Persuasion with Verifiable Information*

Maria Titova[†]

September 2022

Abstract

This paper studies how an informed sender with state-independent preferences can use verifiable information to persuade a receiver to approve his proposal. I find that the equilibrium outcomes can be characterized as the sets of approved states that satisfy the receiver's obedience constraint and the sender's incentive-compatibility constraint. In the sender-worst equilibrium, information unravels, and the receiver acts as if fully informed. In the sender-preferred equilibrium, the sender reaches the full-commitment payoff using evidence only, as long as the state space is sufficiently rich.

KEYWORDS: Persuasion, Evidence, 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, Arjada Bardhi, Simone Galperti, Remy Levin, Aleksandr Levkun, Elliot Lipnowski, Denis Shishkin, Joel Watson, Alex Weiss, and the audience members at Amsterdam, Bern, BSE, CERGE-EI, 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 a sender is privately informed about the state of the world and would like to convince a receiver to take his favorite action. The sender does not have commitment power, but his messages are verifiable. 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. For example, in a courtroom, a prosecutor tries to persuade a judge to convict a defendant by selectively presenting inculpatory evidence. In an electoral campaign, a politician carefully chooses which campaign promises he can credibly make 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 board members in order to obtain higher compensation. In a labor market, a job candidate lists specific certifications in order to make his application more attractive to an employer.

I consider the following formal model of persuasion with verifiable information. There is an underlying space of possible states of the world. 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 receiver to approve his proposal. The receiver is uninformed about the state of the world, which to her is payoff-relevant. The sender communicates with the receiver using verifiable messages. Each messages is a statement about the state of the world. Verifiability means that the message contains the truth (the true state of the world), but not necessarily the whole truth (it may contain other states, as well). The receiver chooses between two options: to approve or reject the proposal.

How does the sender convince the receiver using verifiable information? Rather than looking at the sender's messages and the receiver's beliefs, I focus on what the receiver does in each state of the world. Since she chooses between two options, we can partition the state space into two subsets: the states in which she approves the proposal and the states in which she rejects it. My first result says that a subset of the state space is an equilibrium set of approved states if and only if it satisfies two constraints. The first is the *sender's incentive-compatibility (IC) constraint*, which ensures that the sender does not wish to deviate toward a fully informative strategy that induces the receiver to act as if she knows the state of the world. Conveniently, that is the only deviation of the sender that needs to be checked. The second constraint is the *receiver's obedience constraint*, which ensures that the receiver approves the proposal whenever her expected net payoff from 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 she would approve under complete information, 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. Specifically, the sender pools the "good" states, in which the receiver prefers to approve, with some of the "bad" states, in which the receiver prefers to reject. The solution is characterized by a cutoff value of the receiver's net payoff from approval: she approves whenever it 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 in 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 disclosure literature introduced by Milgrom (1981) and Grossman (1981), and the Bayesian persuasion literature pioneered by Kamenica and Gentzkow (2011). The sender does not need ex-ante commitment power; he can persuade the receiver with evidence alone.

Motivating Example

Suppose that a prosecutor (he/him) wishes to persuade a judge (she/her) to convict a defendant. The prosecutor knows that the defendant committed $\theta \in \{0, 1, 2\}$ violations, and has hard inculpatory evidence to prove each of them. The judge does not observe the number of violations (or any of the evidence) and thinks that each number between 0 and 2 is equally likely.

The judge wishes to convict the defendants who committed 2 violations and acquit others. Her net payoff from conviction equals 1 if $\theta = 2$ and -1 otherwise. The prosecutor wants the judge to convict every defendant and receives a payoff of 1 if the judge convicts, and 0 otherwise. The prosecutor communicates with the judge using verifiable messages. Specifically, if the defendant committed θ violations, the prosecutor can present up to θ pieces of inculpatory evidence.

As with other games of verifiable information, there always exists a fully informative equilibrium, in which unraveling takes place. Specifically, in his least preferred equilibrium, the prosecutor always presents all the evidence he has. When the judge sees two pieces of evidence, she concludes that the defendant committed 2 violations. When the judge sees one piece of evidence, she knows that the defendant committed *at least* 1 violation. However, the judge also knows that the number of violations cannot be 2, or she would have seen two pieces of evidence. Thus, when seeing one piece of evidence, the judge concludes that the defendant committed exactly 1 violation. Similarly, when seeing no evidence, the judge concludes that the defendant committed no violations. In this equilibrium, the judge makes a fully informed choice. The prosecutor's ex-ante utility is $1/3$ because the judge only convicts the defendants who committed 2 violations.

Now, consider the following strategy of the prosecutor. For any defendant who committed 1 or more violations, present exactly one piece of evidence. Otherwise, present nothing. When the judge sees no evidence, she acquits. When she sees one piece of evidence, she concludes that it is equally likely that the defendant committed 1 or 2 violations. On average, the judge's net payoff from conviction equals zero, which makes her indifferent, and she breaks the tie in favor of conviction. The prosecutor's ex-ante utility is now $1/3$ from convicting the defendants who committed 2 violations, plus $1/3$ from convicting the defendants who committed 1 violation.

Several things are important here. Firstly, observe that if the defendant committed 2 violations, the prosecutor could present two pieces of evidence instead of one. However, he does not have a strict incentive to do so, because presenting just one piece already convinces the judge to convict. At the same time, when the defendant committed no violations, the prosecutor cannot get the judge to convict, because he does not have any evidence to present. As a result, the proposed strategies of the prosecutor and the judge constitute an equilibrium. In fact, it is the sender-preferred equilibrium, in which the prosecutor's ex-ante utility is maximized across all equilibria.

Let us now endow the prosecutor with commitment power and solve the corresponding Bayesian persuasion problem. With commitment, the prosecutor designs an experiment that signals a recommendation to convict or acquit as a function of the state $\theta \in \{0, 1, 2\}$. The prosecutor designs this experiment to maximize his ex-ante utility subject to the constraint that ensures that the judge wishes to convict when recommended to do so. Suppose that the prosecutor recommends conviction with probability 1 whenever $\theta \in \{1, 2\}$, which he effectively does in the sender-preferred equilibrium. Then, his ex-ante utility is $2/3$, the judge is indifferent when recommended to convict, and convicts. Notice that there is no way for the prosecutor to improve his ex-ante payoff any further: the judge is already indifferent, and if the prosecutor recommends conviction any more frequently (that is, some fraction of the time when $\theta = 0$) the judge would strictly prefer to acquit. Hence, the sender-preferred equilibrium outcome is also a commitment outcome.

