PERSUASION WITH VERIFIABLE INFORMATION*

Maria Titova†

May 2021

LATEST VERSION

Abstract

This paper studies how an informed sender with state-independent preferences persuades receivers to approve his proposal with verifiable information. I find that every equilibrium outcome is characterized by each receiver's set of approved states that satisfies this receiver's obedience and the sender's incentive-compatibility constraints. That allows me to describe the complete equilibrium set. In the sender-worst equilibrium, information unravels, and receivers act as if fully informed. 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 with public disclosure. The more polarized the electorate, the more likely it is that the challenger swings an unwinnable election with targeted advertising.

KEYWORDS: Persuasion, Value of Commitment, Targeted Advertising, Elections

JEL CLASSIFICATION: D72, D82, D83

*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 SWET, YES, and EGSC 2019, 6th Annual Conference on Network Science and Economics, and seminar audiences at UC San Diego. All errors are my own.

†Department of Economics, University of California, San Diego. E-mail: motitova@gmail.com.

1. Introduction

Suppose the sender attempts to convince a group of receivers to take his favorite action. The only tool available to him is hard evidence. What he can do is choose how much of it to reveal. On average, what is the best outcome that the sender can hope for?

Persuasion with verifiable information plays an essential role in electoral campaigns, product advertising, financial disclosure, and job market signaling, among many other economic situations. In politics, a challenger convinces voters to elect him over the status quo by sending fact-checked ads about his policy position on some relevant socio-economic issues, saying nothing about other issues. In business, a firm convinces consumers to adopt its product by advertising some product characteristics, not mentioning others. In finance, a CEO convinces the board of directors to approve managerial compensation by presenting some financial indicators and statements, omitting others. In labor markets, a job candidate convinces committee members to offer him a job by attaching to his application selected evidence of his qualifications.

I consider the following formal model of persuasion with verifiable information. There is an underlying continuous space of possible states of the world, which is a unit interval. The sender is fully informed about the state of the world, but his preferences do not depend on it. Receivers are uninformed about the state of the world, which to them is payoff-relevant. The sender sends a verifiable message to each receiver. Verifiability means that the message contains the truth (hard evidence is presented), but it could be vague (not all the evidence is presented). Each receiver independently chooses between two options: to approve the proposal or to reject it. There are no information spillovers between the receivers: each receiver only hears her own private message.

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, we can partition the state space into two subsets: the set of approved states and the set of rejected states. My first result states that a subset of the unit interval is an equilibrium set of approved states if and only if it satisfies two constraints. Firstly, 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. Secondly, the *receiver's obedience constraint* ensures that the receiver approves the proposal whenever her expected net payoff of approval is non-negative.

In the sender's least preferred equilibrium, his ex-ante odds of approval are minimized across all equilibria. The receiver learns whether the state of the world is within her complete information approval set 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. In the sender-preferred equilibrium, the receiver approves the proposal whenever her net payoff of approval is sufficiently high, but possibly negative. That is, the sender improves his odds of approval upon full disclosure by convincing the receiver to approve when she prefers not to.

In his most preferred equilibrium, the sender pulls the "good" states that the receiver prefers to approve and the "bad" states that the receiver prefers to reject. The solution is characterized by a cutoff value: the receiver approves every state that is not too "bad". When the receiver approves, her obedience constraint binds, and she is indifferent between approval and rejection. The sender improves his ex-ante payoff over full disclosure because the receiver approves some of the "bad" states. In fact, in his most preferred equilibrium, the sender reaches the commitment payoff. This observation bridges the gap between the verifiable information literature and the Bayesian persuasion Kamenica and Gentzkow, 2011. The sender need not benefit from having ex-ante commitment power and can persuade the receiver with verifiable messages.

With many receivers, I get similar results. Every receiver makes a fully informed choice in the sender-worst equilibrium, and the sender-preferred equilibrium outcome is a commitment outcome.

SWINGING ELECTIONS

Targeted advertising played an important role in the recent US Presidential Elections. In 2016, the Trump campaign used voter data from Cambridge Analytica to target voters via Facebook and Twitter. In 2008, the Obama campaign pioneered the use of social media to communicate with the electorate. Even before social media, in 2000, The Bush campaign targeted voters via direct mail. Given that the winning candidate had access to better technology or better voter data in all these cases, one may wonder whether targeted advertising was why these candidates won.[1] In other words, can targeted advertising swing electoral outcomes?

To answer that question, I apply my model to study elections. The state space is now a one-dimensional policy space with positions ranging from ultra-left (0) to ultra-right (1). The voters choose between the challenger, whose policy is unknown, and the status quo policy, which is fixed and known. Each voter prefers to vote in favor of the policy that is closest to her bliss point. In his electoral campaign, the challenger sends verifiable messages to the voters to inform them about his policy and convince them to elect him.

Suppose that winning an election requires convincing 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 privately advertise to each of these voters, 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. I call this election unwinnable for the challenger. Whether an election is unwinnable depends on the institution (the social choice function) and the ideology of the electorate (bliss point of the voters). For example, under the majority rule, I show that an election is unwinnable if and only if the status quo is the median voter's bliss point.

When the challenger has access to targeted advertising, he can tell different things to different voters. Recall that in his most preferred equilibrium, the sender improves his odds of approval upon full disclosure. In particular, the challenger manages to convince voter L (R) even when his policy is slightly to the right (left) of the status quo. Consequently, he can convince both voters at the same time and win unwinnable elections with positive probability. That said, the challenger only benefits from private communication 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).

When a voter's bliss point moves away from the status quo, she becomes less sat-

[1]For comparison of advertising strategies between the candidates, see Kim et al. (2018) and Wylie (2019) for the 2016 election, Harfoush (2009) and Katz, Barris, and Jain (2013) for 2008, and Hillygus and Shields (2014) for 2008.

isfied with the status quo, and that makes her more persuadable. Consequently, when the electorate becomes more polarized, which happens when one of the voters' positions becomes more extreme, the challenger has higher odds of swinging an unwinnable election. As voter R's position moves further to the right, she becomes more persuadable also by policies further to the left of the status quo. Consequently, when voter R's bliss point shifts to the right, the challenger-preferred set of approved policies shifts to the left, toward the policies preferred by the less extreme voter L.

RELATED LITERATURE

I assume that the sender uses hard evidence to communicate with the receivers. This verifiable information communication protocol was introduced by Milgrom (1981) and Grossman (1981). Other communication protocols include cheap talk by Crawford and Sobel (1982) and Bayesian persuasion by Kamenica and Gentzkow (2011). 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) 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.[2] In contrast to their result, I show that the sender does not necessarily benefit from commitment if he possesses the hard evidence to verify his messages.

There is extensive literature on applications of Bayesian persuasion models. It includes settings 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); matching platforms persuade sellers to match with buyers (Romanyuk and Smolin, 2019); politicians persuade voters (Alonso and Câmara, 2016; Bardhi and Guo, 2018); governments persuade citizens through media (Gehlbach and Sonin, 2014; Egorov and Sonin, 2019). My contribution states that in all these applications, one can replace the assumption that the sender has commitment

[2] 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.

power with the assumption that the sender has hard evidence.

The leading application contributes to the growing literature on voter persuasion. My results are in line with the recent findings in the information design literature on the private persuasion of strategic voters. In particular, Chan et al. (2019) confirm that the politician does better when private disclosure is allowed, and Heese and Lauermann (2019) confirm that the politician needs very little commitment power to achieve the desired outcome. In the verifiable information literature, electoral competition usually results in the full unraveling of information (Board, 2009; Janssen and Teteryatnikova, 2017; Schipper and Woo, 2019) because the candidates play a zero-sum game, and that pushes them to disclose all information voluntarily. In contrast to these papers, I consider a non-symmetric model in which one candidate has a significant advantage over his opponent in that he is the only one who can communicate with the voters. Unraveling does not necessarily occur, and the challenger can improve his odds of winning over full disclosure.

The leading application sheds more light on how political advertising, especially targeted advertising, affects electoral outcomes and why it has become widespread. DellaVigna and Gentzkow (2010) and Prat and Strömberg (2013) 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 (Oberholzer-Gee and Waldfogel, 2009; Enikolopov, Petrova, and Zhuravskaya, 2011). 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 growing literature on polarization and targeted political advertising through media. As the number of media outlets increases, they become more specialized and target voters with more extreme preferences, which leads to social disagreement (Perego and Yuksel, 2018). If the electorate is polarized to begin with, so are the candidates' chosen policy platforms (Hu, Li, and Segal, 2019; Prummer, 2020). 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.

This paper is organized as follows. Section 2 introduces the model. Section 3 describes equilibrium outcomes in the game with one receiver. Section 4 generalizes the model to many receivers. Section 5 studies targeted advertising in elections. Section 6 is a conclusion.

