraveling of information)** For any game $G$ where senders have access to exclusive messages, $M_t = M \cup M^*_t$:

- A separating equilibrium where $\text{sup}(\pi^*_t) \cap \text{sup}(\pi^*_s) = \emptyset$ for all $t \neq s$ always exists. Let $\bar{\pi}^s$ be its expected signal.

- if a semi-separating equilibrium exists with expected signal $\bar{\pi}^{ss}$, then $\bar{\pi}^s \not\succ \bar{\pi}^{ss}$.

- If senders are single-issue, the separating equilibrium is preferred to any semi-separating equilibrium, for at least one of the senders that is partially pooled, i.e. for some sender with $\text{sup}(\pi^*_t) \cap M \neq \emptyset$. For generic prior, $\mu_0$, the separating equilibrium is strictly preferred.

The first part of the last result implies that there exists an equilibrium where uncertainty unravels and the receiver learns the type of the sender, and the informativeness of the message received. Further, the expected signal is at least as informative as in any semi-separating equilibrium. Notice that for any sender $t$, if $\bar{\pi}^s_{-t}$ is supported on $M$, the choice between supporting their signal in $M$ or to use their exclusive messages, $M^*_t$, has the following trade-offs. On the one hand, if $\pi_t$ is supported in $M^*_t$, the sender has full control of the posterior beliefs of the receiver, and so can choose a signal the is globally optimal for him. However, it has no control on the posteriors induced when nature chooses a different sender. On the other hand, if $\pi_t$ is supported in $M$, the sender can partially control the posteriors of the receiver even if nature chooses the message of some other sender. Therefore, for a semi-separating equilibrium to exist, the utility received by pooling with other senders, like public figures or well-known institutions, but to every individual that chooses to verify their account.

The purpose of this section is to study the equilibrium in a game where senders can choose to disclose their identities or to post anonymously, by using the messages that all senders have access to. I conclude the section discussing the implications of the results as well as the way it can be implemented. I argue that even though verifying the identity can solve the problem of impersonation, if a receiver is still uncertain about the motives of a source, knowing its identity might not be enough. I propose a way in which current policies can be enhanced in order to allow for sources to self-disclose their motives.

Consider the message space for each sender, $M_t$, to include a common part, $M$, and an exclusive part, $M^*_t$. The former consists of messages that do not necessarily disclose their type, since every sender can use them, while the second set of messages can only be used by sender $t$. That is $M_t = M \cup M^*_t$ for all $t \in \mathcal{T}$, where $\{M, \{M^*_t\}_{t \in \mathcal{T}}\}$ are $|\mathcal{T}| + 1$ pair-wise disjoint sets. For simplicity lets assume that there are no restrictions on the cardinality of these sets. From lemma ?? we can assume that $|M| = |\mathcal{A}|$ and $|M^*_t| = |\mathcal{A}|$.

**Theorem 4 (Exclusive messages and unraveling of information)** For any game $G$ where senders have access to exclusive messages, $M_t = M \cup M^*_t$:

- A separating equilibrium where $\text{sup}(\pi^*_t) \cap \text{sup}(\pi^*_s) = \emptyset$ for all $t \neq s$ always exists. Let $\bar{\pi}^s$ be its expected signal.

- if a semi-separating equilibrium exists with expected signal $\bar{\pi}^{ss}$, then $\bar{\pi}^s \not\succ \bar{\pi}^{ss}$.

- If senders are single-issue, the separating equilibrium is preferred to any semi-separating equilibrium, for at least one of the senders that is partially pooled, i.e. for some sender with $\text{sup}(\pi^*_t) \cap M \neq \emptyset$. For generic prior, $\mu_0$, the separating equilibrium is strictly preferred.

The first part of the last result implies that there exists an equilibrium where uncertainty unravels and the receiver learns the type of the sender, and the informativeness of the message received. Further, the expected signal is at least as informative as in any semi-separating equilibrium. Notice that for any sender $t$, if $\bar{\pi}^s_{-t}$ is supported on $M$, the choice between supporting their signal in $M$ or to use their exclusive messages, $M^*_t$, has the following trade-offs. On the one hand, if $\pi_t$ is supported in $M^*_t$, the sender has full control of the posterior beliefs of the receiver, and so can choose a signal the is globally optimal for him. However, it has no control on the posteriors induced when nature chooses a different sender. On the other hand, if $\pi_t$ is supported in $M$, the sender can partially control the posteriors of the receiver even if nature chooses the message of some other sender. Therefore, for a semi-separating equilibrium to exist, the utility received by pooling with other}

---

19 In (year), Facebook announced new disclosure requirements and implemented a feature that makes it easy for consumers to have more information about a source and its motives.

