PERSUASION WITH VERIFIABLE INFORMATION[*]

Maria Titova[†]

September, 2020

LATEST VERSION

Abstract

This paper studies persuasion with verifiable information. An informed sender with state-independent preferences sends verifiable messages to multiple receivers attempting to convince them to approve a proposal. I first find that every equilibrium is outcome equivalent to a direct equilibrium, in which the sender tells each receiver what to do, and receivers obediently follow their recommendations. This allows me to characterize the full equilibrium set. The sender-worst equilibrium outcome is one in which information unravels, and receivers act as if under complete information. The sender-preferred equilibrium outcome is the commitment outcome of the Bayesian persuasion game. In the leading application, I study targeted advertising in elections and show that by communicating with voters privately, a challenger may win elections that are unwinnable otherwise. When the challenger swings an election, voters regret their choices, and the amount of regret increases as the electorate becomes more polarized.

[*]I am very grateful to my advisors Renee Bowen and Joel Sobel for their guidance and support. I thank Simone Galperti, Remy Levin, Aleksandr Levkun, Denis Shishkin, Joel Watson, Alex Weiss, and the audiences of the 2019 meetings of the Southwest Economic Theory Conference, Young Economists Symposium, and the Economics Graduate Student Conference at Washington University in St. Louis, as well as seminar audiences at UC San Diego. All errors are my own.

[†]Department of Economics, University of California, San Diego. E-mail: mtitova@ucsd.edu.

1. Introduction

I study environments in which a sender communicates private information to multiple receivers. The sender's preferences depend on the set of receivers who approve his proposal and do not depend on the state of the world. Each receiver independently compares the option of approving the sender's proposal, the value of which is unknown, to a known status quo. The sender privately communicates with the receivers by sending each of them a private verifiable message about the state of the world. I explore the set of outcomes of this game and show that this model can make predictions about the effectiveness of targeted advertising in electoral campaigns.

There are many real-life examples of such situations. A politician challenges the incumbent, thus forcing the voting body to choose between his expected policy and the status quo; a firm advertises its product to consumers; a job candidate convinces his interviewers to make him an offer; a company CEO attempts to convince the board of directors to decide on managerial compensation. In all these scenarios, the sender can prove any true claim, but he is not forced to tell the full truth. In his electoral campaign, the challenger may fail to mention his policy on some of the issues, but once elected, he must deliver on the promises he made, or else he would bear extreme reputational costs. The firm may emphasize some of the product characteristics and fail to mention others, but it cannot advertise falsely. The job candidate may choose which certifications to attach to his application, but cannot forge documents. The CEO may omit some of the financial indicators in his report to the board, but those he does present must be legitimate.

Preview of the results

How does the sender convince one receiver with verifiable information? Rather than looking at the sender's messages and the receiver's beliefs, I focus on what the receiver does in every state of the world. Since she chooses between two options, one can partition the state space, which is a unit interval, into two subsets: the set of states in which the proposal is approved, and its complement. The first result states that if some set of approval states is induced in equilibrium, then there exists a direct equilibrium, in which the sender simply tells the receiver what to do. More specifically, if the state of the world is within the set of approval states, he tells the receiver to approve the proposal; otherwise, he tells her to reject it. Every direct equilibrium is characterized by the set of approval states that satisfies two constraints. The *receiver's obedience constraint* ensures that when the sender tells the receiver to approve the proposal, the receiver must rationally expect that her net payoff of approval is non-negative. The *sender's incentive-compatibility constraint* (IC) ensures that the sender does not wish to deviate toward a fully informative strategy that induces the receiver to act as if fully informed. Equipped with the direct implementation result, I characterize the full equilibrium set by restricting attention to sets of approval states that satisfy these two constraints.

In the sender's least preferred equilibrium, his ex-ante odds of approval are minimized across all equilibria. The sender voluntarily discloses whether the state of the world is in the receiver's complete information approval set; the receiver learns everything relevant for her decision and makes a fully-informed choice.

In the sender's most preferred equilibrium, his odds of approval are maximized subject to the receiver's obedience constraint. The solution features a cutoff state: if the net payoff of approval is greater than the cutoff value, then the receiver approves the proposal, and if not, then she rejects it. The cutoff value is negative: the sender convinces the receiver to approve his proposal when she prefers not to approve it under complete information, thus improving his odds upon full disclosure. The sender obfuscates the state of the world so long as the receiver follows through on the recommendation. Consequently, the receiver is indifferent between approval and rejection when she is recommended to approve. This observation makes a connection between verifiable disclosure and Bayesian persuasion (Kamenica and Gentzkow, 2011). The challenger-preferred equilibrium outcome is also a commitment outcome, meaning that the challenger need not benefit from having commitment power when he has the evidence to back up his claims.

With many receivers, one can also restrict attention to direct equilibria, in which the sender privately tells each receiver what to do, and each receiver obediently follows her recommendation. The sender-worst equilibrium features full disclosure, and the sender-preferred equilibrium outcome is a commitment outcome.

The leading application of the model allows me to make predictions about the effectiveness of targeted advertising in electoral campaigns. The voters choose whether to elect the challenger or to reelect the incumbent. Voters prefer to elect the candidate whose (expected) policy position is closest to their bliss points. The incumbent's status quo policy is fixed and known. The challenger is privately-informed and sends verifiable messages to the voters about his policy. For example, the challenger could say "my policy is to the left of the status quo," and the voter would know that this statement is true, but she would not know how far to the left of the status quo his policy is.

Suppose that winning an election requires the unanimous approval of two voters, L and R, whose bliss points are located to the left and the right of the status quo policy, respectively. Observe that unless the challenger can advertise to each of these voters privately, he always loses this election. As long as these voters hold a common belief, which they do under full disclosure, no disclosure, or public disclosure by the challenger, only one of these voters expects the challenger's policy to be closer to her bliss point than the status quo. Consequently, the challenger loses this election with probability one if he fully discloses his policy, if he reveals no information, and in every equilibrium of the verifiable-information game with public disclosure. I call such elections *unwinnable* for the challenger.

When the challenger has access to targeted advertising, he can tell different things to different voters. Recall that in his most preferred equilibrium of a one-voter election, the challenger convinces the voter to elect him even when his policy is further away from the voter's bliss point than the status quo. In particular, he manages to convince voter L (R) to vote for him when his policy is slightly to the right (left) of the status quo. Thus, if the challenger can advertise privately, then he can convince both voters at the same time, and win elections that are unwinnable otherwise. That said, the challenger only benefits from targeting if his policy is sufficiently close to the status quo: the further to the right (left) his policy is, the harder it becomes to convince voter L (R).

Voter Regret and Polarization

There is a sense in which the challenger "tricks" the voters when he swings an unwinnable election by targeted advertising. In particular, when voter L (R) votes for the challenger whose policy is to the right (left) of the status quo, she makes a mistake, which she later regrets. When the challenger wins an unwinnable election, one of the voters always regrets it. Things worsen when the electorate becomes more polarized, which happens when one of the voters' positions becomes more extreme by moving away from the status quo. Suppose, for example, that the bliss point of voter R moves further to the right. That makes this voter less satisfied with the status quo and allows the challenger to convince her to elect him more often. As a result, the challenger's ex-ante odds of winning increase, and he can swing unwinnable elections with a more polarized electorate more often. With that, both voters also regret their choices more often. Finally, as voter R's position becomes more extreme, the challenger-preferred set of winning policies shifts to the left, toward policies preferred by voter L.

Related Literature

I assume that the sender communicates with the receivers by sending them verifiable messages. This verifiable information communication protocol was introduced by Milgrom (1981) and Grossman (1981). Other communication protocols include cheap talk by Crawford and Sobel (1982), Bayesian persuasion by Kamenica and Gentzkow (2011), and signaling by Spence (1973). Relative to these other models of communication, Bayesian persuasion makes the sender better off because it endows him with ex-ante commitment power. Lipnowski and Ravid (2020) consider a setting similar to ours, except the sender communicates with the receivers via cheap talk. They find that the sender's maximal equilibrium payoff from cheap talk is generally strictly lower than his payoff under commitment. Consequently, a cheap-talk sender values commitment.[1] In contrast to their result, I show that the sender need not benefit from commitment if he possesses the hard

[1] Lipnowski (2020) also notes that the sender reaches the commitment outcome with cheap talk if his value function is continuous in the receiver's posterior belief. That assumption is very restrictive: when receivers choose between two options and the sender's preferences are state-independent, the sender's value function must be constant, meaning that no communication takes place under cheap talk, verifiable information, and Bayesian persuasion. I thank Elliot Lipnowski for this insight.

evidence that allows him to verify his messages.

The result that the sender need not benefit from having commitment power when he has hard evidence to back up his claims allows us to replace the assumption of commitment with the more plausible assumption of verifiable information in the existing applications of Bayesian persuasion models. Equivalence holds as long as *(i)* the sender's utility is state-independent and *(ii)* each receiver is choosing independently between two options. This includes applications in which schools persuade employers to hire their graduates (Ostrovsky and Schwarz, 2010; Boleslavsky and Cotton, 2015); pharmaceutical companies persuade the FDA to approve their drug (Kolotilin, 2015); and matching platforms persuade sellers to match with buyers (Romanyuk and Smolin, 2019). In the political economy setting, politicians persuade voters (Alonso and Câmara, 2016; Bardhi and Guo, 2018), and governments persuade citizens through media (Gehlbach and Sonin, 2014; Egorov and Sonin, 2019).

The leading application sheds more light on how political advertising, and especially targeted advertising, affects electoral outcomes and why it has become so widespread. Prat and Strömberg (2013) and DellaVigna and Gentzkow (2010) provide excellent surveys of the evidence of voter persuasion. First, candidates target their ads based on voters' positions on the political spectrum (George and Waldfogel, 2006; DellaVigna and Kaplan, 2007). Second, one can make a case that an increase in the availability of information catered toward certain electoral groups also counts as targeted advertising because these are the messages intended for and heard by these groups (Enikolopov et al., 2011; Oberholzer-Gee and Waldfogel, 2009). I show that targeted political advertising may be so widespread because it allows politicians to win elections that are unwinnable otherwise.

I also contribute to the literature on voter polarization and targeted political advertising. As the number of media outlets increases, each of them becomes more specialized and targets voters with more extreme preferences, which leads to social disagreement (Perego and Yuksel, 2018). If the electorate is polarized to begin with, then so are the candidates' chosen policy platforms (Prummer, 2020; Hu et al., 2019). Abstracting away from candidates choosing their policies, I find that as the electorate becomes more polarized, more challengers can swing elections that are unwinnable otherwise. That comes at the cost of voters regretting their choices with an increasing probability.

The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 characterizes the set of equilibrium outcomes in the game with one receiver. Section 4 describes the range of the sender's equilibrium odds of approval when there are many receivers. Section 5 studies targeted advertising in elections. Section 6 concludes.

2. MODEL

SETUP

There is a state space $\Omega := [0,1]$ and a finite set of receivers $[I] := \{1,\dots,I\}$. The game begins with the sender (him) observing the realization of the random state $\omega \in \Omega$, which is drawn from an atomless common prior distribution $p > 0$ over $\Delta\Omega$.[2] Having observed the state, the sender transmits a verifiable message $m_i \in \mathbb{M} := 2^{|\Omega|}$, such that $\omega \in m_i$, to each receiver (her) $i \in [I]$.[3]

Receiver i observes her private message m_i, but not the state ω, and decides whether to approve the sender's proposal (take action 1) or reject it (take action 0). Receiver i's preferences are described by a utility function $u_i : \{0,1\} \times \Omega \to \mathbb{R}$, and her net payoff of approval of the sender's proposal in state ω is $\delta_i(\omega) := u_i(1,\omega) - u_i(0,\omega)$.