2. Model

There is a state space $\Omega := [0,1]$ and a finite set of receivers $I := \{1,\dots,n\}$. 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 Ω.[3] Having observed the state, the sender sends a verifiable message $m_i \subseteq \Omega$, such that $\omega \in m_i$, to each receiver (her) $i \in I$.[4]

The sender's payoff $u_s : 2^n \to \mathbb{R}$ depends only on the subset of receivers who approve his proposal. I assume that if all receivers reject the proposal, then the sender gets the lowest payoff, which is normalized to 0. If every receiver approves the proposal, then the sender gets the highest payoff, which is normalized to 1. Also, I assume that u_s weakly increases in every receiver's action.

Assumption 1. *The sender's payoff u_s satisfies*

1. $u_s(\varnothing) = 0$ *and* $u_s(I) = 1$;
2. *given two sets of receivers* $I_1, I_2 \subseteq I$, $u_s(I_1) \le u_s(I_2)$ *if* $I_1 \subseteq I_2$.

Receiver $i \in I$ chooses between approval (action 1) and rejection (action 0). Receiver i's preferences are described by a utility function $u_i : \{0,1\} \times \Omega \to \mathbb{R}$. Receiver i approves (the proposal in) state ω if her net payoff of approval $\delta(\omega) := u_i(1,\omega) - u_i(0,\omega)$ is non-

[3]For a compact metrizable space S, ΔS denotes the set of all Borel probability measures over S. For any $q \in \Delta\Omega$ and any measurable subset of the state space $W \subseteq \Omega$, $Q(W) = \int_W q(\omega)d\omega$ is the probability measure and $q(\cdot \mid \cdot)$ is the conditional probability distribution: $q(\omega \mid W) = 1$ if $W = \{\omega\}$ and $q(\omega \mid W) = \frac{q(\omega)}{Q(W)}$ if $Q(W) > 0$.

[4]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). This method satisfies normality of evidence (Bull and Watson, 2004), which means that it is consistent with both major ways of modeling hard evidence in the literature.

negative.[5] Define receiver i's <u>approval set</u> as

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

Example 1 (Receiver with Spatial Preferences). This example introduces the receivers with spatial preferences à la Downs (1957). Receiver i has a bliss point $v_i \in \Omega$ and compares the sender's position ω to the status quo $\omega_0 \in (0,1)$. Her net payoff of approval is $\delta_i(\omega) = -|v_i - \omega| + |v_i - \omega_0|$ and her approval set is $\mathcal{A}_i = \left\{\omega \in \Omega \text{ s.t. } |v_i - \omega| \leq |v_i - \omega_0|\right\}$. That is, she approves ω if and only if it is closer to her bliss point than the status quo.

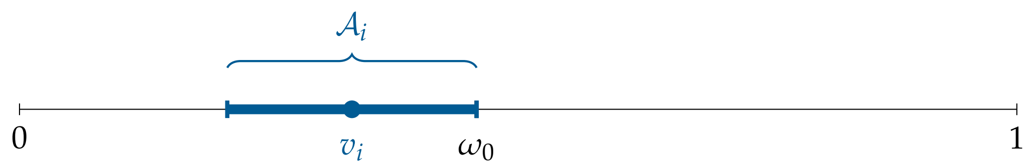


Figure 1. *Receiver i with spatial preferences: her approval set $\mathcal{A}_i$ (solid blue) consists of points on the unit interval that are closer to her bliss point v_i than the status quo ω_0.*

Under incomplete information, define receiver i's <u>set of approval beliefs</u> as

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

I assume that every receiver rejects the proposal under prior belief.

Assumption 2. *For every receiver $i \in I$, $p \notin \mathcal{B}_i$.*

Assumptions 1 and 2 ensure that without any additional information, all receivers reject the proposal and the sender gets the lowest possible payoff. The rest of the paper studies how the sender persuades the receivers with verifiable information.

Equilibrium Outcomes

I consider Perfect Bayesian Equilibria (henceforth just *equilibria*) of this game. The sender's strategy is a probability distribution $\sigma(\cdot \mid \omega)$ over message collections $\{m_i\}_{i \in I}$,

[5] I assume that the receiver breaks ties in favor of approval when she is indifferent, i.e. when $\delta(\omega) = 0$.

where $m_i \subseteq \Omega$ for each $i \in I$. Receiver i's approval strategy $a_i(m)$ specifies which action she takes depending on message m she receives. Receiver i's posterior belief over Ω after message m is $q_i(\cdot \mid m)$. Profiles of receivers' actions and posterior beliefs are $a := \{a_i\}_{i \in I}$ and $q := \{q_i\}_{i \in I}$, respectively.

DEFINITION 1. *A triple (σ, a, q) is an* equilibrium *if*

(i) $\forall \omega \in \Omega$, $\sigma(\cdot \mid \omega)$ *is supported on* $\arg\max_{m_1,\dots,m_n} u_s\big(\{i \in I \mid a_i(m_i) = 1\}\big)$, *s.t.* $\omega \in m_i$, $\forall i \in I$.

The following conditions hold for every receiver $i \in I$:

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

(iii) $\forall m \subseteq \Omega$ *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) $\forall m \subseteq \Omega$, $\operatorname{supp} q_i(\cdot \mid m) \subseteq m$.

In words, *(i)* states that the sender sends a collection of messages with positive probability only if it maximizes his payoff; *(ii)* states that each receiver approves the proposal whenever her expected net payoff of approval 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.

An outcome of the game specifies what action receivers take in every state of the world.

DEFINITION 2.

- *An* outcome $\alpha = \{\alpha_i\}_{i \in I}$ *specifies $\forall i \in I$ and $\forall \omega \in \Omega$ the probability $\alpha_i(\omega) \in [0,1]$ that receiver i approves the sender's proposal in state ω.*
- *An outcome is an* equilibrium outcome *if it corresponds to some equilibrium.*[6]

Some outcomes are deterministic, meaning that in every state ω each receiver either

[6]Specifically, if there exists equilibrium (σ, a, q) such that $\forall i \in I$ and $\forall \omega \in \Omega$, $\alpha_i(\omega) = \int_{\mathcal{M}_i} \sigma_i(m \mid \omega) dm$, where $\mathcal{M}_i := \{m \subseteq \Omega \mid a_i(m) = 1\}$ is the set of messages that convince receiver i to approve.

approves or rejects the proposal with certainty.[7] Consequently, for each receiver, we can partition Ω into states of approval and states of rejection.

DEFINITION 3.

- *An outcome α is deterministic if $\alpha_i(\omega) \in \{0,1\}$ for every $i \in I$ and $\omega \in \Omega$.*
- *The set of approved states W_i of receiver $i \in I$ in deterministic outcome α is*

$$W_i := \{\omega \in \Omega \mid \alpha_i(\omega) = 1\}.$$

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. By Assumption 1, the sender gets 1 if the receiver approves and 0 otherwise. By Assumption 2, the receiver rejects the proposal under the prior belief.

Direct Implementation

Consider a deterministic equilibrium outcome with a set of approved states W. Suppose that the sender learns that $\omega \in \mathcal{A}$. One message that is available to the sender in this state (and unavailable in every other state) is $\{\omega\}$. Since that message is verifiable, upon receiving it, the receiver learns with certainty that the state is ω. Since ω is in the receiver's approval set, she approves the proposal after hearing that message. Then, for every $\omega \in \mathcal{A}$, the receiver should be approving every $\omega \in \mathcal{A}$ in every deterministic equilibrium, or else the sender has a profitable deviation towards full disclosure. That gives rise to the sender's incentive-compatibility constraint

$$\mathcal{A} \subseteq W. \tag{IC}$$

Next, if the receiver approves every state in W, then she expects that on average,

[7] Although each receiver breaks ties in favor of approval, the sender may be playing a mixed strategy in state ω, and then in that state the receiver may be approving the proposal with a probability between 0 and 1.

her net payoff of approval is non-negative. Thus, we obtain the receiver's obedience constraint

$$p(\cdot \mid W) \in \mathcal{B}. \tag{obedience}$$

The first result of this paper allows us to restrict attention to sets of approved states $W \subseteq \Omega$ that satisfy these two constraints.

THEOREM 1. *Suppose $n = 1$. Then, every equilibrium outcome is deterministic. Furthermore, $W \subseteq \Omega$ is an equilibrium set of approved states if and only if it satisfies the sender's* (IC) *and the receiver's* (obedience) *constraints.*

The proofs of Theorem 1 and other results are in the appendix. Here I describe the intuition behind this result. First, in every equilibrium outcome, the receiver either approves or rejects the proposal in every state of the world. Suppose, on the contrary, that in some state, the receiver approves and rejects with positive probability. Since the receiver approves sometimes, the sender has access to at least one message that convinces the receiver to approve. Then, the sender can deviate and send that message with certainty so that the receiver approves with probability one. Hence, all equilibrium outcomes are deterministic.

Next, if W is an equilibrium set of approved states, it satisfies the sender's (IC) constraint, or else the sender can deviate to full disclosure. To see why W also satisfies the receiver's (obedience) constraint, implement this set of approved states directly. Specifically, let the sender send message W from $\omega \in W$ and message $\Omega \setminus W$ from $\omega \notin W$. Intuitively, the (obedience) constraint states that the receiver interprets message W as a recommendation to approve. If the sender induces approval in every state in W in the original equilibrium, he also induces approval with the pooling message W.

