Persuasion with Verifiable Information[*]

Maria Titova[†]

April 2022

LATEST VERSION

Abstract

The large and growing applied Bayesian persuasion literature is sometimes criticized for assuming that the sender has the power to commit to a disclosure strategy that reveals a signal based on the realized state of the world. This paper suggests that when *(i)* the sender's preferences are state-independent, *(ii)* the receivers choose between two actions, *(iii)* the state space is sufficiently rich, then the sender can reach the full commitment outcome with (some) evidence to back up his claims. The latter assumption is more natural in the applications to judicial systems, electoral campaigns, product advertising, financial disclosure, and job market signaling.

KEYWORDS: Persuasion, Value of Commitment

JEL CLASSIFICATION: D72, D82, D83, M37

[*]I am very grateful to Renee Bowen and Joel Sobel for their guidance and support. I thank S. Nageeb Ali, Simone Galperti, Remy Levin, Aleksandr Levkun, Elliot Lipnowski, Denis Shishkin, Joel Watson, Alex Weiss, and the audience members at Amsterdam, Bern, BSE, CIDE, Princeton, Manchester, Vanderbilt, UCSD, SWET 2018, YES 2019, EGSC 2019, CSQIEP 2020, and SITE 2021. All errors are my own.

[†]Department of Economics, Vanderbilt University. E-mail: maria.titova@vanderbilt.edu.

1. Introduction

Suppose the sender would like to convince a group of receivers to take his favorite action. The sender has access to hard evidence, and the only tool at his disposal is to choose how much of this evidence to reveal. On average, what is the best outcome that the sender can hope for?

Persuasion with verifiable information plays an essential role in courtrooms, electoral campaigns, product advertising, financial disclosure, job market signaling, and many other economic situations. In a courtroom, a prosecutor persuades the jury to convict the defendant by selectively presenting inculpatory evidence. In an electoral campaign, the candidates carefully choose the policy issues on which to credibly disclose their positions in order to win over voters. In advertising, a firm convinces consumers to purchase its product by highlighting only specific product characteristics. In finance, a CEO divulges only certain financial statements and indicators to the board of directors in order to obtain higher compensation. In labor markets, a job candidate makes his job application more attractive to employers by listing only his most impressive qualifications.

I consider the following formal model of persuasion with verifiable information. There is an underlying continuous space of possible states of the world, which is a unit interval. The sender is fully informed about the state of the world, but his preferences do not depend on it. Metaphorically, I say that the sender wants the receivers to approve his proposal. Receivers are uninformed about the state of the world, which to them is payoff-relevant. The sender communicates with the receivers by sending them private verifiable messages. Verifiability means that the message contains the truth (hard evidence is presented), but not necessarily nothing but the truth (some evidence may be omitted). Each receiver independently chooses between two options: to approve the proposal or reject it. There are no information spillovers between the receivers: each receiver only hears her own private message.

How does the sender convince one receiver with verifiable information? Rather than looking at the sender's messages and the receiver's beliefs, I focus on what the receiver does in every state of the world. Since she chooses between two options, we can partition the state space into two subsets: the states she approves and the states she rejects. My first result states that a subset of the unit interval is an equilibrium set of approved states if and only if it satisfies two constraints. Firstly, the *sender's incentive-compatibility constraint* (IC) ensures that the sender does not wish to deviate toward a fully informative strategy that induces the receiver to act as if fully informed. Secondly, the *receiver's obedience constraint* ensures that the receiver approves the proposal whenever her expected net payoff of approval is non-negative.

In the sender's least preferred equilibrium, his ex-ante odds of approval are minimized across all equilibria. The receiver learns whether the state of the world is within her complete information approval set and makes a fully informed choice. This is the equilibrium in which *full unraveling* takes place.

In the sender's most preferred equilibrium, his odds of approval are maximized subject to the receiver's obedience constraint. In the sender-preferred equilibrium, the receiver approves the proposal whenever her net payoff of approval is sufficiently high, but possibly negative. That is, the sender improves his odds of approval upon full disclosure by convincing the receiver to approve when she prefers not to.