Assuming that the receiver approves the proposal when she is indifferent, which I do throughout the paper, her preferences under complete information are summarized by her <u>approval set</u>

$$\mathcal{A}_i := \big\{\omega \in \Omega \mid \delta_i(\omega) \geq 0\big\}.$$

The sender's state-independent preferences are described by a utility function $u_s : 2^I \to \mathbb{R}$. I assume that the sender weakly prefers that each receiver approves his proposal: given

[2] For a compact metrizable space S, ΔS denotes the set of all Borel probability measures over S. For any distribution $q \in \Delta\Omega$ and for any subset of the state space $W \subseteq \Omega$, I denote by $Q(W)$ the probability measure $Q(W) := \int_W q(\omega)d\omega$ and by $q(\cdot \mid W) \in \Delta\Omega$ the conditional probability distribution $q(\omega \mid W) := \frac{q(\omega)}{Q(W)}$, if $Q(W) > 0$.

[3] I borrow the definition of a verifiable message as a subset of the state space that includes the true realization from Milgrom and Roberts (1986).

two sets of receivers $T, S \subseteq [I]$, $u_s(T) \leq u_s(S)$ if $T \subseteq S$.

The following example introduces the leading application of the model.

EXAMPLE 1 (ELECTIONS). The sender is a politician (the challenger) who challenges the status quo, the receivers are voters, and the state space Ω is a space of policies (e.g. on one socio-economic issue), with 0 being the utmost "left" position, and 1 being the most "right." The challenger gets a payoff of 1 if he wins the election, which is decided by some social choice function (e.g., unanimity or simple majority). Voters have spatial preferences à la Downs (1957): voter $i \in [I]$ has ideal policy $v_i \in \Omega$ and evaluates all other policies based on how far they are from that bliss point.[4]

Specifically, voter i compares the challenger's unknown policy ω to a known status quo policy $\omega_0 \in \Omega$. Her net payoff of voting for the challenger with policy ω is $\delta_i(\omega) = |v_i - \omega_0| - |v_i - \omega|$. Her approval set comprises of policies that are weakly closer to her bliss point than the status quo: $\mathcal{A}_i = \{\omega \in \Omega \text{ s.t. } \delta_i(\omega) \geq 0\}$. This setup is illustrated in Figure 1.

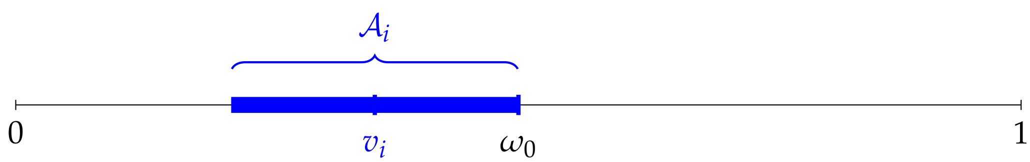


Figure 1: *policy space $\Omega = [0,1]$, status quo policy ω_0, voter i's ideal policy v_i, and her approval set $\mathcal{A}_i$. Under complete information, voter i prefers to vote for the challenger if his policy is in the solid blue region.*

EQUILIBRIUM OUTCOMES

I consider Perfect Bayesian Equilibria (henceforth just *equilibria*) of this game. Note that under incomplete information, receiver i approves the proposal if her expected net payoff of approval is non-negative. Thus, her set of approval beliefs is

$$\mathcal{B}_i := \{q \in \Delta\Omega \mid \mathbb{E}_q[\delta_i(\omega)] \geq 0\}.$$

[4]In this application, voting is assumed to be expressive (also known as sincere). That is, voters derive utility from expressing support for one of the candidates (the challenger or the status quo), and not from the policy that is implemented by the elected candidate. The theory was pioneered by Brennan and Lomasky (1993), Brennan and Hamlin (1998) and reviewed by Hamlin and Jennings (2011). There is a large body of evidence that the behavior of voters is consistent with sincere voting, e.g., in U.S. national elections (Kan and Yang, 2001; Degan and Merlo, 2007), Spanish General elections (Artabe and Gardeazabal, 2014), Israeli General elections (Felsenthal and Brichta, 1985).

DEFINITION 1. *An equilibrium* $(\sigma, \boldsymbol{a}, \boldsymbol{q})$ *consists of the messaging strategy of the sender* $\sigma : \Omega \to \Delta(\mathbb{M}^I)$ *and the profiles of approval strategies* $\boldsymbol{a} := \{a_i : \mathbb{M} \to \{0,1\}\}_{i\in[I]}$ *and posterior beliefs* $\boldsymbol{q} := \{q_i : \mathbb{M} \to \Delta\Omega\}_{i\in[I]}$ *of the receivers, such that*

(i) $\forall \omega \in \Omega$, $\sigma(\cdot \mid \omega) > 0$ *only if* $\arg\max_{m_1,\dots,m_I} u_s\big(\{i \in [I] \mid a_i(m_i) = 1\}\big)$, *s.t.* $\omega \in m_i$, $\forall i \in [I]$.

The following conditions must hold for every receiver $i \in [I]$*:*

(ii) $\forall m \in \mathbb{M}$, $a_i(m) = \mathbb{1}\big(q_i(\cdot \mid m) \in \mathcal{B}_i\big)$;

(iii) $\forall m \in \mathbb{M}$ *such that* $\int_\Omega \sigma_i(m \mid \omega) d\omega > 0$, $q_i(\omega \mid m) = \frac{\sigma_i(m \mid \omega)\cdot p(\omega)}{\int_\Omega \sigma_i(m \mid \omega')\cdot p(\omega') d\omega'}$, *where* σ_i *is the marginal distribution of messages heard on the equilibrium path by receiver* i*;*

(iv) $\forall m \in \mathbb{M}$, $\operatorname{supp} q_i(\cdot \mid m) \subseteq m$.

In words, *(i)* states that the sender puts positive probability only on collections of messages that maximize his utility; *(ii)* states that each receiver is sequentially rational and approves the proposal if and only if her expected net payoff of approving is non-negative under her posterior belief; *(iii)* states that receivers' posterior beliefs are Bayes-rational on the equilibrium path; *(iv)* states that the receivers' posterior beliefs on and off the path are concentrated on the states in which the message is available to the sender.

I analyze the set of equilibrium outcomes of this game via a *state-based approach*. In each state of the world ω, I am interested in which action each receiver takes, and with what probability.

DEFINITION 2.

- *An outcome* $\alpha_i(\omega) \in [0,1]$, *for every* $i \in [I]$ *and* $\omega \in \Omega$, *specifies the probability that receiver* i *approves the sender's proposal in state* ω.
- *An outcome is an equilibrium outcome if it corresponds to some equilibrium.*[5]

Some outcomes are simpler than others in that every receiver takes a particular action

[5]Specifically, if there exists equilibrium $(\sigma, \boldsymbol{a}, \boldsymbol{q})$ such that for every receiver $i \in [I]$ and state $\omega \in \Omega$, $\alpha_i(\omega) = \int_{\mathcal{M}_i(\omega)} \sigma_i(m \mid \omega) dm$, where $\mathcal{M}_i(\omega) := \{m \in \mathbb{M} \mid \sigma_i(m \mid \omega) > 0 \text{ and } a_i(m) = 1\}$ is the set of messages transmitted from state ω that convince receiver i to approve the proposal.

with probability one in each state.[6] Put differently, each receiver essentially partitions the state space into two subsets, the states of approval and the states of rejection. Formally,

DEFINITION 3.

- *an outcome is deterministic if* $\alpha_i(\omega) \in \{0,1\}$ *for every* $i \in [I]$ *and* $\omega \in \Omega$*;*
- *if an outcome is deterministic, then for every receiver* $i \in [I]$, $W_i := \{\omega \in \Omega \mid \alpha_i(\omega) = 1\}$ *is the set of approval states.*

Some equilibria are simpler than others in that the sender plays a pure strategy of directly recommending an action to each voter.

DEFINITION 4. *Equilibrium* $(\sigma^D, \boldsymbol{a}^D, \boldsymbol{q}^D)$ *is a direct equilibrium characterized by a collection of sets of approval states* $\{W_1, \ldots, W_I\}$*, if for every receiver* $i \in [I]$ *and message* $m_i \in \{W_i, \Omega \setminus W_i\}$

$$\sigma_i^D(m_i \mid \omega) = \mathbb{1}(\omega \in m_i), \qquad a^D(m_i) = \mathbb{1}(m_i = L_i).$$

In a direct equilibrium, the sender sends message W_i to receiver i if the realized state is $\omega \in W_i$, and message $\Omega \setminus W_i$ otherwise. In the first case, he effectively recommends that the receiver approves the proposal, while in the second case, he recommends rejection. Notice that W_i is both the *set of states* in which the proposal is approved and the *message* that convinces the receiver to approve the proposal. Given the sender's simple pure strategy, each receiver's Bayesian updating is straightforward. It takes the form of conditioning the prior distribution on the set of states from which each message could have come from, which coincides with the message itself. Thus, for each message on the equilibrium path, receiver i's posterior belief is $q_i^D(\cdot \mid m) = p(\cdot \mid m)$ for all $m \in \{W_i, \Omega \setminus W_i\}$. Consequently, for receiver $i \in [I]$ to correctly interpret message W_i as a recommendation to approve the proposal, message W_i must satisfy the following constraint in any direct equilibrium:

$$p(\cdot \mid W_i) \in \mathcal{B}_i. \qquad \text{(obedience)}$$

---

[6] Even though each receiver always plays a pure strategy and breaks ties in favor of approval, the sender may be playing a mixed strategy in some state, and then in that state the receiver may be approving the proposal with probability between 0 and 1.

Notice that the sender always has the option of fully revealing the state of the world $\omega \in \Omega$ by sending message $\{\omega\}$. According to the equilibrium condition *(iv)*, when a receiver hears this message, her posterior belief places probability one on ω, meaning that she learns the true state of the world. Consequently, the sender can guarantee that receiver i approves the proposal in states in which she approves it under complete information. That gives rise to the sender's incentive-compatibility constraint $\forall i \in [I]$:

$$\mathcal{A}_i \subseteq W_i. \tag{IC}$$

VALUE OF COMMITMENT

I compare the *equilibrium outcomes* of the described verifiable-information game to the *commitment outcomes* of the Bayesian persuasion game. In the Bayesian persuasion game, the sender commits to a signaling policy (that is known to receivers) $\sigma^{BP} : \Omega \to \Delta(\Theta_1, \ldots, \Theta_I)$, where Θ_i is the private signal set of receiver i. Once state $\omega \in \Omega$ is realized, a collection of signals $\{\theta_i\}_{i \in [I]}$ is generated according to σ^{BP}, and receiver i observes her private signal realization θ_i.

DEFINITION 5.

- $(\sigma^{BP}, a^{BP}, q^{BP})$ *is a* <u>*commitment protocol*</u> *if it satisfies equilibrium conditions (ii) and (iii).*
- *An outcome is a* <u>*commitment outcome*</u> *if it corresponds to some commitment protocol.*

Since a commitment protocol satisfies two out of four equilibrium conditions, the sender must do weakly better in the Bayesian persuasion game relative to the game with verifiable information.

