Persuasion with Verifiable Information

by

Maria (Masha) Titova

vanderbilt university

September 29, 2022

- persuasion games with verifiable information
  - privately informed sender
    - wants receiver to approve his proposal
    - sends verifiable messages
  - uninformed receiver who chooses choosing between
    - approving and rejecting proposal
- many applications
  - prosecutor convinces judge to convict, selects evidence
  - politician convinces voter to elect him, chooses campaign promises
  - job market candidate convinces employer to offer job, lists qualifications

- **persuasion games with verifiable information**
  - direct implementation: can restrict attention to direct equilibria
    - sender tells receiver what to do
  - ranking of equilibrium outcomes (ex-ante utility of sender):
    - worst: equivalent to full disclosure
    - best: commitment outcome (Kamenica and Gentzkow, 2011)

- prosecutor knows that defendant committed $\theta \in \{0, 1, 2\}$ violations
  - possesses hard inculpatory evidence for each violation
  - communicates with judge by presenting evidence
  - wants to convince judge to convict
- judge thinks each $\theta \in \{0, 1, 2\}$ is equally likely
  - wants to convict if $\theta = 2$
  - *net payoff from conviction* is 1 id $\theta = 2$ and -1 otherwise

one equilibrium:

- prosecutor presents all the evidence he has
- when judge sees
  - 2 pieces of evidence $\rightarrow \theta$ must be 2
  - 1 piece of evidence $\rightarrow \theta$ could be 1 or 2 $\rightarrow \theta$ must be 1
  - 0 pieces of evidence $\rightarrow \theta$ must be 0
- outcome: defendants with 2 violations are convicted, rest acquitted
  - ex-ante probability of conviction is $1/3$

another equilibrium:

- prosecutor presents 1 piece of evidence whenever $\theta \geq 1$
- when judge sees
  - 2 pieces of evidence $\rightarrow$ θ must be 2 (this is off path / never happens)
  - 1 piece of evidence $\rightarrow$ θ could be 1 or 2
    - $\theta = 1$ with prob. $1/2$ and $\theta = 2$ with prob. $1/2$
    - net payoff from conviction is 0, on average
    - judge convicts
- <u>outcome</u>: defendants with 1 and 2 violations are convicted
  - ex-ante probability of conviction is $2/3$
  - happens to also be a *commitment outcome*

- **communication**:
  - **Milgrom (1981)** and **Grossman (1981)**; Kamenica and Gentzkow (2011); Crawford and Sobel (1982); Spence (1973); Lipnowski and Ravid (2020)

  my contribution: sender reaches commitment outcome with evidence

- **mechanisms with evidence**:
  - Glazer and Rubinstein (2004, 2006); Sher (2011); Hart, Kremer, and Perry (2017); Ben-Porath, Dekel, and Lipman (2019)

  my contribution: upper bound of sender's equilibrium payoff is solution to optimal information design problem

- **applied Bayesian persuasion**:
  - Kolotilin (2015); Ostrovsky and Schwarz (2010); Boleslavsky and Cotton (2015); Romanyuk and Smolin (2019); Alonso and Câmara (2016); Bardhi and Guo (2018); Gehlbach and Sonin (2014); Egorov and Sonin (2019)

  my contribution: sender has commitment → sender's messages are verifiable

MODEL

$$\Theta := \left\{0, \frac{1}{T}, \ldots, \frac{T-1}{T}, 1\right\} \text{ – \underline{state space}}$$

$$\text{if } T = \infty, \text{ then } \Theta = [0,1] \text{ is \textit{rich}}$$

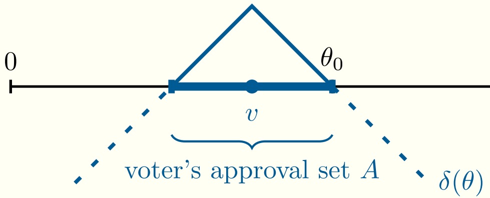
- **sender (he/him)**
  - privately observes state of the world $\theta \in \Theta$
    - θ drawn from common prior $\mu_0 \in \Delta\Theta$ with full support
  - gets 1 if receiver approves, 0 otherwise
    - state-independent preferences
  - sends verifiable message $m \in M := \mathcal{B}(\Theta)$ to receiver
    - message m is verifiable in state θ if $\theta \in m$