Notice that we obtained the same outcome with evidence (and without commitment) as with commitment. Each of these outcomes is characterized by the threshold $\theta = 1$, the lowest number of violations that gets the judge to convict the defendant. In Bayesian persuasion, this threshold comes from calculating the optimal signal that convinces the judge to convict. In the verifiable-information game, it comes from calculating the sender-preferred equilibrium message that convinces the judge to convict. Regardless of the setup, the constraints are the same and boil down to the judge's interpretation of the *signal realization* (in Bayesian persuasion) or the *message* (in the verifiable disclosure game) as a recommendation to convict.

The motivating example is special in that there exists a *deterministic* commitment outcome, in which, conditional on any state $\theta \in \{0, 1, 2\}$, conviction is recommended with probability 0 or 1. This is not always the case.[1] In general, to achieve receiver indifference in the persuasion game with verifiable information, we will need the state space to be sufficiently rich.[2]

Related Literature

I assume that the sender communicates with the receiver using verifiable messages. This communication protocol was introduced by Milgrom (1981) and Grossman (1981). Other communication protocols include cheap talk (Crawford and Sobel,

[1]For instance, in the seminal example of Kamenica and Gentzkow (2011), the sender recommends conviction with probability $\frac{3}{7}$ in the state when the defendant is innocent.

[2]While many Bayesian persuasion papers feature a finite (often binary) state space Θ, the space of signals must be *sufficiently rich*. A common way to model a sufficiently rich signal space is to let each signal be a partition of $\Theta \times [0, 1]$ (see, e.g., Gentzkow and Kamenica, 2017 and Frankel and Kamenica, 2019).

1982) and Bayesian persuasion (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 the result of Lipnowski and Ravid (2020), I show that the sender does not necessarily benefit from commitment if his messages are verifiable.

This paper is related to the literature that compares equilibrium outcomes to the optimal mechanism outcomes for the sender-receiver games with verifiable information. When the sender's preferences are state-independent, Glazer and Rubinstein (2004, 2006) and Sher (2011) find that the receiver does not need commitment to obtain the optimal mechanism outcome. Hart, Kremer, and Perry (2017) establish the conditions for the equivalence of the equilibrium and optional mechanism outcomes. Ben-Porath, Dekel, and Lipman (2019) confirm that the receiver does not need commitment even if the sender's preferences depend on the state of the world. I focus on the verifiable-disclosure game between the sender with state-independent preferences and the receiver with a binary choice set. I conclude that the equilibrium set of these games ranges from the solution to the optimal *mechanism design* problem (when the *receiver* has commitment power) to the solution to the optimal *information design* problem (when the *sender* has commitment power).

There is an extensive applied Bayesian persuasion literature. It includes settings in which pharmaceutical companies persuade the Food and Drug Administration to approve their drug (Kolotilin, 2015); schools persuade employers to hire their graduates (Ostrovsky and Schwarz, 2010; Boleslavsky and Cotton, 2015); matching platforms persuade sellers to match with buyers (Romanyuk and Smolin, 2019); politicians persuade voters (Alonso and Câmara, 2016; Bardhi and Guo, 2018); and governments persuade citizens (Gehlbach and Sonin, 2014; Egorov and Sonin, 2019). My contribution is to show that in all these applications, one can replace the assump-

[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 the receiver is choosing between two options and the sender's preferences are state-independent, the sender's value function can be continuous only if it is constant, meaning that the sender provides no information under cheap talk, verifiable information, and Bayesian persuasion. I thank Elliot Lipnowski for this insight.

tion that the sender has commitment power with the assumption that the sender's messages are verifiable.

2. Model

There is one sender (he/him), and one receiver (she/her). The state space is $\Theta := \{0, {}^1\!/_T, \ldots, {}^{T-1}\!/_T, 1\}$, where $T \geq 2$; if $T = \infty$, then $\Theta = [0,1]$, in which case I refer to the state space as *rich*. The game begins with the sender observing the state of the world $\theta \in \Theta$, which is drawn from a common prior distribution $\mu_0 \in \Delta\Theta$ with $\operatorname{supp} \mu_0 = \Theta$.[4]

The receiver chooses between approval (action 1) and rejection (action 0). Her preferences are described by an integrable utility function $u : \{0,1\} \times \Theta \to \mathbb{R}$. The receiver approves (the proposal in) state θ if her *net payoff from approval* $\delta(\theta) := u(1,\theta) - u(0,\theta)$ is non-negative. I define the receiver's complete-information *approval set* $A := \{\theta \in \Theta \mid \delta(\theta) \geq 0\}$ to include all the states of the world she wishes to approve under *complete information*. Throughout the paper, I assume that

1. the receiver strictly prefers to approve a positive measure of states under complete information, or $\int\limits_A \delta(\theta) d\mu_0(\theta) > 0$;
2. the receiver rejects the proposal under the prior belief, or $\int\limits_\Theta \delta(\theta) d\mu_0(\theta) < 0$.

In the model with one receiver, we can also assume without loss of generality that the states are ranked from the receiver's point of view.[5]

Assumption 1. *The receiver's net payoff of approval δ is weakly increasing in θ.*

The sender's preferences depend only on the receiver's action: his payoff u_s is 1 if the receiver approves and 0 if she rejects. Having observed θ, the sender chooses a message $m \in M := \mathcal{B}(\Theta)$ from the Borel σ-algebra of Θ. Simply speaking, each

---

[4] For a compact metrizable space Y, I let ΔY denote the set of all Borel probability measures over Y, endowed with the weak* topology. For $\gamma \in \Delta Y$, I let $\operatorname{supp} \gamma$ denote the support of γ.

[5] Assumption 1 is no longer without loss of generality in a model with multiple heterogeneous receivers considered in Section 4, which is why I state most results without it.

message is a subset of the state space. Each message of the sender must be verifiable:[6]

Definition 1. *Message $m \in M$ is* verifiable in state $\theta \in \Theta$ *if $\theta \in m$.*

Example 1. Suppose that the sender is a prosecutor, the receiver is a judge, T is finite, and the prior is uniform on Θ. The prosecutor knows that the defendant committed $t \leq T$ violations, and has hard inculpatory evidence to prove each of them. His goal is to convince the judge to convict the defendant, regardless of what t is.

The judge wishes to convict those defendants who committed at least $\tau \in (T/2, T]$ violations, and acquit others. To be specific, her net payoff from approval $\delta(\theta)$ equals 1 if $\theta \geq \tau/T$, and -1 otherwise. Thus, the judge's complete-information approval set is $[\tau/T, 1]$. Figure 1 illustrates the state space and the judge's preferences.