Finally, suppose that $W \subseteq \Omega$ satisfies (IC) and (obedience). Then, we can construct an equilibrium that directly implements the set of approved states W. Let the sender send message W from every state within W and message $\Omega \setminus W$ from every state outside of W. Then, the receiver interprets message W as a recommendation to approve by the (obedience) constraint. Off the equilibrium path, let the receiver be "skeptical" and assume that any unexpected message comes from the worst possible state. Then, the sender does not have profitable deviations: if $\omega \in W$, he is getting the highest possible payoff;

if the state is not in W, the sender cannot replicate message W because $\omega \notin W$, and the receiver rejects after every other message.

Note that Theorem 1 is a version of the communication revelation principle for games with verifiable information. According to Myerson (1986) and Forges (1986), any equilibrium outcome of a mediated sender-receiver game may be implemented truthfully and obediently. In the present context, it translates into *(i)* the sender truthfully 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, Theorem 1 also provides the necessary and sufficient conditions for a set of approved states to be implementable in equilibrium.

Equilibrium Range and Value of Commitment

For the purposes of characterizing equilibrium outcomes, Theorem 1 allows us to restrict attention to sets $W \subseteq \Omega$ satisfying (IC) and (obedience). I rank equilibria in terms of the sender's ex-ante utility, which is the same as his ex-ante odds of approval and equals $P(W)$, the prior measure of the set of approved states.

In the sender-worst equilibrium, the set of approved states $\underline{W}$ minimizes the sender's ex-ante utility across all equilibria. Thus, the (IC) constraint binds and $\underline{W} = \mathcal{A}$. In this equilibrium, the receiver approves the proposal if and only if she approves it under complete information. Hence, the sender-worst equilibrium is outcome-equivalent to *full disclosure* (also known as *full unraveling*), salient in the verifiable information literature.[8]

In the sender-preferred equilibrium, the set of approved states $\overline{W}$ maximizes the sender's ex-ante utility across all equilibria. Mathematically,

$$\overline{W} = \arg\max_{W \subseteq \Omega} P(W), \quad \text{subject to} \quad \begin{array}{c} \mathcal{A} \subseteq W, \\ p(\cdot \mid W) \in \mathcal{B}. \end{array} \tag{1}$$

To find the sender-preferred equilibrium, we would increase the ex-ante measure of

[8]See, e.g., Milgrom (1981), Grossman (1981), Milgrom and Roberts (1986) and review by Milgrom (2008).

the set of approved states W so long as the receiver, when approving, expects that her net payoff of approval is non-negative, on average. Because the state space is continuous, $\overline{W}$ makes the receiver exactly indifferent between approval and rejection, binding her (obedience) constraint.

THEOREM 2. *When $n = 1$, the sender-preferred set of approved states $\overline{W}$ is characterized by a cutoff value $c^* > 0$ such that*

- *the receiver almost surely approves the proposal if $\delta(\omega) > -c^*$ and rejects it if $\delta(\omega) < -c^*$;*[9]
- *whenever the receiver approves the proposal, her expected net payoff of approval is zero: $\mathbb{E}_p\left[\delta(\omega) \mid \overline{W}\right] = 0$.*

Furthermore, the sender-preferred equilibrium outcome is a commitment outcome.

First, notice that the receiver's (obedience) constraint binds, or else we could increase the value of the objective while still satisfying that constraint. I prove the first part of Theorem 2 by contradiction. Suppose that the sender-preferred set of approved states $\overline{W}$ in not characterized by a cutoff value of the receiver's net payoff of approval. Then, there exist two sets $X, Y \subseteq \Omega$ of positive and equal measure, such that $\overline{W}$ includes X, $\overline{W}$ does not include Y, yet the receiver has a higher net payoff of approving any state in Y over any state in X. Consider an alternative set of approved states W^* that replaces X with Y, i.e. $W^* = (\overline{W} \setminus X) \cup Y$. The sender has the same ex-ante payoff at W^* and $\overline{W}$ because sets X and Y have the same measure. Yet, the (obedience) constraint for W^* is loose, while for $\overline{W}$ it is binding. That happens because every state in Y is "cheaper" in terms of the constraint than each state in X. Thus, we can improve upon both $\overline{W}$ and W^*, which is a contradiction.

Next, let us compare the problems of *(i)* finding the sender-preferred equilibrium outcome and *(ii)* finding the commitment outcome. In *(i)*, we maximize the ex-ante measure of the set of approved states subject to (IC) and (obedience) constraints. In *(ii)*, the sender maximizes his ex-ante utility subject to an obedience-like constraint of the receiver. Crucially, under commitment, the sender does not face an incentive-compatibility constraint. Also, a commitment outcome may not be deterministic.

[9] Almost surely with respect to the prior distribution p of the state of the world ω.

A commitment outcome is characterized by a cutoff value of the receiver's net payoff of approval for the same reason $\overline{W}$ is.[10] That is, the receiver certainly approves (rejects) the states with a net payoff of approval above (below) some threshold. Furthermore, that threshold is negative, and the receiver certainly approves every state in her approval set. Hence, any commitment outcome satisfies the sender's incentive-compatibility constraint.

In a non-deterministic commitment outcome, the sender induces both actions of the receiver with positive probabilities on some set $\mathcal{D} \subseteq \Omega$. Since any commitment outcome is characterized by a cutoff value, the receiver's net payoff of approval must be the same for every state in $\mathcal{D}$. Rather than making a mixed recommendation, partition the set of these states in two and let the sender recommend one action on each subset with certainty. Due to the continuity of the state space, such partitioning does not affect the objective function or the obedience constraint of the receiver. As a result, there exists deterministic commitment outcome. Since this commitment outcome satisfies the sender's incentive-compatibility constraint, it is an equilibrium outcome.

EXAMPLE 2 (RECEIVER WITH SPATIAL PREFERENCES: EQUILIBRIUM RANGE). Suppose that the receiver has spatial preferences described in Example 1. In the sender-worst equilibrium, the set of approved states is $\underline{W} = \mathcal{A}$, and the receiver approves the proposal if and only if the sender's position is closer to her bliss point than the status quo.

To find the sender-preferred set of approved states $\overline{W}$, we maximize the measure of set $W \subseteq \Omega$ subject to the receiver's (obedience) constraint. According to Theorem 2, in the sender-preferred equilibrium, the receiver approves some states outside of her approval set. However, there is a cutoff for how far the sender's position could be to be approved. When approving, the receiver expects that the sender's position and the status quo are equidistant from her bliss point. Furthermore, the sender-preferred equilibrium outcome is a commitment outcome, meaning that the sender need not benefit from having ex-ante commitment power. Figure 2 illustrates the equilibrium range.

[10] Alonso and Câmara (2016) prove that if the state space is finite, then the solution under commitment features a cutoff state.

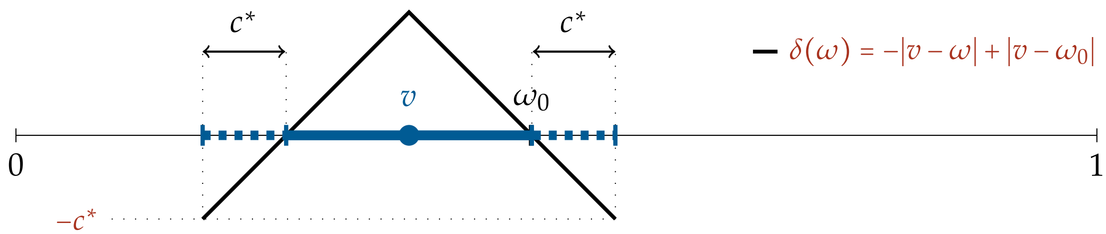


Figure 2. *The sender-worst set of approved states $\underline{W} = \{\omega \in \Omega \text{ s.t. } |v-\omega| \leq |v-\omega_0|\}$ (solid blue) and the sender-preferred set of approved states $\overline{W} = \{\omega \in \Omega \text{ s.t. } |v-\omega| \leq |v-\omega_0| + c^*\}$ (solid plus dotted blue), where c^* solves $\mathbb{E}_p[|v-\omega| \mid \overline{W}] = |v-\omega_0|$.*

4. Many Receivers

Having assumed that the receivers solve independent problems, I get similar results in the many-receiver case.[11]

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

1. *$\overline{u}_s$ is reached in equilibrium;*
2. *$\overline{u}_s$ is given by*

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

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

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

The proof of the theorem follows the same steps as the proof of Theorem 1. The only substantial difference is that Theorem 3 characterizes the sender's equilibrium ex-ante utility, while Theorem 1 characterizes the equilibrium sets of approved states. The reason is that with many receivers, some equilibrium outcomes are not deterministic. That happens because the sender may not try his hardest to convince the receivers whose approval does not strictly increase his payoff.