3. ONE RECEIVER

Let us first focus on the case with one receiver, i.e. $[I] = \{1\}$. For ease of exposition, I drop all receiver-relevant subscripts i. I assume that the sender's utility is strictly increasing in the receiver's action: the sender gets a payoff of 1 if he convinces the receiver to approve the proposal, and 0 otherwise. The first result of this paper establishes an equivalence between the set of equilibrium outcomes and the set of direct equilibrium outcomes. It also provides a simple characterization for both in terms of two constraints.

THEOREM 1. *Suppose $[I] = \{1\}$. Then, every equilibrium outcome is deterministic. Moreover, the following statements about $W \subseteq \Omega$ are equivalent:*

(1) W is an equilibrium set of approval states;

(2) W is a set of approval states in a direct equilibrium;

(3) W satisfies the receiver's (obedience) *and the sender's* (IC) *constraints.*

The formal proof of Theorem 1 can be found in the appendix, along with all other proofs. Here I describe the intuition behind this result. First, every equilibrium outcome is deterministic and induces a partition of the state space into W, on which the proposal is approved, and $\Omega \setminus W$, on which the proposal is rejected. To see why this has to be true, suppose that in some state the sender induces both approval and rejection with positive probability. But since the sender has access to a message that induces approval, nothing prevents him from deviating to that message and inducing approval with certainty. Consequently, the sender always plays a pure strategy of recommending one of the actions with probability one.[7]

First, observe that *(1)* implies *(3)*. If W is an equilibrium set of approval states, it has to satisfy the sender's (IC) constraint, or else the sender has a profitable deviation to full disclosure. To see why W also satisfies the receiver's (obedience) constraint, imagine that this set of approval states is implemented directly. Specifically, rather than playing the (possibly mixed) strategy prescribed in the original equilibrium, the sender uses a direct pure strategy of sending message W from states $\omega \in W$, and message $\Omega \setminus W$ from states $\omega \in \Omega \setminus W$. Since in the original equilibrium the receiver interprets every message coming from state $\omega \in W$ as a recommendation to approve, in the direct equilibrium, she also interprets the pooling message W the same way.

Next, notice that *(3)* implies *(2)*. An outcome partition that satisfies the two constraints in itself describes a path of a direct equilibrium. Because of (IC), the sender does not have profitable deviations towards full disclosure. Due to verifiability, the sender

[7] This is a purification argument – when the state space is continuous and the information structure is atomless the existence of pure strategy equilibria is often guaranteed (Radner and Rosenthal, 1982; Aumann et al., 1983; Milgrom and Weber, 1985).

cannot deviate to sending the convincing message W from states outside of W. Because of (obedience), the receiver is best responding. To complete the equilibrium characterization, one must specify the off-the-path beliefs of the receiver. To prevent other deviations of the sender, let the receiver be "skeptical" and assume that any unexpected message comes from the worst possible state.

Finally, since a direct equilibrium is an equilibrium, *(2)* implies *(1)*.

Note that the equivalence between statements *(1)* and *(2)* of Theorem 1 can be viewed as a version of the communication revelation principle for games with verifiable information. The communication revelation principle for mediated sender-receiver games, introduced by Myerson (1986) and Forges (1986), states that any equilibrium outcome may be implemented truthfully and obediently. In the present context, it translates into *(i)* the sender fully revealing the state of the world to the mediator, *(ii)* the mediator translating this report into an action recommendation for the receiver, and *(iii)* the receiver obediently following her recommendation. Which equilibrium outcome is implemented is decided by the mediator at step *(ii)*. Conveniently, statement *(3)* of Theorem 1 provides the necessary and sufficient conditions for a set of approval states to be implementable in equilibrium.

EQUILIBRIUM RANGE AND VALUE OF COMMITMENT

When characterizing the full equilibrium set of the game with one receiver, one may restrict attention to sets of approval states that satisfy (IC) and (obedience) constraints, according to Theorem 1. The advantage of the state-based approach lies in that the sender's ex-ante utility in a direct equilibrium equals $P(W)$ and depends only on the set of approval states W and the prior measure $P(\cdot)$.

In the sender-worst equilibrium, the set of approval states is the smallest in terms of the sender's ex-ante utility, and is $\underline{W} = \mathcal{A}$. In this equilibrium, the receiver effectively learns all the relevant information (whether the true state is within her approval set or not), thus making her decision as if under complete information. This equilibrium is outcome-equivalent to *full disclosure* (also known as *full unraveling*), salient in the verifiable information literature.

In the sender-preferred equilibrium, the set of approval states $\bar{W}$ maximizes the sender's ex-ante utility across all equilibria. Mathematically, $\bar{W}$ solves the following problem:

$$\max_{W \subseteq \Omega} P(W)$$

subject to

- the receiver's obedience constraint $p(\cdot \mid W) \in \mathcal{B}$;
- the sender's incentive-compatibility constraint $\mathcal{A} \subseteq W$.

To find the sender-preferred equilibrium, one would increase the ex-ante measure of the set W for as long as message W convinces the receiver to approve the proposal. Because the state space is continuous, the set of approval states is chosen to make the receiver exactly indifferent between approval and rejection, thus binding her obedience constraint. The resulting sender-preferred set of approval states $\bar{W}$, which also solves the sender's problem in the Bayesian persuasion game, is characterized in the following theorem.

THEOREM 2. *When $[I] = \{1\}$, the sender-preferred equilibrium outcome is a commitment outcome. The sender-preferred set of approval states $\bar{W}$ is characterized by a cutoff value $c^* < 0$ such that*

- *sender's proposal is approved if $\delta(\omega) > c^*$ and rejected if $\delta(\omega) < c^*$;*
- *whenever the sender's proposal is approved, the receiver's expected net payoff of approval equals zero: $\mathbb{E}_p\big[\delta(\omega) \mid \bar{W}\big] = 0$.*

Recall that an outcome specifies probability $\alpha(\omega)$ that the receiver approves the proposal in state ω. According to Kamenica and Gentzkow (2011), when solving the sender's problem in the game of Bayesian persuasion, one may restrict attention to *straightforward* signals with realizations that are interpreted by the receiver as recommendations to take particular actions. Consequently, $\alpha(\omega)$ becomes the probability that the recommendation to approve came from state ω, and the sender's problem rewrites itself in terms of finding the optimal outcome subject to an obedience-like constraint of the receiver.

Because the receiver ranks states according to her net payoff of approval $\delta(\omega)$, the sender-preferred outcome of the Bayesian persuasion game features a set of cutoff states. To see why, observe that by adding state ω to the set of approval states W, one increases the value of the objective by $p(\omega)$, at the cost (in terms of the obedience constraint) of $-\delta(\omega) \cdot p(\omega)$. States that are approved under complete information have a non-negative net payoff of approval (and thus a negative cost). These are added first, automatically satisfying the sender's (IC) constraint. If some state ω such that $\delta(\omega) < 0$ is added, it must be the case that every state with a higher (but possibly also negative) net payoff of approval is added as well, since those states are "cheaper" in terms of the constraint. States are added until the obedience constraint of the receiver binds, or until $\bar{W} = \Omega$.

While the logic above applies to both an equilibrium and a commitment outcome, it is possible that a commitment outcome is not deterministic. Suppose that there is a set of states from which the sender is sending a mixed recommendation, thus inducing both actions with positive probabilities. By the logic above, all those states must provide the same net payoff of approval to the receiver, meaning that these states have the same cost in terms of the obedience constraint. Rather than sending a mixed recommendation, the sender can split that set into two subsets and recommend one action on each of those subsets with certainty. Continuity of the state space allows him to perform such partitioning without altering the constraints or the value of the objective function. Now that the commitment outcome is also deterministic, it coincides with the sender-preferred equilibrium outcome because the receiver's obedience constraint is the same with and without commitment and imposes that the receiver correctly interprets the *straightforward signal* or the *convincing message* as a recommendation to take the appropriate action.

EXAMPLE 2 (DICTATORIAL ELECTIONS). Consider the setting from Example 1 with just one voter (call him the dictator), convincing whom is necessary and sufficient for the challenger to win the election. According to Theorem 1, one can restrict attention to direct equilibria, which are characterized by the set of approval states W.

In the present context, W is the <u>set of winning policies</u> for the challenger as well as the <u>message that convinces</u> the dictator to vote for the challenger. To be induced in a direct equilibrium, W has to satisfy two conditions. First, there is the sender's (IC) constraint $\mathcal{A} \subseteq W$, which ensures that the challenger does not have a profitable deviation toward fully revealing his policy. Second, there is the receiver's (obedience) constraint $p(\cdot \mid W) \in \mathcal{B}$, which ensures that the dictator interprets message W as a recommendation to vote for the challenger.

In the sender-worst equilibrium, the set of winning policies coincides with the dictator's approval set, $\underline{W} = \mathcal{A}$. Since the voter effectively learns all the information relevant for her decision, i.e. whether the challenger's policy is preferred to the status quo under complete information, this equilibrium outcome is equivalent to *full disclosure*.

In the sender-preferred equilibrium, the challenger's odds of winning $P(W)$ are maximized subject to the receiver's (obedience) constraint. According to Theorem 2, the resulting sender-preferred set of winning policies $\bar{W}$ is the same as in the game of Bayesian persuasion, implying that the challenger need not benefit from having commitment power. The sender-preferred set of winning policies is characterized by a cutoff value c^* of the dictator's net payoff of voting for the challenger, such that

- challenger with policy ω is elected if and only if $\delta(\omega) \geq c^*$;
- $c^* < 0$: some challengers with policies that the dictator considers worse than the status quo are elected;
- when voting for the challenger, the dictator expects the challenger's policy to be exactly as far from her bliss point as the status quo: $\mathbb{E}_p\left[|v - \omega| \,\middle|\, \delta_v(\omega) \geq c^*\right] = |v - \omega_0|$.

This solution is illustrated in Figure 2.

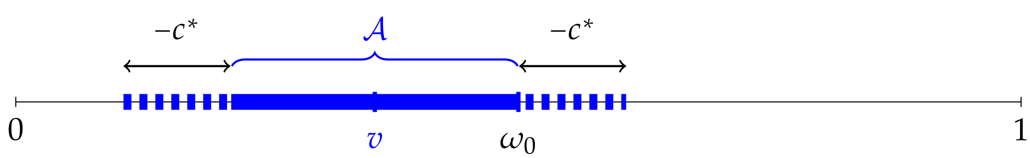


Figure 2: *the challenger-preferred set of winning policies $\bar{W}$ (in blue) consists of the dictator's approval set (solid area) and policies outside of the approval set (dashed area).*

4. Many Receivers

Recall that with one receiver, every equilibrium outcome is deterministic, because the sender has a strict incentive to convince the receiver in every state. As a result, Theorem 1 established the equivalence between the sets of direct equilibrium outcomes and all equilibrium outcomes. With many receivers, there may exist non-deterministic equilibria, which means that the outcome equivalence breaks down. However, an equilibrium may only be non-deterministic when the sender does not have a strict incentive to induce some receiver's approval in some states. Consequently, to characterize the range of the sender's ex-ante utility, one may restrict attention to direct equilibria.