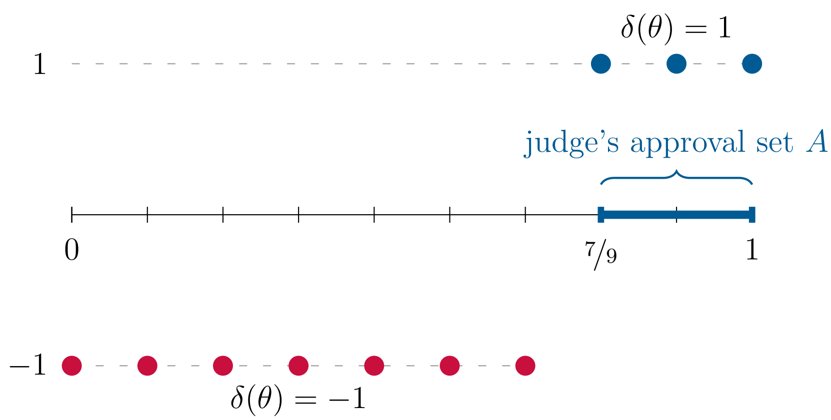

Figure 1. *The state space Θ with $T = 9$. The judge prefers to convict the defendants who committed at least $\tau = 7$ violations. Her approval set is $\{7/9, 8/9, 1\}$.*

The prosecutor's message is a statement about the number of violations. We will be able to implement both the sender-worst and the sender-preferred equilibria using the messages of the form "the defendant committed at least t' violations." For these messages, verifiability simply ensures that the prosecutor can present the evidence of t' violations only if $t \geq t'$.

[6] I borrow from Milgrom and Roberts (1986) the definition of a verifiable message as a subset of the state space that includes the true realization. This method satisfies normality of evidence (Bull and Watson, 2007), which means that it is consistent with both major ways of modeling hard evidence in the literature.

Equilibrium Outcomes

I consider perfect Bayesian equilibria of this game. Firstly, the sender's strategy $\sigma : \Theta \times M \to [0,1]$ is a regular conditional probability, such that *1)* for every $\theta \in \Theta$, $\sigma(\cdot \mid \theta)$ is a probability measure over M, and *2)* for every $m \in M$, $\sigma(m \mid \cdot)$ is measurable. Secondly, the receiver's strategy is a function $a : M \to \{0,1\}$. Finally, the receiver's posterior is a regular conditional probability $q : M \times \Theta \to [0,1]$, such that *1)* for every $m \in M$, $q(\cdot \mid m)$ is a probability measure over Θ, and *2)* for every $\theta \in \Theta$, $q(\theta \mid \cdot)$ is measurable.

Definition 2. *A triple (σ, a, q) is an* equilibrium *if*

(i) $\forall \theta \in \Theta$, $\sigma(\cdot \mid \theta)$ *is supported on* $\arg\max_{m \in M} u_s\big(a(m)\big)$*, subject to* $\theta \in m$*;*

(ii) $\forall m \in M$, $a(m) = \mathbb{1}\left(\int_\Theta \delta(\theta) dq(\theta \mid m) \geq 0\right)$*;*[7]

(iii) q *is obtained from* μ_0*, given* σ*, using Bayes rule;*[8]

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

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

An outcome of the game is a description of what action the receiver takes realization of the state of the world.

Definition 3.

- *An* outcome $\alpha : \Theta \to [0,1]$ *specifies* $\forall \theta \in \Theta$ *the probability* $\alpha(\theta)$ *that the receiver approves the proposal in state* θ*.*

---

[7] I restrict attention to the pure strategies of the receiver for ease of exposition as it does not influence the results. The receiver is never indifferent in the sender-worst equilibrium, and breaks ties in favor of approval in the sender-preferred equilibrium.

[8] That is, $\int\limits_{\widehat{\Theta}} \sigma(\widehat{M} \mid \theta) d\mu_0(\theta) = \int\limits_{\Theta} \int\limits_{\widehat{M}} q(\widehat{\Theta} \mid m) d\sigma(m \mid \theta) d\mu_0(\theta)$ for every Borel $\widehat{\Theta} \subseteq \Theta$ and $\widehat{M} \subseteq M$.

- *An outcome α is an* equilibrium outcome *if it corresponds to some equilibrium.*[9]
- *An outcome α^c is a* commitment outcome *if it solves*[10]

$$\max_{\alpha} \int_{\Theta} \alpha(\theta) d\mu_0(\theta), \qquad \textit{subject to} \qquad \begin{array}{l} \forall \theta \in \Theta,\ 0 \leq \alpha(\theta) \leq 1, \\ \displaystyle\int_{\Theta} \alpha(\theta)\delta(\theta) d\mu_0(\theta) \geq 0. \end{array} \tag{1}$$

Some outcomes are deterministic, meaning that in every state θ the receiver either approves or rejects the proposal with certainty. For such outcomes, we can partition Θ into the states that the receiver approves and those that she rejects.

Definition 4.

- *An outcome α is* deterministic *if $\alpha(\theta) \in \{0, 1\}$ for every $\theta \in \Theta$.*
- *The* set of approved states *W in a deterministic outcome α is $W := \{\theta \in \Theta \mid \alpha(\theta) = 1\}$.*

3. Analysis

Direct Implementation

Consider a deterministic equilibrium outcome with a set of approved states W. Suppose that the sender learns that $\theta \in A$. One message that is available to the sender in this state (and unavailable in every other state) is $\{\theta\}$. Since that message is verifiable, upon receiving it, the receiver learns with certainty that the state is θ, and, since θ is in the receiver's complete-information approval set, she approves the proposal. Thus, in every equilibrium, the receiver should be approving every $\theta \in A$; otherwise, the sender has a profitable deviation toward full disclosure. This observation gives

[9]Specifically, α is an equilibrium outcome if there exists an equilibrium (σ, a, q) such that $\forall \theta \in \Theta$, $\alpha(\theta) = \int_{\mathcal{M}} d\sigma(\cdot \mid \theta)$, where $\mathcal{M} := \{m \subseteq \Theta \mid a(m) = 1\}$ is the set of messages that convince the receiver to approve.

