PERSUASION WITHOUT COMMITMENT[*]

Maria Titova[†]

September, 2020

LATEST VERSION

Abstract

This paper shows how a sender with state-independent preferences can persuade receivers with verifiable messages and without having commitment power if the state space is continuous. When the sender's messages are verifiable, we 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. That allows us to characterize the full equilibrium set. Unsurprisingly, 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, we study targeted advertising in elections and show that by communicating with voters privately, the politician may win elections that are unwinnable otherwise. That is inefficient because voters regret their choices, and the size of the inefficiency rises as voters' positions become more extreme.

[*]I am very grateful to my advisors Renee Bowen and Joel Sobel for their continuous guidance and support. I thank Simone Galperti, Remy Levin, Aleksandr Levkun, Denis Shishkin, Joel Watson, and 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

We study environments in which a sender communicates his private information to the receivers. The receivers rely on the sender's messages to make choices, knowing two things: the sender's preferences do not depend on his private information, and his messages may contain lies of omission but not lies of commission.

There are many real-life examples of such situations. A politician challenges the incumbent, thus forcing the voting body to choose between his 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 had 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 exaggerate his skillset, but claiming to have qualifications he does not possess will result in him not being to do the job well and eventually getting fired. The CEO may omit some of the financial indicators in his report to the board, but those that he does present must be legitimate.

In all the applications above, the assumption of verifiable disclosure is much more plausible than the assumption of disclosure with commitment.[1] The former states that the sender possesses hard evidence to prove any true claim, meaning that he can be vague and obfuscate information, but he cannot make false statements. Full commitment allows the sender to condition his signals on the realized state of the world. In the context of elections, the former assumption means that the politician knows his policy, can lie by omission but not by commission. The latter assumption allows the politician to condition his political ads on any policy that he may end up being endowed with. The central result of this paper states that the sender does not benefit from commitment, and can persuade the receivers with verifiable information only.

[1]The assumption of ex-ante commitment is central in the Bayesian persuasion literature pioneered by Kamenica and Gentzkow (2011).

MOTIVATING EXAMPLE

Suppose a politician challenges the status quo and attempts to convince the electorate to vote in his favor. He does so by sending private verifiable messages to each voter. Each message is a subset of the policy space, which is a unit interval. He could, for example, fully reveal his policy, in which case the message is a point. He could say nothing, in which case the message is the whole policy space. He could also send an interval (or any other subset of the policy space), and that would contain a range of possible policies that he may have. The only restriction on each message that the challenger may send is that it contains his true policy.

How does the challenger convince just one voter, if that voter is a dictator? The outcome of any equilibrium is a partition of the policy space into winners and losers. The winners (losers) are the policies of the challenger that win (lose) this election in this equilibrium. Rather than studying what strategy of the challenger may have lead to each partition, we implement this outcome directly, instead. In the direct equilibrium, all the winners (losers) pool together and send the same winning (losing) message. Our first result states that direct equilibrium implementation is possible if and only if two constraints on the outcome partition hold. One is the sender's incentive-compatibility (IC) constraint, which states that the sender does not wish to deviate toward a fully informative strategy that induces the receiver to act as if fully informed. Two is the voter's obedience constraint, which forces the voter to interpret the winning (losing) message as a recommendation to vote for the challenger (status quo). Equipped with the direct implementation result, we can characterize the full equilibrium set by restricting attention to outcome partitions that satisfy the two constraints.

In the sender-worst equilibrium, the sender's IC constraint binds, meaning that he does exactly as well as under full disclosure. In other words, information unravels, and the voter learns everything that is relevant for her decision and makes a fully-informed choice.

To find the sender-preferred equilibrium, we maximize the ex-ante measure of the set of winning policies subject to the voter's obedience constraint. In simple words, the challenger obfuscates his policy as much as possible, for as long as the voter still wishes to vote for him after hearing the winning message. The resulting sender-preferred outcome is also a commitment outcome. *This is the main result of this paper: the sender does not benefit from having ex-ante commitment power and can persuade the receiver with verifiable information only.* This result holds in a more general setting with many receivers and many actions of each receiver, for as long as the sender's utility does not depend on the state of the world.

VALUE OF COMMITMENT

To see why the sender does not benefit from commitment, revisit the canonical example from Kamenica and Gentzkow (2011), in which the prosecutor persuades the judge to convict the defendant. The judge prosecutes if the probability that the defendant is guilty is above 0.5. The prior that the defendant is guilty is 0.3. Rather than looking at the signals of the sender or posteriors of the receiver, let us focus on what the receiver does in every state. The authors write that "the judge knows 70 percent of defendants are innocent, yet she convicts 60 percent of them!"[2] How does this happen? When the defendant is guilty (with prior probability of 0.3), the judge convicts them for sure. When they are innocent (with prior probability of 0.7), the judge convicts them with probability 3/7.[3] Can we replicate this outcome if the prosecutor knows the true state, does not have ex-ante commitment power, but can send verifiable messages, which have to include the true state of the world? The answer is no. When the defendant is guilty, the prosecutor has two available messages: "the defendant is guilty" (the fully revealing message) and "the defendant could be innocent or guilty" (the fully uninformative message). Notice that the second message is available in both states. If the prosecutor sends this message when the defendant is guilty and the judge convicts, then nothing prevents the prosecutor from sending the same message when the defendant is innocent. As a result, there is no incentive-compatible way for the prosecutor to get the judge to convict the defendant when they are innocent with a positive probability that is less than one.

Now consider the following continuous interpretation of the same story. Instead of being innocent or guilty, suppose that the defendant could fit on a range from innocent (0% guilty) to 100% guilty. The judge convicts if the defendant is at least 50% guilty.

[2]See Kamenica and Gentzkow (2011), p. 2591.

[3]The signal realization that recommends conviction comes from the state in which the defendant is guilty with probability 1 and from the state in which the defendant is innocent with probability 3/7.

In the solution to the Bayesian persuasion game, the judge convicts 3/7 of the innocent defendants. In other words, the judge acquits defendants who are less than $2/7 \cdot 100 \approx 28.6\%$ guilty. The geometric interpretation can be observed in Figure 1.

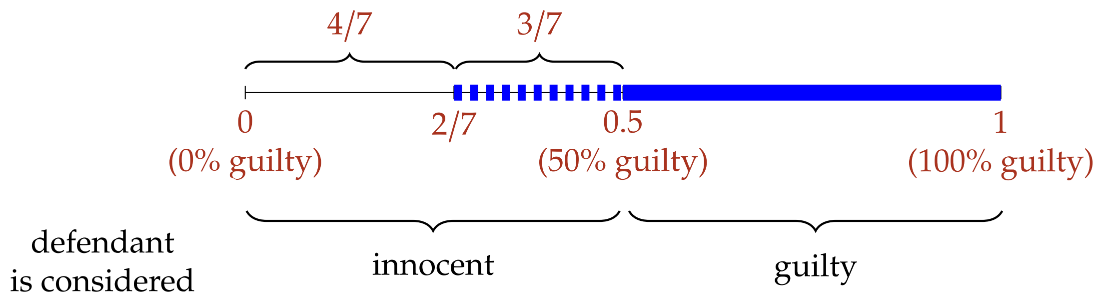


Figure 1: *the continuous interpretation of the judge-prosecutor example. In blue are the defendants who are convicted: all of those who are guilty (solid-blue interval $[0.5,1]$) and 3/7 of those who are innocent (dashed blue interval $[2/7,0.5]$).*

The same outcome can be implemented without commitment but with verifiable information. The prosecutor can say "the defendant is at least 28.6% guilty" whenever that statement is true. Upon receipt of this message, the judge updates her prior belief using the Bayes rule and her response remains the same – she convicts. The prosecutor does not have profitable deviations. When the defendant is indeed at least 28.6% guilty, the prosecutor receives the largest possible payoff. When the defendant is less guilty than that threshold, the prosecutor cannot verifiably claim otherwise. Notice how the prosecutor-preferred threshold of conviction at 28.6% is the same with or without commitment. In Bayesian persuasion, it comes from calculating the optimal *signal* that persuades the judge to convict. In the verifiable-information game, it comes from calculating the sender-preferred *message* that convinces the judge to convict. Regardless of the setup, the constraints to either maximization problem are the same and boil down to the judge interpreting the *signal realization* or the *message* as a recommendation to convict.

SWINGING ELECTIONS

With many receivers, the same theoretical results hold. When characterizing the possible outcomes of an election, we can restrict attention to direct equilibria. In a direct equilibrium, the challenger makes a private recommendation to every voter. Each voter hears one of two messages on the equilibrium path: one convinces her to vote for the challenger, one recommends that she votes for the status quo. Two types of conditions must hold. First, the challenger must guarantee at least the full disclosure payoff. Second, each voter must rationally interpret the convincing message as a recommendation to vote for the challenger. In the sender-worst equilibrium, the challenger fully reveals his policy to every voter, and voters make fully informed choices. In the sender-preferred equilibrium, the challenger's ex-ante odds of winning are the same as under commitment.

The advantage of assuming that the challenger communicates with the voters via verifiable messages, and not by committing to a distribution of signals conditional on his policy, is twofold. First, the former is a more plausible assumption in the context of elections. Second, the problem of maximizing the challenger's ex-ante odds of winning has a simple and intuitive solution that we obtain using the tools developed in this paper. Suppose there are just two voters, the Left voter and the Right voter, whose positions are to the left and the right (of the status quo), respectively. Suppose that the election is decided by unanimity rule and the challenger wins if and only if he convinces both voters to vote for him. Because the voters are located on the opposite sides of the status quo, they never both vote for the challenger at the same time when they hold a common belief. Consequently, this election is unwinnable for the challenger under public disclosure, meaning that he loses it with probability one in every equilibrium when he is restricted to advertising his policy publicly (e.g. in a televised debate).

This is where targeted advertising makes a difference. When communicating his information to each voter privately, the challenger can make the voters believe different things. The Left voter prefers positions to the left (of the status quo), while the Right voter prefers positions to the right. To convince both voters, the challenger must convince the Left voter to vote for some right policies, and vice versa. Using the direct implementation result, we iteratively construct the message that convinces the Left voter. First, we include in this message the left policies that this voter prefers to the status quo under complete information. Next, we start adding policies that are to the right of the status quo. As we add more right policies to the message that convinces the Left voter, and the further to the right those policies are, the less enthusiastic this voter becomes about voting for the challenger after hearing that message. At some point, she becomes indifferent between voting for the challenger and voting for the status quo. At that point, if we add any more right policies, the message will seize to convince the voter, so we stop. The Right voter's convincing message is designed similarly.

As a result, when the challenger's policy is not too far to the left and not too far to the right from the status quo, he convinces both voters and wins the election. Note that any winning policy of the challenger must be located sufficiently close to the status quo, or else one of the voters will not be convinced.

POLARIZATION AND WELFARE

There is a sense in which winning elections that are unwinnable under public disclosure is inefficient. In particular, when the Left (Right) voter votes for the right (left) policies, she is tricked by the challenger into making a mistake which she later regrets. When the challenger wins an unwinnable election, one of the voters always regrets it.

Things get worse 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 Right voter moves further to the right. That makes her more persuadable because she becomes less satisfied with the status quo. 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, the more persuadable the right voter becomes, the more policies to the left of the status quo can be added to the message that convinces her to vote for the challenger. Hence, as the right voter moves to the right, the set of winning policies of the challenger shifts to the left.

RELATED LITERATURE

We 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. In the most common formulation, Bayesian persuasion allows the sender to commit to sending any distribution of messages as a function of the state of the world. Lipnowski and Ravid (2020) consider a setting similar to ours, except their 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. In contrast to their result, we show that the sender never benefits from commitment if he possesses the hard evidence that allows him to send verifiable messages.