According to Theorem 3, when characterizing the sender's equilibrium ex-ante utility, we can restrict attention to collections of sets of approved states $(W_1, \ldots, W_n)$, each of

[11] That is, receiver i's utility does not depend on other receivers' actions, and receiver i's message is private and observed by her only.

which satisfies the IC and obedience constraints for each receiver. Moreover, the sender's ex-ante utility only depends on $(W_1, \dots, W_n)$ 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 approved states of receiver $i \in I$ is $\underline{W}_i = \mathcal{A}_i$, and each receiver makes her decision as if under complete information.

The sender-preferred equilibrium outcome is characterized by the collection of sets of approved 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's ex-ante payoff in the sender-preferred equilibrium is the commitment payoff.*

The proof of Theorem 4 follows the same steps as the proof of Theorem 2. That is, I show that if we take an arbitrary commitment outcome, we can find a deterministic commitment outcome with the same payoff of the sender. That deterministic commitment outcome satisfies every (IC) constraint of the sender, meaning that it is also an equilibrium outcome.

In general, the problem of finding the sender-preferred equilibrium outcome is computationally hard.[12] In the following section, I make additional assumptions on the sender's payoff and study elections.

5. Targeted Advertising in Elections

In this section, I show that targeted advertising helps politicians swing elections. I compare communication via targeted advertising to public disclosure. In the first case, the politician sends a private message to each voter, for example, through social media. Tar-

[12] Babichenko and Barman (2016) show that the problem of finding the commitment outcome is NP-hard when the sender's utility is submodular; Arieli and Babichenko (2019) find the commitment outcome for the case of supermodular utility.

geted advertising is an application of the main model. In the second case, the politician sends a public message to all voters. Public disclosure is not an application of the main model. However, analysis of that case is simple because the voters share a common prior belief, and if they receive the same message, they will also share a common posterior belief.

In this application, Ω is the policy space, with positions ranging from far-left (0) to far-right (1). The sender is a politician who challenges the status quo. The challenger is privately informed about his policy $\omega \in \Omega$, while the receivers hold a prior belief p. The challenger receives 1 if he wins the election and 0 otherwise. The outcome of the election is decided by the social choice function u_s that satisfies Assumption 1. For example, the election may be decided by a simple majority: the challenger wins the election if and only if the majority of receivers approve his policy, i.e. $u_s(X) = 1 \iff |X| > n/2$.

The set of receivers I is now the electorate, and the receivers are sincere voters with spatial preferences. Firstly, each voter chooses expressively, and not strategically, between the challenger and the status quo.[13] Secondly, I assume that the approval set of voter $i \in I$ is $\mathcal{A}_i = \{\omega \in \Omega \text{ s.t. } |v_i - \omega| \leq |v_i - \omega_0| - \varepsilon\}$, where $\varepsilon > 0$.[14] That is, voter i approves policies that are closer than the status quo to her bliss point by at least ε.

Observe that the preferences of the electorate can be summarized by the preferences of at most two voters whose bliss points are located closest to the status quo.

DEFINITION 4. *Voter* $L = \arg\max_{i \in I,\, v_i < \omega_0} v_i$ *is the left representative voter and voter* $R = \arg\min_{j \in I,\, v_j > \omega_0} v_j$ *is the right representative voter.*

First, notice that as a voter's bliss point moves away from the status quo, her approval set expands to include more policies of the challenger. Put differently, the further a voter's bliss point is from the status quo, the easier it is for the challenger to convince her. As a

[13] The theory of sincere voting 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 in large elections 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).

[14] ε is the status quo bias; $\varepsilon > 0$ rules out situations wherein the challenger with the status quo policy always wins the election.

result, if the challenger convinces the left (right) representative voter, he also convinces all voters with bliss points further to the left (right). Second, voters with bliss points on opposite sides of the status quo have incompatible preferences. Intuitively, left voters prefer to approve the left policies of the challenger, while right voters prefer to approve policies on the right. These observations are summarized in Corollary 1 and illustrated in Figure 3.

COROLLARY 1. *If L and R are representative voters, then*

1. *if L (R) prefers to approve challenger's policy, then so does every voter with a bliss point to her left (right), i.e.*

$$\mathcal{A}_L \subset \mathcal{A}_i \text{ and } \mathcal{B}_L \subset \mathcal{B}_i, \forall i \in I \text{ such that } v_i < v_L,$$

$$\mathcal{A}_R \subset \mathcal{A}_j \text{ and } \mathcal{B}_R \subset \mathcal{B}_j, \forall j \in I \text{ such that } v_j > v_R;$$

2. *approval sets and sets of approval beliefs of voters L and R do not intersect, i.e.*

$$\mathcal{A}_L \cap \mathcal{A}_R = \varnothing \text{ and } \mathcal{B}_L \cap \mathcal{B}_R = \varnothing.$$

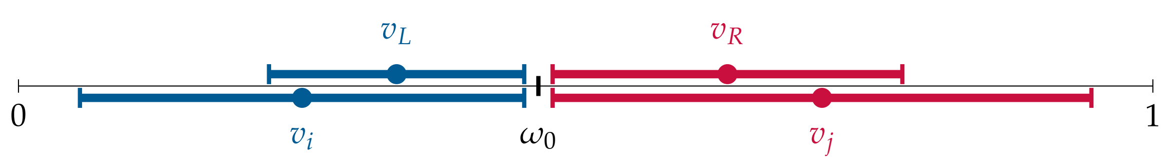


Figure 3. *Voter i is convinced if voter L is convinced: her approval set includes L's approval set (solid blue lines). Voters L and R have incompatible preferences: their approval sets do not intersect.*

SWINGING UNWINNABLE ELECTIONS

Part 2 of Corollary 1 implies that voters L and R *never* both approve the challenger's policy when they hold the same belief. Thus, if representative voters L and R are jointly pivotal, the challenger always loses the election under common belief.

DEFINITION 5. *Election with representative voters L and R is unwinnable for the challenger under common belief if for all $X \subseteq I$, $u_s(X) = 1$ if and only if $\{L, R\} \in X$.*

Whether an election is unwinnable is determined by the institution (the social choice function) and the ideology (bliss points of the voters). For example, under the simple majority rule, we arrive at a version of the median voter theorem.[15] Intuitively, for an election to be unwinnable, there may not be a majority of voters located on either side of the status quo.

COROLLARY 2. *Under the simple majority rule, an election is unwinnable for the challenger under common belief if and only if ω_0 is the median voter's bliss point.*

With targeted advertising, the challenger can say different things to different voters. The voters will no longer hold the same belief, which opens up a possibility of winning (with positive probability) an unwinnable election. Here I show how the challenger can convince representative voters L and R, persuading who is sufficient to win *any* unwinnable election. I focus on the best-case scenario for the challenger and thus consider the sender-preferred equilibrium.

By Theorem 3, we can restrict attention to a pair of sets of approved policies (W_L, W_R). In the sender-preferred equilibrium, we maximize the challenger's odds of convincing the representative voters subject to their obedience constraints:

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

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

The following theorem describes the solution to this problem.

THEOREM 5. *In the sender-preferred equilibrium of an unwinnable election with representative voters L and R, if ε is small enough,*

- *the set of approved policies $\overline{W}_i$ of voter $i \in \{L, R\}$ is an interval $[a_i, b_i] \supset \mathcal{A}_i$;*
- *the challenger wins the election if his policy is in the interval $[a_R, b_L]$ with $a_R < \omega_0 < b_L$;*
- *the challenger's ex-ante odds of winning the election are positive. i.e. $P([a_R, b_L]) > 0$.*

[15] Black (1948) states the median voter theorem as "If Ω is a single-dimensional issue and all voters have single-peaked preferences defined over Ω, then ω_0, the median position, could not lose under majority rule."

To understand the intuition behind this result, recall that when voter L (R) is the only receiver, the challenger can convince her to approve his policy even when his policy is slightly to the right (left) of the status quo. I illustrated that in Figure 2 of Example 2. One thing that the challenger can do under private communication is treat each voter as if she is the only receiver. If his policy is close enough to the status quo and ε is small enough, the challenger convinces both voters at the same time and swings an unwinnable election. However, he can do even better. To convince voter L (R), the challenger needs to make her believe that his policy is on average to the left (right) of the status quo. To induce that belief, the challenger could pull left (right) policies within this voter's approval set with some of the right (left) policies preferred by her counterpart. More precisely, voter L's (R's) message would include her approval set and as many policies to the right (left) of the status quo as this voter's obedience constraint permits. This solution is illustrated in Figure 4.