[10]Under commitment, the model in this paper is a version of the model of Alonso and Câmara (2016) with one receiver and a continuous state space. According to Kamenica and Gentzkow (2011), the optimal straightforward experiment is supported on the set $\{s^+, s^-\}$, where s^+ induces the posterior q^+ and leads to a recommendation of approval, while s^- induces the posterior q^- and leads to a recommendation of rejection. The outcome then takes the form $\alpha^c(\theta) = \Pr(s^+ \mid \theta)$.

rise to the sender's incentive-compatibility (IC) constraint

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

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

$$\int_W \delta(\theta) d\mu_0(\theta) \geq 0. \tag{obedience}$$

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

THEOREM 1. *Every equilibrium outcome is deterministic. Furthermore, $W \subseteq \Theta$ is an equilibrium set of approved states if and only if it satisfies the sender's* (IC) *constraint and the receiver's* (obedience) *constraint.*

The first part of the theorem says that every equilibrium outcome is deterministic: the receiver either approves or rejects the proposal with probability one in every state of the world. To verify this, suppose instead that in some state, the receiver approves and rejects with positive probability. Since the receiver sometimes approves, the sender has access to at least one message that convinces the receiver to approve. But then the sender can deviate and send that message with certainty so that the receiver approves with probability one. Hence, every equilibrium outcome is deterministic.

Secondly, consider equilibrium (σ, a, q) with the set of approved states W. Then, W must satisfy the sender's IC constraint, or else the sender can deviate to full disclosure. It remains to show that if the sender induces approval in every state in W in the original equilibrium, then W satisfies the receiver's obedience constraint.

Let $\mathcal{M} := \{m \in M \mid a(m) = 1\}$ be the set of messages that convince the receiver to approve. Notice that if the sender has access to a convincing message in state θ, then he has to convince the receiver with probability 1 in that state. Thus, for every state that the receiver approves, the sender must be sending convincing messages only, or $\theta \in W$, $\int_{\mathcal{M}} d\sigma(\cdot \mid \theta) = 1$. Next, consider a convincing message $m \in \mathcal{M}$. Notice that $\operatorname{supp} q(\cdot \mid m) \subseteq m \subseteq W$. Here, the first inclusion holds by the equilibrium condition *(iv)*. The second inclusion holds because $\theta \in m \in \mathcal{M}$ implies that $\theta \in W$. In words, the receiver believes that every convincing message in $\mathcal{M}$ comes from the states in W, since those messages are not available outside of W.

Now, since every message in $\mathcal{M}$ convinces the receiver to approve, we have

$$\forall m \in \mathcal{M}, \int\limits_{\Theta} \delta(\theta) dq(\theta \mid m) = \int\limits_{W} \delta(\theta) dq(\theta \mid m) \geq 0.$$

Using Bayes rule, we conclude that W satisfies (obedience):

$$\forall m \in \mathcal{M}, \int\limits_{W} \delta(\theta) dq(\theta \mid m) \geq 0 \iff \int\limits_{W} \delta(\theta) \sigma(m \mid \theta) d\mu_0(\theta) \geq 0$$

$$\implies \int\limits_{W} \delta(\theta) \underbrace{\sigma(\mathcal{M} \mid \theta)}_{=1,\ \forall \theta \in W} d\mu_0(\theta) = \int\limits_{W} \delta(\theta) d\mu_0(\theta) \geq 0.$$

Now, suppose that $W \subseteq \Theta$ satisfies (IC) and (obedience). Then we can construct an equilibrium that directly implements the set of approved states W. Let the sender send the message W for every state within W and the message $W^c := \Theta \smallsetminus W$ for every state outside of W. Formally, let $\sigma(W \mid \theta) = \mathbb{1}(\theta \in W)$ and $\sigma(W^c \mid \theta) = \mathbb{1}(\theta \in W^c)$. On the path, the receiver only hears two messages, W and W^c. By (obedience), she approves after message W because her expected net payoff from approval is non-negative. On the other hand, she rejects after message W^c because her net payoff from approval is negative for every $\theta \in W^c$. In words, the sender sends two messages and the receiver interprets them as a recommendation to approve or reject. Off the equilibrium path, let the receiver be "skeptical" and assume that any unexpected message $m \notin \{W, \Theta \smallsetminus W\}$ comes from outside of her complete-information approval set, when possible. Formally,

$$\forall m \subseteq A, \quad \text{supp}\, q(\cdot \mid m) \subseteq m, \text{ so that } \int_{\Theta} \delta(\theta) dq(\theta \mid m) \geq 0,$$

$$\forall m \nsubseteq A, m \neq W, \quad \text{supp}\, q(\cdot \mid m) \subseteq m \smallsetminus A, \text{ so that } \int_{\Theta} \delta(\theta) dq(\theta \mid m) < 0.$$

Observe that the sender has no profitable deviations: if $\theta \in W$, he is getting the highest possible payoff; if $\theta \notin W$, then he cannot replicate the message W, and the receiver rejects after every other message. Hence, (σ, a, q) is an equilibrium with the set of approved states W. The proof of Theorem 1 is now complete.

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, this means that *(i)* the sender truthfully reveals the state of the world to the mediator, *(ii)* the mediator translates this report into an action recommendation for the receiver, and *(iii)* the receiver obediently follows her recommendation. Which equilibrium outcome is to be implemented is decided by the mediator at Step *(ii)*. Conveniently, Theorem 1 also provides necessary and sufficient conditions for a set of approved states to be implementable in equilibrium.

Theorem 1 is the most general result of this paper and relies only on the sender's preferences being state-independent. As I discuss in Section 4, Theorem 1 is easily generalizable to multiple actions of the receiver, and multiple independent receivers.

Equilibrium Range and Value of Commitment

For the purposes of characterizing equilibrium outcomes, Theorem 1 allows us to restrict attention to sets of approved states $W \subseteq \Theta$ 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 $\mu_0(W)$, 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* (or *full unraveling*), which is salient in the verifiable-information literature.[11]

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

$$\overline{W} = \arg\max_{W \subseteq \Theta} \mu_0(W), \qquad \text{subject to} \qquad \begin{array}{c} A \subseteq W, \\ \displaystyle\int_W \delta(\theta) d\mu_0(\theta) \geq 0. \end{array}$$

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

The second result of this paper describes the solution to this problem when the state space is continuous.