20 See theorem 3 for the definition of “single-issue”

21 In general sender’s might support their signals in the same messages, but to the extent that they assign different probabilities to each message, the equilibrium is semi-separating. A pooling equilibrium exists only if all senders agree in their preferred expected signal.
senders and receiving a lower payoff (though with higher probability), must be relatively close to the first-best payoff that can be achieved by separating. The content of the second statement in theorem 4 is that when senders are single-issue, their payoff is proportional to the probability that their favorite action is chosen by the receiver. Since the probabilities with which actions are played are bound to add up to the unity, senders face a zero-sum game. If a sender strictly benefits from pooling with other sender, that other sender must strictly benefit from separating.

**Example 4 (Bias verification in Social Media)** A key component of the last result is the exclusivity of messages $M_i^t$. In digital platforms, where the source of information is decentralized, the problem of impersonation of reputable sources, and in general of “Fake News” is large (REFERENCE). This makes it difficult for the receivers of information to assess the quality of the statements and, in general, undermines the overall usefulness of these platforms as sources of information. This is of great concern both for the private companies as for politicians (See, ADD REFERENCE TO OFFICIAL STATEMENTS).

Most of the big social platforms have adopted, in response, a series of measures to address these concerns. The “Identity verification” tool, for example, has been very successful at reducing the ability to impersonate the identity of a public figure (REFERENCE). More recently, (October, 2017) Facebook launched the “i button”, a feature that shows more information about the source of an article even if it was shared by a third-party. The information they currently provide includes the Wikipedia entry of the publisher (if available). It shows if the identity of the publisher has been verified and the tenure of the source as a registered site. They also include a summary of other articles published by the same source, or by a different source on the same topic, etc. To the extent that this information cannot be mimicked by other sources, it effectively reduces the uncertainty about the motives of the source.

The previous result suggests that these measures could be complemented by tools that would give the option for senders to verify their sponsorships and political endorsements; similar to the disclosure policies that are in place for radio and television.

6 Literature review

In the classic model of cheap talk (Crawford and Sobel, 1982), persuasion occurs between an informed sender and an uninformed receiver. The first one chooses a message to communicate their information to the latter, the decision maker, but the sender has no commitment power to be truthful. The lack of commitment and freedom to use any message to convey information implies that communication is only possible when the preferences of the two parties are sufficiently aligned. That is, if the sender is not too biased\(^\text{22}\) In contrast, when the sender is constrained to only use statements that are verifiable, Grossman (1981) and Milgrom (1981) showed that even a single biased sender will fully reveal the state of the world. By restricting a sender to, at worst, send no

\(^{\text{22}}\)The bias of a sender is given by the gap between his preferred action and the optimal action for the receiver in each state of the world.
information, the misalignment of incentives between the sender and the receiver unravels and full information arises as equilibrium. In contrast, Kamenica and Gentzkow (2011) (henceforth KG11) study the case when a sender can disclose their methods to learn about the state of the world and commit to reveal the results. In this setting, persuasion can occur even for arbitrarily high biases, so long as the sender has some incentives to provide information. In equilibrium a biased sender does not fully reveal the state of the world: not by misrepresenting the results of their experimentation process, but by strategically choosing methods that partially obfuscate information. In these three benchmarks, however, the receiver has no uncertainty about the information process that generates the information received.

Various papers have studied setting where the receiver is uncertain about the sender’s type in the context of cheap talk and persuasion games with verifiable information. This paper contributes to the literature by studying this uncertainty in the context of Bayesian Persuasion. Within the cheap talk literature, Morgan and Stocken (2003), Dimitrakas, and Sarafidis (2005), and Li and Madarasz (2008), study the effect of having multiple types of senders that differ in the size or direction of their biases. They illustrate how, in cheap talk, uncertainty may soften the problem of misaligned incentives and improve the amount of information communicated in equilibrium, specially when senders have opposite biases. In a similar way Shin (1994), and Wolinsky (2003) consider multiple types of senders communicating through verifiable messages. In the first one, senders differ in how informed they are about the state of the world while in the second one, they differ in their bias. They find that the uncertainty about the source prevents the unraveling mechanism to operate and information is only fully revealed in the state where the sender’s preferences are aligned with the receivers, otherwise information is obfuscated.

Within the framework of Bayesian persuasion, source uncertainty has been studied in one of three ways. First, Le Treust and Tomala (2017) and Tsakas and Tsakas (2018) consider a receiver that with some probability receives some exogenous noise. That is, the receiver observes some exogenously given garbling of the signal designed by the sender. In the latter they find that some noisy communication channels are beneficial to the sender while others are not. The former restricts the message space available to the sender and shows how multiple opportunities for persuasion allow the sender to overcome these restrictions. Second, Parego et. al (2017) and Min (2017) consider a sender with partial commitment. There is an exogenous probability for the sender to privately modify the signal after the message is realized. That is, with some probability the information comes from an agent with commitment to reveal their information and with complementary probability the message is cheap talk; the sender is uncertain about what type of sender is sending the message. Third, several recent papers have studied Bayesian persuasion with a privately informed sender. Perez-Richet (2014) and Kosenko (2018) consider the problem of partially informed senders with some exogenous restrictions on the space of signals they can use to persuade. Piermont (2016) allows for arbitrary signals and in a dynamic setting explores the incentives for disclosing their private information. Finally, Hedlung (2017) considers a sender that is privately informed and has monotonic preferences over the posteriors beliefs of the receiver. In contrast, I will allow for multiple types of senders that can arbitrarily differ in their information and their preferences over
the actions of the receiver, and have access to a rich space of signals\(^{23}\).