It is important to understand why there exist equilibria in which all information does not unravel like it usually happens in the verifiable information literature.[4] In the most common formulation, the sender is a seller who wishes to maximize the quantity sold to the buyer, who is the receiver. The state of the world, known only to the seller, is the quality of the good, which for simplicity we assume is uniform. The buyer wishes to purchase more if the quality is high, and the seller always wants to sell as much as possible. Suppose the sellers with above-average quality pull together and send message $[0.5, 1]$. Under a uniform prior, the buyer rationally expects that the quality is 0.75. In Milgrom and Roberts (1986), her utility is strictly concave, and that implies that she optimally purchases a little less than if she knew that the quality truly was 0.75. That forces the seller to shrink the disclosed interval towards higher average quality until his true type unravels. Grossman (1981) points out that the seller also wants to deviate from the pooling message $[0.5, 1]$. Even if the buyer's preferences were non-concave and best responded to the average expected quality 0.75, sellers whose quality is above that level would strictly benefit from deviating to full disclosure. Both these arguments fail in this paper's setup because the receiver cannot continuously adjust her best response because her action space is finite. In other words, if the report is $[0.5, 1]$, then, even if the receiver's utility is strictly concave, she cannot purchase a slightly lower quantity. Similarly, the good-quality sellers may not profit from deviating to full disclosure because they are already selling the highest possible (discrete) quantity. While the assumption of the discrete action space of the receiver may seem oddly specific, it holds in many applications outlined below.

The main conclusion of this paper allows us to replace the assumption of commitment with a more plausible assumption of verifiable information in the existing applications of Bayesian persuasion models while keeping all the results obtained in this fruitful literature. Equivalence holds as long as *(i)* the sender's utility is state-independent and *(ii)* the receiver's action space is finite or can be made finite. This includes applications

[4] See, e.g., a review by Milgrom (2008).

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); contest organizers persuade participants to exert a higher effort (Feng and Lu, 2016; Zhang and Zhou, 2016). 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 two Bayesian persuasion papers closest to ours are Alonso and Câmara (2016) and Arieli and Babichenko (2019). In the former paper, the sender designs a public signal to reach his objective of persuading a group of voters. According to our results, every election that is unwinnable in their setting could be won (with positive probability) by targeting, provided that the voters are sincere and do not share their private signals. Arieli and Babichenko (2019) provide optimal signals from senders with super- or submodular utilities. According to our equivalence results, the outcome of their game with commitment will also arise in the verifiable-information setting. While voters in our main application are expressive, two papers in the applied information design literature provide insight as to what happens under commitment when voters are strategic. Chan et al. (2019) show that when receivers' voting costs are heterogeneous, the sender's preferred action is adopted more often when private disclosure is allowed, relative to when it is not. Heese and Lauermann (2019) show that when receivers' preferences are heterogeneous and private, the sender needs very little commitment power to achieve his desired outcome.

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, there is literature on how candidates explicitly target voters based on voters' positions on the political spectrum. For example, George and Waldfogel (2006) argue that candidates who target educated voters who care more about global issues post their ads in the *New York Times*, while candidates with more local agenda post in local newspapers. DellaVigna and Kaplan (2007) show that the introduction of Fox News in the local markets in the US increased the Republican vote in the

2000 presidential election. Second, a case can be made that an increase in the availability of information catered toward certain electoral groups also counts as targeted advertising, because these are the messages that are intended for and heard by these groups. Enikolopov et al. (2011) study the 1999 parliamentary election in Russia and show that the introduction of a government-independent TV channel increased the vote for major opposition parties. Oberholzer-Gee and Waldfogel (2009) document that an increase in the availability of Spanish-language news positively affected voter turnout in the US in 1994-2002, suggesting that the reason is that these news outlets are much more likely to report on issues that are of interest to Hispanics. We claim that targeted political advertising may be so widespread because it allows politicians to win elections that are unwinnable otherwise.

The leading application of the model also contributes to the literature on disclosure in electoral competition. When candidates send verifiable messages about their policy, competition usually results in full unraveling of information. The reason is that the candidates play a zero-sum game, and it pushes them to voluntarily disclose all information, however unfavorable, as first shown by Board (2009). Expanding the candidates' options beyond disclosing information about their own position does not help, either. Schipper and Woo (2019) examine the effects of allowing candidates to microtarget voters (who may be unaware that some issues even exist), and, separately, allow candidates to "negatively campaign" by sending verifiable messages about each other. In both cases, the voters end up learning everything and making the same choices they would have made under complete information. Janssen and Teteryatnikova (2017) come to the same conclusion after studying a model of electoral competition in which the candidates are allowed to send (comparative) statements about their position relative to their opponent. While political candidates are symmetric is all these papers, in our application, the challenging candidate has a significant advantage over his opponent in that he is the only one who can send (targeted) messages to voters, perhaps because his budget is much higher, or because he has access to better technology. As a result, the challenger can improve (in some equilibria significantly) his ex-ante odds of winning, relative to the full disclosure benchmark.

Another strand of literature focuses on the inefficiencies that arise from targeted political advertising through media. Perego and Yuksel (2018) observe that the more media outlets there are, the more specialized each of them becomes, ultimately tailoring the news to almost perfectly meet the needs of individual voters. A rise in social disagreement occurs because learning more about ideology comes at a higher cost (to the voters) of learning less about non-policy aspects. Prummer (2020) asserts that more media fragmentation leads to candidates targeting voters with more extreme preferences. If the population of voters is polarized to begin with, so will be the candidates' chosen policy platforms. Hu et al. (2019) also find policy polarization, this time caused by rationally inattentive voters consuming news aggregated by an infomediary. The infomediary prefers to target extreme voters with skewed (towards own party) signals, which is only credible when the candidates actually choose more extreme positions. Polarization does not arise endogenously in our model, but we show that as the electorate becomes more polarized, targeted advertising exacerbates a new kind of inefficiency. Challenging politicians swing unwinnable elections more often, and that comes at the cost of voters regretting their choices with an increasing probability.

2. Model

There is a state space $\Omega := [0, 1]$ and a finite set of receivers $[I] := \{1, \ldots, I\}$. 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 $\Delta\Omega$.[5] After observing 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) $i \in [I]$.[6] Then, receiver i observes her private message m_i (but not the state ω) and decides which action $j_i \in \mathbb{A}$ to take. The receiver's preferences (potentially) depend on the state, while the sender's do not.

I impose the following assumptions. The receiver's action set is $\mathbb{A} := \{0, 1, \ldots, J\}$, where $j \geq 1$ reflects this receiver's approval of proposal j, and $j = 0$ may be thought of as the default action. The sender's utility function is $u_s : \mathbb{A}^I \to \mathbb{R}$, and it is weakly increas-

[5] For any distribution $q \in \Delta\Omega$ and for any subset of the state space $W \subseteq \Omega$, we denote by capital $Q(W)$ the probability measure $Q(W) := \int_W q(\omega) d\omega$ and by $q(\cdot \mid W) \in \Delta\Omega$ the conditional probability distribution $q(\omega \mid W) := \frac{q(\omega)}{Q(W)}$.

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

ing in j_i, $\forall i \in [I]$, meaning that the sender weakly prefers that each receiver approves a proposal with a higher label. Receiver i has a utility function $u_i : \mathbb{A} \times \Omega \to \mathbb{R}$, and her preferences under complete information can be summarized by a partition of the state space Ω into approval sets of each action $j \in \mathbb{A}$:

$$\mathcal{A}_i^j := \Big\{\omega \in \Omega \mid j = \max\big[\arg\max_{j' \in \mathbb{A}} u_i(j', \omega)\big]\Big\}.$$

In words, in each state, receiver i prefers to take action that maximizes her utility, and if she is indifferent, then she approves the proposal with a higher label.[7]

Similarly, under incomplete information, receiver i takes action $j \in \mathbb{A}$ if it maximizes her expected utility, which defines her sets of approval beliefs:

$$\mathcal{B}_i^j := \Big\{q \in \Delta\Omega \mid j = \max\big[\arg\max_{j' \in \mathbb{A}} \mathbb{E}_q u_i(j', \omega)\big]\Big\}.$$

I consider Perfect Bayesian Equilibria (henceforth just *equilibria*) of this game.

Definition 1. *The 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 strategies $\boldsymbol{a} := \{a_i : \mathbb{M} \to \mathbb{A}\}_{i \in [I]}$ and posterior beliefs $\boldsymbol{q} := \{q_i : \mathbb{M} \to \Delta\Omega\}_{i \in [I]}$ of the receivers, such that*

(i) $\forall \omega \in \Omega$, $\sigma(\cdot \mid \omega)$ *is supported on* $\arg\max_{m_1, \dots, m_I} u_s\big(a_1(m_1), \dots, a_I(m_I)\big)$, *s.t.* $\omega \in m_i$, $\forall i \in [I]$.

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

(ii) $\forall m \in \mathbb{M}$, $a_i(m) = j$ *such that* $q_i(\cdot \mid m) \in \mathcal{B}_i^j$;

(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;*[8]

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

In words, *(i)* says that the sender puts positive probability only on collections of messages that maximize his utility; *(ii)* states that each receiver takes the action that maxi-

[7] Throughout the paper, we assume that each receiver breaks ties in favor of the sender-preferred actions.

[8] That is, $\sigma_i(m_i \mid \omega) = \int_{\mathbb{M}^{I-1}} \sigma(\{m_i, m_{-i}\} \mid \omega) dm_{-i}$, where $m_{-i} \in \mathbb{M}^{I-1}$ is the collection of messages sent to every receiver, but i.

mizes her expected utility given her posterior belief, breaking ties in favor of the sender-preferred actions; *(iii)* asserts that receivers' posterior beliefs are Bayes-rational on the equilibrium path, i.e. following any message that is heard in equilibrium with positive probability; *(iv)* says that because all messages are verifiable, the receivers' posterior beliefs must be concentrated on the states in which each message is available to the sender.

The following example introduces the leading application of the present model.

EXAMPLE 1 (TARGETED ADVERTISING IN 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, the outcome of which is decided by some social choice function. 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.[9]

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 for voting for the challenger 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 2.

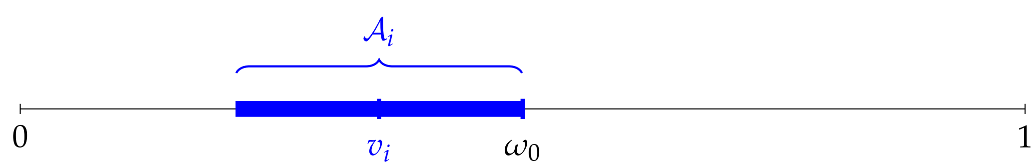


Figure 2: *policy space $\Omega = [0,1]$, status quo policy ω_0, voter i's ideal policy v_i, and her approval set $\mathcal{A}_i$. Under complete information, voter i votes for the challenger if and only if his policy is in the solid blue region.*

I analyze the set of equilibrium outcomes of this game, and we do so via a *state-based*

[9] 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).

approach. That is, in each state of the world ω, we are interested in which actions each receiver takes, and with what probability.

DEFINITION 2.

- *An* <u>*outcome*</u> *is a family of distributions* $\alpha_i(\cdot \mid \omega) \in \Delta\mathbb{A}$ *for every* $i \in [I]$ *and* $\omega \in \Omega$*, where* $\alpha_i(j \mid \omega)$ *is the probability that receiver* $i \in [I]$ *take action* $j \in \mathbb{A}$ *in state* ω*.*
- *An outcome is an* <u>*equilibrium outcome*</u> *if it corresponds to some equilibrium.*[10]

Some outcomes are simpler than others in that every receiver takes a certain action with probability one in each state. Put differently, each receiver essentially partitions the state space into subsets on each of which she takes the same action with certainty. Formally,

DEFINITION 3.