DIRECT IMPLEMENTATION

THEOREM 3. *The following statements about the sender's ex-ante utility $\bar{u}_s$ are equivalent:*

(1) $\bar{u}_s$ *is reached in equilibrium;*

(2) $\bar{u}_s$ *is reached in a direct equilibrium;*

(3) $\bar{u}_s$ *is given by*

$$\bar{u}_s = \int_{\Omega} u_s\big(\{i \in [I] \mid \omega \in W_i\}\big) \cdot p(\omega) d\omega,$$

where for every receiver $i \in [I]$, $W_i \subseteq \Omega$ *is her set of approval states, which satisfies*

- *receiver's obedience constraint constraint* $p(\cdot \mid W_i) \in \mathcal{B}_i$*;*
- *sender's incentive-compatibility constraint* $\mathcal{A}_i \subseteq W_i$.

When studying the multiple-receiver case, I rely on the implicit assumptions of *(i)* no information spillovers between the receivers, and *(ii)* receivers' utilities being independent of other receivers' actions. Receivers are essentially solving independent utility maximization problems and are thus considered individually. The proof of the theorem follows the same steps as the proof of Theorem 1. The only substantial difference arises in proving that *(1)* implies *(3)*, because not all equilibria are deterministic. If an equilibrium outcome is not deterministic, then the sender induces multiple actions of some receiver i in some state ω, and it is unclear whether ω should be added to the approval recommendation W_i or not. Notice that if the sender's utility in state ω strictly increases in receiver i's action, then the sender would never convince her with probability between 0 and 1 in equilibrium: once a receiver is "convincible," the sender should be convincing her with certainty. Consequently, removing all the states in which the sender makes a mixed recommendation to receiver i from this receiver's set of approval states is inconsequential to the sender's ex-ante utility.

EQUILIBRIUM RANGE AND VALUE OF COMMITMENT

According to Theorem 3, when characterizing the sender's equilibrium ex-ante utility, one can restrict attention to collections of sets of approval states $(W_1, \ldots, W_I)$, each of which satisfies two constraints. Moreover, the sender's ex-ante utility only depends on the sets of approval states and the prior distribution.

Once again, in the sender-worst equilibrium, in which the sender's ex-ante utility is minimized across all equilibria, the sender does as well as under full disclosure. The set of approval states of receiver $i \in [I]$ is $\underline{W}_i = \mathcal{A}_i$, each receiver learns all the relevant information, and she makes her decision as if under complete information.

The sender-preferred (direct) equilibrium outcome is characterized by the collection of sets of approval states that maximizes the sender's ex-ante utility across all equilibria, i.e. subject to every receiver's obedience constraint and every incentive-compatibility constraint of the sender. When there are many receivers, the sender need not benefit from having commitment power, either.

THEOREM 4. *The sender-preferred equilibrium outcome is a commitment outcome.*

Once again, the proof boils down to showing that there exists a deterministic commitment outcome, and that every receiver's set of approval beliefs satisfies the sender's incentive-compatibility constraint. The former is possible because of the continuity of the state space, and the latter is true because adding state $\omega \in \mathcal{A}_i$ to the induced set of approval states loosens receiver i's obedience constraint and weakly increases the value of the objective.

The problem of finding the sender-preferred equilibrium is computationally hard when the sender's preferences are not separable in receivers' actions.[8] In some special cases, particularly when the state space is binary, solutions to the problem under commitment are available. Most notably, Arieli and Babichenko (2019) solve the case of the supermodular utility of the sender, and Babichenko and Barman (2016) show that when the sender's utility is submodular, the problem is NP-hard and provide an approximation. In the following section, I return to the setup of the spatial model of elections. That setup is simple enough to characterize the solution, make a meaningful comparison between public and private disclosure, and analyze comparative statics.

[8]If sender's utility is separable in receivers actions, then the sender can determine the optimal signal receiver by receiver and faces a set of independent problems of a single-receiver variety, as observed by Kamenica (2019).

5. Targeted Advertising in Elections

We now revisit the setting of Example 1. Suppose there are two voters and the outcome of the election is decided by unanimity rule: the challenger gets a payoff of 1 if and only if he convinces both voters to vote in his favor. With a slight abuse of notation, I let $[I] = \{L, R\}$ to reflect that the voters L and R are located on the opposite sides of the status quo, $v_L < \omega_0 < v_R$. Notice that L and R are incompatible in that their approval sets and sets of approval beliefs do not intersect:

Remark 1. *Approval sets and sets of approval beliefs of voters L and R with ideal policies $v_L < \omega_0 < v_R$ do not intersect a.s.*

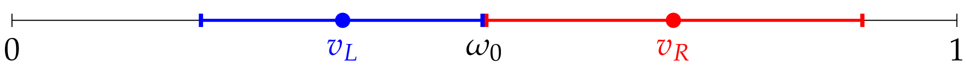


Figure 3: *voters L and R are incompatible and their approval sets do not intersect a.s.*

An interesting implication of Remark 1 is that the election under consideration is going to be unwinnable for the challenger when he cannot communicate privately and is restricted to sending a public message to all voters. Formally,

Definition 6. *An election is unwinnable for the challenger under public disclosure if his ex-ante odds of winning are zero in every equilibrium.*

The voters have a common prior, and if they hear a public message (or signal), they all form the same posterior. But that belief is not in both voters' set of approval beliefs a.s.! Note that this observation is not limited to elections with just two voters. *Any election in which two voters on the opposite sides of the status quo are jointly pivotal is unwinnable for the challenger under public disclosure.*

Swinging unwinnable elections

According to Theorem 3, when characterizing the sender's ex-ante utility, one may restrict attention to a pair of sets of convincing policies (W_L, W_R), one for each voter, where W_i is also the message that convinces voter i (to vote for the challenger). Each W_i must satisfy receiver i's (obedience) and the sender's (IC) constraints. The set of winning policies, in which the sender wins the election by securing both votes, is $W = W_L \cap W_R$.

The sender-worst equilibrium is outcome-equivalent to full disclosure and the message that convinces voter i is simply voter i's approval set, $\underline{W}_i = \mathcal{A}_i$, $i \in \{L, R\}$. By Remark 1, the voters' approval sets do not intersect, implying that the challenger's ex-ante odds of winning are zero.

In the sender-preferred equilibrium, the outcome partition maximizes to sender's odds of winning subject to each voter's (obedience) constraint:

$$\max_{W_L, W_R} P(W_L \cap W_R)$$

$$\text{subject to } p(\cdot \mid W_i) \in \mathcal{B}_i, \text{ for } i \in \{L, R\}.$$

The solution to this problem is described in the following theorem.

THEOREM 5. *In the sender-preferred equilibrium of an election with voters L and R whose ideal policies are $v_L < \omega_0 < v_R$,*

- *message $\bar{W}_i$ that convinces voter $i \in \{L, R\}$ is an interval $[a_i, b_i] \supset \mathcal{A}_i$;*
- *challenger has positive ex-ante odds of winning this election; the set of winning policies is $[a_R, b_L]$ with $a_R < \omega_0 < b_L$.*

To understand the intuition behind this result, recall that when voter L (R) is a dictator, the challenger can convince her to elect him even when his policy is slightly to the right (left) of the status quo. This notion was illustrated in Figure 2 of Example 2. One thing that the challenger can do when private communication is allowed is treat each voter as a dictator. If he does that, he convinces both voters at the same time provided that his policy is close enough to the status quo. At this point, the challenger is already winning an unwinnable election with a positive probability. However, he can do even better. Rather than sending to each voter a message that is centered around this voter's bliss point, skew her message toward the other voter. More precisely, start with voter L's (R's) approval set and then add policies to the right (left) of the status quo, until the voter's obedience constraint binds. This solution is illustrated in Figure 4.

VOTER REGRET

Voter i prefers to vote for the challenger under complete information if and only if $\omega \in \mathcal{A}_i$. When she votes for the challenger, and later finds out that his policy is $\omega \in W_i \setminus \mathcal{A}_i$, she

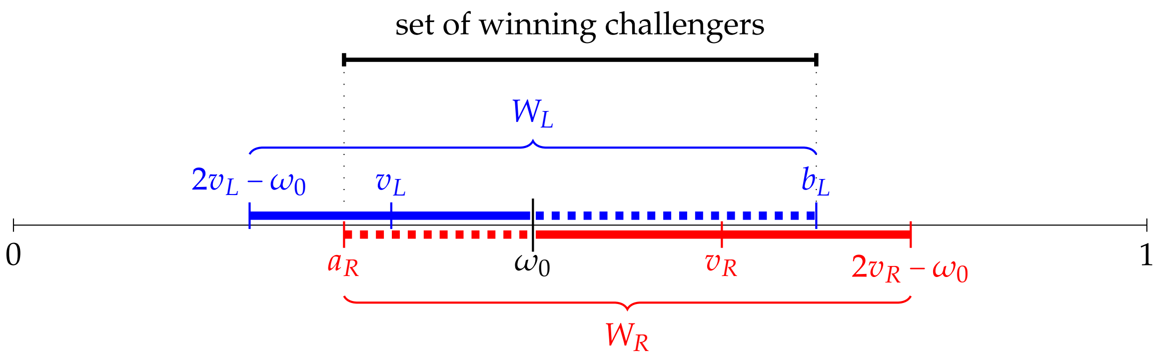


Figure 4: *sender-preferred messages W_L (in blue) and W_R (in red) that convince voters L and R. W_i consists of voter i's approval set (solid) and policies preferred by $j \neq i$ (dashed). Challenger wins the election by convincing both voters when his policy is in $W = W_L \cap W_R$ and sufficiently close to the status quo, $\omega_0 \in W$.*

realizes she made a mistake and regrets it. The amount of regret is the ex-ante measure of how often the voter is "tricked" into voting for the challenger.

DEFINITION 7. *The region of regret of voter $i \in [I]$ when she is convinced by message $W_i \subseteq \Omega$ is $W_i \setminus \mathcal{A}_i$ and the amount of regret is $P(W_i \setminus \mathcal{A}_i)$.*

COROLLARY 1. *In the sender-preferred equilibrium with voters L and R with ideal policies $v_L < \omega_0 < v_R$, both voters experience a positive amount of regret. Their total amount of regret coincides with the challenger's ex-ante odds of winning.*

COMPARATIVE STATICS

If at least one of the obedience constraints does not bind when $b_L = 2v_R - \omega_0$ or when $a_R = 2v_L - \omega_0$, then the solution may look different, although both convincing messages will still be intervals. To better understand comparative statics, assume for the rest of this section that the prior belief is uniform.[9] Notice that the distance from a voter's bliss point to the status quo measures how persuadable this voter is.

DEFINITION 8. *Assume uniform prior. Then,*

- *voter i is more persuadable than voter j if $|v_i - \omega_0| > |v_j - \omega_0|$, for $i \in \{L', R'\}$ and $j \in \{L, R\}$;*
- *electorate $\{L', R'\}$ is more polarized than electorate $\{L, R\}$ if $v'_L \leq v_L < \omega_0 < v_R \leq v'_R$, with at least one of the non-strict inequalities being strict.*

[9] The prior is chosen to be uniform for ease of exposition. Similar results hold for any prior distribution.

In words, voter L' is more persuadable than voter R (voter L) if her bliss point is located further away from the status quo than the bliss point of R (L). I say voter i becomes more persuadable if $|v_i - \omega_0|$ increases. Of course, according to the definitions above, when either voter becomes more persuadable, the electorate becomes more polarized. Figure 5 illustrates the dynamics of the numerical solution to the problem of finding sender-preferred equilibrium as voter R becomes more persuadable (voters L and R become more polarized). Theorem 6 summarizes the comparative statics.