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

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

THE 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!"[1] How does this happen? When the defendant is guilty (with the prior probability of 0.3), the judge convicts them for sure. When they are innocent (with the prior probability of 0.7), the judge convicts them with probability $3/7$.[2]

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's guiltiness ranges from 0 (0% guilty) to 1 (100% guilty). The prior is uniform on $[0, 1]$. The judge wishes to convicts the defendants who are at least 70% guilty. Note that from the judge's point of view, the prior that the defendant is *sufficiently* (over 70%) guilty is 0.3. At the same time, the judges best response under incomplete information is to convict whenever the probability that the defendant is sufficiently guilty exceeds 0.5. Thus, the continuous interpretation and Kamenica and Gentzkow (2011)'s example tell the same story.

With a richer state space comes a richer message space: now, the prosecutor can send any message that includes the actual state. In particular, the prosecutor can send the message $[0.4, 1]$ whenever the defendant is at least 40% guilty and the message $[0, 0.4]$ otherwise. When

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

[2]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 $3/7$.

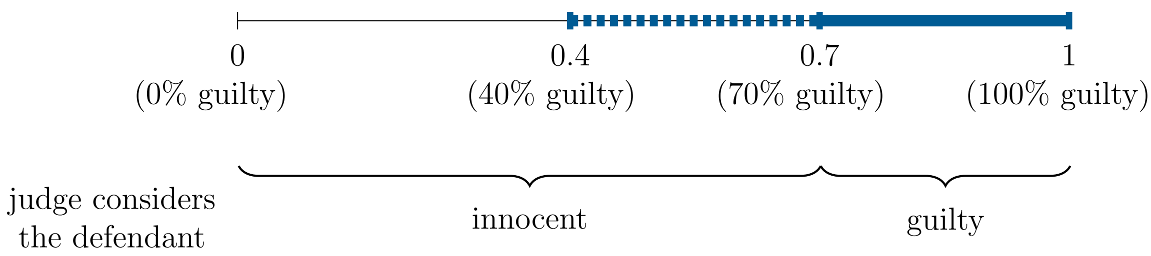

Figure 1. *The sender-preferred equilibrium of the verifiable-information game. The prosecutor sends message* $[0.4, 1]$ *(blue) when the defendant is at least 40% guilty, and* $[0, 0.4]$ *otherwise. The judge convicts after former message and acquits after latter. The judge convicts 60% of all defendants even though her prior is that 70% are innocent.*

the judge hears the former message, she concludes that the defendant is, on average, 70% guilty and convicts. The prosecutor does not have profitable deviations from the proposed strategy. When the defendant is indeed at least 40% guilty, the prosecutor receives the largest possible payoff. The prosecutor cannot credibly claim otherwise when the defendant is less than 40% guilty.

Notice that the judge convicts all the defendants who are at least 70% guilty, as well as the fraction $3/7$ of the defendants she considers innocent (those less than 70% guilty). In other words, the prosecutor-preferred threshold of conviction at 40% 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 are the same and boil down to the judge interpreting the *signal realization* (in Bayesian persuasion) or the *message* (in the verifiable disclosure game) as a recommendation to convict. The richness of the message space allows for the exact same solution.

RELATED LITERATURE

I assume that the sender uses hard evidence to communicate with the receivers. This verifiable information communication protocol was introduced by Milgrom (1981) and Grossman (1981). Other communication protocols include cheap talk by Crawford and Sobel (1982) and Bayesian persuasion by Kamenica and Gentzkow (2011). Relative to these other models of communication, Bayesian persuasion makes the sender better off because it endows him with ex-ante commitment power. Lipnowski and Ravid (2020) find that the sender's maximal equilibrium payoff from cheap talk is generally strictly lower than his payoff under commitment. Consequently, a cheap-talk sender values commitment.[3] In contrast to their result, I show that the sender does not necessarily benefit from commitment if he possesses the hard evidence to verify his messages.