- *an outcome is* <u>*deterministic*</u> *if* $\alpha_i(j \mid \omega) \in \{0, 1\}$ *for every* $i \in [I]$*,* $j \in \mathbb{A}$*,* $\omega \in \Omega$*;*
- *an* <u>*outcome partition*</u> *is a collection of sets* $\{\mathcal{W}_i\}_{i\in[I]} \in \Omega^{I(J+1)}$*, such that*
  - *for every receiver* $i \in [I]$*,* $\mathcal{W}_i := \{W_i^0, \dots, W_i^J\}$ *is a partition of the state space* Ω*;*[11]
  - W_i^j *is the set of states on which receiver* $i \in [I]$ *takes action* $j \in \mathbb{A}$*;*
- *an outcome partition is an* <u>*equilibrium outcome partition*</u> *if it corresponds to some deterministic equilibrium outcome.*[12]

Some equilibria are simpler than others in that the sender directly implements the outcome partition.

DEFINITION 4. *Equilibrium* $(\sigma^D, \boldsymbol{a}^D, \boldsymbol{q}^D)$ *with outcome partition* $\{\mathcal{W}_i\}_{i\in[I]} = \{W_i^0, \dots, W_i^J\}_{i\in[I]}$ *is* <u>*direct*</u> *if for every receiver* $i \in [I]$ *and every action* $j \in \mathbb{A}$*,*

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

[10] Specifically, if there exists equilibrium $(\sigma, \boldsymbol{a}, \boldsymbol{q})$, such that for every receiver $i \in [I]$ and state $\omega \in \Omega$, $\alpha_i(j \mid \omega) = \int_{\mathcal{M}_i^j} \sigma_i(m \mid \omega) dm$, where $\mathcal{M}_i^j(\omega) := \{m \in \mathbb{M} \mid \sigma_i(m \mid \omega) > 0 \text{ and } a_i(m) = j\}$ is the set of messages transmitted from state ω that induce action j. Note that $\alpha_i(\cdot \mid \omega) \in \Delta\mathbb{A}$ because $\int_{\cup_{j\in\mathbb{A}} \mathcal{M}_i^j(\omega)} \sigma_i(m \mid \omega) dm = 1$.

[11] $\{W^0, \dots, W^J\} \in \Omega^J$ is a partition of Ω if $W^j \cap W^{j'} = \varnothing$ for all $j, j' \neq j \in \mathbb{A}$ and $\cup_{j\in\mathbb{A}} W^j = \Omega$.

[12] That is, $W_i^j = \{\omega \in \Omega \mid \alpha_i(j \mid \omega) = 1\}$, for every receiver $i \in [I]$ and action $j \in \mathbb{A}$.

In a direct equilibrium, the sender sends message W_i^j to receiver i if and only if the realized state is $\omega \in W_i^j$. In other words, W_i^j is both the *set of states* in which receiver i takes action j and the *message* that convinces this receiver to play this action. Given the sender's simple pure strategy, each receiver's Bayesian updating is straightforward and 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 $m \in \mathcal{W}_i$. Consequently, for receiver $i \in [I]$ to correctly interpret message W_i^j as a recommendation to play action $j \in \mathbb{A}$, set W_i^j must satisfy the following constraint in any direct equilibrium:

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

When calculating the value of commitment, we compare equilibrium outcomes of the described game with verifiable information 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.

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

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

One Receiver

Direct Implementation

Let us first focus on the case with one receiver, i.e. $[I] = \{1\}$. For ease of exposition, we drop all receiver-relevant subscripts i. Without loss of generality, we assume that the sender's utility is strictly increasing in $j \in \mathbb{A}$.[13] The first technical result of this paper establishes an equivalence between the set of equilibrium outcomes, the set of direct equilibrium outcomes, and provides a simple characterization for both in terms of two types of constraints.

Theorem 1. *Suppose $[I] = \{1\}$. Then, every equilibrium outcome is deterministic. Moreover, the following statements about outcome partition $\mathcal{W} = \{W^0, \ldots, W^J\}$ are equivalent:*

(1) $\mathcal{W}$ is an equilibrium outcome partition;

(2) $\mathcal{W}$ is a direct equilibrium outcome partition;

(3) every $W^j \in \mathcal{W}$ satisfies the receiver's (obedience) constraint; partition $\mathcal{W}$ satisfies the sender's incentive-compatibility constraint

$$\mathcal{A}^j \subseteq W^j \cup \cdots \cup W^J, \quad \forall j \in \mathbb{A}. \tag{IC}$$

Note that the latter constraint simply says that in any (equilibrium) outcome partition, the sender does not have profitable deviations towards full disclosure. That is, if the actual state is $\omega \in \mathcal{A}^j$, then in this state, the sender has to induce action j or higher. If he does not, he has a profitable deviation to the fully revealing message $\{\omega\}$, to which the receiver always responds by playing action j.

The formal proof of Theorem 1 can be found in the appendix, along with all other proofs. Here we describe the intuition behind this result. First of all, every equilibrium outcome is deterministic and the state space can be partitioned into at most $J+1$ sets, and on each of those sets, the receiver plays action $j \in \mathbb{A}$ with certainty. Such partitioning is

[13] If two actions j and j' lead to the same level of sender's utility, i.e. $u_s(j) = u_s(j')$, then one can unite them into one action $\tilde{j}$ and let the corresponding approval set be $\mathcal{A}^{\tilde{j}} = \mathcal{A}^j \cup \mathcal{A}^{j'}$ and the set of approval beliefs be $\mathcal{B}^{\tilde{j}} = \mathcal{B}^j \cup \mathcal{B}^{j'}$.

possible because the sender's utility is strictly increasing in the receiver's action, meaning that he would like to send messages that induce higher actions whenever possible. Thus, if he is inducing two or more actions in the same state, he has a profitable deviation to the message that induces the highest action. As a result, the sender always plays a pure strategy of recommending one of the actions with probability one.[14]

(1) $\Rightarrow$ *(3)*: observe that every outcome partition has to satisfy the sender's (IC) constraint, or else the sender can profitably deviate to full disclosure. To see that any outcome partition also satisfies the obedience constraint for every action, imagine that this partition is implemented directly. Specifically, rather than playing the (possibly mixed) equilibrium strategy, the sender uses a direct pure strategy of sending message W^j if and only if $\omega \in W^j$. Since in the original equilibrium, the receiver interprets every message coming from state $\omega \in W^j$ as a recommendation to take action j, in the direct equilibrium, she also interprets the pooling message W^j the same way.

(3) $\Rightarrow$ *(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 cannot deviate to other messages on the path. 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, we can let the receiver be "skeptical" and assume that any unexpected message comes from the worst possible state.

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

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 his private information (the state of the world) to the mediator, *(ii)* the mediator translating this report into an action recommendation for the receiver,

[14] 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).

and *(iii)* the receiver obediently following her recommendation. Which outcome is implemented is decided by the mediator at step *(ii)*. Conveniently, statement *(3)* of Theorem 1 provides the necessary and sufficient conditions for an outcome partition to be implementable.

EQUILIBRIUM RANGE AND VALUE OF COMMITMENT

When characterizing the full equilibrium set of the game with one receiver, one may restrict attention to outcome partitions 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 $\sum_{j \in \mathbb{A}} u_s(j) \cdot P(W^j)$ and depends only on the outcome partition $\{W^0, \ldots, W^J\}$ and the prior measure $P(\cdot)$.

In the sender-worst equilibrium, the outcome partition minimizes the sender's ex-ante utility subject to the two constraints. The solution results in the partition that binds the sender's incentive-compatibility constraint and is $W^j = \mathcal{A}^j$ for all $j \in \mathbb{A}$. In this equilibrium, the receiver effectively learns all the relevant information for her decision (which approval set has been realized), thus making her decision as if under complete information. This equilibrium is outcome-equivalent to *full disclosure* (also known as *full unraveling*), which is salient in the verifiable information literature.

In the sender-preferred equilibrium, the outcome partition maximizes the sender's ex-ante utility

$$\max_{\{W^0, \ldots, W^J\}} \sum_{j \in \mathbb{A}} u_s(j) \cdot P(W^j)$$

subject to

- the receiver's obedience constraints $p(\cdot \mid W^j) \in \mathcal{B}^j$, $\forall j \in \mathbb{A}$;
- the sender's incentive-compatibility constraints $\mathcal{A}^j \subseteq W^j \cup \cdots \cup W^J$, $\forall j \in \mathbb{A}$.

Consider the binary setup that is common in applications: the receiver only has two actions – to approve the sender's proposal (sender gets $u_s(1) = 1$) or to reject it (sender gets $u_s(0) = 0$). An outcome partition becomes $\{W, \Omega \setminus W\}$, where W denotes the set of approval states and $\Omega \setminus W$ denotes the set of rejection states. To find the sender-preferred equilibrium, one would maximize the ex-ante measure of the set W for as long as message W convinces the receiver to approve. 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 for action "approve". Once the set of approval states is pinned down, the leftover states, in which it is not optimal for the sender to recommend approval, are pooled into recommending rejection. Notice that the receiver's (obedience) constraint for action "reject" is not binding because the sender does not aspire to induce that action as often as possible. Also, neither of the sender's (IC) constraints are binding and the sender does strictly better than under full disclosure. This solution is formalized in the following corollary.

COROLLARY 1. *Suppose* $[I] = \{1\}$, $\mathbb{A} = \{0,1\}$ *and let* $\delta(\omega) := u(1,\omega) - u(0,\omega)$ *be the receiver's net payoff of approval. 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:* $\mathbb{E}_p\big[\delta(\omega) \mid \bar{W}\big] = 0$.

When the receiver has more than two actions, the solution is not as straightforward. The following theorem, which is the main result of this paper, shows that the sender-preferred equilibrium outcome partition of the verifiable-information game solves the sender's problem in the Bayesian persuasion game. Put differently, the sender does not benefit from commitment when his preferences are state-independent and can persuade the receiver with verifiable information only.

THEOREM 2. *When* $[I] = \{1\}$, *the sender-preferred equilibrium outcome is a commitment outcome.*

Recall that an outcome is a distribution of actions that the receiver takes in every state, and $\alpha(j \mid \omega)$ denotes the probability that she takes action $j \in \mathbb{A}$ in state ω. According to Kamenica and Gentzkow (2011), when solving the sender's problem in the Bayesian persuasion game, one may restrict attention to *straightforward* signals that are interpreted by the receiver as recommendations to take particular actions. Consequently, each $\alpha(j \mid \omega)$ becomes the probability that the straightforward signal to take action j came from state ω. Thus, the sender's problem in the Bayesian persuasion game rewrites itself in terms of finding the optimal outcome subject to an obedience-like constraint of the receiver for every action. Of course, there is no guarantee that the commitment outcome results in a partition. Suppose it does not and that on some set of states (call it X) the sender recommends that the receiver plays multiple actions. Why does he do that? He would like to recommend the higher action more often, but cannot do so because the obedience constraint for that action is already binding. Rather than sending the mixed recommendation, the sender could split set X into subsets, and on each of these subsets recommend one action with certainty.[15] Continuity of the state space allows us to perform such partitioning for every action without altering the constraints or the value of the objective function. Finally, the receiver's (obedience) constraints are the same with and without commitment and impose that the receiver correctly interprets the straightforward signal or the convincing message as a recommendation to take 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, we can restrict attention to equilibrium partitions $\{W, \Omega \setminus W\}$, with W being both the convincing message and the set of winning policies for the challenger. The outcome partition must satisfy two conditions. One is the sender's (IC) constraint, $\mathcal{A} \subseteq W$, which ensures that the sender cannot deviate to full disclosure. Two 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, $W = \mathcal{A}$ (thus binding the sender's (IC) constraint). 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* (full unraveling occurs).