THEOREM 2. *Let $\Theta = [0,1]$. The sender-preferred equilibrium set of approved states $\overline{W}$ is characterized by a cutoff value $c^* > 0$ for the receiver's net payoff from approval such that*

- *the receiver a.s. approves the proposal if $\delta(\theta) > -c^*$ and rejects it if $\delta(\theta) < -c^*$;*[12]
- *whenever the receiver approves the proposal, her expected net payoff from approval is zero: $\int\limits_{\overline{W}} \delta(\theta) d\mu_0(\theta) = 0$.*

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

Let us prove Theorem 2. First, notice that the receiver's obedience constraint binds; otherwise we could increase the value of the objective while still satisfying that constraint. Secondly, suppose that the sender-preferred equilibrium set of approved states $\overline{W}$ is not characterized by a cutoff value for the receiver's net payoff from approval. Then there exist two sets $X, Y \subseteq \Theta$, of positive and equal measure, such that $\overline{W}$ includes X, $\overline{W}$ does not include Y, yet the receiver's net payoff from approval is higher for any state in Y than for any state in X. Consider an alternative set of approved states $\widetilde{W}$ formed by replacing X with Y, i.e. $\widetilde{W} = (\overline{W} \setminus X) \cup Y$. The sender has the same ex-ante payoff at $\widetilde{W}$ and $\overline{W}$, because the sets X and Y have the same measure. But the (obedience) constraint for $\widetilde{W}$ is loose, while for $\overline{W}$ it is binding. This is 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 $\widetilde{W}$, which is a contradiction.

Next, let us compare the problems of *(i)* finding the sender-preferred equilibrium outcome and *(ii)* finding the commitment outcome.[13] In *(i)*, we maximize the ex-ante measure of the set of approved states subject to the (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 IC constraint. Also, a commitment outcome may not be deterministic.

[12] Almost surely with respect to the prior measure μ_0.

[13] Note that I do not prove the existence of a commitment outcome as it was established in, e.g., Arieli et al. (2022).

A commitment outcome is characterized by a cutoff value for the receiver's net payoff from approval, for the same reason $\overline{W}$ is.[14] That is, the receiver certainly approves (rejects) the states with a net payoff from approval above (below) some threshold. Furthermore, that threshold is negative, since the receiver certainly approves every state in her complete-information approval set. Hence, any commitment outcome satisfies the sender's IC constraint.

Next, consider a (possibly non-deterministic) commitment outcome α^c with the cutoff value $\widetilde{c}$ of the receiver's net payoff from approval. Let $\mathcal{D} := \{\theta \in \Theta \mid 0 < \alpha^c(\theta) < 1\}$ be the set of states the receiver approves and rejects with a positive probability. Note that since α^c is characterized by the cutoff value $\widetilde{c}$, $\delta(\theta) = -\widetilde{c}$ for every $\theta \in \mathcal{D}$.

The next step shows that if α^c is non-deterministic (whenever $P(\mathcal{D}) > 0$), we can construct a deterministic commitment outcome $\widetilde{\alpha}$ by partitioning $\mathcal{D}$ into two subsets and letting the sender recommend one action on each subset with certainty.[15] Specifically, let $\widetilde{\alpha}(\theta) = \alpha^c(\theta)$ for all $\theta \notin \mathcal{D}$ and $\widetilde{\alpha}(\theta) = \mathbb{1}(\theta \in X)$ for all $\theta \in \mathcal{D}$, where $X \subseteq \mathcal{D}$ solves

$$\int_{\mathcal{D}} \alpha^c(\theta) d\mu_0(\theta) = \int_{\mathcal{D}} \widetilde{\alpha}(\theta) d\mu_0(\theta) = \mu_0(X).$$

Thanks to the continuity of the state space, such partitioning does not affect the objective function of the sender or the obedience constraint of the receiver.[16] Consequently, there always exists a deterministic commitment outcome as long as the state space is rich. Since this commitment outcome satisfies the sender's IC constraint, it is also an equilibrium outcome. This completes the proof of Theorem 2.

Note that Theorem 2 characterizes $\overline{W}$ even if the states in Θ are not ordered. Us-

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

[15] This is a standard purification argument introduced by Dvoretzky, Wald, and Wolfowitz (1951) (Theorem 4, which is a generalization of an earlier result by Liapounoff, 1940). It says that, for any mixed strategy with finite actions, there exists a pure strategy with identical integrals with respect to a finite set of atomless measures on a measurable space.

[16] The value of the sender's objective function is the same, because $\int_{\mathcal{D}} \alpha^c(\theta) d\mu_0(\theta) = \int_{\mathcal{D}} \widetilde{\alpha}(\theta) d\mu_0(\theta) = \mu_0(X)$. The receiver's constraint is also the same: $\int_{\mathcal{D}} \delta(\theta)\alpha^c(\theta) d\mu_0(\theta) = \int_{\mathcal{D}} \delta(\theta)\widetilde{\alpha}(\theta) d\mu_0(\theta)$ since $\forall \theta \in \mathcal{D}$, $\delta(\theta) = -\widetilde{c}$.

ing the fact that δ is increasing with θ, we can obtain an even simpler characterization of the sender-preferred equilibrium in terms of a cutoff *state*.

Corollary 1. *Let $\Theta = [0,1]$ and suppose that Assumption 1 is satisfied. Then, there exists $\theta^* \in A^c$ such that $\int_{\theta^*}^1 \delta(\theta) d\mu_0(\theta) = 0$. Furthermore, the set $[\theta^*, 1]$ is a sender-preferred equilibrium set of approved states and the sender-preferred equilibrium outcome $\overline{\alpha}(\theta) = \mathbb{1}(\theta \geq \theta^*)$ is a commitment outcome.*

First observe that for any integrable δ, the function $G(t) := \int_t^1 \delta(\theta) d\mu_0(\theta)$ is continuous and increasing in t. We have $G(0) < 0$ and $G(\tau) > 0$, where τ is the lower bound of the receiver's approval set A. By the intermediate value theorem, there exists $\theta^* \in (0, \tau) \subseteq A^c$ such that $G(\theta^*) = 0$.

To show that $[\theta^*, 1]$ is a sender-preferred set of approved states, we apply the purification argument again. Suppose that $\overline{W}$ is a sender-preferred equilibrium outcome and is not an interval. Then, we can represent it as a union of an interval $[\overline{\theta}, 1]$, where $\overline{\theta} := \inf\{\theta \in \Theta \mid \delta(\theta) > -c^*\}$ and some set $Z \subseteq \{\theta \in \Theta \mid \delta(\theta) = -c^*\}$. From the binding obedience constraint, we have $\mu_0(Z) = \frac{1}{c^*} \int_{\overline{\theta}}^1 \delta(\theta) d\mu_0(\theta)$. If $\mu_0(Z) = 0$, then $[\overline{\theta}, 1]$ is a sender-preferred equilibrium set of approved states. If $\mu_0(Z) > 0$, then there exists $\theta^* \in Z$ such that $\mu_0(Z) = \mu_0([\theta^*, \overline{\theta}])$. Then, $[\theta^*, 1]$ is a sender-preferred equilibrium set of approved states. The sender-preferred equilibrium outcome $\overline{\alpha}(\theta) = \mathbb{1}(\theta \geq \theta^*)$ is a commitment outcome by Theorem 2.