There is extensive literature on applications of Bayesian persuasion models. It includes settings in which schools persuade employers to hire their graduates (Ostrovsky and Schwarz, 2010; Boleslavsky and Cotton, 2015); pharmaceutical companies persuade the FDA to approve their drug (Kolotilin, 2015); matching platforms persuade sellers to match with buyers (Romanyuk and Smolin, 2019); politicians persuade voters (Alonso and Câmara, 2016; Bardhi and Guo, 2018); governments persuade citizens through media (Gehlbach and Sonin, 2014; Egorov and Sonin, 2019). My contribution states that in all these applications, one can replace the assumption that the sender has commitment power with the assumption that the sender has hard evidence.

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

2. Model

There is a state space $\Omega := [0,1]$ and a finite set of receivers $I := \{1, \dots, n\}$. The game begins with the sender (he/him) observing the realization of the random state $\omega \in \Omega$, which is drawn from an atomless common prior distribution $p > 0$ over Ω.[4] Having observed the state of the world, the sender decides which message m_i from the set M of all (Borel) intervals of Ω to send to each receiver $i \in I$. The verifiability restriction states that the sender cannot

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

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

communicate messages that contain lies of commission, i.e. $\omega \in m_i$, for all $i \in I$.[5]

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

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

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

Receiver $i \in I$ chooses between approval (action 1) and rejection (action 0). Receiver i's preferences are described by a utility function $u_i : \{0, 1\} \times \Omega \to \mathbb{R}$. Receiver i approves (the proposal in) state ω if her net payoff of approval $\delta(\omega) := u_i(1, \omega) - u_i(0, \omega)$ is non-negative.[6] Define receiver i's approval set as

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

Under incomplete information, define receiver i's set of approval beliefs as

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

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

Assumption 2. *For every receiver* $i \in I$, $p \notin B_i$.

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

[5] I borrow the definition of a verifiable message as a subset of the state space that includes the true realization from Milgrom and Roberts (1986). This method satisfies normality of evidence (Bull and Watson, 2004), which means that it is consistent with both major ways of modeling hard evidence in the literature.

[6] For simplicity, I assume that the receiver breaks ties in favor of approval when she is indifferent, i.e. when $\delta(\omega) = 0$. This tiebreaker is necessary for the existence of the sender-preferred equilibrium, and is inconsequential in all other equilibria.

Example 1 (The Prosecutor and the Jury). Here I introduce an extended version of the seminal example from Kamenica and Gentzkow (2011). There is a prosecutor and a jury $I = \{1, \ldots, n\}$. The state of the world reflects how guilty the defendant is, with values ranging from 0% guilty (0) to 100% guilty (1). The prosecutor observes the state of the world drawn from the common prior $U[0, 1]$.

Juror $i \in I$ has a (known) threshold of conviction $v_i \in \Omega$ and wishes to convict if and only if $\omega \geq v_i$. Specifically, i's utility function is $u_i(a, \omega) = \mathbb{1}(a = 1 \text{ and } \omega \geq v_i, \text{ or } a = 0 \text{ and } \omega < v_i)$, and her net payoff of approval is $\delta_i(\omega) = 1$ if $\omega \geq v_i$ and $\delta_i(\omega) = -1$ if $\omega < v_i$. Juror i's approval set is then $A_i = \{\omega \in \Omega \mid \omega \geq v_i\}$, and her set of approval beliefs is $B_i = \left\{q \in \Delta\Omega \mid \int_{v_i}^{1} q(\omega) d\omega \geq \frac{1}{2}\right\}$. Figure 2 illustrates the preferences of juror i with $v_i = 0.7$.