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 under the commitment protocol, suggesting that the challenger does not benefit from having the power of com-

[15]For example, if the prior is uniform and it is optimal for the sender to recommend action $j = 0$ with prob. 0.5 and action $j = 1$ with prob. 0.5 on interval $[0, 1]$, then it is also optimal for him to recommend action $j = 0$ with prob. 1 on $[0, 1/2]$ and action $j = 1$ with prob. 1 on $[1/2, 1]$.

mitting ex-ante to a signaling strategy. Corollary 1 states that the sender-optimal set of winning policies is characterized by a cutoff value c^* of the voter's net payoff of voting for the challenger $\delta(\omega) = |v - \omega_0| - |v - \omega|$, 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 3.

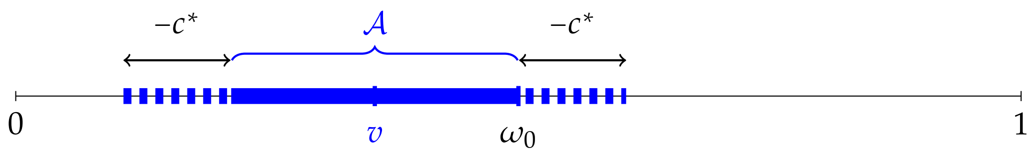


Figure 3: *the challenger-preferred set of winning policies $\bar{W}$ (in blue) consists of the dictator's approval set (solid area) and policies outside of the approval set (dashed area).*

When the receiver's action set is not binary, the problem becomes more complicated and the resulting outcome partition does not feature cutoff states. In the Bayesian persuasion literature, Kolotilin and Wolitzky (2020) assert that it is without loss to focus on pairwise signals so that each induced posterior distribution has at most binary support. Arieli, Babichenko, et al. (2020) focus on the special case where the sender's and receiver's utilities are linear in the unknown state and conclude that all optimal signals are supported on at most two disjoint intervals. Going back to our setup, we can conclude that in the sender-preferred equilibrium outcome partition, the receiver will be taking the same action on disjoint intervals of the state space.

Many Receivers

Direct Implementation

Recall that with many receivers, the outcome partition is the collection $\{\mathcal{W}_i\}_{i \in [I]}$ of partitions $\mathcal{W}_i = \{W_1^0, \ldots, W_i^J\}$, one for each receiver $i \in [I]$, where W_i^j is the set of states in which receiver i takes action $j \in \mathbb{A}$.

When there are many receivers, assuming that the sender's utility is strictly increasing in every receiver's action is no longer without loss of generality.[16] Unfortunately, when that assumption does not hold, there exist non-deterministic equilibria, in which some receivers take multiple actions with positive probability. That happens because the sender's utility does not depend on this receiver's action, and hence the sender does not care to recommend the highest action with certainty.

The argument above implies that we lose the equivalence between sets of equilibrium outcomes and direct equilibrium outcomes. However, suppose we are only interested in the range of the ex-ante utilities that arise in equilibria. In that case, we can still restrict our attention to direct equilibria with outcome partitions that satisfy two types of constraints.

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

$$\bar{u}_s = \sum_{j_1 \in \mathbb{A}} \cdots \sum_{j_I \in \mathbb{A}} u_s(j_1, \ldots, j_I) \cdot P\Big(\bigcap_{i \in [I]} W_i^{j_i} \Big),$$

where $\{\mathcal{W}_i\}_{i \in [I]}$ is an outcome partition such that for every receiver $i \in [I]$

- *every $W_i^j \in \mathcal{W}_i$ satisfies the receiver's (obedience) constraint for action $j \in \mathbb{A}$;*
- *partition $\mathcal{W}_i$ satisfies the sender's incentive-compatibility constraint*

$$\mathcal{A}_i^j \subseteq W_i^j \cup \cdots \cup W_i^J, \quad \forall j \in \mathbb{A}. \tag{IC}$$

The intuition behind this result is similar to that of Theorem 1. When studying the multiple-receiver case, we 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. Another significant difference from the one-receiver case is that now the focus is on the sender's ex-ante utility and not the equilibrium partition, which may not exist.

[16] It does not hold in many important applications, e.g., in most voting models the sender's utility does not depend on the choices of non-pivotal coalitions of voters.

(1) ⇒ (3) states that in terms of the ex-ante utility of the sender, any equilibrium outcome is equivalent to a deterministic outcome with a partition that satisfies two types of constraints. Take any state ω in which in equilibrium, some receiver $i \in [I]$ plays multiple actions with positive probability. Make the following adjustment: let the sender recommend action j, such that $\omega \in \mathcal{A}_i^j$, in this state with certainty. This adjustment must be inconsequential for the sender's ex-ante utility or his incentive-compatibility constraint, or else he would have induced some action with certainty in the original equilibrium. For receiver i, the adjustment loosens her obedience constraint for action j, because the new recommendation to take action j now includes a state in this action's approval set. Having adjusted all the states with previously mixed recommendations, we end up with an outcome partition that satisfies all the sender's (IC) and the receivers' (obedience) constraints.

(3) ⇒ (2): an outcome partition that satisfies receivers' (obedience) and sender's (IC) constraints in itself describes the play in a direct equilibrium. Two things remain to be specified. One is the sender's full messaging strategy, which for every state of the world indicates which action to recommend to each receiver. Two is the "skeptical" beliefs of receivers following unexpected messages that ensure that the sender does not have deviations to off-the-path messages.

(2) ⇒ (1) is trivial because a direct equilibrium is an equilibrium.

EQUILIBRIUM RANGE AND VALUE OF COMMITMENT

Here we characterize the range of ex-ante utilities of the sender that he may achieve in the verifiable-information game. According to Theorem 3, we can restrict our attention to outcome partitions that satisfy the receivers' obedience and the sender's incentive-compatibility constraints. Furthermore, the sender's ex-ante utility only depends on the outcome partition 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 is doing as well as under full disclosure. The corresponding equilibrium partition is $\{\mathcal{A}_i^0, \dots, \mathcal{A}_i^J\}_{i \in [I]}$, and each receiver learns which of her approval sets has been realized. Consequently, every receiver makes her decision as if under complete information.

In the sender-preferred equilibrium, the outcome partition maximizes the sender's ex-ante utility

$$\max_{\mathcal{W}_1,\ldots,\mathcal{W}_I} \sum_{j_1\in\mathbb{A}} \cdots \sum_{j_I\in\mathbb{A}} u_s(j_1,\ldots,j_I)\cdot P\Big(\bigcap_{i\in[I]} W_i^{j_i}\Big),$$

subject to $\forall i \in [I]$:

- $\forall j \in \mathbb{A}$, recommendation W_i^j satisfies receiver's (obedience) constraint $p(\cdot \mid W_i^j) \in \mathcal{B}_i^j$;
- partition $\mathcal{W}_i = \{W_i^0,\ldots,W_0^J\}$ satisfies the sender's (IC) constraint

$$\mathcal{A}_i^j \subseteq W_i^j \cup \cdots \cup W_i^J, \quad \forall j \in \mathbb{A}.$$

When there are many receivers, the sender does not benefit from having commitment power, either.

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

The problem of finding the sender-preferred equilibrium is computationally hard when the sender's preferences are not separable in receivers' actions.[17] In some special cases, particularly when each receiver chooses between two options, 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 the problem is NP-hard and provide an approximation for when it is submodular. In the following section, we 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.

3. APPLICATION: 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, we let $[I] = \{L, R\}$ to reflect that the voters L and R are located on the opposite sides of the

[17] 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).

status quo, $v_L < \omega_0 < v_R$. Notice immediately that L and R are incompatible in that their approval sets and sets of approval beliefs do not intersect:

COROLLARY 2. *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.*

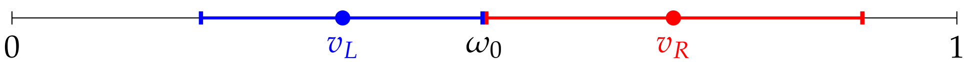


Figure 4: *voters L and R are incompatible and their approval sets do not intersect a.s.*

An interesting implication of Corollary 2 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 of 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, we may restrict attention to outcome partitions $\mathcal{W}_i = \{W_i, \Omega \setminus W_i\}$, one for each voter $i \in \{L, R\}$, where W_i is 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 each voter is simply this voter's approval set, $W_i = \mathcal{A}_i$, $i \in \{L, R\}$. By Corollary 2, 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, consider the following iterative process of designing the challenger-preferred convincing messages $\bar{W}_L$ and $\bar{W}_R$. First, from the challenger's (IC) constraint, $\mathcal{A}_i \subseteq \bar{W}_i$, so we start by letting $W_L = \mathcal{A}_L = [a, \omega_0]$ and $W_R = \mathcal{A}_R = [\omega_0, b]$.[18] At that point, neither voter's (obedience) constraint is binding, and the challenger wins with probability 0. Next, add interval $(\omega_0, \omega_0 + \varepsilon_L]$ to W_L and $[\omega_0 - \varepsilon_R, \omega_0)$ to W_R, for some small $\varepsilon_L, \varepsilon_R > 0$. The set of winning policies becomes $[\omega_0 - \varepsilon_R, \omega_0 + \varepsilon_L]$, meaning that the challenger is now winning the election with positive, albeit small, probability. Notice that by adding $(\omega_0, \omega_0 + \varepsilon_L]$ to W_L, we are only using voter L's (obedience) constraint, since those policies are already in W_R. In that sense, we have increased the challenger's objective (the measure of set $W_L \cap W_R$) at the lowest cost reflected by the two constraints. We can keep skewing the message that convinces voter L (R) towards positions preferred by voter R (L) by increasing ε_L (ε_R), until voter L's (R's) (obedience) constraint binds. If both constraints bind before ε_L (ε_R) reaches the right boundary b of voter R's approval set (the left boundary a of voter L's approval set), we stop, let $b_L = \omega_0 + \varepsilon_L$ and $a_R = \omega_0 - \varepsilon_R$, and conclude that the resulting set of winning policies $[a_R, b_L]$ maximizes the challenger's ex-ante odds of winning. Notice that this set

[18] $a := 2v_L - \omega_0$ denotes the left boundary of L's approval set, $b := 2v_R - \omega_0$ denotes the right boundary of R's approval set

describes policies that are "sufficiently close to the status quo" in the sense that the challenger can be at most ε_R to the left of the status quo and at most ε_L to the left of the status quo, to win. This solution is illustrated in Figure 5.

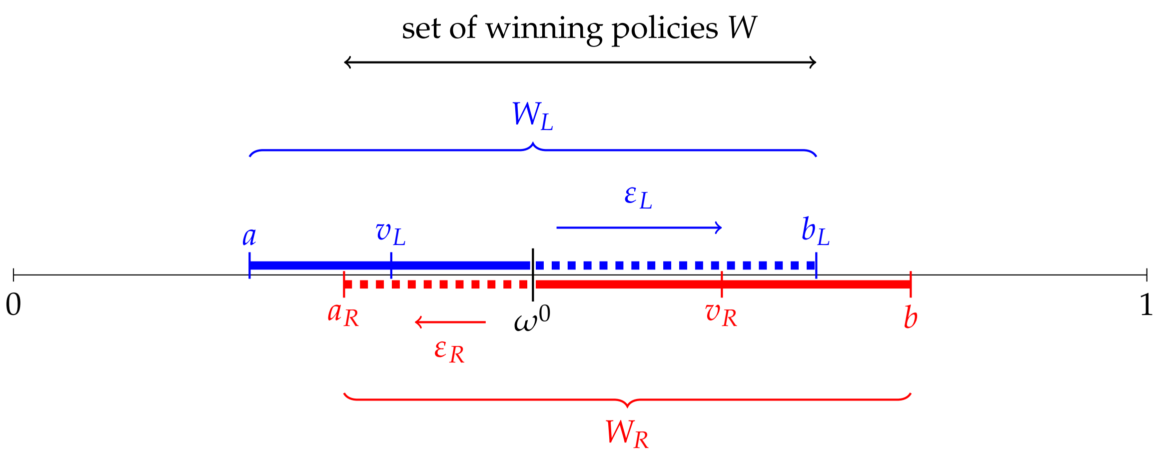


Figure 5: *sender-preferred messages W_L (in blue) and W_R (in red) that convince voters L and R. W_i consists of voter i's approval set (solid) and policies preferred by $j \neq i$ (dashed). Challenger wins the election by convincing both voters when his policy is in $W = W_L \cap W_R$ and sufficiently close to the status quo, $\omega_0 \in W$.*

VOTER WELFARE

We just observed that in the challenger wins with positive probability some elections that are unwinnable under public disclosure. In particular, he swings elections that he would have lost under complete information. In other words, we observe an inefficiency that arises as a result of targeted advertising.

