Persuasion with Verifiable Information

by

Maria (Masha) Titova

Vanderbilt University

May 19, 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, presents 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)

MOTIVATING EXAMPLE, KAMENICA AND GENTZKOW (2011)

state space:	(defendant is)	*innocent*	*guilty*
common prior:		0.7	0.3

prosecutor wants judge to convict; judge wants to convict iff $Pr(guilty) \geq \frac{1}{2}$

- commitment outcome:
  - prosecutor:
    - if *guilty*, send g with prob. 1
    - if *innocent*, send g with prob. α
  - judge:

$$Pr(guilty \mid g) = \frac{1 \cdot 0.3}{1 \cdot 0.3 + \alpha \cdot 0.7} = \frac{1}{2} \Longrightarrow \alpha = \frac{3}{7}$$

  - judge convicts 60% of defendants (all *guilty* and $\frac{3}{7}$ of *innocent*)

state space:	(defendant is)	*innocent*	*guilty*
common prior:		0.7	0.3

prosecutor wants judge to convict; judge wants to convict iff $Pr(guilty) \geq \frac{1}{2}$

- **prosecutor is informed**, does not have commitment power
- message space: $\{g, i, \{g, i\}\}$
  - verifiability: cannot say g (i) if *innocent* (*guilty*)
- can judge convict $\frac{3}{7}$ of *innocent* in equilibrium?
  - NO: in every equilibrium, *guilty* are convicted, *innocent* are acquitted

continuous state space allows sender

to reach commitment outcome with verifiable messages!

- **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 verifiable information
- **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

$$\Omega := [0,1] - \underline{\text{state space}}$$

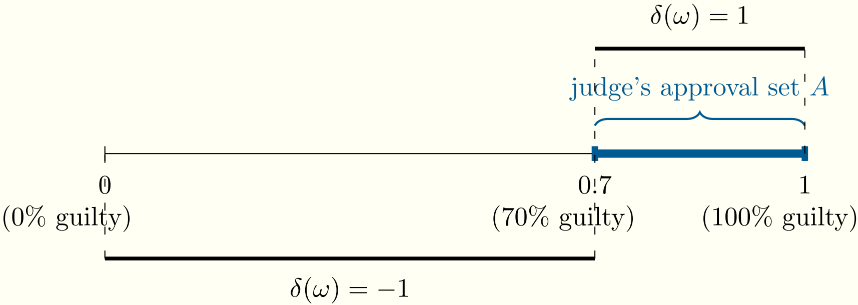
- **sender (he/him)**
  - privately observes state of the world $\omega \in \Omega$
    - ω drawn from common prior $p > 0$ over Ω
  - gets 1 if receiver approves, 0 otherwise
    - state-independent preferences
  - sends verifiable message $m \subseteq \Omega$ to receiver
    - *grain of truth*: $\omega \in m$

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

$$A := \{\omega \in \Omega \mid \delta(\omega) \geq 0\}$$

  - assume $\mathbb{E}_p[\delta(\omega)] < 0$

- **state space**: how guilty defendant is (0% to 100% guilty)
  - prior is uniform on $[0, 1]$
- **sender**: prosecutor (wants judge to convict)
- **receiver**: judge (wants to convict defendants who are $\geq 70\%$ guilty)



EQUILIBRIUM OUTCOMES

- (Perfect Bayesian) Equilibrium (σ, a, q)
  - $\sigma(m \mid \omega)$ – prob. that sender sends m is state ω
    - maximizes sender's utility $\forall \omega \in \Omega$ subject to $\omega \in m$, $\forall m \subseteq \Omega$
  - $a(m) \in \{0, 1\}$ – approval strategy of receiver
    - best response $a(m) = \mathbb{1}\big(\mathbb{E}_{q(\cdot \mid m)}[\delta(\omega)] \geq 0\big)$
  - $q(\cdot \mid m) \in \Delta\Omega$ – posterior belief of receiver
    - Bayes-rational on equilibrium path
    - $\operatorname{supp} q(\cdot \mid m) \subseteq m$, $\forall m \subseteq \Omega$

- outcome α specifies $\forall\omega \in \Omega$ probability $\alpha(\omega)$ that receiver approves
  - outcome α is equilibrium outcome if it corresponds to some equilibrium
  - outcome α^c is commitment outcome if it solves[1]

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

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

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

$$W := \{\omega \in \Omega \mid \alpha(\omega) = 1\}$$

EQUILIBRIUM ANALYSIS

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

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

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

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

$$\mathbb{E}_p[\delta(\omega) \mid W] \geq 0 \tag{obedience}$$

Theorem 1

(1) every equilibrium outcome is deterministic

(2) $W \subseteq \Omega$ 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 $\Longrightarrow$ exists ω s.t. $\alpha(\omega) \in (0, 1)$
  - since $\alpha(\omega) > 0$, there exists message m s.t. $a(m) = 1$ and $\omega \in m$
  - <u>profitable deviation</u>: send m with certainty when state is ω