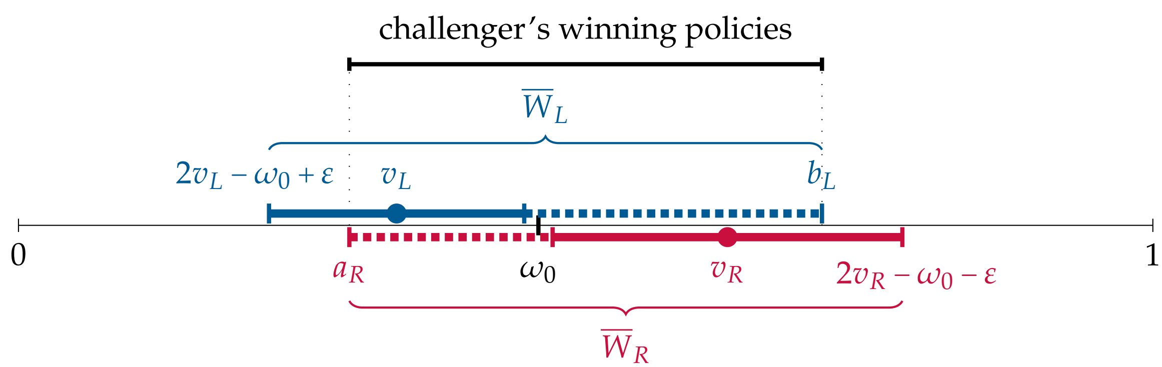


Figure 4. *The sender-preferred sets of approved policies $\overline{W}_L$ (in blue) and $\overline{W}_R$ (in red). $\overline{W}_i$ consists of voter i's approval set (solid) and policies preferred by voter $j \neq i$ (dotted). The challenger wins the election by convincing both voters when his policy is in $\overline{W}_L \cap \overline{W}_R = [a_R, b_L]$.*

Comparative Statics

Assume for the rest of this section that the prior is uniform.[16] Notice that the distance from a voter's bliss point to ω_0 measures this voter's persuadability.

Definition 6. *Suppose that $p \sim U[0,1]$. Then,*

- *voter i is more persuadable than voter j if $|v_i - \omega_0| > |v_j - \omega_0|$, where $i, j \in I$;*

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

- *consider electorates I and I' with representative voters $\{L, R\}$ and $\{L', R'\}$. I' is <u>more polarized</u> than I if $v'_L \leq v_L < \omega_0 < v_R \leq v'_R$.*

In words, the further from the status quo the voter's bliss point is, the less satisfied she is with the status quo policy, and that makes her more persuadable. I say voter $i \in I$ *becomes more persuadable* if $|v_i - \omega_0|$ increases. The electorate *becomes more polarized* when either representative voter becomes more persuadable. Figure 5 illustrates the dynamics of the numerical solution to the problem of finding the sender-preferred equilibrium as voter R becomes more persuadable (and the electorate becomes more polarized). Theorem 6 summarizes the comparative statics.

THEOREM 6. *Suppose that $p \sim U[0,1]$. In the sender-preferred equilibrium of an unwinnable election with representative voters L and R,*

- *as R becomes more persuadable, the challenger's ex-ante odds of winning $P([a_R, b_L])$ increase;*
- *suppose $|v_L - \omega_0| = |v_R - \omega_0|$, meaning that neither voter is more persuadable than the other. Then, as R becomes more persuadable, the set of challenger's winning policies $[a_R, b_L]$ shifts to the left, i.e. a_R and b_L decrease.*

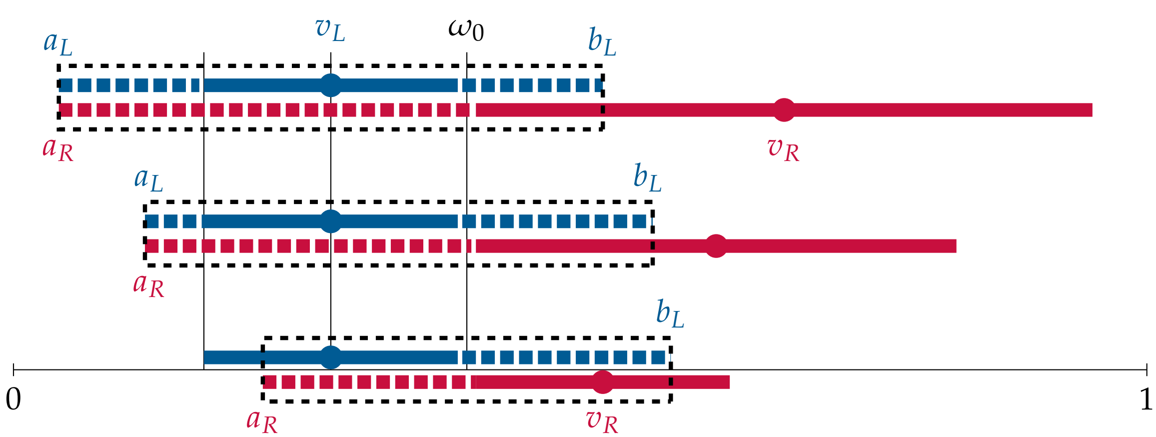


Figure 5. *Comparative statics as voter R moves to the right (bottom to top): her approval set (solid red area) expands; she is convinced by more policies on the left (dashed red area); the set of challenger's winning policies (dashed black area) moves to the left and expands.*

In words, as voter R becomes more persuadable, it becomes easier for the challenger to swing the election by targeting, in the sense that his ex-ante odds of winning increase.

Furthermore, R becomes more persuadable by policies further to the left, meaning that the set of winning policies shifts to the left, also. When voter R is far enough to the right, her obedience constraint no longer binds (as in the top exhibit of Figure 5), and the sender-preferred set of approved policies is the same as if voter L was the only receiver.

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 is more plausible than assuming that the sender has commitment power.

While illustrated in the simplified framework, the observation that targeted advertising helps challenger swing elections holds for more than one dimension and any social choice 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. (pp. 5, 14.)

ARIELI, ITAI and YAKOV BABICHENKO (2019), "Private Bayesian Persuasion", *Journal of Economic Theory*, 182 (July 2019), pp. 185-217. (p. 16.)

ARTABE, ALAITZ and JAVIER GARDEAZABAL (2014), "Strategic Votes and Sincere Counterfactuals", *Political Analysis*, 22, 2 (Jan. 2014), pp. 243-257. (p. 17.)

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

BARDHI, ARJADA and YINGNI GUO (2018), "Modes of Persuasion toward Unanimous Consent", *Theoretical Economics*, 13, 3, pp. 1111-1149. (p. 5.)

BLACK, DUNCAN (1948), "On the Rationale of Group Decision-making", *Journal of Political Economy*, 56, 1, pp. 23-34. (p. 19.)

Board, Oliver (2009), "Competition and Disclosure", *The Journal of Industrial Economics*, 57, 1 (Feb. 2009), pp. 197-213. (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. 5.)

Brennan, Geoffrey and Alan Hamlin (1998), "Expressive Voting and Electoral Equilibrium", *Public Choice*, 95, 1-2, pp. 149-175. (p. 17.)

Brennan, Geoffrey and Loren Lomasky (1993), *Democracy and Decision: The Pure Theory of Electoral Preference*, Cambridge University Press. (p. 17.)

Bull, Jesse and Joel Watson (2004), "Evidence Disclosure and Verifiability", *Journal of Economic Theory*, 118, 1 (Sept. 2004), pp. 1-31. (p. 7.)

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang (2019), "Pivotal Persuasion", *Journal of Economic Theory*, 180 (Mar. 2019), pp. 178-202. (p. 6.)

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. 17.)

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. 5.)

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. 17.)

Forges, Françoise (1986), "An Approach to Communication Equilibria", *Econometrica*, 54, 6 (Nov. 1986), p. 1375. (p. 12.)

Gehlbach, Scott and Konstantin Sonin (2014), "Government Control of the Media", *Journal of Public Economics*, 118 (Oct. 2014), pp. 163-171. (p. 5.)

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. (pp. 5, 12.)

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. 17.)

HARFOUSH, RAHAF (2009), *Yes We Did! An Inside Look at How Social Media Built the Obama Brand*, New Riders, p. 199. (p. 4.)

HEESE, CARL and STEPHAN LAUERMANN (2019), "Persuasion and Information Aggregation in Elections", *Working Paper*. (p. 6.)

HILLYGUS, D. SUNSHINE and TODD G. SHIELDS (2014), *The Persuadable Voter: Wedge Issues in Presidential Campaigns*, Princeton University Press, p. 267. (p. 4.)

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

JANSSEN, MAARTEN C. W. and MARIYA TETERYATNIKOVA (2017), "Mystifying but not Misleading: when does Political Ambiguity not Confuse Voters?", *Public Choice*, 172, 3-4 (Sept. 2017), pp. 501-524. (p. 6.)

KAMENICA, EMIR and MATTHEW GENTZKOW (2011), "Bayesian Persuasion", *American Economic Review*, 101, 6 (Oct. 2011), pp. 2590-2615. (pp. 3, 5, 28, 31.)

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. 17.)

KATZ, JAMES EVERETT., M. BARRIS, and A. JAIN (2013), *The Social Media President : Barack Obama and the Politics of Digital Engagement*, Palgrave Macmillan, p. 215. (p. 4.)

KIM, YOUNG MIE, JORDAN HSU, DAVID NEIMAN, COLIN KOU, LEVI BANKSTON, SOO YUN KIM, RICHARD HEINRICH, ROBYN BARAGWANATH, and GARVESH RASKUTTI (2018), "The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook", *Political Communication*, pp. 1-29. (p. 4.)