When the state space Θ is finite, the purification argument no longer applies. Consequently, the sender-preferred equilibrium outcome is not always characterized by a cutoff value for the receiver's net payoff from approval. To see this, suppose that $\Theta = \{0, 1/2, 1\}$, the prior belief is $(\mu(0), \mu(1/2), \mu_0(1)) = (1/6, 1/2, 1/3)$, and the receiver's net payoff from approval is $(\delta(0), \delta(1/2), \delta(1)) = (-2, -1, 1)$. Then, in the sender-preferred equilibrium, the set of approved states is $\overline{W} = \{0, 1\}$, because it maximizes $\sum_{W \subseteq \Theta} \mu_0(\theta)$ subject to $\sum_{W \subseteq \Theta} \delta(\theta)\mu_0(\theta) \geq 0$ and $\{1\} \in W$. Here, $\overline{W}$ is not characterized by a cutoff value: 0 and 1 are included, but $1/2$ is not. Including state $1/2$ instead of 0 does improve the objective by $\mu(1/2) > \mu_0(0)$, but is unaffordable in terms of the obedience constraint, because $\delta(1/2)\mu_0(1/2) + \delta(1)\mu_0(1) < 0$. Consequently, when the state space is finite, a sender-preferred equilibrium outcome is a commitment outcome only if $\overline{W}$ happens to *(1)* be characterized by a cutoff state, and *(2)* bind the obedience constraint.

Theorem 3. *Suppose that* Θ *is finite, Assumption 1 is satisfied, and there exists* $\theta^* \in A^c$ *such that* $\sum_{\theta=\theta^*}^{1} \delta(\theta)\mu_0(\theta) = 0$*. Then, the set* $[\theta^*, 1]$ *is a sender-preferred equilibrium set of approved states. Furthermore, the sender-preferred equilibrium outcome* $\overline{\alpha}(\theta) = \mathbb{1}(\theta \geq \theta^*)$ *is a commitment outcome.*

Let us first show that $\overline{W} = [\theta^*, 1]$ solves

$$\max_{W \subseteq \Theta} \mu_0(W) \quad \text{subject to} \quad \sum_{\theta \in W} \delta(\theta)\mu_0(\theta) = 0 \text{ and } A \subseteq W. \tag{2}$$

Consider an alternative candidate solution $\widetilde{W} \neq \overline{W}$. Let $\widetilde{\Theta} := \widetilde{W} \setminus \overline{W} \subseteq \widetilde{W}$ and $\overline{\Theta} := \overline{W} \setminus \widetilde{W} \subseteq \overline{\Theta}$. Note that $\overline{\Theta}$ must be non-empty to satisfy $\widetilde{W} \neq \overline{W}$, or else $\widetilde{W} \subset \overline{W}$ and $\mu_0(\widetilde{W}) < \mu_0(\overline{W})$. Now, for every $\overline{\theta} \in \overline{\Theta}$ and $\widetilde{\theta} \in \widetilde{\Theta}$, we have $\delta(\widetilde{\theta}) \leq \delta(\overline{\theta})$ and $\delta(\widetilde{\theta}) < 0$ because $\widetilde{\theta} \leq \theta^* \leq \overline{\theta}$ and $\theta^* \in A^c$. If $\widetilde{W}$ is a solution and $\overline{W}$ is not, then $\mu_0(\widetilde{W}) > \mu_0(\overline{W}) \iff \mu_0(\widetilde{\Theta}) > \mu_0(\overline{\Theta})$. From the last two observations,

$$\sum_{\widetilde{\theta} \in \widetilde{\Theta}} \delta(\widetilde{\theta})\mu_0(\widetilde{\theta}) < \sum_{\overline{\theta} \in \overline{\Theta}} \delta(\overline{\theta})\mu_0(\overline{\theta}),$$

which, in turn, implies that $\widetilde{W}$ violates the obedience constraint and thus cannot be a solution to (2).

Next, let us show that the sender-preferred equilibrium outcome $\overline{\alpha}$ is a commitment outcome. Specifically, let us show that $\overline{\alpha}$ solves

$$\max_{\alpha} \sum_{\theta \in \Theta} \alpha(\theta)\mu_0(\theta) \;\; \text{subject to} \;\; \sum_{\theta \in \Theta} \alpha(\theta)\delta(\theta)\mu_0(\theta) \geq 0 \text{ and } \alpha(\theta) \in [0,1], \forall \theta \in \Theta. \tag{3}$$

Consider an alternative solution, $\widetilde{\alpha} \neq \overline{\alpha}$. Let $\widetilde{W} := \{\theta \in \Theta \mid \widetilde{\alpha}(\theta) > 0\}$. If $\widetilde{W} \subseteq \overline{W}$, then $\mu_0(\widetilde{W}) < \mu_0(\overline{W})$ and $\widetilde{\alpha}$ cannot be a solution to (3). Otherwise, let $\widetilde{\Theta} := \widetilde{W} \setminus \overline{W}$ and $\overline{\Theta} := \overline{W} \setminus \widetilde{W}$, and apply the same argument we used to prove that $\overline{W}$ solves (2). This completes the proof of Theorem 3.

Example 2. Consider the setting from Example 1, and suppose that the judge's threshold for conviction is $\tau = 7$ violations. Assume that the prior is uniform on Θ, so that $\mu_0(\theta) = 1/10$ for all $\theta \in \Theta$. Note that under these parameters, the judge considers 70% of the defendants ex-ante innocent, in the sense that she wishes to acquit everyone who committed 0 to 6 violations, and that event has a prior measure of $^7/_{10}$.

In the sender-worst equilibrium, the set of approved states is $\{^7/_{10}, ^8/_{10}, ^9/_{10}, 1\}$, which coincides with the judge's complete-information approval set. One way to implement this equilibrium is to have the prosecutor reveal all the evidence he has by sending message $\{^t/_T, \dots, 1\}$ for every $t \leq T$. An alternative way to implement the sender-worst equilibrium to to have the prosecutor present 7 pieces of hard evidence for every defendant who committed at least 7 violations, and say nothing/present no evidence otherwise. Either way, the prosecutor does not have profitable deviations because he gets the judge to convict if and only if $t \geq 7$, and there is no way he can replicate a convincing messages when he does not possess the evidence.