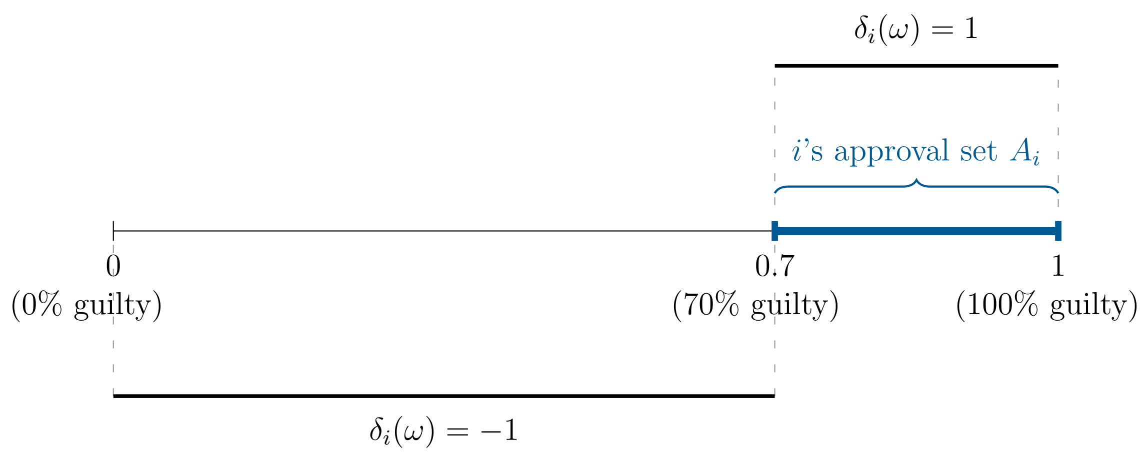

Figure 2. *The state space Ω and a juror $i \in I$ with a threshold of conviction $v_i = 0.7$. Her net payoff of approval $\delta_i(\omega)$ equals 1 if the defendant is at least 70% guilty, and -1 otherwise. She prefers to convict whenever her net payoff of approval is positive.*

The sender's utility function u_s plays the role of the social choice function that aggregates the juror's votes into the collective decision to convict or not. Assumption 1 establishes that the prosecutor receives 0 if every juror votes innocent, 1 if every juror votes guilty, and that the prosecutor's payoff is monotone in the number of guilty votes. For instance, the social choice function could be *dictatorial*, if $u_s(T) = 1 \iff i \in T$; *simple majority*, if $u_s(T) = 1 \iff |T| > n/2$; *unanimity*, if $u_s(T) = 1 \iff T = I$.

Note that if $I = \{1\}$, then only have one juror, whom we can call the judge. In that case, the prosecutor's utility function is particularly simple: $u_s(\varnothing) = 0$ and $u_s(\{1\}) = 1$.

Equilibrium Outcomes

I consider Perfect Bayesian Equilibria (henceforth just *equilibria*) of this game. The sender's strategy is a probability distribution $\sigma(\cdot \mid \omega)$ over message collections $\{m_i\}_{i \in I}$, where $m_i \in M$ for each $i \in I$. Receiver i's approval strategy $a_i(m)$ specifies which action she takes depending on message m she receives. Receiver i's posterior belief over Ω after message m is $q_i(\cdot \mid m)$. Profiles of receivers' actions and posterior beliefs are $a := \{a_i\}_{i \in I}$ and $q := \{q_i\}_{i \in I}$, respectively.

Definition 1. *A triple* (σ, a, q) *is an* <u>*equilibrium*</u> *if*

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

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

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

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

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

In words, *(i)* states that the sender sends a collection of messages with positive probability only if it maximizes his payoff; *(ii)* states that each receiver approves the proposal whenever her expected net payoff of approval is non-negative under her posterior belief; *(iii)* states that receivers' posterior beliefs are Bayes-rational on the equilibrium path; *(iv)* states that the receivers' posterior beliefs on and off the path are concentrated on the states in which the message is available to the sender.

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

Definition 2.

- *An* <u>*outcome*</u> $\alpha = \{\alpha_i\}_{i \in I}$ *specifies* $\forall i \in I$ *and* $\forall \omega \in \Omega$ *the probability* $\alpha_i(\omega) \in [0, 1]$ *that receiver* i *approves the sender's proposal in state* ω.

- *An outcome is an equilibrium outcome if it corresponds to some equilibrium.*[7]

Some outcomes are deterministic, meaning that in every state ω each receiver either approves or rejects the proposal with certainty.[8] Consequently, for each receiver, we can partition Ω into states of approval and states of rejection.

Definition 3.