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 $P(W_i \setminus \mathcal{A}_i)$.*

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 realizes she had 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.

COROLLARY 3. *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.*

If at least one of the obedience constraints does not bind when $\varepsilon_L = b$ or when $\varepsilon_R = a$, then the solution may look different, although both convincing messages will still be intervals. To better understand the dynamics, assume for the rest of this section that the prior belief is uniform. In that case, the distance from a voter's bliss point to the status quo measures how persuadable this voter is.[19]

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.*

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). We 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 6 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.*

---

[19] Voter R's (obedience) constraint is $\int_{W_R} \delta_R(\omega)p(\omega)d\omega \geq 0 \iff -\int_{W_R \setminus \mathcal{A}_R} \delta_R(\omega)p(\omega)d\omega \leq \int_{\omega_0}^{b} \delta_R(\omega)p(\omega)d\omega$. Under uniform prior, as v_R increases, so does b, so does the right-hand side of the obedience constraint, thus loosening it and making the voter more persuadable.

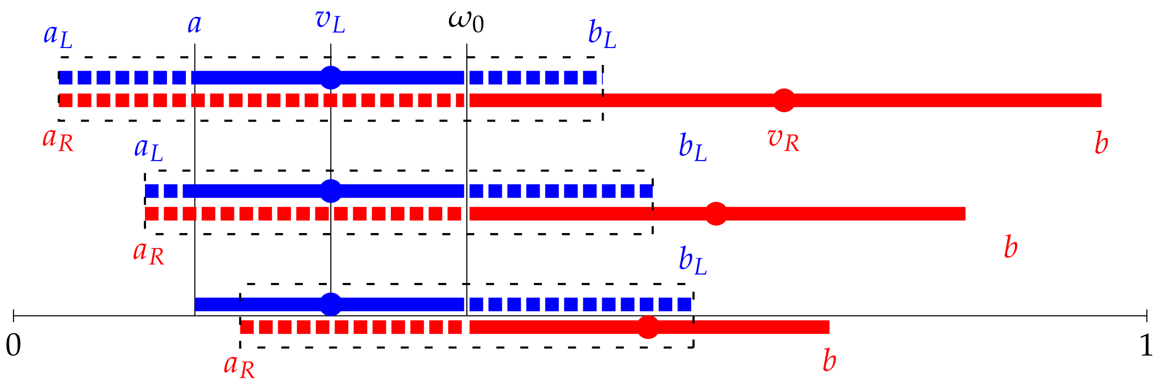


Figure 6: *comparative statics as voter R moves to the right: her approval set (solid red) expands, thus making her more persuadable; regions of regret (dashed line), convincing messages (solid plus dotted lines), and winning states (black dashed area) move to the left.*

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 6), 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 an example, 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.

4. Conclusion and Discussion

This paper argued that the sender can persuade the receivers with verifiable information and without commitment. The assumptions that somewhat restrict the applicability of this result are

(*i*) state-independent preferences of the sender;

(*ii*) no information spillovers or strategic interactions between the receivers.

We need *(i)* to write down the sender's (IC) constraint. If the sender's preferences depend on the state, even with one receiver who chooses between two actions, types in the receiver's approval set may want to be part of the losing message, while types outside of the approval set may want to be part of the winning message. A version of Theorem 1 may still hold for state-dependent preferences of the sender with a more complex incentive-compatibility constraint, but that characterization is left for future research.

When *(ii)* does not hold, the model under commitment becomes a problem of information design (Bergemann and Morris, 2016; Taneva, 2019). It is unclear whether the verifiable-information outcomes will be commitment outcomes in this more general setting, and answering that question is also left for future research.

While we prove that verifiable-information outcome is a commitment outcome, we do not provide the solution to the problem of finding the sender-preferred equilibrium outcome partition. A promising new area of research on the duality approach to information design may shed some light on how to characterize the solution. Particularly close formulations to ours include Kolotilin (2018) and Galperti et al. (2020), whose dual problem is formulated in terms of choosing the "price" of each state to minimize the total expenditure of the information designer.[20]

CONTINUITY OF THE STATE SPACE

The last point we would like to make is concerned with the applicability of this model to games with a finite state space. Because there are many readily-available solutions for games of Bayesian persuasion with a finite state space, it is important to understand how these solutions can be interpreted in terms of verifiable information models with a continuous state space. Furthermore, the conversion of the state space from finite to continuous may become useful in case the problem of finding an optimal partition presented in this paper is easier to solve than the sender's problem in Bayesian persuasion games.

We need the state space to be continuous to make the purification argument that

[20] Another relevant reference is Dworczak and Martini (2019), except the prices in their persuasion economy are assigned to the posterior beliefs induced by signals, rather than the states.

allows us to restrict attention to strategies of the sender that make a deterministic recommendation to each receiver in every state. Once we narrow down the sender's scope of action to partitioning the state space into sets that recommend the same action, we can find the sender-preferred partition that maximizes the sender's utility under receivers' obedience constraints. Under commitment, the continuity of the state space allows us to modify any commitment outcome to take the form of a partition, as well. The receivers' obedience constraints are the same with and without commitment: they simply state that recommendations (whether in the form of a message or a signal realization) are interpreted and followed as intended. Consequently, both problems lead to the same solution.

Revisit the judge-prosecutor example from Kamenica and Gentzkow (2011). The state space is how guilty the defendant really is, and $\omega = 0$ is code for "defendant is absolutely innocent", and $\omega = 1$ is code for "defendant is absolutely guilty". The judge (receiver) wishes to match the state and convict the defendant if and only if they are sufficiently guilty, the threshold for guiltiness being $\omega = 1/2$. The prosecutor (sender) receives 1 if the defendant (them) is convicted, and 0 if acquitted. The full setup for the finite and the continuous state spaces is summarized in the table below.

To solve the Bayesian persuasion problem on the left, one would restrict attention to straightforward signals such that each signal realization is interpreted by the receiver as a recommendation to take a certain action. The realization that recommends conviction ($s = 1$) comes from state $\omega = 1$ with probability 1 (in this state the judge is easily convinced) and with probability α from state $\omega = 0$. By Bayes rule, the receiver's posterior is $q_1(s = 1) = Prob(\omega = 1 \mid s = 1) = \frac{p_1}{p_1 + \alpha p_0}$. In optimum, the realization that recommends conviction ($s = 1$) makes the judge indifferent between convicting and acquitting. That happens when $q_{s=1}(\omega = 1) = 1/2$, so that $\alpha = p_1/p_0$.

To solve the verifiable-information problem on the right, one would restrict attention to direct equilibrium outcome partitions that satisfy the sender's incentive-compatibility and the receiver's obedience constraints. The sender-optimal convincing message includes the receiver's approval set $[1/2, 1]$ plus fraction α of the interval $[0, 1/2]$.[21] Recall that the judge's set of approval beliefs consists of beliefs that place more measure on the

[21] Since the prior conditional on set $[0, 0.5]$ is uniform, it is irrelevant which part of that interval is included in the message, what matters is the length of the included interval.

	Kamenica and Gentzkow (2011)	this paper
State Space	$\Omega = \{0,1\}$	$\Omega = [0,1]$
JUDGE		
action space $\mathbb{A}$	$\{0,1\}$ $j=0$ is acquitting, $j=1$ is convicting	
payoff	1 if $j=\omega$ -1 if $j \neq \omega$	1 if $j = \mathbb{1}\{\omega \geq 1/2\}$ -1 if $j = \mathbb{1}\{\omega < 1/2\}$
	the judge wishes to match the state	
approval set (for conviction)	$\omega = 1$	$[1/2, 1]$
set of approval beliefs (for conviction)	$q(\omega=1) \geq q(\omega=0)$	$Q(\omega \geq 1/2) \geq Q(\omega < 1/2)$
PROSECUTOR		
payoff	$\mathbb{1}\{j=1\}$	
	prosecutor always wants the judge to convict	
Prior Belief	$p(\omega) = \begin{cases} p_0, & \text{if } \omega = 0, \\ p_1, & \text{if } \omega = 1. \end{cases}$	$p(\omega) = \begin{cases} 2p_0, & \text{if } \omega \geq 1/2, \\ 2p_1, & \text{if } \omega < 1/2. \end{cases}$

interval $[0.5, 1]$ than $[0, 1/2]$. Consequently, the judge is exactly indifferent when her prior conditional on the message coming from $[0, 0.5]$, which is αp_0, equals the prior conditional on the message coming from $[0.5, 1]$, which is p_1. Once again, $\alpha = p_1/p_0$.

REFERENCES

ALONSO, RICARDO and ODILON CÂMARA (2016), "Persuading Voters", *American Economic Review*, 106, 11 (Nov. 2016), pp. 3590-3605. (p. 9.)

ARIELI, ITAI and YAKOV BABICHENKO (2019), "Private Bayesian Persuasion", *Journal of Economic Theory*, 182 (July 2019), pp. 185-217. (pp. 9, 24.)

ARIELI, ITAI, YAKOV BABICHENKO, RANN SMORODINSKY, and TAKURO YAMASHITA (2020), "Optimal Persuasion via Bi-Pooling", *SSRN Electronic Journal*. (p. 21.)

ARTABE, ALAITZ and JAVIER GARDEAZABAL (2014), "Strategic Votes and Sincere Counterfactuals", *Political Analysis*, 22, 2 (Jan. 2014), pp. 243-257. (p. 13.)

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. 17.)

Babichenko, Yakov and Siddharth Barman (2016), "Computational Aspects of Private Bayesian Persuasion", pp. 1-16. (p. 24.)

Bardhi, Arjada and Yingni Guo (2018), "Modes of Persuasion toward Unanimous Consent", *Theoretical Economics*, 13, 3, pp. 1111-1149. (p. 9.)

Bergemann, Dirk and Stephen Morris (2016), "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium", *American Economic Review*, 106, 5 (May 2016), pp. 586-591. (p. 30.)

Board, Oliver (2009), "Competition and Disclosure", *The Journal of Industrial Economics*, 57, 1 (Feb. 2009), pp. 197-213. (p. 10.)

Boleslavsky, Raphael and Christopher Cotton (2015), "Grading Standards and Education Quality", *American Economic Journal: Microeconomics*, 7, 2 (May 2015), pp. 248-279. (p. 9.)

Brennan, Geoffrey and Alan Hamlin (1998), "Expressive Voting and Electoral Equilibrium", *Public Choice*, 95, 1-2, pp. 149-175. (p. 13.)

Brennan, Geoffrey and Loren Lomasky (1993), *Democracy and Decision: The Pure Theory of Electoral Preference*, Cambridge University Press. (p. 13.)

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang (2019), "Pivotal Persuasion", *Journal of Economic Theory*, 180 (Mar. 2019), pp. 178-202. (p. 9.)