From the receiver’s perspective, there are two dimensions of uncertainty: the state of the world and the type of sender that is providing the information. This paper considers senders that have commitment power to reveal the results of their experiment, but not to reveal their type. In a closely related work, Jain (2018) also considers a two dimensional space of uncertainty, where senders have commitment power in only one dimension. In Jian’s work, however, both dimensions are relevant for the receiver’s payoff. Here, the type of the sender has no direct effect on the receiver’s utility, rather it is important to the extent that different types of senders provide information of different quality.

Finally, this paper relates to the competitive persuasion literature. Gentzkow and Kamenica (2017a,b) (henceforth GK17) consider a receiver that observes a collection of signals, and their realizations. Without source uncertainty about the signals, they show that, through competition, increasing the number of senders cannot reduce the amount of information provided. Li and Normal (2018) showed how this result is not robust to the assumptions about how senders compete. They showed that assuming that senders only play pure strategies, move simultaneously, and have access to a sufficiently rich signal space\(^{24}\) are both sufficient and necessary assumptions to establish that competition leads to better information. I show that the receiver’s ability to process all the available information and to have certainty about the source are necessary assumptions as well. In their absence, more sources of information do not necessarily lead to better information.

6.1 Persuasion by Multiple Senders: Discussion of Assumptions

In the problem of strategic transmission of information, the party that provides the information has, in general, incentives to obfuscate some of it. An intuition that is common place in the literature prescribes competition as the means to improve the transmission of information in equilibrium. The intuition is simple, once information is out there, it cannot be destroyed. Thus, having more sources of information, especially with contrary incentives, is expected to increase the amount of information that decision makers receive\(^{25}\). Gentzkow and Kamenica explore this idea in two recent papers, GK17. They find sufficient conditions for rationalizing this intuition that more competition in persuasion leads to more informative equilibria. Some of the key assumptions they make in their model are that:

1. The receiver observes the message from all signals.
2. The receiver knows which signal generated each message.
3. Senders can choose signals that are arbitrarily correlated with other senders’ signals.

---

\(^{23}\)In this paper, the richness comes from the fact that the number of messages will be chosen to not be binding as in Le Treust and Tomala (2017) and, in contrast with Perez-Richet (2014) and Kosenko (2018), senders can choose a signal with any desired level of the precision.

\(^{24}\)In KG17’s setting, this assumption means that signals that can be arbitrarily correlated.

\(^{25}\)Similar conclusions are obtained by classic papers in the \textit{cheap talk} literature, see Krishna and Morgan (2001a,b), Battaglini (2002) and Ambrus and Takahashi (2008), and in the literature of \textit{verifiable information}, see Milgrom and Roberts (1986)
4. Senders play pure strategies.

5. Senders choose their signal simultaneously.

Li and Norman (2008) showed that assumptions 3 – 5 are also necessary. They provide examples where, in the absence of any of these assumptions, all equilibria with two senders are less informative than with a single sender. This paper contributes to the discussion of these assumptions by showing that assumptions 1 and 2 are jointly necessary as well. Li and Norman (2008) shed light on how the way in which senders compete is relevant for having a more informative equilibrium. I instead focus on the ability of the receiver to process all the information that is produced. While in previous papers with multiple senders, the receiver is assumed to update their beliefs based on the joint probability of observing all the messages from each sender, in this paper the receiver updates her beliefs based on the average signal.

GK17 showed that when these assumptions hold, the information environment is Blackwell-connected and no sender can decrease the amount of information provided by other senders. The receiver is able to understand and process all the information they receive, so the worst any sender can do is to provide no further information. On the other hand, any sender can arbitrarily increase the amount of information regardless of the actions of other senders. Therefore, their framework assumes that given the distribution of posteriors induced by the signals of other senders, say \( \tau_{-t} \), sender \( t \) can unilaterally induce any mean preserving spread of this distribution, in particular even a distribution over posteriors that are perfectly informative. Likewise, it assumes that no mean preserving contraction of \( \tau_{-t} \) is available to sender \( t \). In contrast, these two properties fail in this framework. Given the strategies of other senders, summarized by \( \bar{\pi}_{-t} \) and the distribution over posteriors it induces, \( \tau_{-t} \), sender \( t \) can unilaterally induce expected signals in a neighborhood of this signal. Some of these expected signals are weakly more informative and some are weakly less informative, but not arbitrarily so (see Theorem 3 and ??).