Theorem 1

(1) every equilibrium outcome is deterministic

(2) $W \subseteq \Omega$ 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 $\mathcal{M} := \{m \subseteq \Omega \mid a(m) = 1\}$ be *set of convincing messages*
    - if $\omega \in W$, sender convinces w. prob. 1: $\sum_{m \in \mathcal{M}} \sigma(m \mid \omega) = 1$
    - every $m \in \mathcal{M}$ convinces receiver: $\int_W \delta(\omega)\sigma(m \mid \omega)p(\omega)d(\omega) \geq 0$
    - take sum over all $m \in \mathcal{M}$, get $\mathbb{E}_p[\delta(\omega) \mid W] \geq 0$

Theorem 1

(1) every equilibrium outcome is deterministic

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

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

$$\forall m \subseteq A, \quad \operatorname{supp} q(\cdot \mid m) \subseteq m, \text{ so that } \mathbb{E}_q[\delta(\omega)] \geq 0$$

$$\forall m \nsubseteq A, m \neq W, \quad \operatorname{supp} q(\cdot \mid m) \subseteq m \setminus A, \text{ so that } \mathbb{E}_q[\delta(\omega)] < 0$$

- **Theorem 1** allows us to restrict attention to sets of approved states $W \subseteq \Omega$ satisfying (IC) and (obedience)
- rank equilibria by sender's ex-ante utility
  - same as his ex-ante odds of approval
  - equals $P(W) := \int_W p(\omega) d\omega$, 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

$\overline{W}$ is characterized by a cutoff value $c^* > 0$ such that

- receiver approves a.s. if $\delta(\omega) > -c^*$ and rejects if $\delta(\omega) < -c^*$
- whenever receiver approves, her expected net payoff from approval is zero: $\mathbb{E}_p\left[\delta(\omega) \mid \overline{W}\right] = 0$

Furthermore, SP equilibrium outcome is a commitment outcome

PROOF OF THEOREM 2, PART I

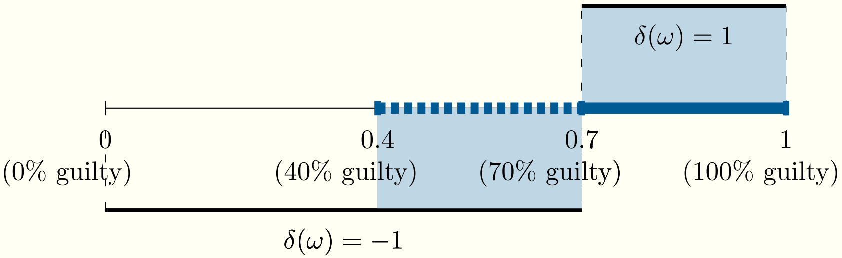
- $\overline{W}$ solves $\max_{W \subseteq \Omega} \int_W p(\omega) d\omega$ subject to $A \subseteq W$ and $\int_W \delta(\omega) p(\omega) d\omega \geq 0$
  - ◇ adding ω to $\overline{W}$ has "benefit" $p(\omega)$ and "cost" $-\delta(\omega)p(\omega)$
    - add $\omega \in A$ to $\overline{W}$ because $\delta(\omega) \geq 0 \implies$ <u>(IC) holds</u>
    - if $\delta(\omega_2) < \delta(\omega_1) < 0$, add ω_1 first
  - ◇ <u>(obedience) binds</u>, or else can increase objective

PROOF OF THEOREM 2, PART II

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

- commitment outcome α^c may not be deterministic
  - let $\mathcal{D} := \{\omega \in \Omega \mid 0 < \alpha^c(\omega) < 1\}$ – notice that $\delta(\omega) = const$, $\forall \omega \in \mathcal{D}$
  - partition $\mathcal{D}$ into X and Y, where $\int_{\mathcal{D}} \alpha^c(\omega)p(\omega)d\omega = \int_X p(\omega)d\omega$

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



- sender-preferred equilibrium: $\max_{W \subseteq \Omega} |W|$ s.t. $\int_W \delta(\omega) d\omega = 0 \;\rightarrow\; \boxed{\overline{W} = [0.4, 1]}$
  - ◇ judge convicts 60% of defendants even though 70% are innocent
  - ◇ <u>implementation</u>: if $\omega \geq 0.4$, send $[0.4, 1]$; if $\omega < 0.4$, send $[0, 0.4)$
    - judge: $U[0,1] \xrightarrow{[0.4,1]} U[0.4, 1] \rightarrow$ posterior mean is $0.7 \rightarrow$ convict

ROBUSTNESS

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

p is common prior

- **sender**:
  - has state-independent utility $u_s : 2^I \to \mathbb{R}$
  - u_s weakly increases in every receiver's action
- **receiver** $i \in I$:
  - observes private verifiable message $m_i \subseteq \Omega$ chosen by sender
  - solves independent problem: approves if $\omega \in A_i := \{\omega \in \Omega \mid \delta_i(\omega) \geq 0\}$

many (independent) receivers: results

- $(W_1, \ldots, W_n)$ is an equilibrium collection of sets of approved states $\iff$
  - $A_i \subseteq W_i$, for all $i \in I$
  - $\mathbb{E}_p[\delta_i(\omega) \mid W_i] \geq 0$
- 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 $(W_0, W_1, \ldots, W_k)$
  - $W_j \subseteq \Omega$ are states in which receiver plays action $j \in J$
- (IC): if $\omega \in A_j$ then $\omega \in W_j \cup \cdots \cup W_k$
  - **may be violated in all commitment outcomes**

- 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

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 ω 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*