Crawford, Vincent P. and Joel Sobel (1982), "Strategic Information Transmission", *Econometrica*, 50, 6 (Nov. 1982), p. 1431. (p. 7.)

Degan, Arianna and Antonio Merlo (2007), "Do Voters Vote Sincerely?", *NBER Working Paper 12922*. (p. 13.)

DellaVigna, Stefano and Matthew Gentzkow (2010), "Persuasion: Empirical Evidence", *Annual Review of Economics*, 2, 1 (Sept. 2010), pp. 643-669. (p. 9.)

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. 9.)

Downs, Anthony (1957), "An Economic Theory of Political Action in a Democracy", *Journal of Political Economy*, 65, 2 (Apr. 1957), pp. 135-150. (p. 13.)

Dworczak, Piotr and Giorgio Martini (2019), "The Simple Economics of Optimal Persuasion", *Journal of Political Economy*, 127, 5 (Oct. 2019), pp. 1993-2048. (p. 30.)

Egorov, Georgy and Konstantin Sonin (2019), "Persuasion on Networks", *SSRN Electronic Journal*. (p. 9.)

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. 10.)

Felsenthal, Dan S. and Avraham Brichta (1985), "Sincere and Strategic Voters: An Israeli Study", *Political Behavior*, 7, 4, pp. 311-324. (p. 13.)

Feng, Xin and Jingfeng Lu (2016), "The Optimal Disclosure Policy in Contests with Stochastic Entry: A Bayesian Persuasion Perspective", *Economics Letters*, 147 (Oct. 2016), pp. 103-107. (p. 9.)

Forges, Françoise (1986), "An Approach to Communication Equilibria", *Econometrica*, 54, 6 (Nov. 1986), p. 1375. (p. 17.)

GALPERTI, SIMONE, ALEKSANDR LEVKUN, and JACOPO PEREGO (2020), “The Price of Data”, *Working Paper*. (p. 30.)

GEHLBACH, SCOTT and KONSTANTIN SONIN (2014), “Government Control of the Media”, *Journal of Public Economics*, 118 (Oct. 2014), pp. 163-171. (p. 9.)

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. 9.)

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. 7, 8.)

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. 13.)

HEESE, CARL and STEPHAN LAUERMANN (2019), “Persuasion and Information Aggregation in Elections”, *Working Paper*. (p. 9.)

HU, LIN, ANQI LI, and ILYA SEGAL (2019), “The Politics of Personalized News Aggregation”, *Working Paper* (Oct. 2019). (p. 11.)

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. 10.)

KAMENICA, EMIR (2019), “Bayesian Persuasion and Information Design”, *Annual Review of Economics*, 11, pp. 249-272. (p. 24.)

KAMENICA, EMIR and MATTHEW GENTZKOW (2011), “Bayesian Persuasion”, *American Economic Review*, 101, 6 (Oct. 2011), pp. 2590-2615. (pp. 2, 4, 7, 19, 31, 38, 43.)

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. 13.)

KOLOTILIN, ANTON (2015), “Experimental Design to Persuade”, *Games and Economic Behavior*, 90 (Mar. 2015), pp. 215-226. (p. 9.)

— (2018), “Optimal Information Disclosure: a Linear Programming Approach”, *Theoretical Economics*, 13, 2 (May 2018), pp. 607-635. (p. 30.)

KOLOTILIN, ANTON and ALEXANDER WOLITZKY (2020), “Assortative Information Disclosure”, *SSRN Electronic Journal*, pp. 1-62. (p. 21.)

LIPNOWSKI, ELLIOT and DORON RAVID (2020), “Cheap Talk With Transparent Motives”, *Econometrica*, 88, 4, pp. 1631-1660. (p. 7.)

MILGROM, PAUL R. (1981), “Good News and Bad News: Representation Theorems and Applications”, *The Bell Journal of Economics*, 12, 2, p. 380. (p. 7.)

— (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. 8.)

MILGROM, PAUL R. and JOHN ROBERTS (1986), “Relying on the Information of Interested Parties”, *The RAND Journal of Economics*, 17, 1, p. 18. (pp. 8, 11.)

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. 17.)

MYERSON, ROGER B. (1986), “Multistage Games with Communication”, *Econometrica*, 54, 2 (Mar. 1986), p. 323. (p. 17.)

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. 10.)

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. 9.)

PEREGO, JACOPO and SEVGI YUKSEL (2018), “Media Competition and Social Disagreement”, *Working Paper*. (p. 10.)

PRAT, ANDREA and DAVID STRÖMBERG (2013), “The Political Economy of Mass Media”, *Advances in Economics and Econometrics*, Cambridge University Press, pp. 135-187. (p. 9.)

PRUMMER, ANJA (2020), “Micro-Targeting and Polarization”, *Journal of Public Economics*, 188 (Aug. 2020), p. 104210. (p. 11.)

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. 17.)

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. 9.)

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. 10.)

SPENCE, MICHAEL (1973), “Job Market Signaling”, *The Quarterly Journal of Economics*, 87, 3 (Aug. 1973), p. 355. (p. 7.)

TANEVA, INA (2019), “Information Design”, *American Economic Journal: Microeconomics*, 11, 4 (Nov. 2019), pp. 151-185. (p. 30.)

ZHANG, JUN and JUNJIE ZHOU (2016), “Information Disclosure in Contests: A Bayesian Persuasion Approach”, *The Economic Journal*, 126, 597 (Nov. 2016), pp. 2197-2217. (p. 9.)

APPENDIX: OMITTED PROOFS

LEMMATA A.1. *Let $i \in [I]$, $\mathcal{M} \subseteq \mathbb{M}$ and $W \subseteq \Omega$. If*

- *on set W, the sender only sends messages from set $\mathcal{M}$, i.e. $\forall \omega \in W$, $\int_{\mathcal{M}} \sigma_i(m \mid \omega) = 1$;*
- *each message in $\mathcal{M}$ convinces the receiver to take action j, i.e. $\forall m \in \mathcal{M}$, $q_i(\cdot \mid m) \in \mathcal{B}_i^j$,*

then set W satisfies the receiver's (obedience) constraint $p(\cdot \mid W) \in \mathcal{B}_i^j$.

Proof.

$$\forall m \in \mathcal{M}, \int\limits_W u_i(j,\omega)\cdot q_i(\omega \mid m)d\omega \geq \max_{j' \in \mathbb{A}\setminus\{j\}} \int\limits_W u_i(j',\omega)\cdot q_i(\omega \mid m)d\omega \iff$$

$$\int\limits_W u_i(j,\omega)\cdot \frac{\sigma_i(m \mid \omega)\cdot p(\omega)}{\int_W \sigma_i(m \mid \omega')\cdot p(\omega')d\omega'}d\omega \geq \max_{j' \in \mathbb{A}\setminus\{j\}} \int\limits_W u_i(j',\omega)\cdot \frac{\sigma_i(m \mid \omega)\cdot p(\omega)}{\int_W \sigma_i(m \mid \omega')\cdot p(\omega')d\omega'}d\omega$$

$$\iff \int\limits_W u_i(j,\omega)\cdot \sigma_i(m \mid \omega)\cdot p(\omega)d\omega \geq \max_{j' \in \mathbb{A}\setminus\{j\}} \int\limits_W u_i(j',\omega)\cdot \sigma_i(m \mid \omega)\cdot p(\omega)d\omega.$$

Integrate the above inequality over all $m \in \mathcal{M}$, using the observation that

$$\int\limits_{\mathcal{M}} \int\limits_W f(\omega)\cdot \sigma_i(m \mid \omega)d\omega dm = \int\limits_W f(\omega) \underbrace{\int\limits_{\mathcal{M}} \sigma_i(m \mid \omega)dm}_{=1,\ \forall \omega \in W} d\omega = \int\limits_W f(\omega)d\omega,$$

arrive at

$$\int\limits_W u_i(i,\omega)\cdot p(\omega)d\omega \geq \max_{j' \in \mathbb{A}\setminus\{j\}} \int\limits_W u_i(j',\omega)\cdot p(\omega)d\omega$$

$$\iff \int\limits_W u_i(j,\omega)\cdot \frac{p(\omega)}{P(W)}d\omega \geq \max_{j' \in \mathbb{A}\setminus\{j\}} \int\limits_W u_i(j',\omega)\cdot \frac{p(\omega)}{P(W)}d\omega.$$

Consequently, $q_i(\cdot \mid m) \in \mathcal{B}_i^j, \forall m \in \mathcal{M} \implies p(\cdot \mid W) \in \mathcal{B}_i^j$. Even if the receiver is "indifferent" after any $m \in \mathcal{M}$ and chose action j because it has the highest label, the same can be said about message W.

□

PROOF OF THEOREM 1.

Part I, an outcome partition arises in every equilibrium. Consider equilibrium $(\sigma, \boldsymbol{a}, \boldsymbol{q})$. Suppose, on the contrary, that $\alpha(j \mid \omega) \in (0,1)$ for some $j \in \mathbb{A}$ and $\omega \in \Omega$. In words, in this state, the sender is inducing more than one action, one of the induced actions being j. Then, either *(i)* the sender could deviate to message m that induces action j to convince the receiver to take this action with certainty, or *(ii)* there exists $m' \ni \omega$ that induces action $j' > j$, in which case the sender should deviate to m' and induce a better action j' with certainty. Formally, every message transmitted from state $\omega \in \Omega$ has to induce the same action, i.e. $\forall m \in \mathbb{M}$ such that $\sigma(m \mid \omega) > 0$, $a(m) = j$ for some action $j \in \mathbb{A}$. Let

- $\mathcal{M}^j(\omega) := \{m \in \mathbb{M} \mid \sigma(m \mid \omega) > 0 \text{ and } a(m) = j\}$ be the set of messages transmitted (on-the-path) from state ω that induce action j.

If $\mathcal{M}^j(\omega)$ is non-empty, then action j must be induced from state ω with probability one by the argument above. Mathematically, $\int_{\mathcal{M}^j(\omega)} \sigma(m \mid \omega) dm = 1$, so that $\alpha(j \mid \omega) = 1$. Consequently, $W^0, \ldots, W^J\}$, where $W^j = \{\omega \in \Omega \mid \exists \mathcal{M}^j(\omega) \neq \varnothing\}$ is an outcome partition.

Part II, *(1) ⇒ (3)*: consider equilibrium $(\sigma, \boldsymbol{a}, \boldsymbol{q})$ with outcome partition $\{W^0, \ldots, W^J\}$. This outcome partition must satisfy the sender's (IC) constraint, or else the sender can deviate to full disclosure. To show that it satisfies (obedience), for every action $j \in \mathbb{A}$, let

- $\mathcal{M}^j := \{m \in \mathbb{M} \mid \int_\Omega \sigma(m \mid \omega) d\omega > 0 \text{ and } a(m) = j\}$ be the set of messages that induce receiver action j.

Then, the set of states in which the receiver is convinced to take action j is

$$W^j = \{\omega \in \Omega \mid \exists m \in \mathcal{M}^j \text{ s.t. } \omega \in m\}.$$

To sum up, *(i)* for all $\omega \in W^j$, $\int_{\mathcal{M}^j} \sigma(m \mid \omega) dm = 1$ and *(ii)* $\forall m \in \mathcal{M}^j$, $a(m) = j$. By Lemmata A.1, W^j satisfies the (obedience) constraint for every $j \in \mathbb{A}$.