7 Conclusion

This paper explores the question of how information is strategically transmitted to a decision maker who is uncertain about the sender’s motives to provide information. The receiver understands the quality of information for each type of sender, but upon receiving a message may not be able to distinguish its source, and thus learns about the state of the world from the average signal conditional on the message received. In general, each sender has partial influence on the information used to make decisions. If there is a sender with strong incentives to obfuscate information, it can disproportionally limit the ability of every other sender to convey information. In fact, if the frequency with which the receiver observes messages from this sender surpasses a threshold, no relevant information can be communicated in equilibrium. As a consequence, if information is transmitted in a way thatsender

\footnote{Using Blackwell’s (1953) partial order on distributions of beliefs, distribution \( \tau \) is said to be more informative than distribution \( \tau' \) if \( \tau \) is a mean preserving spread of \( \tau' \). If this is the case, we also say that \( \tau' \) is a mean preserving contraction of \( \tau \).}
platforms - such as social media - wish to remain influential, editorial work must be done to limit the prevalence of noisy signals.

I moreover argue that increasing the variety of senders with opposed biases can not only increase the uncertainty about the source of a message, but also reduce the incentives for senders to provide information. Each sender leverages the fact that when other senders send a message that is favorable to them it is highly credible. This relaxes how informative their own signal should be in order to persuade the receiver, so they recommend their preferred action more often, thus lowering the informativeness of messages. I also show that if senders have access to exclusive messages, the uncertainty can unravel so that the receiver endogenously learns the quality of the information received. Social media platforms can improve the quality of information by verifying self-reported claims about the motives of a sender, rather than directly assessing the quality of the information they provide. For example the current efforts of social media platforms to verifying the identity of content sources can be improved by allowing this content producers to also have their financial and political sponsorships verified.

References


104(5), 469-74.


8 Appendix

(Incomplete)

8.1 Simplifying the common message space

In many real-world situations the headline of a news article, or the post in a social platform contains information about the source and fully discloses their incentives to provide information. For example, institutional affiliations, the name of a well-known news media, etc. This is akin to having some senders with access to messages that no other sender can use, and therefore being able to disclose their type as described in the previous section. However, some of the growing concern about fake information in digital media stems from the relative ease with which a message can be made to look like coming from a respectable source, or how easy it is to steal the identity of a person or organization to deceive the receivers of information. When every sender can mimic the way in which other senders convey information, we can assume that all senders have access to the same set of messages. I’ll start by characterizing this extreme and symmetric case; i.e. $M_t = M$ for all $t \in T$. Though a restrictive assumption, it will clarify how the receiver’s uncertainty about the source affects the way in which senders choose the information to convey to the receiver. All of the results derived in this section can be generalized to asymmetric settings. In section ??, I explain how this generalization is done.

As we have seen in lemma 2 the receiver updates her beliefs after receiving a message by averaging the signals of all the senders that use this message. Therefore, the incentives to use a message depend on how and whether other senders are using it. One might expect the set of equilibria to be sensitive to the number of messages available (the cardinality of $M$), but this is not the case.

Proof of Proposition 2

Supposed each sender chooses a pure strategy.

$$\text{marg}_{\Theta} \mu(\theta, t|m) = \sum_{t \in T} \frac{\pi_t(m|\theta) \mu_0(\theta, t)}{\sum_{s \in T} \sum_{\omega \in \Theta} \pi_s(m|\omega) \mu_0(\omega, s)}$$

$$= \frac{\sum_{t \in T} \left( \pi_t(m|\theta) \mu_0(t|\theta) \mu_0^a(\theta) \right)}{\sum_{\omega \in \Theta} \sum_{s \in T} \pi_s(m|\omega) \mu_0(s|\omega) \mu_0^a(\theta)}$$

$$= \frac{\bar{\pi}(m|\theta) \mu_0^a(\theta)}{\sum_{\omega \in \Theta} \bar{\pi}(m|\omega) \mu_0^a(\omega)}$$

Proof of Corollary ??
\[
\mu_1^\Theta(\theta_0|m_0) = \frac{\bar{\pi}(m_0|\theta_0)\mu_0^\Theta(\theta_0)}{\sum_{\theta \in \Theta} \bar{\pi}(m_0|\theta)\mu_0^\Theta(\theta)} = \frac{\left[\sum_{t \in T} \pi_t(m_0|\theta_0)\mu_0(t|\theta_0)\right] \mu_0^\Theta(\theta)}{\sum_{\theta \in \Theta} \sum_{t \in T} \pi_t(m_0|\theta)\mu_0(t|\theta) \mu_0^\Theta(\theta)}
\]

The first two equalities follow from proposition 2, the third one reflects that \(\pi_t(m_0|\theta_0) = 0\) for all \(t \notin T_0\), and divides numerator and denominator by the positive constant \(\sum_{t \in T_0} \sum_{\theta \in \Theta} \mu_0(t, \theta)\). Finally, the last equality comes from the fact that \(\mu_0(t|\theta)\sum_{t \in T_0} \sum_{\theta \in \Theta} \mu_0(t, \theta) = \mu_0(t|T_0) = \mu_0(t|T_0)\mu_0^\Theta(\theta|T_0)\).