- **receiver (she/her)**
  - net payoff from approval is $\delta(\theta)$
    - she approves in state θ if only if $\delta(\theta) \geq 0$
  - her complete-information approval set is

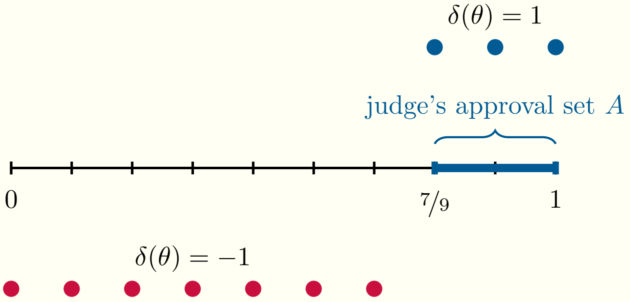
$$A := \{\theta \in \Theta \mid \delta(\theta) \geq 0\}$$

  - I assume that
    - δ is integrable
    - receiver's approval set has positive prior measure: $\int_A \delta(\theta) d\mu_0(\theta) > 0$
    - receiver rejects under prior: $\int_\Theta \delta(\theta) d\mu_0(\theta) < 0$

- **state space**: $[0, 1]$, range of policies
  - status quo policy $\theta_0 \in [0, 1]$
- **sender**: challenger who wants voter to approve his policy
- **receiver**: voter with spatial preferences
  - ideal policy $v \in [0, 1]$, net payoff from approval is $\delta(\theta) = |v - \theta_0| - |v - \theta|$



- **state space**: fraction of violations committed by defendant (out of 9)
- **sender**: prosecutor who wants judge to convict
- **receiver**: judge who wants to convict defendants with at least 7 violations
  - net payoff from approval is $\delta(\theta) = 1$ if $\theta \geq 7/9$ and $\delta(\theta) = -1$ if $\theta < 7/9$



EQUILIBRIUM OUTCOMES

- (Perfect Bayesian) Equilibrium (σ, a, q)
  - $\sigma : \Theta \times M \to [0,1]$ – sender's strategy
    - $\sigma(m \mid \theta)$ is probability of sending message m from state θ
    - $\forall \theta \in \Theta$, $\sigma(\cdot \mid \theta)$ is supported on $\arg\max_{m \in M} u_s\big(a(m)\big)$, subject to $\theta \in m$
  - $a : M \to \{0,1\}$ – approval strategy of receiver
    - $\forall m \in M$, best response is $a(m) = \mathbb{1}\left(\int_\Theta \delta(\theta) dq(\theta \mid m) \geq 0\right)$
  - $q : \Theta \times M \to [0,1]$ – posterior belief of receiver
    - $q(\cdot \mid m) \in \Delta\Theta$ is posterior belief after message $m \in M$
    - Bayes-rational on equilibrium path
    - $\operatorname{supp} q(\cdot \mid m) \subseteq m$, $\forall m \subseteq \Theta$

OUTCOMES: DEFINITIONS

- outcome $\alpha : \Theta \to [0,1]$ specifies probability $\alpha(\theta)$ that receiver approves proposal in state θ
- outcome α is equilibrium outcome if it corresponds to some equilibrium
- outcome α^c is commitment outcome if it solves[1]

$$\max_{\alpha} \int_{\Theta} \alpha(\theta) d\mu_0(\theta), \quad \text{subject to} \quad \begin{array}{c} \forall \theta \in \Theta,\ 0 \leq \alpha(\theta) \leq 1 \\ \displaystyle\int_{\Theta} \alpha(\theta)\delta(\theta) d\mu_0(\theta) \geq 0 \end{array}$$

[1] *Kamenica and Gentzkow (2011), Alonso and Câmara (2016)*

- outcome α is deterministic if $\alpha(\theta) \in \{0, 1\}$ for every $\theta \in \Theta$
- set of approved states W in the deterministic outcome α is

$$W := \{\theta \in \Theta \mid \alpha(\theta) = 1\}$$

EQUILIBRIUM ANALYSIS

- consider deterministic equilibrium outcome with set of approved states W. What conditions does W satisfy?