Part III, *(3) ⇒ (2)*: assume that partition $\{W^0, \ldots, W^J\}$ satisfies (IC) and (obedience). To complete the characterization of the direct equilibrium that induces this outcome partition, one needs to specify off-path beliefs of the receiver, i.e. following any message $m \in \mathbb{M} \setminus \{W^0, \ldots, W^J\}$. 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 toward inducing higher-labeled actions. One way to ensure that does not happen is to impose "skeptical beliefs"

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

$$\forall m \nsubseteq \mathcal{A}^j, \forall j \in \mathbb{A}, \ \operatorname{supp} q^D(\cdot \mid m) \subseteq S, \text{ where } S = \min_{j \in \mathbb{A}} m \cap \mathcal{A}^j \text{ s.t. } S \text{ is not empty},$$

which assign positive probability mass to the states that under complete information induce the lowest possible action, thus preventing the sender from deviating.

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

PROOF OF THEOREM 2.

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 $\{s^0, s^1, \ldots, s^J\}$. Signal realization s^j is then interpreted by the receiver as a recommendation to take action $j \in \mathbb{A}$, leading to an appropriate obedience-like constraint on the posterior belief $q^{BP,j}$ induced by this realization. Letting $\alpha(j \mid \omega) := \text{Prob}(s^j \mid \omega)$, one can rewrite the sender's problem in the Bayesian persuasion game as

$$\max_{\alpha} \sum_{j \in \mathbb{A}} \int_{\Omega} u_s(j) \cdot \alpha(j \mid \omega) p(\omega) d\omega$$

subject to

- $\forall \omega \in \Omega, \alpha(\cdot \mid \omega) \in \Delta\mathbb{A}$;
- receiver's obedience constraint $q^{BP,j}(\cdot) \in \mathcal{B}^j$, for all $j \in \mathbb{A}$, which expands to

$$\int_{\Omega} u(j, \omega) \cdot \underbrace{\frac{\alpha(j \mid \omega) \cdot p(\omega)}{\int_{\Omega} \alpha(j \mid \omega') \cdot p(\omega') d\omega'}}_{= \text{Prob}(\omega \mid s^j) =: q^{BP,j}(\omega)} d\omega \geq \max_{k \in \mathbb{A} \setminus \{j\}} \int_{\Omega} u(k, \omega) \cdot \frac{\alpha(j \mid \omega) \cdot p(\omega)}{\int_{\Omega} \alpha(j \mid \omega') \cdot p(\omega') d\omega'} d\omega.$$

Notice that if we let $\alpha(j \mid \omega) = \mathbb{1}\{\omega \in W^j\}$ for all $j \in \mathbb{A}$, then the the problem above would exactly coincide with the problem of finding the sender-preferred equilibrium partition $\mathcal{W}$. Thus, it remains to show that in the Bayesian persuasion game, the sender cannot do better than setting $\alpha(j \mid \omega) = \mathbb{1}\{\omega \in W^j\}$ and then solving for $\{W^0, \ldots, W^J\}$.

Suppose that on some set $\mathcal{D} \subseteq \Omega$ of positive measure, the sender is making a mixed recommendation, and action $j \in \mathbb{A}$ is the best action among those that he is recommending:

$$\forall \omega \in \mathcal{D}, \alpha(j \mid \omega) \in (0, 1)$$

$$\text{and } \forall j' \in \mathbb{A} \text{ s.t. } \alpha(j' \mid \omega) \in (0, 1) \text{ on a positively-measured subset of } \mathcal{D}, j > j'.$$

Let $g^j(\omega) := u(j,\omega) - \max_{k\in\mathbb{A}\setminus\{j\}} u(k,\omega)$, receiver's obedience constraint for action $j \in \mathbb{A}$ becomes

$$\int_{\mathcal{D}} g^j(\omega)\alpha(j \mid \omega)p(\omega)d\omega \le \mathcal{I}^j := -\int_{\Omega\setminus\mathcal{D}} g^j(\omega)\alpha(j \mid \omega)p(\omega)d\omega.$$

First observe that the sender's "benefit" of increasing $\alpha(j \mid \omega)$ on $\mathcal{D}$ is $u_s(j)\cdot p(\omega)$ (from the objective function), while the "cost" is $g^j(\omega)\cdot p(\omega)$ (from the obedience constraint). Intuitively, the sender would like to increase $\alpha(j \mid \omega)$ on $\mathcal{D}$, since it's the best recommended action on this set, but he cannot because the constraint is already binding. Unless the "cost" of the constraint is constant, the sender can do better by recommending action j more often. This idea is formalized in the following claim.

CLAIM 1. *$g^j(\omega)$ is constant almost everywhere on $\mathcal{D}$.*

Proof. By contradiction, suppose there are sets $X \subset \mathcal{D}$ and $Y \subseteq \mathcal{D}\setminus X$ of positive measure such that

$$\exists g^X, g^Y \in \mathbb{R}, \text{ such that } g^j(\omega) \le g^X < g^Y \le g^j(\omega'),$$

$$\forall \omega \in X, \forall \omega' \in Y.$$

Because $\alpha(j \mid \omega) \in (0,1)$ on $\mathcal{D}$, make the following adjustment to it: increase it by $\varepsilon^X > 0$ on X and decrease it by $\varepsilon^Y > 0$ on Y, while keeping the value of the objective function constant:

$$\int_{X\cup Y} \alpha(j \mid \omega)p(\omega)d\omega = \int_X \alpha(j \mid \omega)p(\omega)d\omega + \int_Y \alpha(j \mid \omega')p(\omega')d\omega'$$

$$= \int_X \left[\alpha(j \mid \omega) + \varepsilon^X\right]\cdot p(\omega)d\omega + \int_Y \left[\alpha(j \mid \omega') - \varepsilon^Y\right]\cdot p(\omega')d\omega',$$

where $\varepsilon^X P(X) = \varepsilon^Y P(Y)$. Notice that the latter equation only specifies the relative value of $\varepsilon^X/\varepsilon^Y$, and hence this adjustment can always be made while keeping $\alpha(j \mid \omega) + \varepsilon^X \le 1$ for all $\omega \in X$ and $\alpha(j \mid \omega') - \varepsilon^Y \ge 0$ for all $\omega' \in Y$.

Next observe that

$$\varepsilon^Y \int_Y g^j(\omega')p(\omega')d\omega' \ge g^Y\cdot\varepsilon^Y P(Y) = g^Y\cdot\varepsilon^X P(X) > g^X\cdot\varepsilon^X P(X) \ge \varepsilon^X \int_X g^j(\omega)\cdot p(\omega)d(\omega),$$

meaning that

$$\int_X g^j(\omega)\cdot\left[\alpha(j\mid\omega)+\varepsilon^X\right]\cdot p(\omega)d\omega+\int_Y g^j(\omega')\cdot\left[\alpha(j\mid\omega')-\varepsilon^Y\right]\cdot p(\omega')d\omega'$$

$$+\int_{\mathcal{D}\setminus(X\cup Y)} g^j(\omega)\alpha(j\mid\omega)p(\omega)d\omega=\mathcal{I}^j+\varepsilon^X\int_X g^j(\omega)\cdot p(\omega)d(\omega)-\varepsilon^Y\int_Y g^j(\omega')p(\omega')d\omega'<\mathcal{I}^j.$$

Thus, by increasing $\alpha(j\mid\omega)$ by ε^X on the low-cost region X and by decreasing it by ε^Y on the high-cost region Y, we kept the value of the objective function constant, while loosening the constraint. There is a contradiction and the original $\alpha(j\mid\omega)$ on $\mathcal{D}$ is not a solution to the sender's problem under commitment if $g^j(\omega)$ is not constant almost everywhere on $\mathcal{D}$. □

Next partition $\mathcal{D}$ into $\mathcal{D}^j\subset\mathcal{D}$ and $\mathcal{D}^{-j}=\mathcal{D}\setminus\mathcal{D}^j$. Let $\tilde{\alpha}(j\mid\omega)=\mathbb{1}\{\omega\in\mathcal{D}^j\}$, where $\mathcal{D}^j$ solves

$$\int_{\mathcal{D}}\alpha(j\mid\omega)p(\omega)d\omega=\int_{\mathcal{D}}\tilde{\alpha}(j\mid\omega)p(\omega)d\omega=P(\mathcal{D}^j).$$

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

$$\int_{\mathcal{D}} u_s(j)\alpha(j\mid\omega)p(\omega)d\omega=\int_{\mathcal{D}} u_s(j)\tilde{\alpha}(j\mid\omega)p(\omega)d\omega=u_s(j)P(\mathcal{D}^j);$$

the obedience constraint does not change, either:

$$\mathcal{I}^j=\int_{\mathcal{D}} g^j(\omega)\alpha(j\mid\omega)p(\omega)d\omega=\int_{\mathcal{D}} g^j(\omega)\tilde{\alpha}(j\mid\omega)p(\omega)d\omega=\bar{g}P(\mathcal{D}^j),$$

since $g^j(\omega)$ is constant and equals $\bar{g}$ almost everywhere on $\mathcal{D}$.

As a result, when solving the sender's problem in the Bayesian persuasion game, we can restrict attention to $\alpha(j\mid\omega)=\mathbb{1}\{\omega\in W^j\}$ for all $j\in\mathbb{A}$ and solve for $\{W^0,\dots,W^J\}$, which coincides with the problem of finding the sender-preferred equilibrium in the verifiable information game. Consequently, the sender-preferred outcome is a commitment outcome.

Proof of Corollary 1.

Using Theorem 4, the sender's problem under commitment power is

$$\max_{\alpha(\cdot)} \int_{\Omega} \alpha(\omega)\cdot p(\omega)d\omega, \quad \text{subject to} \quad \begin{aligned} &\int_{\Omega} \alpha(\omega)\cdot\delta(\omega)\cdot p(\omega)d\omega \ge 0, \\ &0 \le \alpha(\omega) \le 1 \quad \forall \omega \in \Omega. \end{aligned}$$

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 constraint and hence should be added the set of approved states first. Consequently, there must exist a threshold value c^* of the receiver's net payoff of approving the sender's proposal, and $c^* = \delta(\omega)$ for all ω such that $\alpha(\omega) \in (0,1)$.

In the last proof, we established that when the receiver is indifferent (that is, when $g^1(\omega)$ is constant almost everywhere on $\mathcal{D}$), then one can partition the set $\mathcal{D}$ into two subsets $\mathcal{D}^1$ and $\mathcal{D}^0$, and recommend action $j = 1$ on $\mathcal{D}^1$ and action $j = 0$ on $\mathcal{D}^0$.

Overall, the solution will take the following form:

- $\{\omega \in \Omega \text{ s.t. } \delta(\omega) < c^*\}$ is in the *set of rejection states* W;
- $\{\omega \in \Omega \text{ s.t. } \delta(\omega) > c^*\}$ is in the *set of approval states* $\Omega \setminus W$;
- $\{\omega \in \Omega \text{ s.t. } \delta(\omega) = c^*\}$ is split into $\mathcal{D}^1 \subset W$ and $\mathcal{D}^0 \subset \Omega \setminus W$ according to the obedience constraint.

Proof of Theorem 3.

(1) ⇒ (3): consider equilibrium $(\sigma, \boldsymbol{a}, \boldsymbol{q})$ with the ex-ante utility of the sender $\bar{u}_S$. For each receiver $i \in [I]$, let

- $X_i^j := \{\omega \in \Omega \mid \exists j \in \mathbb{A}, \exists \mathcal{M}_i^j \subseteq \mathbb{M} \text{ s.t. } \int_{\mathcal{M}_i^j} \sigma_i(m \mid \omega)dm = 1 \text{ and } a_i(m) = j, \forall m \in \mathcal{M}_i^j\}$ be the set of states in which the sender convinces this receiver to take action j with certainty;
- $Y_i := \Omega \setminus \left(\bigcap_{j \in \mathbb{A}} X_i^j\right)$ be the set of states in which this receiver is *payoff-irrelevant* for the sender. If there is no action $j \in \mathbb{A}$ that the sender finds it optimal to induce in the considered equilibrium with certainty, then the sender's utility does not depend on

this receiver's action in this state;

- $Y_i^j := Y_i \cap \mathcal{A}_i^j$;
- $W_i^j := X_i^j \cup Y_i^j$.