Using the law of iterated expectations.

**Proof of Proposition 2**

Claim: Any equilibrium is outcome equivalent to an equilibrium with signals restricted to be have their support contained on a message space, \(M'\) with cardinality \(|M'| \leq |A|\).

Proof: Let \(\{\pi_t\}_{t \in T}\) be an equilibrium set of signals. Define \(M(a) = \{m \in M|a^*(\mu^\Theta(m) = a\}\) for every \(a \in A\). Let \(M'\) be the collection of actions for which \(M(a)\) has positive measure. Clearly, \(|M'| \leq |A|\). Now, define a set of signals \(\{\pi'_t\}_{t \in T}\) with support in \(M'\) such that \(\pi'_t(a|\theta) = \sum_{m \in M(a)} \pi_t(m|\theta)\) for all \(a \in M', \theta \in \Theta\) and for each \(t \in T\). The posterior induced by the new average signal, \(\bar{\pi}'\) after message \(a\) must be in the convex hull of the posteriors induced by the original expected signal \(\bar{\pi}\) after each of the messages in \(M(a)\). Since expected utility is linear in beliefs, \(\mu_1^\Theta(a)\) also induces \(a\). Therefore, the distribution over actions is the same under the original and the new set of signals. If signals are restricted to be supported in \(M'\), there is no profitable deviation for any sender.

Claim: Any equilibrium with signals, \(\{\pi'_t\}_{t \in T}\), restricted to have their support contained on a set \(M'\), is outcome equivalent to an equilibrium with signals whose support must be contained in a finite set \(M^*\), whenever \(|M'| \leq |M^*|\).

Proof: Since \(|M'| \leq |M^*|\), there exists a partition of \(M^*\) with \(|M'|\) elements. Since \(M^*\) is finite, every partition is measurable, so we can construct signals \(\{\pi'_t\}\) supported on \(M^*\) such that each sender distributes uniformly the mass in each point of \(M'\) into a different element in the partition of \(M^*\). The distribution over posteriors, thus over actions will be identical and since there are no un-used messages, it constitutes an equilibrium. By letting \(M^* = A\), we show that any equilibrium restricted on a smaller set of messages is equivalent to an equilibrium with \(|A|\) messages available, and by letting \(M^* = M\) we complete the proof.

**Proof of Theorem ??**

A key difference with classic games is that each sender’s action space depends on the profile of
actions of other senders. Debreu (1952), however, overcomes this hurdle by, in the terms of this model, requiring \( \Pi \) to be a contractible polyhedron, \( \Pi'(\bar{\pi}_{-t}) \) to have a closed graph, \( \hat{v}_t(\cdot) \) to be continuous and the best-response correspondence contractible. The first two conditions are guaranteed, given that \( \Pi \) is a closed set and \( \Pi'(\bar{\pi}_{-t}) \) is an upper hemi-continuous correspondence on a closed domain, hence with a closed graph. Though the functions \( \hat{v}_t(\cdot) \) may not be continuous, their upper semi-continuity ensures that the game is better reply secure. Given the quasi-concavity of \( \hat{v}_t(\cdot) \) we can complete the proof of existence using arguments similar to Reny (1999).

Proof of Proposition ??

1. We have

\[
\hat{a}^*(\mu) = \arg\max_{a \in A} \sum_{\theta \in \Theta} \sum_{t \in T} u(a, \theta)\mu(\theta, t)
\]

\[
= \arg\max_{a \in A} \sum_{\theta \in \Theta} \sum_{t \in T} u(a, \theta)\mu(\theta, t)
\]

\[
= \arg\max_{a \in A} \sum_{\theta \in \Theta} u(a, \theta)\mu^\Theta(\theta)
\]

\[
= \hat{a}^*(\mu^\Theta)
\]

Therefore, \( \hat{a}^*(\mu) \) does not depend on the belief about \( T \).

2. From proposition ?? the distribution over posteriors induced by \( \{\pi_t\}_{t \in T} \) is given by \( \tau^\Theta(\bar{\pi}) \), and equation ?? defines implicitly the expected signal as a function of other senders’ signals and its own signal. By equation ?? \( E_{\tau} v_t(a^*(\tau), \theta) = E_{\tau} \hat{v}_t(\pi_t, \bar{\pi}_{-t}) \). Finally, \( \pi_t \in \Pi'(\bar{\pi}_{-t}) \) denotes that signal \( \pi_t \) uses messages that are congruent across senders.