In the sender-preferred equilibrium, the prosecutor maximizes the odds of conviction subject to the judge's obedience constraint, which states that, given the sender-preferred equilibrium message $\overline{W}$, the judge's net payoff from approval must be non-negative. Recall that the judge's net payoff from approval equals 1 if $\theta \geq {}^7/_{10}$, and -1 otherwise. Thus, to maximize the odds of conviction, the prosecutor pools the "guilty" (7 or more violations) defendants with as many "innocent" (6 or less violations) ones as possible, while on average across this pool, the judge still wants to convict. Mathematically, the prosecutor selects $\theta^* \in \Theta$ that maximizes $\sum_{\theta=\theta^*}^{1} \mu_0(\theta)$ subject to $\sum\limits_{\theta=\theta^*}^{1} \delta(\theta)\mu_0(\theta) \geq 0$. For the uniform prior, we get $\theta^* = {}^4/_{10}$. Fortunately, $\theta^* = {}^4/_{10}$ binds the obedience constraint, so by Theorem 3 we conclude that $\{^4/_9, \dots, ^8/_9, 1\}$ is a sender-preferred equilibrium set of approved states, and $\overline{\alpha}(\theta) = \mathbb{1}(\theta \geq {}^4/_{10})$ is a commitment outcome.

In the sender-preferred equilibrium, the prosecutor presents exactly 4 pieces of inculpatory evidence for each defendant who committed at least 4 violations. When the judge observes 4 pieces of evidence, she learns that the defendant committed at least 4 violations. Beyond that, she also knows that the defendant is equally likely to have committed 4, 5, ..., 9 violations. On average, her net payoff from approval is exactly zero, so she convicts. The sender's ex-ante utility is ${}^4/_9 \cdot {}^1/_{10} + \cdots + {}^8/_9 \cdot {}^1/_{10} + 1 \cdot {}^1/_{10} = {}^6/_{10}$.

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!" In the setup of this paper, the prosecutor reaches the same outcome without having to commit to an experiment. He gets the judge to convict 60% of defendants by presenting 4 pieces of inculpatory evidence whenever he possesses it.

4. Robustness

Here I discuss how the *direct implementation* result of Theorem 1 and the *no benefit from commitment* result of Theorem 2 generalize to more complex environments. Specifically, I extend the model in two directions. Firstly, I show that both results hold in the model with many (independent) receivers. Secondly, I show that if a receiver is choosing among three or more actions, the direct implementation result still holds, but the sender-preferred equilibrium outcome may or may not reach the commitment payoff.

Many Receivers

Suppose there is a set $I = \{1, \dots, n\}$, $n \geq 2$, of receivers. Each receiver's payoff depends only on the state θ and her own action. More specifically, receiver $i \in I$ has some net payoff from approval $\delta_i(\theta)$ that defines her complete-information approval set A_i. The sender's preferences are state-independent and his payoff $u_s : \{0,1\}^n \to \mathbb{R}$ is strictly increasing in each receiver's action.[17] Assume that the sender communicates with each receiver in private: he chooses $m_i \in M$ such that $\theta \in m_i$ for every $i \in I$, and receiver i observes only her own message m_i.

As in the one-receiver case, we can define a deterministic outcome via a set of approved states $W_i \subseteq \Theta$ for each receiver $i \in I$. Since the receivers are independent, we can claim that *in equilibrium*, receiver $i \in I$ approves W_i if and only if it satisfies her obedience constraint $\int\limits_{W_i} \delta(\theta) d\mu_0(\theta) \geq 0$. At the same time, since the sender's preferences are monotone in every receiver's action, W_i should also satisfy the sender's IC constraint for each $i \in I$. We thus obtain a generalization of the recommendation principle that accommodates multiple receivers.

Corollary 2. *In the model with a set of receivers $I = \{1, \dots, n\}$ with $n \geq 2$,*

[17] Similar results can be obtained for the sender's utility that *weakly* increases in every receiver's action. In that case, not every equilibrium outcome is deterministic, however, every equilibrium is payoff-equivalent to a deterministic one.

every equilibrium outcome is deterministic. Furthermore, $(W_1, \ldots, W_n) \subseteq \Theta^n$ *is an equilibrium collection of sets of approved states if and only if, for every receiver* $i \in I$, W_i *satisfies*

- *the sender's IC constraint* $A_i \subseteq W_i$ *for receiver* i*;*
- *receiver* i*'s obedience constraint* $\int_{W_i} \delta_i(\theta) d\mu_0(\theta) \geq 0$.

Using this, we can characterize the set of equilibrium payoffs of the sender. In his least preferred equilibrium, the sender reveals the state of the world to each receiver, i.e. $\underline{W}_i = A_i$ for every $i \in I$. In his most preferred equilibrium, the sender chooses $(\overline{W}_1, \ldots, \overline{W}_n)$ so as to maximize his objective subject to each receiver's obedience constraint. In that equilibrium, we see by the same argument as in the one-receiver case that the sender reaches the full-commitment payoff.

Corollary 3. *Let* $\Theta = [0, 1]$. *In the model with a set of receivers* $I = \{1, \ldots, n\}$ *with* $n \geq 2$*, the sender-preferred equilibrium outcome is a commitment outcome.*

One Receiver Choosing among 3+ Actions

Suppose now that there is one receiver, who chooses her action from a set $J = \{0, 1, \ldots, k\}$ with $k \geq 2$. The receiver's utility function is $u : J \times \Theta \to \mathbb{R}$. Define the receiver's *complete-information approval set for action* j as the set A_j consisting of all states of the world in which she prefers to take action j when she is fully informed. Assume that the receiver's approval sets $A_0, \ldots, A_k$ form a partition of the state space. Also assume that the sender's preferences are state-independent, and his payoff is increasing in the receiver's action.

An outcome is now a partition of the state space into $k + 1$ subsets, $(W_0, W_1, \ldots, W_k)$, some of which may be empty, where $W_j \subseteq \Theta$ consists of the states in which the receiver takes action $j \in J$. Then, the receiver's obedience constraint states that j is the best action to take if the state is, on average, in W_j. Additionally, the sender's IC constraint states that he cannot profitably deviate to full disclosure. As a result, in environments in which the receiver has multiple actions, the recommendation principle generalizes as follows:

Corollary 4. *In the model with one receiver who chooses an action from the set* $J = \{0, 1, \ldots, k\}$*, every equilibrium outcome is deterministic. Furthermore, a partition*

$(W_0, \ldots, W_J)$ *of the state space* Θ *is an equilibrium collection of action sets if and only if, for every action* $j \in J \setminus \{0\}$, W_j *satisfies*

- *the sender's IC constraint:* $\theta \in A_j \implies \theta \in W_j \cup \cdots \cup W_k$*;*
- *the receiver's' obedience constraint: action* j *solves* $\max\limits_{l \in J} \int\limits_{W_j} u(l, \theta) d\mu_0(\theta)$.