Next, make the following observations for every receiver $i \in [I]$:

- $\forall j \in \mathbb{A}$, $p(\cdot \mid X_i^j) \in \mathcal{B}_i^j$ by Lemmata A.1;
- $\forall j \in \mathbb{A}$, $p(\cdot \mid W_i^j) \in \mathcal{B}_i^j$ because adding $Y_i^j \subseteq \mathcal{A}_i^j$ to X_i^j loosens the receiver's (obedience) constraint for action j;
- $\mathcal{W}_i := \{W_i^0, \dots, W_i^J\}$ is a partition of Ω. Clearly, X_i^j, Y_i^j, $\forall j \in \mathbb{A}$ are all disjoint. Furthermore, their union adds up to Ω because *(i)* Ω is partitioned into $\{X_i^j\}_{j \in \mathbb{A}}$ and Y_i, while *(ii)* Y_i itself is partitioned into $\{Y_i^j\}_{j \in \mathbb{A}}$;
- $\mathcal{W}_i$ satisfies (IC). First, consider $\omega \in \mathcal{A}_i^J$. The receiver could be payoff-relevant in this state, in which case $\omega \in X_i^J$, or not, in which case $\omega \in Y_i^J$. Either way, $\mathcal{A}_i^J \subseteq W_i^J$. Next, consider $\omega \in \mathcal{A}_i^{J-1}$. If the receiver is payoff relevant, then $\omega \in X_i^{J-1} \cup X_i^J$, and if not, then $\omega \in Y_i^{J-1}$. Either way, $\mathcal{A}_i^{J-1} \subseteq W_i^{J-1} \cup W_i^{J-1}$, and so on for every action $j \in \mathbb{A}$.

Overall, the outcome partition $\{\mathcal{W}_i\}_{i \in [I]}$ satisfies all (obedience) constraints and all (IC) constraints and leads to the same level of ex-ante utility $\bar{u}_s$ as the original equilibrium.

(3) $\Rightarrow$ *(2)*: consider outcome partition $\{\mathcal{W}_i\}_{i \in [I]}$ that satisfies all (obedience) constraints and all (IC) constraints. To complete the characterization of the direct equilibrium with this outcome partition, we need to specify the strategy of the sender:

$$\sigma^D\big(\{m_i\}_{i \in [I]} \mid \omega\big) = 1 \iff \omega \in m_i, \quad \forall m_i \in \mathcal{W}_i, \quad \forall i \in [I].$$

In words, to send collection $\{m_i\}$, where message m_i is intended for receiver $i \in [I]$, the realized state ω must lie within every m_i, whereas message m_i has to come from $\mathcal{W}_i = \{W_i^0, \dots, W_i^J\}$. Since each $\mathcal{W}_i$ is a partition, the marginal probability of receiver i getting recommendation W_i^j is $\sigma_i^D(W_i^j \mid \omega) = 1$ if and only if $\omega \in W_i^j$.

From the (IC) and (obedience) constraints it follows that neither the sender nor the receivers have profitable deviations on the described path. Off-the-path, the same "skeptical beliefs" apply as in the one-receiver case.

(2) $\Rightarrow$ *(1)* is trivially true.

PROOF OF THEOREM 4.

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 $S_i = \{s_i^0, \ldots, s_i^J\}$ is the private set of *straightforward* signal realizations of receiver $i \in [I]$. Signal realization s_i^j is then interpreted by receiver i as a recommendation to take action $j \in \mathbb{A}$, leading to an appropriate obedience-like constraint on the posterior belief $q_i^{BP,j}$ induced by this realization. Letting $\alpha_i(j \mid \omega) := \text{Prob}(s_i^j \mid \omega)$, one can rewrite the sender's problem in the Bayesian persuasion game as

$$\max_{\alpha_1,\ldots,\alpha_I} \sum_{j_1 \in \mathbb{A}} \cdots \sum_{j_I \in \mathbb{A}} u_s(j_1, \ldots, j_I) \int_\Omega \prod_{i \in [I]} \alpha_i(j_i \mid \omega) p(\omega) d\omega$$

subject to

- $\forall \omega \in \Omega$ and $\forall i \in [I]$, $\alpha_i(\cdot \mid \omega) \in \Delta\mathbb{A}$;
- each receiver's obedience constraint $q_i^{BP,j}(\cdot) \in \mathcal{B}_i^j$, for all $j \in \mathbb{A}$, which expands to

$$\int_\Omega u_i(j, \omega) \cdot \underbrace{\frac{\alpha_i(j \mid \omega)(\omega) \cdot p(\omega)}{\int_\Omega \alpha_i(j \mid \omega') \cdot p(\omega') d\omega'}}_{=\text{Prob}(\omega \mid s_i^j) =: q_i^{BP,j}(\omega)} d\omega \geq \max_{k \in \mathbb{A} \setminus \{j\}} \int_\Omega u_i(k, \omega) \cdot \frac{\alpha_i(j \mid \omega) \cdot p(\omega)}{\int_\Omega \alpha_i(j \mid \omega') \cdot p(\omega') d\omega'} d\omega.$$

Notice that if we let $\alpha_i(j \mid \omega) = \mathbb{1}\{\omega \in W_i^j\}$ for all $i \in [I]$ and $j \in \mathbb{A}$, then the problem above would exactly coincide with the problem of finding the sender-preferred equilibrium partition.

We have already established in the proof of Theorem 2 that if the sender is sending a mixed recommendation on some set, then he could partition that set into subsets on each of which he recommends an action with probability one. This adjustment does not alter the sender's ex-ante utility or the receiver's obedience constraint. Consequently, a solution that features a deterministic recommendation in every state always exists.

With many receivers, the same logic applies. Suppose on some set $\mathcal{D}$, the sender recommends two or more actions (highest one being k) to receiver 1 and two or more actions (highest one being l) to receiver 2.

Suppose that receiver i's obedience constraint for action $j \in \{k, l\}$ does not bind. Then the sender's utility does not change when this receiver changes action on set $\mathcal{D}$. This is because his utility is weakly increasing in each receiver's action, which means he could induce action j but chooses not to. Let $\tilde{\alpha}_i(j \mid \omega) = \mathbb{1}\{\omega \in \mathcal{D} \cap \mathcal{A}_i^j\}$. Since action j is the highest-labeled action among those recommended to that receiver, making this adjustment weakly improves the sender's ex-ante utility. By adding states within the approval set for an action, we also loosen the receiver's obedience constraint for that action. Notice that we just "purified" the outcome on set $\mathcal{D}$. After this process is complete for every action of receiver i, she will be receiving deterministic recommendations only from every $\omega \in \mathcal{D}$.

Next suppose the obedience constraint of receiver $i \in \{1, 2\}$ for action $j \in \{k, l\}$ is binding. Let $g_i^j(\omega) := u_i(j, \omega) - \max_{j' \in \mathbb{A} \setminus \{j\}} u_i(j', \omega)$ for $i \in \{1, 2\}$ and $j \in \{k, l\}$. Because the sender is making a mixed recommendation to both these receivers, by Claim 1, $g_i^j(\omega)$ must be constant almost everywhere on $\mathcal{D}$, for $i \in \{1, 2\}$. Next, for each $i \in \{1, 2\}$ and $j \in \{k, l\}$, partition $\mathcal{D}$ into $\mathcal{D}_i^j \subset \mathcal{D}$ and $\mathcal{D}_i^{-j} = \mathcal{D} \setminus \mathcal{D}_i^j$. Let $\tilde{\alpha}_i(j \mid \omega) = \mathbb{1}\{\omega \in \mathcal{D}_i^j\}$, where $\mathcal{D}_i^j$ solves

$$\int_{\mathcal{D}} \alpha_i(j \mid \omega) p(\omega) d\omega = \int_{\mathcal{D}} \tilde{\alpha}_i(j \mid \omega) p(\omega) d\omega = P(\mathcal{D}_i^j).$$

Once again, imposing $\tilde{\alpha}_i(\cdot \mid \omega)$ instead of $\alpha_i(\cdot \mid \omega)$ for $\omega \in \mathcal{D}$ and for $i \in \{1, 2\}$ does not alter the value of the objective function or violate either receiver's obedience constraint.

This is how any outcome $\alpha_i(\cdot \mid \omega) \in \Delta\mathbb{A}$, $\forall i \in [I]$ and $\forall \omega \in \Omega$, that solves the sender's problem in the Bayesian persuasion game can be made deterministic. Once there is a commitment outcome partition, the sender's problem and all the obedience constraints coincide with those in the problem of finding the sender-preferred equilibrium in the verifiable information game.

Proof of Corollary 2

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 $i \in \{L, R\}$,

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

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

Proof of Theorem 5

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

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

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

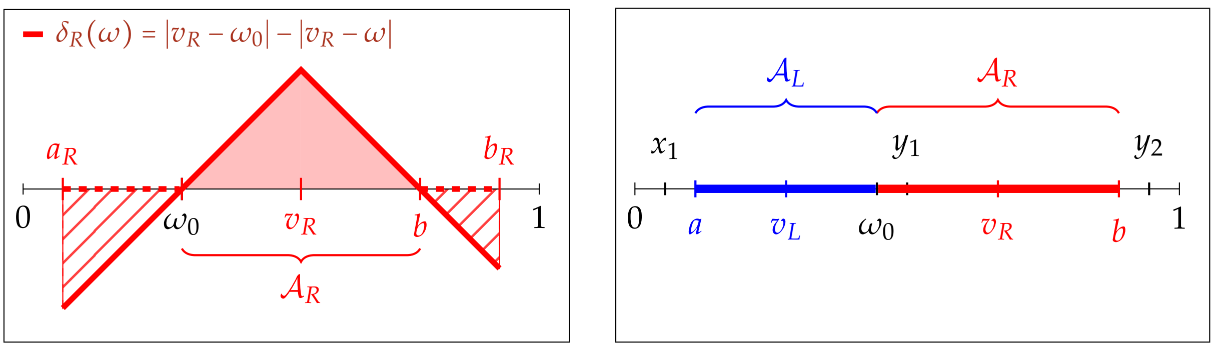
Notice that when $\omega \notin \mathcal{A}_i$, $-\delta_i(\omega)$ reflects the distance from point ω to the approval set of voter i. The voter's obedience constraint states that the expected distance from the challenger to the voter's approval set must not exceed a known quantity $\mathcal{I}_i$, which reflects how persuadable this voter is. For example, Figure 7 – 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 $\delta_i(x) \ge \delta_i(y)$. Points in the approval set of the voter are "free" in terms of the obedience constraint of that voter.

Next we show that $W_i = [a_i, b_i] \supset \mathcal{A}_i$. For ease of exposition, 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 $a_L \le a$ and $b_L \ge \omega_0$, and Figure 7, 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 $|a - x_1| \ge |y_1 - \omega_0|$, $y_1 \in W_1$,

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 7: *why challenger-preferred convincing messages are intervals.*

PROOF OF THEOREM 6.

First, observe that increasing v_R loosens voter R's (obedience) constraint and does not affect voter L's (obedience) constraint or the objective function. Hence, the solution, specifically, the challenger's ex-ante odds of winning, can only improve.

Next notice that 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 $|v_L| = |v_R|$. This means that 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 $|b_L - \omega_0| = |\omega_0 - a_R|$.[22] In other words, b_L solves voter L's obedience

[22]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 7, part *(a)*, that $|\omega_0 - a_R| < |b - \omega_0| = |\omega_0 - a|$, meaning that $a < a_R$.

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;
- 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 Corollary 1, and no further dynamics are observed.