  - sender cannot deviate to full disclosure:
    - if $\theta \in A$, message $\{\theta\}$ convinces receiver to approve

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

  - receiver's expected net payoff from approval is non-negative:

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

Theorem 1

(1) every equilibrium outcome is deterministic

(2) $W \subseteq \Theta$ is an equilibrium set of approved states $\iff$ W satisfies (IC) and (obedience)

- **Proof** of $\boxed{\textit{(1)}}$, by contradiction:
  - consider equilibrium (σ, a, q) with outcome α
  - α is not deterministic $\implies$ exists θ s.t. $\alpha(\theta) \in (0, 1)$
  - since $\alpha(\theta) > 0$, there exists message m such that $a(m) = 1$ and $\theta \in m$
  - <u>profitable deviation</u>: send m with certainty when state is θ

Theorem 1

(1) every equilibrium outcome is deterministic

(2) $W \subseteq \Theta$ is an equilibrium set of approved states $\iff$ W satisfies (IC) and (obedience)

- **Proof** of $\boxed{\textit{(2)}, \implies}$: W is set of approved states in equilibrium (σ, a, q)
  - W satisfies (IC), or else sender can deviate to full disclosure
  - W satisfies (obedience):
    - let $\mathcal{M} := \{m \in M \mid a(m) = 1\}$ be *set of convincing messages*
    - if $\theta \in W$, sender convinces with prob. 1: $\sigma(\mathcal{M} \mid \theta) = 1$
    - every $m \in \mathcal{M}$ convinces receiver: $\int_W \delta(\theta)\sigma(m \mid \theta) d\mu_0(\theta) \geq 0$
    - true for $\mathcal{M}$: $\int_W \delta(\theta)\sigma(\mathcal{M} \mid \theta) d\mu_0(\theta) = \int_W \delta(\theta) d\mu_0(\theta) \geq 0$

Theorem 1

(1) every equilibrium outcome is deterministic

(2) $W \subseteq \Theta$ is an equilibrium set of approved states $\iff$ W satisfies (IC) and (obedience)

- **Proof** of $\boxed{(2), \Longleftarrow}$: direct implementation of W:
  - sender: $\sigma(W \mid \theta) = \mathbb{1}(\theta \in W)$ and $\sigma(\Theta \setminus W \mid \theta) = \mathbb{1}(\theta \notin W)$
  - receiver:
    - on path, approves after W by (obedience), rejects after $\Theta \setminus W$
    - off path is "skeptical"

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

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

- **Theorem 1** allows us to restrict attention to sets of approved states $W \subseteq \Theta$ satisfying (IC) and (obedience)
- rank equilibria by sender's ex-ante utility
  - same as his ex-ante odds of approval
  - equals $\mu_0(W) = \int_W d\mu_0(\theta)$, prior measure of set of approved states

- minimize sender's ex-ante utility across all equilibria
  - smallest (in terms of ex-ante utility) set of approved states $\underline{W}$
  - sender's (IC) constraint binds: $\underline{W} = A$
- receiver makes fully informed choice
- outcome-equivalent to full disclosure AKA full unraveling
  - (Grossman, 1981); (Milgrom, 1981); (Milgrom and Roberts, 1986); reviewed by (Milgrom, 2008)

- maximize sender's ex-ante utility across all equilibria
  - largest (in terms of ex-ante utility) set of approved states $\overline{W}$
  - receiver's (obedience) constraint binds

Theorem 2

Let $\Theta = [0, 1]$. Then, $\overline{W}$ is characterized by a cutoff value $c^* > 0$ such that