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

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

3. One Receiver

Let us first focus on the case with one receiver, i.e. $I = \{1\}$. For ease of exposition, I drop all receiver-relevant subscripts i. By Assumption 1, the sender gets 1 if the receiver approves and 0 otherwise. By Assumption 2, the receiver rejects the proposal under the prior belief.

Direct Implementation

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

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

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

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

Next, if the receiver approves every state in W, then she expects that on average, her net payoff of approval is non-negative. Thus, we obtain the receiver's obedience constraint

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

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

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

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

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

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

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

Equilibrium Range and Value of Commitment

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

In the sender-worst equilibrium, the set of approved states $\underline{W}$ minimizes the sender's ex-ante utility across all equilibria. Thus, the (IC) constraint binds and $\underline{W} = A$. In this equilibrium, the receiver approves the proposal if and only if she approves it under complete information. Hence, the sender-worst equilibrium is outcome-equivalent to *full disclosure* (also known as *full unraveling*), salient in the verifiable information literature.[9]

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

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

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

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

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

- *the receiver almost surely approves the proposal if $\delta(\omega) > -c^*$ and rejects it if $\delta(\omega) < -c^*$;*[10]
- *whenever the receiver approves the proposal, her expected net payoff of approval is zero: $\mathbb{E}_p\left[\delta(\omega) \mid \overline{W}\right] = 0$.*

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

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

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

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

Also, a commitment outcome may not be deterministic.

A commitment outcome is characterized by a cutoff value of the receiver's net payoff of approval for the same reason $\overline{W}$ is.[11] That is, the receiver certainly approves (rejects) the states with a net payoff of approval above (below) some threshold. Furthermore, that threshold is negative, and the receiver certainly approves every state in her approval set. Hence, any commitment outcome satisfies the sender's incentive-compatibility constraint.

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

Example 2 (The Prosecutor and the Judge). Consider the setting from Example 1 with one receiver, *the judge*, whose threshold of conviction is $v = 0.7$.

In the sender-worst equilibrium, the set of approved states $\underline{W} = [0.7, 1]$, and coincides with the judge's approval set. That is, the judge convicts if and only if the defendant is at least 70% guilty, which is what she would do under complete information.

In the sender-preferred equilibrium, the prosecutor maximizes the odds of conviction subject to the judge's (obedience) constraint. The constraint states that upon receipt of message $\overline{W}$, the judge's average net payoff of approval is non-negative. Recall that the judge's net payoff of approval equals 1 if the defendant is guilty, or $\omega \in [0.7, 1]$, and -1 if the defendant is innocent, or $\omega \in [0, 0.7)$. Thus, to maximize the odds of conviction, the prosecutor pools all the guilty defendants with as many innocents, as possible, so that on average the judge still wants to convict. Mathematically, the prosecutor selects $\beta \in [0, 0.7)$ to solve $\int_{0.7}^{1} 1 \cdot dP + \int_{\beta}^{0.7} dP = 0$. For the uniform prior, we get $\beta = 0.4$. Figure 3 illustrates the

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

sender-preferred equilibrium.