3. This follows from proposition ?? the receiver beliefs only depend on \( \mu^\Theta_t \) and \( \bar{\pi} \).

Proof of Lemma ??

Equation ?? is linear in \( \pi_t, \bar{\pi}_{-t} \) and \( A_t \), thus continuous. Therefore, the correspondence \( \Pi'(\bar{\pi}_{-t}, A_t) \) is also continuous. Since the set \( \Pi'(\bar{\pi}_{-t}) \) is non-empty, compact, and convex, so is the image of the correspondence for any values of \( \bar{\pi}_{-t}, A_t \). Finally, consider two weighting matrices \( A_t < A'_t \), and \( \bar{\pi} \in \Pi'(\bar{\pi}_{-t}, A_t) \). This is, there exists a signal \( \bar{\pi}_t \in \Pi'(\bar{\pi}_{-t}) \) such that \( \bar{\pi} = A_t\bar{\pi}_t + (I - A_t)\bar{\pi}_{-t} \).

Claim: The signal \( \hat{\pi}_t^* = (A'_t)^{-1}A_t\hat{\pi}_t + (I - (A'_t)^{-1}A_t)\bar{\pi}_{-t} \) is in \( \Pi'(\bar{\pi}_{-t}) \) and \( \bar{\pi} = A'_t\hat{\pi}_t^* + (I - A'_t)\bar{\pi}_{-t} \). This implies that \( \bar{\pi} \in \Pi'(\bar{\pi}_{-t}, A'_t) \) which completes the proof. Proof of claim \( (A'_t)^{-1}A_t \) is a square matrix with entries in \( (0, 1) \) since \( A_t < A'_t \). Because \( \hat{\pi}_t \) and \( \bar{\pi}_{-t} \) are in the convex set \( \Pi'(\bar{\pi}_{-t}) \), so is \( \hat{\pi}_t^* \). Using the definition of \( \hat{\pi}_t^* \) and the hypothesis that \( \bar{\pi} = A_t\hat{\pi}_t + (I - A_t)\bar{\pi}_{-t} \), simple calculations show that \( \bar{\pi} = A'_t\hat{\pi}_t^* + (I - A'_t)\bar{\pi}_{-t} \).

Proof of Theorem ??
Let $t_0$ be the sender whose payoff function $v_{t_0}(a, \theta)$ is concave, and consider an uninformative signal $\pi_{t_0}(m|\theta) = 1/|M|$ for all $m \in M$ and all states $\theta \in \Theta$. This sender uses all the messages with positive probability. In the limit when $A_{t_0}$ converges uniformly to the identity matrix $I$, from the continuity of equation $\pi \to \pi_{t_0}$. Therefore, the receiver’s posterior equals the prior, the receiver would then choose its default action, $a^*(\mu_{t_0}^\Theta)$. From the continuity of $\bar{\Pi}(\bar{\pi}_{t-1}, A_t)$ for every $t \in T$, for any $\epsilon > 0$ there exists a $\delta > 0$ such that for for any prior distribution, $\mu_0$ satisfying $A_{t_0} \geq (1 - \delta)I$ the expected signal defined by the average of the signal $\pi_{t_0}$ defined above and identical and perfectly informative signals $\pi_t$ for all $t \neq t_0$, induces posteriors contained in an open ball centered at the prior with radius $\epsilon$. The existence of the uniform $\delta$ can be guarantied given the compactness of $\Theta$, $T$, and using standard continuity arguments.

For generic preferences of the receiver, the default action is optimal for posteriors in a small open ball around the prior. This concludes the proof. To see it, note that a sender with concave preferences has no incentives to provide information (KG11), since the receiver is choosing the default action with probability one, this sender has no profitable deviation. Any sender with incentives to provide information, would want to induce a mean preserving spread of the posteriors by providing more information. However, from the construction of the $\delta$ the receiver is not convinced to change their action even if all other senders were to coordinate to send informative signals. That is, no deviation is profitable.

Now supposed that for the same prior $\mu_0$ satisfying $A_{t_0} \geq (1 - \delta)I$ for some $\delta > 0$ there existed an equilibrium where the receiver does not choose the default action with probability one. Sender $t_0$ could deviate to the uninformative signal $\pi_{t_0}$ defined above and guaranty the receiver to choose $a^*(\mu_{t_0}^\Theta)$ with probability one. Notice that regardless of the signals of other senders, this uninformative signal uses messages that are congruent across players, and so it is feasible.