THEOREM 6. *Assume uniform prior. In the sender-preferred equilibrium of an election with voters L and R whose ideal policies are $v_L < \omega_0 < v_R$,*

- *as v_R becomes more persuadable, the challenger's odds of winning and the total amount of regret increase;*
- *suppose $|v_L| = |v_R|$, meaning that neither voter is more persuadable than the other. Then, as v_R becomes more persuadable,*
  - *the set of winning policies $\bar{W} = [a_R, b_L]$ shifts to the left, i.e. a_R and b_L decrease;*
  - *the amount of regret of each voter increases.*

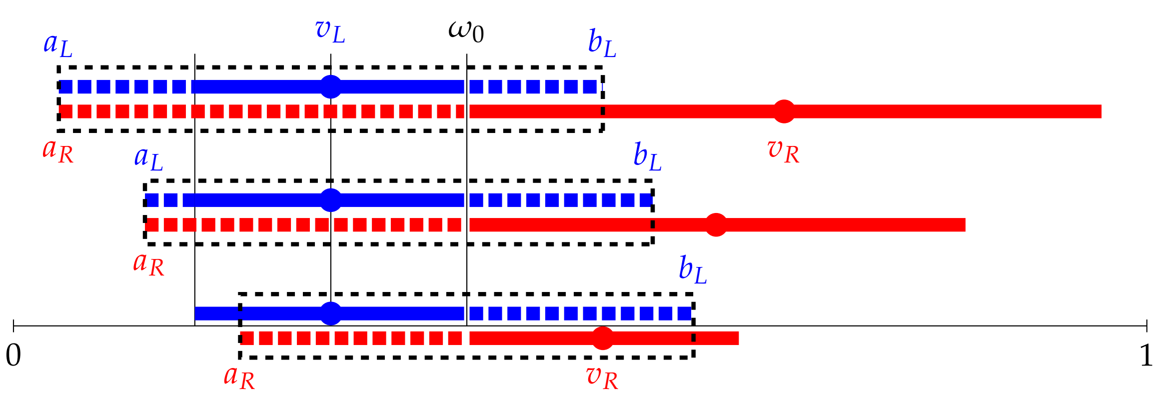


Figure 5: *comparative statics as voter R moves to the right (from bottom to top): her approval set (solid red) expands, thus making her more persuadable; regions of regret (dashed line), convincing messages (solid plus dotted lines), and winning states (black dashed area) move to the left.*

In words, as voter R becomes more persuadable, it becomes easier for the challenger to swing the election by targeting, and his ex-ante odds of winning increase. This comes, of course, at the cost of both voters making mistakes more often, thus increasing the amount of regret for both of them. When voter R is so far to the right that her obedience constraint never binds (as in the top line of Figure 5), the challenger-preferred set of winning states is maximized subject to voter L's obedience constraint only. In that case, interestingly, voter L's amount of regret is the largest.

Overall, this application makes a straightforward prediction that targeted advertising is bad for the voters in that it causes them to regret their choices with positive probability. As the electorate becomes more polarized, things get worse. As the "right" voter moves further to the right, challenger's odds of swinging this election increase; each voter's amount of regret increases; the set of winning states shifts to the left.

6. Conclusion

This paper argued that the sender need not benefit from having commitment power, and can persuade the receivers with verifiable information only. This result is useful in applications, especially in the context of elections, where assuming that the sender has hard evidence to back up his claims is more plausible than assuming that the sender can commit to a signal ex-ante.

While illustrated in the simplified framework, the observation that targeted advertising helps challenger swing elections holds for more than two voters, more than one dimension and for social choice rules other than the unanimity rule. Because targeting leads to election outcomes that are different from the complete-information outcomes, one can argue that targeted advertising is bad for democracy. Certain policy implications, especially concerning restricting the collection and use of personal data by the candidates in their electoral campaigns, should be considered.

References

Alonso, Ricardo and Odilon Câmara (2016), "Persuading Voters", *American Economic Review*, 106, 11 (Nov. 2016), pp. 3590-3605. (p. 6.)

Arieli, Itai and Yakov Babichenko (2019), "Private Bayesian Persuasion", *Journal of Economic Theory*, 182 (July 2019), pp. 185-217. (p. 18.)

Artabe, Alaitz and Javier Gardeazabal (2014), "Strategic Votes and Sincere Counterfactuals", *Political Analysis*, 22, 2 (Jan. 2014), pp. 243-257. (p. 8.)

AUMANN, ROBERT J., YITZHAK KATZNELSON, ROY RADNER, ROBERT W. ROSENTHAL, and BENJAMIN WEISS (1983), "Approximate Purification of Mixed Strategies", *Mathematics of Operations Research*, 8, 3 (Aug. 1983), pp. 327-341. (p. 12.)

BABICHENKO, YAKOV and SIDDHARTH BARMAN (2016), "Computational Aspects of Private Bayesian Persuasion", pp. 1-16. (p. 18.)

BARDHI, ARJADA and YINGNI GUO (2018), "Modes of Persuasion toward Unanimous Consent", *Theoretical Economics*, 13, 3, pp. 1111-1149. (p. 6.)

BOLESLAVSKY, RAPHAEL and CHRISTOPHER COTTON (2015), "Grading Standards and Education Quality", *American Economic Journal: Microeconomics*, 7, 2 (May 2015), pp. 248-279. (p. 6.)

BRENNAN, GEOFFREY and ALAN HAMLIN (1998), "Expressive Voting and Electoral Equilibrium", *Public Choice*, 95, 1-2, pp. 149-175. (p. 8.)

BRENNAN, GEOFFREY and LOREN LOMASKY (1993), *Democracy and Decision: The Pure Theory of Electoral Preference*, Cambridge University Press. (p. 8.)

CRAWFORD, VINCENT P. and JOEL SOBEL (1982), "Strategic Information Transmission", *Econometrica*, 50, 6 (Nov. 1982), p. 1431. (p. 5.)

DEGAN, ARIANNA and ANTONIO MERLO (2007), "Do Voters Vote Sincerely?", *NBER Working Paper 12922*. (p. 8.)

DELLAVIGNA, STEFANO and MATTHEW GENTZKOW (2010), "Persuasion: Empirical Evidence", *Annual Review of Economics*, 2, 1 (Sept. 2010), pp. 643-669. (p. 6.)

DELLAVIGNA, STEFANO and ETHAN KAPLAN (2007), "The Fox News Effect: Media Bias and Voting", *The Quarterly Journal of Economics*, 122, 3 (Aug. 2007), pp. 1187-1234. (p. 6.)

DOWNS, ANTHONY (1957), "An Economic Theory of Political Action in a Democracy", *Journal of Political Economy*, 65, 2 (Apr. 1957), pp. 135-150. (p. 8.)

EGOROV, GEORGY and KONSTANTIN SONIN (2019), "Persuasion on Networks", *SSRN Electronic Journal*. (p. 6.)

ENIKOLOPOV, RUBEN, MARIA PETROVA, and EKATERINA ZHURAVSKAYA (2011), "Media and Political Persuasion: Evidence from Russia", *American Economic Review*, 101, 7 (Dec. 2011), pp. 3253-3285. (p. 6.)

FELSENTHAL, DAN S. and AVRAHAM BRICHTA (1985), "Sincere and Strategic Voters: An Israeli Study", *Political Behavior*, 7, 4, pp. 311-324. (p. 8.)

FORGES, FRANÇOISE (1986), "An Approach to Communication Equilibria", *Econometrica*, 54, 6 (Nov. 1986), p. 1375. (p. 13.)

GEHLBACH, SCOTT and KONSTANTIN SONIN (2014), "Government Control of the Media", *Journal of Public Economics*, 118 (Oct. 2014), pp. 163-171. (p. 6.)

GEORGE, LISA M and JOEL WALDFOGEL (2006), "The New York Times and the Market for Local Newspapers", *American Economic Review*, 96, 1 (Feb. 2006), pp. 435-447. (p. 6.)

GROSSMAN, SANFORD J. (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality", *The Journal of Law and Economics*, 24, 3 (Dec. 1981), pp. 461-483. (p. 5.)

HAMLIN, ALAN and COLIN JENNINGS (2011), "Expressive Political Behaviour: Foundations, Scope and Implications", *British Journal of Political Science*, 41, 3 (July 2011), pp. 645-670. (p. 8.)

HU, LIN, ANQI LI, and ILYA SEGAL (2019), "The Politics of Personalized News Aggregation", *Working Paper* (Oct. 2019). (p. 6.)

KAMENICA, EMIR (2019), "Bayesian Persuasion and Information Design", *Annual Review of Economics*, 11, pp. 249-272. (p. 18.)

KAMENICA, EMIR and MATTHEW GENTZKOW (2011), "Bayesian Persuasion", *American Economic Review*, 101, 6 (Oct. 2011), pp. 2590-2615. (pp. 3, 5, 14, 28, 32.)

KAN, KAMHON and C. C. YANG (2001), "On Expressive Voting: Evidence from the 1988 U.S. Presidential Election", *Public Choice*, 108, 3-4, pp. 295-312. (p. 8.)

KOLOTILIN, ANTON (2015), "Experimental Design to Persuade", *Games and Economic Behavior*, 90 (Mar. 2015), pp. 215-226. (p. 6.)

LIPNOWSKI, ELLIOT (2020), "Equivalence of Cheap Talk and Bayesian Persuasion in a Finite Continuous Model", *note*. (p. 5.)

LIPNOWSKI, ELLIOT and DORON RAVID (2020), "Cheap Talk With Transparent Motives", *Econometrica*, 88, 4, pp. 1631-1660. (p. 5.)

MILGROM, PAUL R. (1981), "Good News and Bad News: Representation Theorems and Applications", *The Bell Journal of Economics*, 12, 2, p. 380. (p. 5.)

MILGROM, PAUL R. and JOHN ROBERTS (1986), "Relying on the Information of Interested Parties", *The RAND Journal of Economics*, 17, 1, p. 18. (p. 7.)

MILGROM, PAUL R. and ROBERT J. WEBER (1985), "Distributional Strategies for Games with Incomplete Information", *Mathematics of Operations Research*, 10, 4 (Nov. 1985), pp. 619-632. (p. 12.)

MYERSON, ROGER B. (1986), "Multistage Games with Communication", *Econometrica*, 54, 2 (Mar. 1986), p. 323. (p. 13.)

OBERHOLZER-GEE, FELIX and JOEL WALDFOGEL (2009), "Media Markets and Localism: Does Local News en Español Boost Hispanic Voter Turnout?", *American Economic Review*, 99, 5 (Dec. 2009), pp. 2120-2128. (p. 6.)

OSTROVSKY, MICHAEL and MICHAEL SCHWARZ (2010), "Information Disclosure and Unraveling in Matching Markets", *American Economic Journal: Microeconomics*, 2, 2 (May 2010), pp. 34-63. (p. 6.)

PEREGO, JACOPO and SEVGI YUKSEL (2018), "Media Competition and Social Disagreement", *Working Paper*. (p. 6.)

PRAT, ANDREA and DAVID STRÖMBERG (2013), "The Political Economy of Mass Media", *Advances in Economics and Econometrics*, Cambridge University Press, pp. 135-187. (p. 6.)

PRUMMER, ANJA (2020), "Micro-Targeting and Polarization", *Journal of Public Economics*, 188 (Aug. 2020), p. 104210. (p. 6.)

RADNER, ROY and ROBERT W. ROSENTHAL (1982), "Private Information and Pure-Strategy Equilibria", *Mathematics of Operations Research*, 7, 3 (Aug. 1982), pp. 401-409. (p. 12.)