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

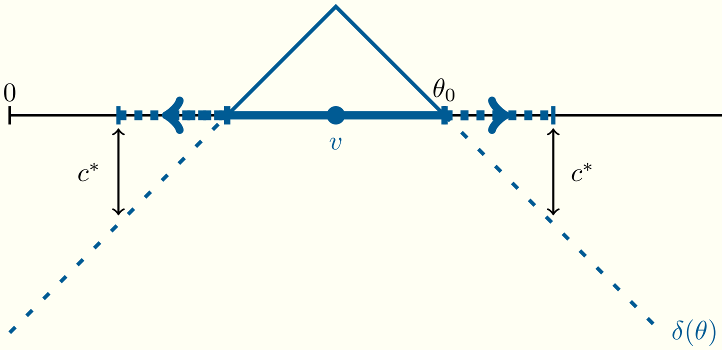
Furthermore, SP equilibrium outcome is a commitment outcome.

- $\overline{W}$ solves $\max_{W \subseteq \Theta} \int_W d\mu_0(\theta)$ subject to $A \subseteq W$ and $\int_W \delta(\theta) d\mu_0(\theta) \geq 0$
  - ◇ adding θ to $\overline{W}$ has "benefit" $\mu_0(\theta)$ and "cost" $-\delta(\theta)\mu_0(\theta)$
    - add $\theta \in A$ to $\overline{W}$ because $\delta(\theta) \geq 0 \implies$ (IC) holds
    - if $\delta(\theta_2) < \delta(\theta_1) < 0$, add θ_1 first
  - ◇ (obedience) binds, or else can increase objective

	SP equilibrium	commitment	
find α to maximize	$\int_\Theta \alpha(\theta) d\mu_0(\theta)$	$\int_\Theta \alpha(\theta) d\mu_0(\theta)$	
subject to	$\int_\Theta \alpha(\theta)\delta(\theta) d\mu_0(\theta) \geq 0$		
	$\alpha(\theta) \in \{0,1\}$	$\alpha(\theta) \in [0,1]$	$\forall \theta \in \Theta$

- if commitment outcome α^c is not deterministic, **purify it**
  - let $\mathcal{D} := \{\theta \in \Theta \mid 0 < \alpha^c(\theta) < 1\}$ – notice that $\delta(\theta) = const$, $\forall \theta \in \mathcal{D}$
  - partition $\mathcal{D}$ into X and Y, where $\int_{\mathcal{D}} \alpha^c(\theta) d\mu_0(\theta) = \int_X d\mu_0(\theta)$

$$\widetilde{\alpha}^c(\theta) = \begin{cases} \alpha^c(\theta), \text{ if } \theta \notin \mathcal{D} \\ 1, \text{ if } \theta \in X \\ 0, \text{ if } \theta \in Y \end{cases} \quad \text{is a deterministic commitment outcome}$$


0
θ_0
1
v
c^*
c^*
$\delta(\theta)$

- suppose receiver weakly prefers higher states
- then, we can characterize sender-preferred equilibrium via cutoff *state*

Corollary 1

Suppose that $\Theta = [0, 1]$ and δ is (weakly) increasing in θ. Then,

- there exists $\theta^* \in \Theta \setminus A$ such that $\int_{\theta^*}^{1} \delta(\theta) d\mu_0(\theta) = 0$
- set $[\theta^*, 1]$ is a SP equilibrium set of approved states
- SP equilibrium outcome $\overline{\alpha}(\theta) = \mathbb{1}(\theta \geq \theta^*)$ is a commitment outcome

- even if $\delta(\theta)$ is increasing, SP equilibrium may not feature cutoff state if Θ is finite
- **example**:

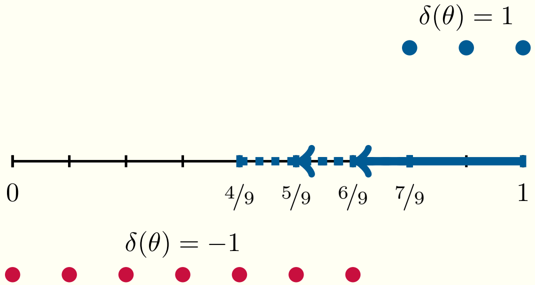
state	0	1/2	1
prior	1/6	1/2	1/3
net payoff from approval	-2	-1	1