KOLOTILIN, ANTON (2015), "Experimental Design to Persuade", *Games and Economic Behavior*, 90 (Mar. 2015), pp. 215-226. (p. 5.)

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. (pp. 5, 12.)

— (2008), "What the Seller Won't Tell You: Persuasion and Disclosure in Markets", *Journal of Economic Perspectives*, 22, 2 (Mar. 2008), pp. 115-131. (p. 12.)

MILGROM, PAUL R. and JOHN ROBERTS (1986), "Relying on the Information of Interested Parties", *The RAND Journal of Economics*, 17, 1, p. 18. (pp. 7, 12.)

MYERSON, ROGER B. (1986), "Multistage Games with Communication", *Econometrica*, 54, 2 (Mar. 1986), p. 323. (p. 12.)

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. 5.)

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.)

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. 5.)

Schipper, Burkhard C. and Hee Yeul Woo (2019), “Political Awareness and Microtargeting of Voters in Electoral Competition”, *Quarterly Journal of Political Science*, 14, 1, pp. 41-88. (p. 6.)

Wylie, Christopher (2019), *Mindf*ck: Cambridge Analytica and the Plot to Break America*, Random House, p. 270. (p. 4.)

Appendix: Omitted Proofs

Definition A.1. *In equilibrium (σ, a, q), for every receiver $i \in I$, let*

- $\mathcal{M}_i := \{m \subseteq \Omega \mid a_i(m) = 1\}$ *be the set of messages that convince receiver i to approve;*
- $\mathcal{W}_i := \{\omega \in \Omega \mid \exists m \in \mathcal{M}_i \text{ s.t. } \omega \in m\}$ *be the set of states in which the sender has access to at least one message that convinces receiver i to approve;*
- $\overline{\mathcal{W}}_i := \{\omega \in \Omega \mid \alpha_i(\omega) = 1\} \subseteq \mathcal{W}_i$ *be the set of states in which this receiver approves the proposal with probability 1.*

Note that $\mathcal{A}_i \subseteq \mathcal{W}_i$: if $\omega \in \mathcal{A}_i$, then $\{\omega\} \in \mathcal{M}_i$ because $q_i(\cdot \mid \{\omega\}) = p(\cdot \mid \{\omega\}) \in \mathcal{B}_i$. Also, $\forall \omega \in \overline{\mathcal{W}}_i$, $\int_{\mathcal{M}_i} \sigma_i(m \mid \omega) dm = 1$, i.e. to convince the receiver in state ω with certainty, the sender must be sending her convincing messages, and convincing messages only.

Lemma A.1. *In equilibrium (σ, a, q), for every receiver $i \in I$, the set $\overline{\mathcal{W}}_i \cup \mathcal{A}_i$ satisfies receiver i's (obedience) constraint, i.e. $p(\cdot \mid \overline{\mathcal{W}}_i \cup \mathcal{A}_i) \in \mathcal{B}_i$.*

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

$$\int_{\operatorname{supp} q_i(\cdot \mid m)} \delta_i(\omega) \cdot q_i(\omega \mid m) d\omega \geq 0.$$

Notice that $\operatorname{supp} q_i(\cdot \mid m) \subseteq m$ because messages are verifiable. Furthermore, $m \subseteq \mathcal{W}_i$ because if $\omega \in m$ such that $m \in \mathcal{M}_i$, then $\omega \in \mathcal{W}_i$. On the equilibrium path, the inequality above 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 $\overline{\mathcal{W}}_i$, $\mathcal{A}_i \setminus \overline{\mathcal{W}}_i$, and $\mathcal{W}_i \setminus (\overline{\mathcal{W}}_i \cup \mathcal{A}_i)$ and observe that

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

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

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

As a result,

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

PROOF OF THEOREM 1 (SEE PAGE 11).

THEOREM 1. *Suppose $n = 1$. Then, every equilibrium outcome is deterministic. Furthermore, $W \subseteq \Omega$ is an equilibrium set of approved states if and only if it satisfies the sender's* (IC) *and the receiver's* (obedience) *constraints.*

PROOF. *Part I*: suppose, on the contrary, that there exists a non-deterministic equilibrium outcome α with $\alpha(\omega) \in (0,1)$ for some $\omega \in \Omega$. Then, $\alpha(\omega) > 0$ implies $\sigma(m_\omega \mid \omega) > 0$ and $q(\cdot \mid m_\omega) \in \mathcal{B}$ for some $m_\omega \subseteq \Omega$. Then, the sender has a profitable deviation to $\widetilde{\sigma}(m_\omega \mid \omega) = 1$. His payoff in state ω increases from $\alpha(\omega) < 1$ to 1.

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

Definition A.1, $\overline{W} = W$, and by Lemma A.1, the (obedience) constraint holds.

Part III: suppose that $W \subseteq \Omega$ satisfies (IC) and (obedience). Let $\sigma(W \mid \omega) = \mathbb{1}(\omega \in W)$ and $\sigma(\Omega \setminus W \mid \omega) = \mathbb{1}(\omega \in \Omega \setminus W)$ be the sender's strategy. On the path, receiver only hears two messages, W and $\Omega \setminus W$, and her posterior belief is $q(\cdot \mid W) = p(\cdot \mid W) \in \mathcal{B}$ by (obedience) and $q(\cdot \mid \Omega \setminus W) = p(\cdot \mid \Omega \setminus W) \notin \mathcal{B}$. In words, the sender sends two messages and the receiver interprets them as a recommendation to approve or reject. Off-the-path, i.e. following any message $m \neq W, \Omega \setminus W$, let the receiver have "skeptical beliefs"

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

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

that assign positive probability to states within the approval set if and only if the message comprises of these states only. Then, the sender does not have profitable deviations.

Proof of Theorem 2 (See page 13).

THEOREM 2. *When $n = 1$, the sender-preferred set of approved states $\overline{W}$ is characterized by a cutoff value $c^* > 0$ such that*

- *the receiver almost surely approves the proposal if $\delta(\omega) > -c^*$ and rejects it if $\delta(\omega) < -c^*$;*[17]
- *whenever the receiver approves the proposal, her expected net payoff of approval is zero: $\mathbb{E}_p\left[\delta(\omega) \mid \overline{W}\right] = 0$.*

Furthermore, the sender-preferred equilibrium outcome is a commitment outcome.

PROOF. Let $\overline{W}$ solve a relaxed problem

$$\max_{W \subseteq \Omega} \int_W p(\omega) d\omega, \quad \text{subject to} \quad \int_W \delta(\omega) p(\omega) d\omega \geq 0. \tag{2}$$

Since $\delta(\omega) \geq 0$ for every $\omega \in \mathcal{A}$, we have $\mathcal{A} \subseteq \overline{W}$. Hence, $\overline{W}$ also solves (1). Furthermore, (obedience) binds, i.e. $\int_W \delta(\omega) p(\omega) d\omega = 0$. If it does not, increase the value of the objective function while satisfying the constraint. Next, suppose that $\overline{W}$ is not character-

[17] Almost surely with respect to the prior distribution p of the state of the world ω.

ized by a cutoff value of $\delta(\cdot)$. Then, there exist $X, Y \subseteq \Omega$ such that *(i)* $P(X) = P(Y) > 0$; *(ii)* $\forall \omega \in X$, $\forall \omega' \in Y$, $\delta(\omega) < \delta(\omega')$; *(iii)* $X \subseteq \overline{W}$ and $Y \subseteq \Omega \setminus \overline{W}$. In words, the sender-preferred set of approved states includes a positive-measure set X, does not include a positive-measure set Y, yet the receiver has a higher net payoff of approving any state in Y over any state in X.

Let $W^* := \left(\overline{W} \setminus X\right) \cup Y$. The value of the objective function is the same for $\overline{W}$ and W^*:

$$P(\overline{W}) = P(\overline{W} \setminus X) + P(X) = P(\overline{W} \setminus X) + P(Y) = P(W^*).$$

The obedience constraint for $\overline{W}$ is

$$\int\limits_{\overline{W} \setminus X} \delta(\omega) p(\omega) d\omega + \int\limits_{X} \delta(\omega) p(\omega) d\omega = 0.$$

The obedience constraint for W^* is

$$\int\limits_{W^* \setminus Y} \delta(\omega) p(\omega) d\omega + \int\limits_{Y} \delta(\omega) p(\omega) d\omega > 0,$$

where the last inequality follows from *(1)* $W^* \setminus Y = \overline{W} \setminus X$, so the first term in both constraints is the same, and *(2)* $\int_X \delta(\omega) p(\omega) d\omega < \int_Y \delta(\omega) p(\omega) d\omega$, so the second term in the second constraint is strictly larger.

We have found that W^* retains the sender's ex-ante utility at the same level as $\overline{W}$. At the same time, the obedience constraint for $\overline{W}$ is binding, whereas for W^* it is loose. Since the obedience constraint is binding at the optimum, W^*, and thus $\overline{W}$, do not maximize the objective function, which brings us to a contradiction. Hence, $\overline{W}$ is characterized by a cutoff value of the receiver's net payoff of approval $\delta(\cdot)$.