ROMANYUK, GLEB and ALEX SMOLIN (2019), "Cream Skimming and Information Design in Matching Markets", *American Economic Journal: Microeconomics*, 11, 2 (May 2019), pp. 250-276. (p. 6.)

SPENCE, MICHAEL (1973), "Job Market Signaling", *The Quarterly Journal of Economics*, 87, 3 (Aug. 1973), p. 355. (p. 5.)

APPENDIX: OMITTED PROOFS

LEMMA A.1. *In equilibrium $(\sigma, \boldsymbol{a}, \boldsymbol{q})$, for every receiver $i \in [I]$, let*

- $\mathcal{M}_i := \{m \in \mathbb{M} \mid a_i(m) = 1\}$ *be the set of all messages that convince this receiver to approve the proposal;*
- $\mathcal{W}_i := \{\omega \in \Omega \mid \exists m \in \mathcal{M}_i \text{ s.t. } \omega \in m\}$ *be the set of all states in which at least one convincing message is available. Note that $\mathcal{A}_i \subseteq \mathcal{W}_i$: if $\omega \in \mathcal{A}_i$, then the fully revealing message $\{\omega\} \in \mathcal{M}_i$ because $a_i(\{\omega\}) = 1$;*
- $X_i := \{\omega \in \Omega \mid \alpha_i(\omega) = 1\} \subseteq \mathcal{W}_i$ *be the set of states in which this receiver approves the proposal with probability 1. To convince the receiver in state ω with certainty, the sender must be sending her convincing messages, and convincing messages only, i.e.*

$$\forall \omega \in X_i, \quad \int_{\mathcal{M}_i} \sigma_i(m \mid \omega) dm = 1.$$

Claim: set $X_i \cup \mathcal{A}_i$ satisfies receiver i's (obedience) constraint $p(\cdot \mid X_i \cup \mathcal{A}_i) \in \mathcal{B}_i$.

Proof. Every message in $\mathcal{M}_i$ convinces the receiver to approve the proposal:

$$\forall m \in \mathcal{M}_i, \ \ q_i(\cdot \mid m) \in \mathcal{B}_i \iff \int_{\text{supp } q_i(\cdot \mid m) \subseteq m \subseteq \mathcal{W}_i} \delta_i(\omega) \cdot q_i(\omega \mid m) d\omega \geq 0,$$

which, on the equilibrium path, becomes

$$\int_{\mathcal{W}_i} \delta_i(\omega) \cdot \frac{\sigma_i(m \mid \omega) \cdot p(\omega)}{\int_{\mathcal{W}_i} \sigma_i(m \mid \omega') \cdot p(\omega') d\omega'} d\omega \geq 0 \iff \int_{\mathcal{W}_i} \delta_i(\omega) \cdot \sigma_i(m \mid \omega) \cdot p(\omega) d\omega \geq 0.$$

Integrate the above inequality over all $m \in \mathcal{M}_i$:

$$\int_{\mathcal{M}_i} \int_{\mathcal{W}_i} \delta_i(\omega) \cdot \sigma_i(m \mid \omega) \cdot p(\omega) d\omega dm \geq 0.$$

Next, partition $\mathcal{W}_i$ into X_i, $\mathcal{A}_i$, and $\mathcal{W}_i \setminus (X_i \cup \mathcal{A}_i)$ and observe that

$$\int_{\mathcal{M}_i}\int_{X_i} \delta_i(\omega)\cdot\sigma_i(m \mid \omega)p(\omega)d\omega dm = \int_{X_i} \delta_i(\omega)p(\omega)\underbrace{\int_{\mathcal{M}_i} \sigma_i(m \mid \omega)dm}_{=1,\forall\omega\in X_i} d\omega = \int_{X_i} \delta_i(\omega)p(\omega)p\omega;$$

$$\int_{\mathcal{M}_i}\int_{\mathcal{A}_i} \delta_i(\omega)\sigma_i(m \mid \omega)p(\omega)d\omega dm = \int_{\mathcal{A}_i} \underbrace{\delta_i(\omega)}_{\geq 0\forall\omega\in\mathcal{A}_i} p(\omega)\underbrace{\int_{\mathcal{M}_i} \sigma_i(m \mid \omega)dm}_{\leq 1} d\omega \leq \int_{\mathcal{A}_i} \delta_i(\omega)p(\omega)p\omega;$$

$$\int_{\mathcal{M}_i}\int_{\mathcal{W}_i\setminus(X_i\cup\mathcal{A}_i)} \underbrace{\delta_i(\omega)}_{\leq 0\forall\omega\notin\mathcal{A}_i} \sigma_i(m \mid \omega)p(\omega)d\omega dm \leq 0.$$

Consequently,

$$\int_{\mathcal{M}_i}\int_{\mathcal{W}_i} \delta_i(\omega)\cdot\sigma_i(m \mid \omega)\cdot p(\omega)d\omega dm \geq 0 \Longrightarrow \int_{X_i\cup\mathcal{A}_i} \delta_i(\omega)p(\omega)p\omega \geq 0 \iff p(\cdot \mid X_i \cup \mathcal{A}_i) \in \mathcal{B}_i.$$

□

Proof of Theorem 1 (See page 12).

Theorem 1. *Suppose $[I] = \{1\}$. Then, every equilibrium outcome is deterministic. Moreover, the following statements about $W \subseteq \Omega$ are equivalent:*

(1) W is an equilibrium set of approval states;

(2) W is a set of approval states in a direct equilibrium;

(3) W satisfies the receiver's (obedience) *and the sender's* (IC) *constraints.*

Proof. Part I, every equilibrium outcome is deterministic. Consider equilibrium (σ, a, q). Suppose, on the contrary, that $\alpha(\omega) \in (0,1)$ for some $\omega \in \Omega$. In words, in this state, the sender is inducing both approval and rejection with positive probabilities. In this scenario, the sender has a profitable deviation to exclusively sending the messages that induce approval.

Part II, (1) ⇒ (3): consider equilibrium (σ, a, q) with set of approval states W. W must satisfy the sender's (IC) constraint, or else the sender can deviate to full disclosure.

Furthermore, since the sender's utility strictly increases in receiver's action, $\alpha(\omega) > 0$ implies $\alpha(\omega) = 1$, or else the sender could deviate to one of the available convincing messages. In terms of Lemma A.1, the set of all states in which at least one convincing message is available is simply W, so one can conclude that set W satisfies the (obedience) constraint.

Part III, *(3) ⇒ (2)*: assume that $W \subseteq \Omega$ satisfies (IC) and (obedience). To complete the characterization of the direct equilibrium that induces this set of approval states, one needs to specify off-path beliefs of the receiver, i.e. following any message $m \in \mathbb{M} \setminus \{W, \Omega \setminus W\}$. There are two restrictions on these beliefs: *(i)* the density must be supported on the set of states in which the message was available to the sender: $\forall m \in \mathbb{M}$, $\operatorname{supp} q^D(\cdot \mid m) \subseteq m$, *(ii)* the sender may not have profitable deviations in states $\Omega \setminus W$. One way to ensure that does not happen is to impose "skeptical beliefs"

$$\forall m \subseteq \mathcal{A}, \ \ \operatorname{supp} q^D(\cdot \mid m) \subseteq m, \text{so that } q^D(\cdot \mid m) \in \mathcal{B},$$

$$\forall m \nsubseteq \mathcal{A}, m \neq \mathcal{W} \ \ \operatorname{supp} q^D(\cdot \mid m) \subseteq m \setminus \mathcal{A}, \text{so that } q^D(\cdot \mid m) \notin \mathcal{B}$$

that assign positive probability to states within the approval set if and only if the message comprises of these states only.

Part IV, *(2) ⇒ (1)* is trivially true. □

PROOF OF THEOREM 2 (SEE PAGE 14).

THEOREM 2. *When $[I] = \{1\}$, the sender-preferred equilibrium outcome is a commitment outcome. The sender-preferred set of approval states $\bar{W}$ is characterized by a cutoff value $c^* < 0$ such that*

- *sender's proposal is approved if $\delta(\omega) > c^*$ and rejected if $\delta(\omega) < c^*$;*
- *whenever the sender's proposal is approved, the receiver's expected net payoff of approval equals zero: $\mathbb{E}_p\big[\delta(\omega) \mid \bar{W}\big] = 0$.*

Proof. Consider the problem of finding the optimal commitment protocol $(\sigma^{BP}, a^{BP}, q^{BP})$. According to Kamenica and Gentzkow (2011), that problem may be simplified to finding an optimal *straightforward* experiment σ^{BP} that is supported on set $\{s^+, s^-\}$, where s^+ induces posterior $q^+ \in \mathcal{B}$ and recommends that the receiver approves the sender's proposal and s^- induces posterior $q^- \notin \mathcal{B}$ and recommends rejection. The outcome takes form of $\alpha(\omega) = \text{Prob}(s^+ \mid \omega)$, and the sender's problem under commitment becomes

$$\max_{\alpha} \int_{\Omega} \alpha(\omega) p(\omega) d\omega, \quad \text{subject to}$$

- $\forall \omega \in \Omega$, $0 \le \alpha(\omega) \le 1$;
- the receiver's obedience constraint $q^+ \in \mathcal{B}$, which becomes

$$\int_{\Omega} \delta(\omega) \cdot q^+(\omega) d\omega = \int_{\Omega} \delta(\omega) \cdot \frac{\alpha(\omega) \cdot p(\omega)}{\int_{\Omega} \alpha(\omega') \cdot p(\omega') d\omega'} d\omega \ge 0$$

$$\iff \int_{\Omega} \delta(\omega) \cdot \alpha(\omega) \cdot p(\omega) d\omega \ge 0.$$

Observe that $\forall \omega \in \mathcal{A}$, $\alpha(\omega) = 1$, or else increasing $\alpha(\omega)$ would relax the constraint (because $\delta(\omega) \ge 0$) and increase the objective function. Next, if the receiver approves the proposal under the prior belief, then the optimal experiment features no disclosure, i.e. $p \in \mathcal{B}$ implies $\alpha(\omega) = 1$ for every $\omega \in \Omega$. If $p \notin \mathcal{B}$, then the receiver's obedience constraint has to bind, or else the value of the objective can be increased. Assume for the rest of this proof that $p \notin \mathcal{B}$.

If $\alpha(\omega) \in (0,1)$ for some $\omega \in \Omega$, then $\alpha(\omega') = 1$ for all $\omega' \in \Omega$ such that $\delta(\omega') > \delta(\omega)$, since these types are "cheaper" in terms of the obedience constraint and hence should be added the set of approved states first. Consequently, there must exist a threshold value $c^* < 0$ of the receiver's net payoff of approving the sender's proposal, and $\delta(\omega) = c^*$ for all ω such that $\alpha(\omega) \in (0,1)$. Let $\bar{\mathcal{D}} := \{\omega \in \Omega \text{ s.t. } \delta(\omega) \ge c^*\}$; $\mathcal{D} := \{\omega \in \Omega \text{ s.t. } \delta(\omega) > c^*\}$; $\partial\mathcal{D} := \bar{\mathcal{D}} \setminus \mathcal{D} = \{\omega \in \Omega \text{ s.t. } \delta(\omega) = c^*\}$.