- however, commitment payoff can always be implemented via cutoff state

Theorem 3

Suppose that Θ is finite, δ is (weakly) increasing in θ, and there exists $\theta^* \in \Theta \setminus A$ such that $\sum_{\theta=\theta^*}^{1} \delta(\theta)\mu_0(\theta) = 0$. Then,

- interval $[\theta^*, 1]$ is a SP equilibrium set of approved states
- SP equilibrium outcome $\overline{\alpha}(\theta) = \mathbb{1}(\theta \geq \theta^*)$ is a commitment outcome

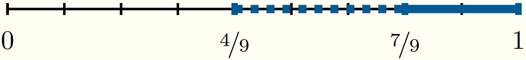


- <u>uniform prior</u>: prosecutor finds smallest θ^+ such that $\sum_{\theta=\theta^+}^{1} \delta(\theta) \geq 0$

  $\rightarrow \boxed{\theta^* = 4/9}$, which binds the constraint

  - apply Theorem 3
    - SP equilibrium set of approved states is $\{4/9, \ldots, 8/9, 1\}$
    - SP equilibrium outcome is a commitment outcome

PROSECUTOR AND JUDGE: SP EQUILIBRIUM



- SP equilibrium set of approved states is $\{4/9, \ldots, 8/9, 1\}$
  - every defendant with 4+ violations is convicted
  - 60% of defendants are convicted / 70% are ex-ante innocent
- implementation: for each defendant with 4+ violations, prosecutor presents exactly 4 pieces of evidence
  - judge sees 4 pieces of evidence → must be 4+ violations
    - 4, 5, … 9 violations are equally likely
    - judge is indifferent so she convicts

ROBUSTNESS

many (independent) receivers

$I := \{1, \ldots, n\}$ – set of receivers

$\mu_0 \in \Delta\Theta$ is common prior

- **sender**:
  - has state-independent utility $u_s : \{0,1\}^n \to \mathbb{R}$
  - u_s weakly increases in every receiver's action
- **receiver** $i \in I$:
  - observes private verifiable message $m_i \subseteq \Theta$ chosen by sender
  - solves independent problem: approves iff $\delta_i(\theta) \geq 0$

many (independent) receivers: results

- $(W_1, \ldots, W_n)$ is an equilibrium collection of sets of approved states $\iff$ for all $i \in I$
  - $A_i \subseteq W_i$
  - $\int\limits_{W_i} \delta_i(\theta) d\mu_0(\theta) \geq 0$
- if $\Theta = [0, 1]$, then SP equilibrium outcome is a commitment outcome

- receiver chooses action from set $J = \{0, 1, \ldots, k\}$ with $k \geq 2$
- receiver's complete-information approval set for action $j \in J$ is A_j
- outcome is a partition $(W_0, W_1, \ldots, W_k)$
  - ◇ $W_j \subseteq \Theta$ are states in which receiver plays action $j \in J$
- (IC): if $\theta \in A_j$ then $\theta \in W_j \cup \cdots \cup W_k$
  - ◇ **may be violated in every commitment outcome**

- I solve persuasion games with verifiable information
  - direct implementation: W is an equilibrium set of approval states $\iff$
    - W satisfies receiver's (obedience) and sender's (IC) constraints
  - set of equilibrium outcomes:

    worst: full disclosure $\rightarrow$ best: commitment outcome*

* *if state space is sufficiently rich*

Thank You!

connection to revelation principle(s)

- Myerson (1986) and Forges (1986):
  - any equilibrium of a *mediated* sender-receiver game is outcome-equivalent to one in which
    - sender truthfully reveals θ to mediator
    - mediator recommends action
    - receiver obediently follows recommendation
  - **Theorem 1** provides necessary and sufficient conditions for W to be implementable in equilibrium
- Kamenica and Gentkow (2011) and Bergemann and Morris (2019):
  - WLOG to let *set of signals* equal *set of actions*