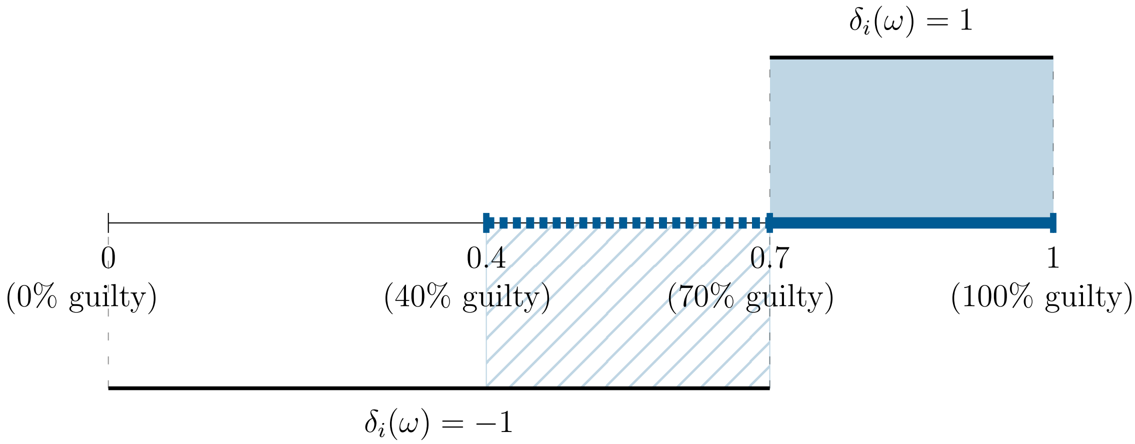

Figure 3. *The sender-preferred set of approved states $\overline{W} = [0.4, 1]$ when the prior is $U[0, 1]$. The receiver's expected net payoff of approval after message $[0.4, 1]$ equals zero because the area under $\delta(\omega)$ taken over the judge's approval set (solid blue) equals to the area above $\delta(\omega)$ outside of the judge's approval set (dashed blue).*

Quoting Kamenica and Gentzkow (2011), "this leads the judge to convict with probability 60 percent. Note that the judge knows 70 percent of defendants are innocent, yet she convicts 60 percent of them!"[12] In this setup, however, the prosecutor reaches the same outcome without having to commit to an experiment. He does so by using an equilibrium strategy of saying "the defendant is guilty" whenever the defendant is at least 40% guilty, and "the defendant is innocent" otherwise. All he needs is some evidence of his claim that the defendant is indeed at least 40% guilty.

4. Many Receivers

Having assumed that the receivers solve independent problems, I get similar results in the many-receiver case.[13]

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

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

[12] In the sender-preferred equilibrium, every defendant in $[0.4, 1]$ is convicted, which is 60% of all defendants. Since the judge considers the defendants in $[0, 0.7]$ innocent, ex-ante 70% of the defendants are innocent.

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

2. $\overline{u}_s$ *is given by*

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

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

- *sender's (IC) constraint* $A_i \subseteq W_i$*,*
- *receiver's obedience constraint* $p(\cdot \mid W_i) \in B_i$*.*

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

According to Theorem 3, when characterizing the sender's equilibrium ex-ante utility, we can restrict attention to collections of sets of approved states $(W_1, \ldots, W_n)$, each of which satisfies the IC and obedience constraints for each receiver. Moreover, the sender's ex-ante utility only depends on $(W_1, \ldots, W_n)$ and the prior distribution.

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

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

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

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

5. Conclusion

This paper studies how an informed sender with state-independent preferences persuades receivers to approve his proposal with verifiable information. I find that every equilibrium outcome is characterized by each receiver's set of approved states that satisfies that receiver's obedience and the sender's incentive-compatibility constraint for that receiver. In the sender-worst equilibrium, information unravels, and receivers act as if fully informed. The sender-preferred equilibrium outcome is the commitment outcome of the Bayesian persuasion game. Consequently, when the sender's preferences are state-independent and the receivers choose between two actions, the sender reaches the full commitment payoff with evidence and without any commitment power, as long as the message space is sufficiently rich.

References

Alonso, Ricardo and Odilon Câmara (2016), "Persuading Voters", *American Economic Review*, 106, 11 (Nov. 2016), pp. 3590-3605. (pp. 6, 14.)

Bardhi, Arjada and Yingni Guo (2018), "Modes of Persuasion toward Unanimous Consent", *Theoretical Economics*, 13, 3, pp. 1111-1149. (p. 6.)

Boleslavsky, Raphael and Christopher Cotton (2015), "Grading Standards and Education Quality", *American Economic Journal: Microeconomics*, 7, 2 (May 2015), pp. 248-279. (p. 6.)

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

Crawford, Vincent P. and Joel Sobel (1982), "Strategic Information Transmission", *Econometrica*, 50, 6 (Nov. 1982), p. 1431. (p. 5.)

Egorov, Georgy and Konstantin Sonin (2019), "Persuasion on Networks", *SSRN Electronic Journal*. (p. 6.)