By the argument above, the solution $\alpha(\cdot)$ takes form

$$\alpha(\omega) = \begin{cases} 1, & \omega \in \mathcal{D}; \\ (0,1), & \omega \in \partial\mathcal{D}; \\ 0, & \omega \in \Omega \setminus \bar{\mathcal{D}}. \end{cases}$$

The (binding) obedience constraint is thus

$$\int_{\mathcal{D}} \delta(\omega)\cdot p(\omega)d\omega + \int_{\partial\mathcal{D}} \alpha(\omega)\cdot \underbrace{\delta(\omega)}_{=c^*,\forall\omega\in\partial\mathcal{D}} \cdot p(\omega)d\omega = 0.$$

If measure of set $\partial\mathcal{D}$ is not zero, partition it into two sets, $X \subseteq \partial\mathcal{D}$ and $Y = \partial\mathcal{D} \setminus X$. Let $\tilde{\alpha}(\omega) = \mathbb{1}(\omega \in X)$, where X solves

$$\int_{\partial\mathcal{D}} \alpha(\omega)\cdot p(\omega)d\omega = \int_{\partial\mathcal{D}} \tilde{\alpha}(\omega)\cdot p(\omega)d\omega = P(X).$$

By changing $\alpha(\omega)$ to $\tilde{\alpha}(\omega)$ on $\partial\mathcal{D}$, the objective function of the sender does not change:

$$\int_{\partial\mathcal{D}} \alpha(\omega)p(\omega)d\omega = \int_{\partial\mathcal{D}} \tilde{\alpha}(\omega)p(\omega)d\omega = P(X);$$

and the obedience constraint does not change, either:

$$\int_{\partial\mathcal{D}} \alpha(\omega)\cdot \underbrace{\delta(\omega)}_{=c^*,\forall\omega\in\partial\mathcal{D}} \cdot p(\omega)d\omega = c^*\cdot \int_{\partial\mathcal{D}} \tilde{\alpha}(\omega)p(\omega)d\omega = c^*\cdot P(X).$$

Hence, the commitment outcome can be expressed as $\alpha(\omega) = \mathbb{1}(\omega \in \mathcal{D}\cup X)$, and the sender's problem under commitment can be stated as

$$\max_{W\subseteq\Omega} \int_{\Omega} \mathbb{1}(\omega\in W)\cdot p(\omega)d\omega \;\text{ subject to } \int_{\Omega} \delta(\omega)\cdot\mathbb{1}(\omega\in W)\cdot p(\omega)d\omega,$$

which is the same problem as finding the sender-preferred equilibrium in the game with verifiable information. By the argument above, the solution is characterized by the cutoff value of receiver's net payoff of approval c^*. □

PROOF OF THEOREM 3 (SEE PAGE 17).

THEOREM 3. *The following statements about the sender's ex-ante utility $\bar{u}_s$ are equivalent:*

(1) $\bar{u}_s$ is reached in equilibrium;

(2) $\bar{u}_s$ is reached in a direct equilibrium;

(3) $\bar{u}_s$ is given by

$$\bar{u}_s = \int\limits_{\Omega} u_s\big(\{i \in [I] \mid \omega \in W_i\}\big) \cdot p(\omega) d\omega,$$

where for every receiver $i \in [I]$, $W_i \subseteq \Omega$ is her set of approval states, which satisfies

- *receiver's obedience constraint constraint $p(\cdot \mid W_i) \in \mathcal{B}_i$;*
- *sender's incentive-compatibility constraint $\mathcal{A}_i \subseteq W_i$.*

Proof. *Part I*, *(1) ⇒ (3)*: consider equilibrium $(\sigma, \boldsymbol{a}, \boldsymbol{q})$ with the ex-ante utility of the sender $\bar{u}_s$. Let $X_i := \{\omega \in \Omega \mid \alpha_i(\omega) = 1\}$ be the set of states in which the sender convinces receiver $i \in [I]$ to approve the proposal with certainty. By Lemma A.1, set $W_i = X_i \cup \mathcal{A}_i$ satisfies receiver i's (obedience) constraint. W_i also obviously satisfies the sender's (IC) constraint, for every $i \in [I]$.

If $(W_1, \ldots, W_I)$ is the collection of the receivers' sets of approval states, then the sender's ex-ante utility equals

$$\int\limits_{\Omega} u_s\big(\{i \in [I] \mid \omega \in W_i\}\big) p(\omega) d\omega,$$

because receiver i approves the proposal if and only if $\omega \in W_i$. What remains to show is that this expression equals $\bar{u}_s$, the ex-ante utility of the sender in the original equilibrium. That is true because if, in the original equilibrium, receiver $i \in [I]$

- is convinced in state ω with certainty, then $\omega \in W_i$;
- is convinced with probability less than 1 in state $\omega \in \mathcal{A}_i$, then this receiver's action does not affect the sender's utility in this state, or else he could deviate to full disclosure. Adding these states to W_i is inconsequential to the sender's ex-ante utility;
- is convinced with probability between 0 and 1 in state $\omega \notin \mathcal{A}_i$, then this receiver's action does not affect the sender's utility in this state, or else he could deviate to one of the messages that induce approval. Not adding this state to W_i is inconsequential to the sender's ex-ante utility.

As a result, $\bar{u}_s$ equals the expression above.

(3) ⇒ (2): consider collection $(W_1, \ldots, W_I)$ of receivers' sets of approval states, each of which satisfies the sender's (IC) and receiver's (obedience) constraints. In the direct equi-

librium, the strategy of the sender is to send message W_i to receiver i if $\omega \in W_i$, and the complementary message $\Omega \setminus W_i$ otherwise. Sender's incentive-compatibility and receivers' obedience constraints ensure that the sender cannot deviate to full disclosure and the receivers interpret the two messages they receive on the path as recommendations to take appropriate actions. Off-the-path, the same "skeptical beliefs" apply as in the one-receiver case, and they ensure that the sender cannot deviate to an off-path message, either.

Part III, *(2)* $\Rightarrow$ *(1)* is trivial because a direct equilibrium is an equilibrium. □

PROOF OF THEOREM 4 (SEE PAGE 18).

THEOREM 4. *The sender-preferred equilibrium outcome is a commitment outcome.*

Proof. Consider the problem of finding the optimal commitment protocol $(\sigma^{BP}, a^{BP}, q^{BP})$. According to Kamenica and Gentzkow (2011), the problem may be simplified to finding an optimal *straightforward* experiment σ^{BP} that is supported on set $(S_1, \ldots, S_I)$, where $S_i = \{s_i^+, s_i^-\}$ is the private set of *straightforward* signal realizations of receiver $i \in [I]$. Signal realization s_i^+ induces posterior $q_i^+ \in \mathcal{B}_i$ and recommends that receiver i approves the proposal and s_i^- induces posterior $q_i^- \notin \mathcal{B}_i$ and recommends rejection. Commitment outcome is then $\alpha_i(\omega) = \text{Prob}(s_i^+ \mid \omega)$, and the sender's problem under commitment becomes

$$\max_{\alpha_i, \forall i \in [I]} \sum_{T \subseteq 2^I} \int_{\Omega} \alpha(T, \omega) \cdot u_s(T) \cdot p(\omega) d\omega,$$

where $\alpha(T, \omega) := \prod_{i \in T} \alpha_i(\omega) \cdot \prod_{j \in [I] \setminus T} \big(1 - \alpha_i(\omega)\big)$ is the probability that receivers in $T \subseteq [I]$ approve the proposal and the receivers in $[I] \setminus T$ reject it, *subject to* $\forall i \in [I]$

- $\forall \omega \in \Omega$, $0 \leq \alpha_i(\omega) \leq 1$;
- receiver i's obedience constraint $q_i^+ \in \mathcal{B}_i$, which is $\int_{\Omega} \delta_i(\omega) \cdot \alpha_i(\omega) \cdot p(\omega) d\omega \geq 0$.

Notice that if $\alpha_i(\omega) = \mathbb{1}\big(\omega \in W_i^j\big)$ for all $i \in [I]$, then $\alpha(T, \omega) = \mathbb{1}\big(T = \{i \in [I] \mid \omega \in W_i\}\big)$, and the sender's problem becomes

$$\max_{W_i \subseteq \Omega, \forall i \in [I]} \int_{\Omega} u_s\big(\{i \in [I] \mid \omega \in W_i\}\big) \cdot p(\omega) d\omega,$$

subject to receiver i's obedience constraint $p(\cdot \mid W_i) \in \mathcal{B}_i$, for all $i \in [I]$. What remains to show is that *(i)* there exists a deterministic commitment outcome, and *(ii)* every set of approval states W_i induced by that outcome satisfies the sender's (IC) constraint.

Suppose that α_i, $\forall i \in [I]$, is a commitment outcome. We construct a deterministic commitment outcome $\tilde{\alpha}_i$, $\forall i \in [I]$, in a sequence of steps.

Step 1: if, for some $i \in [I]$ and $\omega \in \mathcal{A}_i$, $\alpha_i(\omega) < 1$, then let $\tilde{\alpha}_i(\omega) = 1$. This weakly increases the objective, loosens receiver i's obedience constraint, and does not alter other receivers' obedience constraints. Note that this case only arises when the sender's utility does not strictly increase in receiver i's action in state ω;

Step 2: if, for some $i \in [I]$, this receiver's obedience constraint does not bind, then let $\tilde{\alpha}_i(\omega) = 0$ for every ω such that $\alpha_i(\omega) < 1$. In those states, the sender could have increased $\alpha_i(\omega)$ by tightening receiver i's obedience constraint, but did not do so because convincing this receiver in this state is inconsequential to the sender's utility;

Step 3: if, for some receiver $i \in [I]$ and set $\mathcal{D} \subseteq \Omega$, $\alpha_i(\omega) \in (0,1)$ for every $\omega \in \mathcal{D}$, and this receiver's obedience constraint binds, then we follow the steps on the proof of Theorem 2. The (binding) obedience constraint of receiver i becomes

$$\int_{\mathcal{D}} \delta_i(\omega) \cdot \alpha_i(\omega) \cdot p(\omega) d\omega = -\int_{\Omega \setminus \mathcal{D}} \delta_i(\omega) \cdot \alpha_i(\omega) \cdot p(\omega) d\omega := \mathcal{I}_i.$$

Since $\alpha_i(\omega) \in (0,1)$ on $\mathcal{D}$, then $\delta_i(\omega)$ and is constant on $\mathcal{D}$. Otherwise, if $\delta_i(\omega) > \delta_i(\omega')$ for some $\omega, \omega' \in \mathcal{D}$, then increasing $\alpha_i(\omega)$ and decreasing $\alpha_i(\omega')$ increases the value of the objective and loosens receiver i's obedience constraint. Next, partition $\mathcal{D}$ into sets X and $\mathcal{D} \setminus X$, where X solves

$$\int_{\mathcal{D}} \alpha_i(\omega) \cdot p(\omega) d\omega = \int_X p(\omega) d\omega = P(X).$$

Letting

$$\tilde{\alpha}_i(\omega) = \begin{cases} 1, & \omega \in X, \\ 0, & \omega \in \mathcal{D} \setminus X \end{cases}$$

does not alter the value of the objective or any receiver's obedience constraint, similarly to the one-receiver case;

Step 4: if for $i \in [I]$ and $\omega \in \Omega$, $\alpha_i(\omega) \in \{0,1\}$, then let $\tilde{\alpha}_i(\omega) = \alpha_i(\omega)$.