Proof of Theorem 3

Let $a_\theta \in A = \Theta$ be the action that is optimal for the receiver when the state is $\theta$, for all $\theta \in \Theta$, and $t_\theta$ the sender who would like the receiver to choose $a_\theta$ regardless of the state of the world. That is $u_{t_\theta}(a, \theta) = 1\{a = a_\theta\}$. Since $|T| \leq |\Theta|$, let $\Theta' = \{\theta \in \Theta : t_\theta \in T\}$. As for the receiver, to simplify calculations we can normalize the payoff of the default action, $a_{\theta_0}$ to 1, regardless of the state of the world, and the payoff of action $a_\theta$ to $u_\theta > 1$ when the state of the world is $\theta$, and zero otherwise. That is $u(a, \theta) = u_\theta 1\{a = a_\theta\}$ for all $\theta \neq \theta_0$. The normalization allows us to define the decision rule for the receiver to choose any action, $a \neq a_{\theta_0}$, in a simple threshold rule, where $a^*(\mu^\Theta) = a_\theta$ only if $\mu^\Theta(\theta) \geq q_\theta$ for some threshold $q_\theta \in (0, 1)$. This is a necessary, but not a sufficient condition in general. For some beliefs, the receiver could be indifferent between two actions that are different from the default action. In that case, the choice of $a_\theta$ might not only depend on the belief that the state of the world is $\theta$, but also on the belief about other states of the world. This is, however, irrelevant for the proof; for simplicity, lets assume that $a^*(\mu^\Theta) = a_\theta$ if only if $\mu^\Theta(\theta) \geq q_\theta$, i.e. $u_\theta \in (1, 2)$ for all $\theta \neq \theta_0$. For generic a marginal prior, the receiver strictly prefers the default

\footnote{For the general case consider the set of beliefs where action $a_\theta$ is induced and $q_\theta(\{u_\theta\}_{\theta \neq \theta_0})$ the boundary set of beliefs for which the agent is indifferent between action $a_\theta$ and some other action.}
action from any other action, thus \( q_\theta > \mu^*_1(\theta) \).

Let \( \pi^T \) be an equilibrium expected signal with \(|T|\) senders, \( \{\mu^1_\Theta(\theta|m)\}_{m \in M} \) the set of posteriors induced in equilibrium, and \( \hat{\Theta} \subseteq \Theta' \) the set of states for which the receiver-optimal action is induced: \( \hat{\Theta} = \{ a \in \Theta' : \exists m \in M \text{ such that } a^*(\mu^1_\Theta(\theta|m)) = a \} \).

**Claim:** If \( t_{\theta_1} \notin T \), then \( \hat{\Theta} = \Theta' \) for any \( \pi^T \). Proof. Suppose there exists \( \theta^* \in \Theta' \) but \( \theta^* \notin \hat{\Theta} \). Then sender \( t_{\theta^*} \) has a payoff of zero in equilibrium. The \( t_{\theta^*} \)-conditional signal, \( \pi_{-t_{\theta^*}} \), has all messages in its support; otherwise, sender \( t_{\theta^*} \) has a profitable deviation by using a message that is not in the support of \( \pi_{-t_{\theta^*}} \), action \( a_{\theta^*} \) can be induced with some probability and yield a positive payoff to \( \theta^* \).

**Claim:** given the equilibrium \( \pi \), for every \( \theta \in \hat{\Theta} \) there exists an equilibrium posterior, given some message, say \( m_\theta \), such that \( \mu^1_\Theta(\theta|m_\theta) \geq q_\theta \), and \( \mu^1_\Theta(\theta^*|m_{\theta^*}) = q_{\theta^*} \) for at least one \( \theta^* \in \hat{\Theta} \). Proof. The first statement of the claim follows from the definition of \( \Theta \), there must exist a message, \( m \), inducing a posterior \( \mu^1_\Theta(\theta|m) \) that assigns probability to the state of the world being \( \theta \), \( \mu^1_\Theta(\theta|m_\theta) \), of at least \( q_\theta \). For the second claim, suffices to show that \( \mu^1_\Theta(\theta^*|m_{\theta^*}) \neq q_{\theta^*} \), since \( \theta^* \in \hat{\Theta} \). Proceed by contradiction, assume that for each \( \theta \in \hat{\Theta} \) the inequality is strict, \( \mu^1_\Theta(\theta|m_\theta) > q_\theta \), given some \( m \in M \). Then the equilibrium signal for sender \( t_{\theta} \) must send message \( m_\theta \) with probability 1 in all states of the world, otherwise, there would be a profitable deviation. However, if this is true for all senders, then senders have pair-wise disjoint supports of their signals. In fact, after observing \( m_\theta \) the receiver knows it came from \( \pi_\theta \) with probability one and it is a perfectly uninformative signal, so \( \mu^1_\Theta(\theta|m_\theta) = \mu^1_\Theta(\theta) < q_\theta \), a contradiction.