As it turns out, when $k \geq 2$, the sender may or may not reach the commitment payoff in his most preferred equilibrium. Intuitively, in Bayesian persuasion, the sender can commit to a signal that recommends an intermediate action most of the time. In equilibrium, such a solution may violate the sender's IC constraint for the highest action. The determination of necessary and sufficient conditions for the equivalence of payoffs remains an open problem which I leave for further research. However, once a commitment outcome is known, it is easy to check whether it is implementable in equilibrium. The answer is affirmative if the outcome satisfies the sender's IC constraint for every action.

Corollary 5. *Suppose that there exists a deterministic commitment outcome and let* W_j *be the set of states in which the receiver is recommended to take action* $j \in J$*. If* W_j *satisfies the sender's IC constraint for every action* $j \in J$*, then* $(W_0, \ldots, W_k)$ *is a sender-preferred equilibrium collection of action sets.*

To see how to check whether a commitment outcome may be reached in equilibrium, consider the following example from Gentzkow and Kamenica (2016), in which the receiver has three actions available. For each action $a \in \{0, 1, 2\}$, the receiver's utility is $u(a, \theta) = 3a\theta - a^2$; her complete-information approval sets are $A_0 = \left[0, \frac{1}{3}\right)$, $A_1 = \left[\frac{1}{3}, \frac{2}{3}\right)$, $A_2 = \left[\frac{2}{3}, 1\right]$. Given the belief $q \in \Delta\Theta$, the receiver prefers to take action 2 if $\mathbb{E}_q[\theta] \geq \frac{2}{3}$, action 1 if $\mathbb{E}_q[\theta] \in \left[\frac{1}{3}, \frac{2}{3}\right)$, and action 0 otherwise. The sender's payoff is given by $u_s(0) = 0$, $u_s(1) = 1$, and $u_s(2) = 3$. The prior is uniform. The authors find that one way to reach the commitment payoff is by inducing action 0 on $\left[0, \frac{8}{48}\right]$, action 1 on $\left[\frac{11}{48}, \frac{21}{48}\right]$, and action 2 on $\left[\frac{8}{48}, \frac{11}{48}\right] \cup \left[\frac{21}{48}, 1\right]$. Clearly, in this commitment outcome the sender does not have profitable deviations toward full disclosure: if $\theta \in A_2$, then the receiver is playing a_2; if $\theta \in A_1$, then the receiver is playing a_1 or a_2. Thus, we can conclude that this commitment outcome is also an equilibrium outcome.

5. Conclusion

The large and growing applied Bayesian persuasion literature is sometimes criticized for assuming that the sender can commit to an experiment that reveals a signal based on the realized state of the world. This paper shows that if the sender's preferences are state-independent, the receiver is choosing between two actions, and the state space is sufficiently rich, then the sender reaches the full-commitment payoff in an equilibrium of the disclosure game with verifiable information. The latter setup is more natural in the applications to judicial systems, electoral campaigns, product advertising, financial disclosure, and job market signaling.

References

Alonso, Ricardo and Odilon Câmara (2016), "Persuading Voters", *American Economic Review*, 106, 11 (Nov. 2016), pp. 3590-3605. (pp. 6, 10, 15.)

Arieli, Itai, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita (2022), "Optimal Persuasion via Bi-Pooling", *Theoretical Economics* (Forthcoming 2022). (p. 14.)

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

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2019), "Mechanisms With Evidence: Commitment and Robustness", *Econometrica*, 87, 2, pp. 529-566. (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 (2007), "Hard Evidence and Mechanism Design", *Games and Economic Behavior*, 58, 1, pp. 75-93. (p. 8.)

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

Dvoretzky, A., A. Wald, and J. Wolfowitz (1951), "Elimination of Randomization in Certain Statistical Decision Procedures and Zero-Sum Two-Person Games", *The Annals of Mathematical Statistics*, 22 (1 Mar. 1951), pp. 1-21. (p. 15.)

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. 13.)

Frankel, Alexander and Emir Kamenica (2019), “Quantifying Information and Uncertainty”, *American Economic Review*, 109 (10 Oct. 2019), pp. 3650-3680. (p. 5.)

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

Gentzkow, Matthew and Emir Kamenica (2016), “A Rothschild-Stiglitz Approach to Bayesian Persuasion”, *American Economic Review*, 106, 5 (May 2016), pp. 597-601. (p. 21.)

— (2017), “Bayesian Persuasion with Multiple Senders and Rich Signal Spaces”, *Games and Economic Behavior*, 104 (July 2017), pp. 411-429. (p. 5.)

Glazer, Jacob and Ariel Rubinstein (2004), “On Optimal Rules of Persuasion”, *Econometrica*, 72 (6 Nov. 2004), pp. 1715-1736. (p. 6.)

— (2006), “A Study in the Pragmatics of Persuasion: a Game Theoretical Approach”, *Theoretical Economics*, 1, pp. 395-410. (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. 3, 5, 13.)

Hart, Sergiu, Ilan Kremer, and Motty Perry (2017), “Evidence Games: Truth and Commitment”, *American Economic Review*, 107 (3 Mar. 2017), pp. 690-713. (p. 6.)

Kamenica, Emir and Matthew Gentzkow (2011), “Bayesian Persuasion”, *American Economic Review*, 101, 6 (Oct. 2011), pp. 2590-2615. (pp. 3, 5, 6, 10, 18.)

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

Liapounoff, A. (1940), “Sur les Fonctions-Vecteurs Complétement Additives”, *Izv. Akad. Nauk SSSR Ser. Mat.*, 4 (6 1940), pp. 465-478. (p. 15.)

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. 6.)

Milgrom, Paul R. (1981), “Good News and Bad News: Representation Theorems and Applications”, *The Bell Journal of Economics*, 12, 2, p. 380. (pp. 3, 5, 13.)

— (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. 13.)

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, 13.)

Myerson, Roger B. (1986), “Multistage Games with Communication”, *Econometrica*, 54, 2 (Mar. 1986), p. 323. (p. 13.)

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.)

Sher, Itai (2011), “Credibility and Determinism in a Game of Persuasion”, *Games and Economic Behavior*, 71 (2 Mar. 2011), pp. 409-419. (p. 6.)