Next, I show that the sender-preferred equilibrium outcome $\overline{\alpha}(\omega) := \mathbb{1}(\omega \in \overline{W})$ is a commitment outcome. 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

$$\max_{\alpha} \int_{\Omega} \alpha(\omega)p(\omega)d\omega, \quad \text{subject to} \quad \begin{array}{c} \forall \omega \in \Omega, 0 \leq \alpha(\omega) \leq 1, \\ \int_{\Omega} \delta(\omega) \cdot \alpha(\omega)p(\omega)d\omega \geq 0. \end{array} \tag{3}$$

Observe that any commitment outcome α^{BP} is characterized by a cutoff value $c^{BP} > 0$, meaning that

$$\begin{aligned} \alpha^{BP}(\omega) &= 1 && \text{if } \delta(\omega) > -c^{BP}, \\ \alpha^{BP}(\omega) &\in [0,1], && \text{if } \delta(\omega) = -c^{BP}, \\ \alpha^{BP}(\omega) &= 0, && \text{if } \delta(\omega) < -c^{BP}. \end{aligned}$$

α^{BP} is characterized by a cutoff value for the same reason why $\overline{W}$ is. If it was not, then there exist $X, Y \subseteq \Omega$ such that

- $\int_X \alpha^{BP}(\omega)p(\omega)d\omega = \int_Y (1 - \alpha^{BP}(\omega))p(\omega)d\omega$;
- $\forall \omega \in X$, $\forall \omega' \in Y$, $\delta(\omega) < \delta(\omega')$;
- $\forall \omega \in X$, $\alpha^{BP}(\omega) > 0$ and $\forall \omega \in Y$, $\alpha^{BP}(\omega) < 1$.

Then, letting $\alpha^*(\omega) = \alpha^{BP}(\omega)$ for all $\omega \notin X \cup Y$, $\alpha^*(\omega) = 1$ if $\omega \in Y$, $\alpha^*(\omega) = 0$ if $\omega \in X$ leads to the same level of the objective function and a looser constraint.

Notice that the problem of finding the sender-preferred equilibrium set of approved states (2) is the sender's problem under commitment (3) with an additional constraint $\alpha(\omega) \in \{0,1\}$ for every $\omega \in \Omega$. Hence, if there exists a deterministic commitment outcome $\widetilde{\alpha}(\omega) := \mathbb{1}(\omega \in \widetilde{W})$, then $\widetilde{W} = \overline{W}$, meaning that the sender-preferred equilibrium outcome is a commitment outcome.

Next, taking an arbitrary commitment outcome α^{BP}, let $\mathcal{D} := \{\omega \in \Omega \mid 0 < \alpha^{BP}(\omega) < 1\}$ be the set of states the receiver approves and rejects with a positive probability. Since α^{BP} is characterized by the cutoff value c^{BP}, for every $\omega \in \mathcal{D}$, $\delta(\omega) = -c^{BP}$.

Next, let $\widetilde{\alpha}(\omega) = \alpha^{BP}(\omega)$ for all $\omega \notin \mathcal{D}$ and $\widetilde{\alpha}(\omega) = \mathbb{1}(\omega \in X)$ for all $\omega \in \mathcal{D}$, where $X \subseteq \mathcal{D}$

solves

$$\int_{\mathcal{D}} \alpha^{BP}(\omega) \cdot p(\omega) d\omega = \int_{\mathcal{D}} \widetilde{\alpha}(\omega) \cdot p(\omega) d\omega = P(X).$$

Now compare the commitment outcome α^{BP} and the candidate outcome $\widetilde{\alpha}$, keeping in mind that they only differ on $\mathcal{D}$. The value of the sender's objective function is the same:

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

the constraint is also the same:

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

Consequently, $\widetilde{\alpha}(\omega) = \mathbb{1}(\omega \in \mathcal{D}_1 \cup X)$ is a *deterministic* commitment outcome. As a result, the sender-preferred equilibrium outcome $\overline{\alpha}(\omega) = \mathbb{1}(\omega \in \mathcal{D}_1 \cup X)$ is a commitment outcome.

PROOF OF THEOREM 3 (SEE PAGE 15).

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

1. *$\overline{u}_s$ is reached in equilibrium;*
2. *$\overline{u}_s$ is given by*

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

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

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

PROOF. $\Rightarrow$: consider equilibrium outcome α with the ex-ante utility of the sender $\overline{u}_s$. 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. For very $i \in I$, set $W_i = X_i \cup \mathcal{A}_i$ satisfies the sender's (IC) constraint, and by Lemma A.1, W_i also satisfies receiver i's (obedience) constraint.

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

$$\int_{\Omega} u_s(\{i \in I \mid \omega \in W_i\}) \cdot p(\omega) d\omega,$$

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

- with certainty, then $\omega \in W_i$;
- with probability less than 1 and $\omega \in \mathcal{A}_i$, then her action is inconsequential to the sender's utility; adding ω to W_i does not change the sender's utility in state ω;
- with probability less than 1 and $\omega \notin \mathcal{A}_i$, then her action is inconsequential to the sender's utility; removing ω to W_i does not change the sender's utility in state ω.

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

$\Leftarrow$: consider collection $(W_1, \dots, W_n)$ of receivers' sets of approved states, each of which satisfies the sender's (IC) and receiver's (obedience) constraints. Then, let the sender's strategy satisfy $\sigma_i(W_i \mid \omega) = \mathbb{1}(\omega \in W_i)$ and $\sigma_i(\Omega \setminus W_i \mid \omega) = \mathbb{1}(\omega \in \Omega \setminus W_i)$, for every receiver $i \in I$. Then, given the same skeptical off-the-path beliefs of the receivers as in Theorem 1, none of the players have profitable deviations and the direct implementation constitutes an equilibrium.

PROOF OF THEOREM 4 (SEE PAGE 16).

THEOREM 4. *The sender's ex-ante payoff in the sender-preferred equilibrium is the commitment payoff.*

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 $(S_1, \dots, S_n)$, 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. The commitment outcome is

$$\alpha^{BP} = \arg\max_{\alpha_i, \forall i \in I} \int_{\Omega} \sum_{T \subseteq 2^I} \alpha(T,\omega) \cdot u_s(T) \cdot p(\omega) d\omega, \text{ subject to } \forall i \in I$$

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

where $\alpha(T,\omega) := \prod_{i \in T} \alpha_i(\omega) \cdot \prod_{j \in I \setminus T} (1 - \alpha_i(\omega))$ is the probability that receivers in $T \subseteq I$ approve the proposal and the receivers in $I \setminus T$ reject it. Notice that if $\alpha_i(\omega) = \mathbb{1}(\omega \in W_i^j)$ for all $i \in I$, then $\alpha(T,\omega) = \mathbb{1}(T = \{i \in I \mid \omega \in W_i\})$, and the sender's problem becomes

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

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 approved states W_i induced by that outcome satisfies the sender's (IC) constraint. I construct a deterministic commitment outcome $\widetilde{\alpha}$ in a sequence of steps.

Step 0: start with $\widetilde{\alpha} = \alpha^{BP}$;

Step 1: if, for some $i \in I$ and $\omega \in \mathcal{A}_i$, $\alpha_i^{BP}(\omega) < 1$, then let $\widetilde{\alpha}_i(\omega) = 1$. 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 payoff in state ω does not strictly increase in receiver i's action;

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

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

Rewrite receiver i's obedience constraint as

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

Since $\alpha_i(\omega) \in (0,1)$ on $\mathcal{D}_i$, then $\delta_i(\omega)$ is constant on $\mathcal{D}_i$. Next, let $\widetilde{\alpha}_i(\omega) = \mathbb{1}(\omega \in X)$ for all $\omega \in \mathcal{D}$, where $X_i \subseteq \mathcal{D}_i$ solves

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

<u>*Step 4*</u>: if for $i \in I$ and $\omega \in \Omega$, $\alpha_i(\omega) \in \{0,1\}$, then let $\widetilde{\alpha}_i(\omega) = \alpha_i(\omega)$.

At this point, $\widetilde{\alpha}_i$, $\forall i \in I$, is a deterministic commitment outcome that satisfies all of the sender's (IC) constraints. Consequently, it is also the sender-preferred equilibrium outcome.

PROOF OF COROLLARY 1 (SEE PAGE 18).