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

Gehlbach, Scott and Konstantin Sonin (2014), "Government Control of the Media", *Journal of Public Economics*, 118 (Oct. 2014), pp. 163-171. (p. 6.)

Grossman, Sanford J. (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality", *The Journal of Law and Economics*, 24, 3 (Dec. 1981), pp. 461-483. (pp. 5, 12.)

Kamenica, Emir and Matthew Gentzkow (2011), "Bayesian Persuasion", *American Economic Review*, 101, 6 (Oct. 2011), pp. 2590-2615. (pp. 3-5, 8, 15, 22, 25.)

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

Lipnowski, Elliot (2020), "Equivalence of Cheap Talk and Bayesian Persuasion in a Finite Continuous Model", *note*. (p. 6.)

Lipnowski, Elliot and Doron Ravid (2020), "Cheap Talk With Transparent Motives", *Econometrica*, 88, 4, pp. 1631-1660. (p. 5.)

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

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

Milgrom, Paul R. and John Roberts (1986), "Relying on the Information of Interested Parties", *The RAND Journal of Economics*, 17, 1, p. 18. (pp. 7, 12.)

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

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

Romanyuk, Gleb and Alex Smolin (2019), "Cream Skimming and Information Design in Matching Markets", *American Economic Journal: Microeconomics*, 11, 2 (May 2019), pp. 250-276. (p. 6.)

Appendix: Omitted Proofs

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

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

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

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

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

$$\int\limits_{\text{supp } q_i(\cdot \mid m)} \delta_i(\omega) \cdot q_i(\omega \mid m) d\omega \geq 0.$$

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

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

Integrate the above inequality over all $m \in \mathcal{M}_i$: $\int\limits_{\mathcal{M}_i} \int\limits_{\mathcal{W}_i} \delta_i(\omega) \cdot \sigma_i(m \mid \omega) \cdot p(\omega) d\omega dm \geq 0$. Next, partition $\mathcal{W}_i$ into $\overline{\mathcal{W}}_i$, $A_i \setminus \overline{\mathcal{W}}_i$, and $\mathcal{W}_i \setminus (\overline{\mathcal{W}}_i \cup A_i)$ and observe that

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

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

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

As a result,

$$\int\limits_{\overline{\mathcal{W}}_i \cup A_i} \delta_i(\omega) p(\omega) p\omega \geq \int\limits_{\mathcal{M}_i} \int\limits_{\mathcal{W}_i} \delta_i(\omega) \cdot \sigma_i(m \mid \omega) \cdot p(\omega) d\omega dm \geq 0 \Longrightarrow p(\cdot \mid \overline{\mathcal{W}}_i \cup A_i) \in B_i.$$

PROOF OF THEOREM 1 (SEE PAGE 11).

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

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

Part II: consider equilibrium (σ, a, q) with the set of approved states W. W must satisfy the sender's (IC) constraint, or else the sender can deviate to full disclosure. Next, using Definition A.1, $\overline{\mathcal{W}} = W$, and by Lemma A.1, the (obedience) constraint holds.

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

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

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

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

PROOF OF THEOREM 2 (SEE PAGE 13).

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

- *the receiver almost surely approves the proposal if $\delta(\omega) > -c^*$ and rejects it if $\delta(\omega) <$*

$-c^*$;[14]

- *whenever the receiver approves the proposal, her expected net payoff of approval is zero:* $\mathbb{E}_p\big[\delta(\omega) \mid \overline{W}\big] = 0$.

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

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

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

Since $\delta(\omega) \geq 0$ for every $\omega \in A$, we have $A \subseteq \overline{W}$. Hence, $\overline{W}$ also solves (1). Furthermore, (obedience) binds, i.e. $\int\limits_{W} \delta(\omega)p(\omega)d\omega = 0$. If it does not, increase the value of the objective function while satisfying the constraint. Next, suppose that $\overline{W}$ is not characterized by a cutoff value of $\delta(\cdot)$. Then, there exist $X, Y \subseteq \Omega$ such that *(i)* $P(X) = P(Y) > 0$; *(ii)* $\forall\omega \in X$, $\forall\omega' \in Y$, $\delta(\omega) < \delta(\omega')$; *(iii)* $X \subseteq \overline{W}$ and $Y \subseteq \Omega\setminus\overline{W}$. In words, the sender-preferred set of approved states includes a positive-measure set X, does not include a positive-measure set Y, yet the receiver has a higher net payoff of approving any state in Y over any state in X.