At this point, $\tilde{\alpha}_i$, $\forall i \in [I]$, is a deterministic commitment outcome that satisfies all of the sender's (IC) constraints since $\tilde{\alpha}_i(\omega) = 1$ for all $\omega \in \mathcal{A}_i$. Consequently, this outcome also solves the problem of finding the sender-preferred equilibrium in the verifiable information game. $\square$

PROOF OF REMARK 1 (SEE PAGE 19).

REMARK 1. *Approval sets and sets of approval beliefs of voters L and R with ideal policies $v_L < \omega_0 < v_R$ do not intersect a.s.*

Proof. By the definition of the set of approval beliefs, for every $i \in [I]$

$$q \in \mathcal{B}_i \iff \int_{\Omega} |v_i - \omega| \cdot q(\omega) d\omega \leq |v_i - \omega_0|.$$

Adding up the right-hand sides for $i \in \{L, R\}$,

$$q \in \mathcal{B}_L \cap \mathcal{B}_R \implies \int_{\Omega} \big[|v_L - \omega| + |v_R - \omega|\big] \cdot q(\omega) d\omega \leq |v_L - \omega_0| + |\omega_0 - v_R| = |v_L - v_R|.$$

However, the right hand side almost surely violates the triangle inequality, which states that $|v_L - \omega| + |\omega - v_R| \geq |v_L - v_R|$ for every $\omega \in \Omega$. $\square$

PROOF OF THEOREM 5 (SEE PAGE 20).

THEOREM 5. *In the sender-preferred equilibrium of an election with voters L and R whose ideal policies are $v_L < \omega_0 < v_R$,*

- *message $\bar{W}_i$ that convinces voter $i \in \{L, R\}$ is an interval $[a_i, b_i] \supset \mathcal{A}_i$;*
- *challenger has positive ex-ante odds of winning this election; the set of winning policies is $[a_R, b_L]$ with $a_R < \omega_0 < b_L$.*

Proof. Recall that $\delta_i(\omega) = |v_i - \omega_0| - |v_i - \omega|$ is voter i's net payoff from voting for the challenger. The (obedience) constraint is:

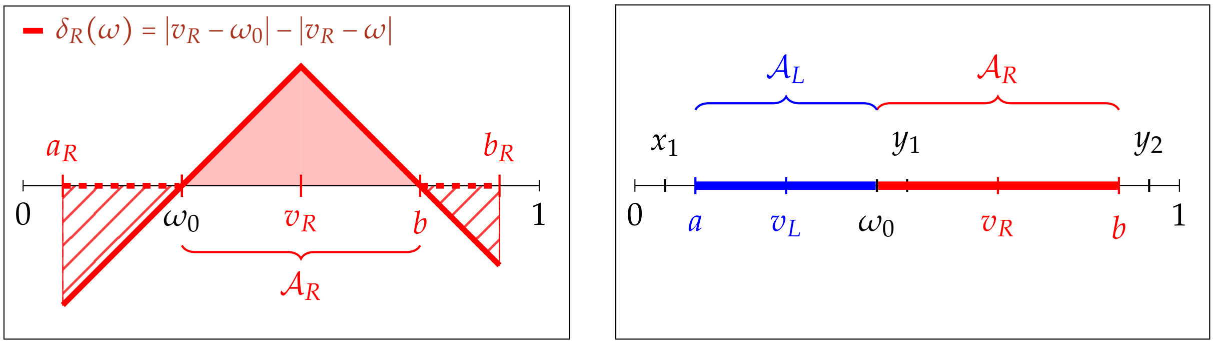
$$p(\cdot \mid W_i) \in \mathcal{B}_i \iff \int\limits_{W_i} \delta_i(\omega) p(\omega) d\omega \geq 0$$

$$\iff \int\limits_{W_i \setminus \mathcal{A}_i} - \underbrace{\delta_i(\omega)}_{<0, \forall \omega \notin \mathcal{A}_i} \cdot p(\omega) d\omega \leq \int\limits_{\mathcal{A}_i} \underbrace{\delta_i(\omega)}_{>0, \forall \omega \in \mathcal{A}_i} p(\omega) d\omega := \mathcal{I}_i.$$

Notice that when $\omega \notin \mathcal{A}_i$, $-\delta_i(\omega)$ reflects the distance from point ω to the approval set of voter i. The voter's obedience constraint states that the expected distance from the challenger to the voter's approval set must not exceed a known quantity $\mathcal{I}_i$, which reflects how persuadable this voter is. For example, Figure 6 – part *(b)* illustrates how under uniform prior, voter R's obedience constraint states that the area under the function $\delta_R(\omega)$ over the approval set (it equals $\mathcal{I}_R$) must exceed the area over the same function outside of the approval set. Observe that adding point x to $W_L \cap W_R$ increases the objective function by $p(x)$ and costs $-\delta_i(x)p(x) \cdot \mathbb{1}(x \notin \mathcal{A}_i)$ to each voter $i \in \{L, R\}$. Consequently, $x \notin \mathcal{A}_i$ is "cheaper" in terms of i's obedience constraint than $y \notin \mathcal{A}_i$ if $\delta_i(x) \geq \delta_i(y)$. Points in the approval set of the voter are "free" in terms of the obedience constraint of that voter.

Next we show that $W_i = [a_i, b_i] \supset \mathcal{A}_i$. Let $a = 2v_L - \omega_0$ be the left boundary of L's approval set, and let $b = 2v_R - \omega_0$ be the right boundary of R's approval set. The chain of arguments below proves that $W_L = [a_L, b_L]$, with $a_L \leq a$ and $b_L \geq \omega_0$, and Figure 6, part *(b)* illustrates:

- $[a, \omega_0] \subseteq W_L$ because it is the approval set of this voter;
- if $x_1 \in [0, a)$ and $x \in W_L$, then $\forall y_1 \in [\omega_0, b]$ such that $|a - x_1| \geq |y_1 - \omega_0|$, $y_1 \in W_1$, because y_1 is cheaper in terms of L's constraint and free in terms of R's constraint;
- if $x_1 \in [0, a)$ and $x \in W_L$, then $\forall x \in (x_1, a]$, $x \in W_1$, because x is cheaper in terms of both voters' constraints;
- if $y_1 \in (\omega_0, b]$ and $y_1 \in W_L$, then $\forall y \in [\omega_0, y_1)$, $y \in W_1$, because y is cheaper in terms of L's constraint and free in terms of R's constraint;
- if $y_2 \in (b, 1]$ and $y_2 \in W_L$, then $\forall y \in [\omega_0, y_2)$, $y \in W_1$, because x is cheaper in terms of both voters' constraints;
- $b_L > \omega_0$ because $\mathcal{I}_L > 0$.



(a) *Voter R's net payoff of voting for the challenger. Under uniform prior, her obedience constraint states that the solid area exceeds the dashed area.*

(b) *Approval sets of the voters and points x_1, y_1, y_2.*

Figure 6: *why challenger-preferred convincing messages are intervals.*

□

PROOF OF THEOREM 6 (SEE PAGE 22).

THEOREM 6. *Assume uniform prior. In the sender-preferred equilibrium of an election with voters L and R whose ideal policies are $v_L < \omega_0 < v_R$,*

- *as v_R becomes more persuadable, the challenger's odds of winning and the total amount of regret increase;*
- *suppose $|v_L| = |v_R|$, meaning that neither voter is more persuadable than the other. Then, as v_R becomes more persuadable,*
  - *the set of winning policies $\bar{W} = [a_R, b_L]$ shifts to the left, i.e. a_R and b_L decrease;*
  - *the amount of regret of each voter increases.*

Proof. Given convincing message $[a_R, b_R]$, voter R's obedience constraint becomes

$$\int_{a_R}^{\omega_0} (\omega_0 - \omega) p(\omega) d\omega + \int_{2v_R - \omega_0}^{b_R} (\omega - 2v_R + \omega_0) p(\omega) d\omega$$

$$\leq \int_{\omega_0}^{v_R} (\omega - \omega_0) p(\omega) d\omega + \int_{\omega_0}^{2v_R - \omega_0} (2v_R - \omega - \omega_0) p(\omega) d\omega.$$

The derivative of the left-hand side of this inequality with respect to v_R is negative and equals $-2P\big([2v_R - \omega_0, b_R]\big)$, while the derivative of the right-hand side with respect to v_R

is positive and equals $2P([v_R, 2v_R - \omega_0])$. Consequently, as v_R increases, voter R's obedience constraint loosens, and that is true for any prior distribution. Hence, the solution, specifically, the challenger's ex-ante odds of winning, can only improve.

Under uniform prior, the ex-ante measure of the set of winning policies $\bar{W} = [a_R, b_L]$ is $b_L - a_R$, which coincides with the total amount of regret ($\omega_0 - a_L$ for voter R plus $b_R - \omega_0$ for voter L). Hence, the total amount of regret increases, as well.

Now suppose $|v_L| = |v_R|$ and let $a = 2v_L - \omega_0$ be the left boundary of L's approval set, and let $b = 2v_R - \omega_0$ be the right boundary of R's approval set. Voters' (obedience) constraints are symmetric about ω_0, implying that the solution is symmetric, as well, with $a_R > a$ and $b_L < b$, and $|b_L - \omega_0| = |\omega_0 - a_R|$.[10] In other words, b_L solves voter L's obedience constraint $\int_a^{\omega_0} \delta_L(\omega)d\omega = -\int_{\omega_0}^{b_L} \delta_L(\omega)d\omega$, while a_R solves voter R's obedience constraint $\int_{\omega_0}^{b} \delta_R(\omega)d\omega = -\int_{a_R}^{\omega_0} \delta_R(\omega)d\omega$.

As v_R increases, voter R's obedience constraint loosens, while voter L's obedience constraint remains the same. An increase in the value of the objective function is thus obtained by decreasing both a_L and b_R.

As a_L decreases, voter R's amount of regret $\omega_0 - a_L$ increases. For voter L,

- b_R cannot increase because it is obtained from the binding obedience constraint that was not affected by the change;
- for high enough v_R, $\int_{\omega_0}^{b} \delta_R(\omega)d\omega > -\int_{a_R}^{\omega_0} \delta_R(\omega)d\omega$, meaning that the optimal message that convinces voter L has to be optimally shifted to the left and becomes $[a_L, b_L']$, with $a_L < a$ and $b_L' < b_L$;
- voter L's obedience constraint becomes $\int_a^{\omega_0} \delta_L(\omega)d\omega \geq -\int_{a_L}^{a} \delta_L(\omega)d\omega - \int_{\omega_0}^{b_L} \delta_L(\omega)d\omega$. Because b_L is further from v_L than a is, removing $b_L - \varepsilon$ from the message that convinces voter L and replacing it with $a - \varepsilon$ (for some $\varepsilon > 0$) loosens voter L's obedience constraint and keeps the value of the objective the same. That means that as b_L decreases, a_L decreases even more. Consequently, the amount of regret of voter L, which is $(b_L - \omega_0) + (a - a_L)$ increases;

[10] For example, $b_L < b$ because R's obedience constraint is $\int_{a_R}^{\omega_0} -\delta_R(\omega)d\omega \leq \int_{\omega_0}^{b} \delta_R(\omega)d\omega$. Consequently, we can infer from Figure 6, part *(a)*, that $|\omega_0 - a_R| < |b - \omega_0| = |\omega_0 - a|$, meaning that $a < a_R$.

- the above argument stops working when $b_L - \omega_0 = a - a_L$. At that point, voter R is so persuadable that only voter L's constraint binds. The problem boils down to persuading just voter L, is characterized in Theorem 2, and no further dynamics are observed.

$\square$