COROLLARY 1. *If L and R are representative voters, then*

1. *if L (R) prefers to approve challenger's policy, then so does every voter with a bliss point to her left (right), i.e.*

$$\mathcal{A}_L \subset \mathcal{A}_i \text{ and } \mathcal{B}_L \subset \mathcal{B}_i, \forall i \in I \text{ such that } v_i < v_L,$$

$$\mathcal{A}_R \subset \mathcal{A}_j \text{ and } \mathcal{B}_R \subset \mathcal{B}_j, \forall j \in I \text{ such that } v_j > v_R;$$

2. *approval sets and sets of approval beliefs of voters L and R do not intersect, i.e.*

$$\mathcal{A}_L \cap \mathcal{A}_R = \varnothing \text{ and } \mathcal{B}_L \cap \mathcal{B}_R = \varnothing.$$

PROOF. 1. By contradiction, suppose $\exists q \in \mathcal{B}_L$ such that $q \notin \mathcal{B}_i$. Notice that because $v_i < v_L < \omega_0$, we have $|v_i - \omega| = |v_i - v_L| + |v_L - \omega_0|$, that is, v_L is located between v_i and ω_0. Then,

$$q \notin \mathcal{B}_i \iff \mathbb{E}_q[|v_i - \omega|] > |v_i - \omega_0| - \varepsilon = |v_i - v_L| + |v_L - \omega_0| - \varepsilon \overset{q \in \mathcal{B}_L}{\geq} |v_i - v_L| + \mathbb{E}_q[|v_L - \omega|].$$

We have arrived at a violation of the triangle inequality, which for every $\omega \in \Omega$ states that $|v_i - \omega| \le |v_i - v_L| + |v_L - \omega|$. $\mathcal{B}_L \subseteq \mathcal{B}_i$ implies that $\mathcal{A}_L \subseteq \mathcal{A}_i$, because $\omega \in \mathcal{A}$ if and only if belief that puts probability 1 on state ω belongs to $\mathcal{B}$. The proof for voter R is analogous.

2. By the definition of the set of approval beliefs, for every $i \in \{L, R\}$

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

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

$$q \in \mathcal{B}_L \cap \mathcal{B}_R \implies \int_{\Omega} [|v_L - \omega| + |\omega - v_R|] \cdot q(\omega) d\omega \le |v_L - \omega_0| + |\omega_0 - v_R| - 2\varepsilon < |v_L - v_R|.$$

The right hand side violates the triangle inequality, which states that $|v_L - \omega| + |\omega - v_R| \ge |v_L - v_R|$ for every $\omega \in \Omega$. This proves that $\mathcal{B}_L \cap \mathcal{B}_R = \varnothing$. Since $\mathcal{B}_i$ includes beliefs that put probability 1 on $\omega \in \mathcal{A}_i$ for $i \in \{L, R\}$, $\mathcal{A}_L \cap \mathcal{A}_R = \varnothing$.

Proof of Theorem 5 (See page 19).

THEOREM 5. *In the sender-preferred equilibrium of an unwinnable election with representative voters L and R, if ε is small enough,*

- *the set of approved policies $\overline{W}_i$ of voter $i \in \{L, R\}$ is an interval $[a_i, b_i] \supset \mathcal{A}_i$;*
- *the challenger wins the election if his policy is in the interval $[a_R, b_L]$ with $a_R < \omega_0 < b_L$;*
- *the challenger's ex-ante odds of winning the election are positive. i.e. $P([a_R, b_L]) > 0$.*

PROOF. Recall that $\delta_i(\omega) = |v_i - \omega_0| - |v_i - \omega| - \varepsilon$ is voter i's net payoff of approval. Her (obedience) constraint is:

$$p(\cdot \mid W_i) \in \mathcal{B}_i \iff \int_{W_i} \delta_i(\omega) p(\omega) d\omega \ge 0$$

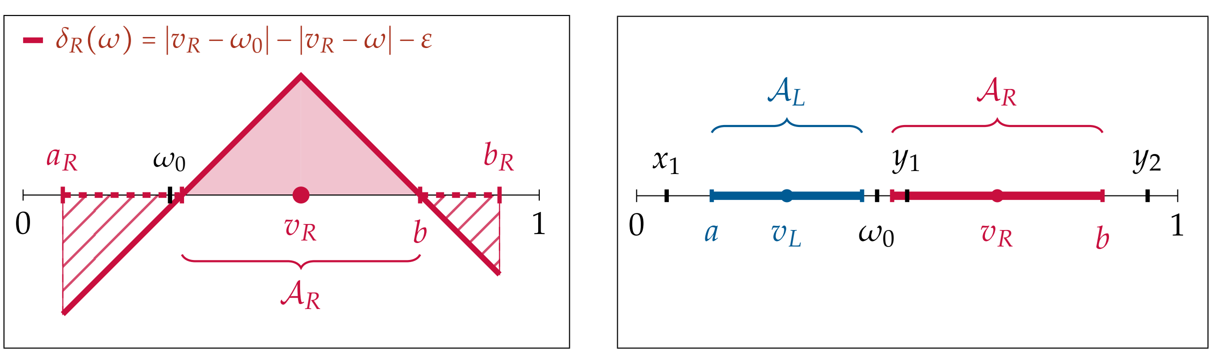
$$\iff \int_{W_i \setminus \mathcal{A}_i} - \underbrace{\delta_i(\omega)}_{<0, \forall \omega \notin \mathcal{A}_i} \cdot p(\omega) d\omega \le \int_{\mathcal{A}_i} \underbrace{\delta_i(\omega)}_{>0, \forall \omega \in \mathcal{A}_i} p(\omega) d\omega := \mathcal{I}_i.$$

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 *(a)* 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. 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.

Relying on these observations, the following arguments, illustrated in Figure 6, part *(b)*, prove that $W_L = [a_L, b_L]$ with $a_L \leq a$ and $b_L > \omega_0 - \varepsilon$. Letting $a = 2v_L - \omega_0 + \varepsilon$ and $b = 2v_R - \omega_0 - \varepsilon$ be the left boundary of L's approval set and right boundary of R's approval set, respectively, we get

- $[a, \omega_0 - \varepsilon] \subseteq W_L$ because it is the approval set of voter L;
- if $x_1 \in [0, a)$ and $x \in W_L$, then $\forall y_1 \in [\omega_0 - \varepsilon, b]$ such that $|a - x_1| \geq |y_1 - \omega_0 + \varepsilon|$, $y_1 \in W_L$;
- if $x_1 \in [0, a)$ and $x \in W_L$, then $\forall x \in (x_1, a]$, $x \in W_L$;
- if $y_1 \in (\omega_0 - \varepsilon, b]$ and $y_1 \in W_L$, then $\forall y \in [\omega_0 - \varepsilon, y_1)$, $y \in W_L$;
- if $y_2 \in (b, 1]$ and $y_2 \in W_L$, then $\forall y \in [\omega_0 - \varepsilon, y_2)$, $y \in W_L$.

Finally, $b_L > \omega_0 - \varepsilon$ because $\mathcal{I}_L > 0$, and for small enough ε, $b_L > \omega_0$.



(a) *Voter R's net payoff of approval. 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 sender-preferred convincing messages are intervals.*

PROOF OF THEOREM 6 (SEE PAGE 21).

THEOREM 6. *Suppose that $p \sim U[0,1]$. In the sender-preferred equilibrium of an unwinnable election with representative voters L and R,*

- *as R becomes more persuadable, the challenger's ex-ante odds of winning $P([a_R, b_L])$ increase;*
- *suppose $|v_L - \omega_0| = |v_R - \omega_0|$, meaning that neither voter is more persuadable than the other. Then, as R becomes more persuadable, the set of challenger's winning policies $[a_R, b_L]$ shifts to the left, i.e. a_R and b_L decrease.*

PROOF. Given convincing message $[a_R, b_R] \supseteq [\omega_0 + \varepsilon, 2v_R - \omega - \varepsilon] = \mathcal{A}_R$, voter R's obedience constraint becomes

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

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

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

Now suppose $|v_L - \omega_0| = |v_R - \omega_0|$ and let $a = 2v_L - \omega_0 + \varepsilon$ be the left boundary of L's approval set, and let $b = 2v_R - \omega_0 - \varepsilon$ be the right boundary of R's approval set. Voters' (obedience) constraints are symmetric about ω_0, implying that the solution $\overline{W}_L \cap \overline{W}_R$ is symmetric, as well, i.e. $|b_L - \omega_0| = |\omega_0 - a_R|$. Here, a_R solves voter R's obedience constraint $-\int_{a_R}^{\omega_0+\varepsilon} \delta_R(\omega)d\omega = \int_{\omega_0+\varepsilon}^{b} \delta_R(\omega)d\omega > 0$. For small enough ε, $a_R < \omega_0 - \varepsilon$ (from obedience, $a_R < \omega_0 + \varepsilon$) and $b_R > \omega_0 + \varepsilon$, implying that $a_R < \omega_0 < b_L$ and the challenger swings the election with a positive probability.

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_R and b_L because

- b_L cannot increase because it is obtained from the binding obedience constraint of voter L that was not affected by the change in v_R;
- for high enough v_R, $\int_{\omega_0+\varepsilon}^{b} \delta_R(\omega)d\omega > -\int_{a_R}^{\omega_0+\varepsilon} \delta_R(\omega)d\omega$, 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 $\int_{a}^{\omega_0-\varepsilon} \delta_L(\omega)d\omega \geq -\int_{a_L}^{a} \delta_L(\omega)d\omega - \int_{\omega_0-\varepsilon}^{b_L} \delta_L(\omega)d\omega$. Because b_L is further from v_L than a is, removing $b_L - d$ from the message that convinces voter L and replacing it with $a - d$ (for some $d > 0$) 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;
- 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 changes in a_R and b_L are observed.