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

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

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

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

The obedience constraint for W^* is

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

where the last inequality follows from *(1)* $W^* \setminus Y = \overline{W} \setminus X$, so the first term in both

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

constraints is the same, and *(2)* $\int_X \delta(\omega)p(\omega)d\omega < \int_Y \delta(\omega)p(\omega)d\omega$, so the second term in the second constraint is strictly larger.

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

Next, I show that the sender-preferred equilibrium outcome $\overline{\alpha}(\omega) := \mathbb{1}(\omega \in \overline{W})$ is a commitment outcome. Consider the problem of finding the optimal commitment protocol $(\sigma^{BP}, a^{BP}, q^{BP})$. According to Kamenica and Gentzkow (2011), that problem may be simplified to finding an optimal *straightforward* experiment σ^{BP} that is supported on set $\{s^+, s^-\}$, where s^+ induces posterior $q^+ \in B$ and recommends that the receiver approves the sender's proposal and s^- induces posterior $q^- \notin B$ and recommends rejection. The outcome takes form of $\alpha(\omega) = \text{prob}(s^+ \mid \omega)$, and the sender's problem under commitment becomes

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

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

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

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

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

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

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

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

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

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

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

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

the constraint is also the same:

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

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

Proof of Theorem 3 (See page 15).

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

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

2. $\overline{u}_s$ *is given by*

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

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

- *sender's (IC) constraint* $A_i \subseteq W_i$,
- *receiver's obedience constraint* $p(\cdot \mid W_i) \in B_i$.

PROOF. $\Rightarrow$: consider equilibrium outcome α with the ex-ante utility of the sender $\overline{u}_s$. Let $X_i := \{\omega \in \Omega \mid \alpha_i(\omega) = 1\}$ be the set of states in which the sender convinces receiver $i \in I$ to approve the proposal with certainty. For very $i \in I$, set $W_i = X_i \cup A_i$ satisfies the sender's (IC) constraint, and by Lemma A.1, W_i also satisfies receiver i's (obedience) constraint.

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

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

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

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

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

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

Proof of Theorem 4 (See page 16).

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

Proof. Consider the problem of finding the optimal commitment protocol $(\sigma^{BP}, a^{BP}, q^{BP})$. According to Kamenica and Gentzkow (2011), the problem may be simplified to finding an optimal *straightforward* experiment σ^{BP} that is supported on set $(S_1, \dots, S_n)$, where $S_i = \{s_i^+, s_i^-\}$ is the private set of *straightforward* signal realizations of receiver $i \in I$. Signal realization s_i^+ induces posterior $q_i^+ \in B_i$ and recommends that receiver i approves the proposal and s_i^- induces posterior $q_i^- \notin B_i$ and recommends rejection. The commitment outcome is

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

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

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

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

subject to receiver i's obedience constraint $p(\cdot \mid W_i) \in B_i$, for all $i \in I$. What remains to show is that *(i)* there exists a deterministic commitment outcome, and *(ii)* every set of approved states W_i induced by that outcome satisfies the sender's (IC) constraint. I construct a deterministic commitment outcome $\widetilde{\alpha}$ in a sequence of steps.

<u>*Step 0*</u>: start with $\widetilde{\alpha} = \alpha^{BP}$;

<u>*Step 1*</u>: if, for some $i \in I$ and $\omega \in A_i$, $\alpha_i^{BP}(\omega) < 1$, then let $\widetilde{\alpha}_i(\omega) = 1$. This weakly increases the objective, loosens receiver i's obedience constraint, and does not alter other receivers' obedience constraints. Note that this case only arises when the sender's payoff in state ω

does not strictly increase in receiver i's action;

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

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

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

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

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

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

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