To see that any equilibrium with 1 sender is more informative. Let \( \theta^* \) be the state of the world where for equilibrium \( \pi^T \) we have that \( \mu^1_\Theta(\theta^*|m_{\theta^*}) = q_{\theta^*} \) and consider the equilibrium signal for \( t_{\theta^*} \), when there is no other sender: \( \pi_{t_{\theta^*}}(m_{\theta^*}|\theta^*) = 1 \), and \( \pi_{t_{\theta^*}}(m_{\theta^*}|\theta) = \left(1 - \frac{q_{\theta^*}}{q_{\theta^*}}\right) \left(\frac{\mu^0_1(\theta^*)}{1 - \mu^0_1(\theta^*)}\right) < 1 \) and send a different message, \( m_{\theta^*} \), with complementary probability in each state \( \theta \neq \theta^* \). Easy algebra shows that this signal maximizes its expected payoff, and since message \( m_{\theta^*} \) is only used in state \( \theta \) for all \( \theta \neq \theta^* \) it perfectly reveals such state of the world. We conclude that this signal must is more informative in the Blackwell order then \( \pi^T \).

**Proof of Lemma ??**

This is a direct result from the second claim made in proposition 4.

**Proof of Theorem ??**

Let \( \pi^* = \arg\max_{\pi \in \Pi} E_{T(\pi)}v(a, \theta) \) given the unconditional prior \( \mu^0_\Theta \). Since \( \hat{v}_t(\cdot) = E_{T(\pi)}v(a, \theta) \) for any \( \pi = \bar{\pi} \), signal \( \pi^* \) is the unconstrained optimal signal for every sender. Clearly, if \( \pi_t = \pi^* \) for all \( t \in T \), then \( \bar{\pi} = \pi^* \) regardless of the prior, \( \mu_0 \), so long as its marginal equals \( \mu^0_\Theta \). No sender has a profitable deviation since the induced expected signal is the unconstrained optimum. Therefore, any set of signals \( \{\pi_t\}_{t \in T} \) that induces the same expected signal, \( \bar{\pi} = \pi^* \) is also an equilibrium. For any such set of signals, every sender is achieving its first best, no profitable deviation exists. Any other equilibrium, must be a local maximum of \( E_{T(\pi)}v(a, \theta) \) because senders can always locally deviate from it. The extent to which they can move the average signal depends on their
weighting matrix $A_t$, thus on the joint prior distribution, $\mu_0$. By the same token, the only equilibria that are robust to the senders’ private information and prevalence, are the ones that induce an expected signal that is a global maximizer of $E_{\tau(\pi)} v(a, \theta)$, hence an equilibrium with a single sender.

**Proof of Lemma ??**

A degenerate posterior after message $m_0$ must put zero probability on at least one state of the world, say $\theta_0$. Therefore, $\sum_{t \in T} \pi_t(m_0|\theta_0)\mu_0(t|\theta_0) = \mu_0(t_0|\theta_0)\pi_{t_0}(m_0|\theta_0) + (1-\mu_0(t_0|\theta_0))\bar{\pi}_{-t_0}(m_0|\theta_0) = 0$. Given the full support of $\mu_0$, $\mu_0(t_0|\theta_0) \in (0,1)$. For the last equation to hold, $\bar{\pi}_{-t_0}(m_0|\theta_0) = 0$ is necessary. This finishes the proof, because if $m_0$ is not used by other senders, then $\bar{\pi}_{-t_0}(m_0|\theta) = 0$ for every $\theta \in \Theta$.

**Proof of Lemma ??**

Message $m_0$ induces a posterior that is equal to the prior, if and only if the expected signal satisfies $\bar{\pi}(m_0|\theta) = \bar{\pi}(m_0|\theta')$ for every $\theta, \theta' \in \Theta$. That is, for any pair of states:

$$\mu_0(t_0|\theta)\pi_{t_0}(m|\theta) - \mu_0(t_0|\theta')\pi_{t_0}(m|\theta') = (1-\mu_0(t_0|\theta))\bar{\pi}_{-t_0}(m|\theta) - (1-\mu_0(t_0|\theta'))\bar{\pi}_{-t_0}(m|\theta').$$

From the congruency of messages across senders, the left hand side and the right hand side have opposite signs. Therefore, the equality holds iff both sides equal zero. This implies the necessity of $\bar{\pi}_{-t_0}(m_0|\theta) = (1-\mu_0(t_0|\theta'))\bar{\pi}_{-t_0}(m_0|\theta')$. For the sufficiency, note that by choosing $\pi_{t_0}(m|\theta) = 0$ for all $\theta \in \Theta$, the left hand side equals zero. Choosing a signal with such a property is feasible to $t_0$ because the fact that the two terms in the right hand side are equal means that there is no restriction imposed by congruency on how sender $t_0$ can use message $m_0$. 

34