On Information Design in Games

Mathevet, Perego and Taneva (2018)

Akira Matsushita

UTGSE M1

October 15, 2018

Overview

1. Summary

Settings

- This paper extends Kamenica and Gentzkow (2011)[1]'s single agent model into a multi-agent Bayesian game setting
- Each agent's utility depends on the unobservable state of the world θ and agents' action profile $\mathbf{a} = (a_i)$
- An information designer commits to a mechanism that observes the true state θ and sends signals to the agents
- The designer would like to find the best mechanism that maximizes the value function depending on θ and $\mathbf{a}$
- e.g. investment game under imcomplete information

[1] Emir Kamenica and Matthew Gentzkow. "Bayesian persuasion". In: *American Economic Review* 101.6 (2011), pp. 2590–2615.

Contributions

- Characterizing the feasible distributions of agents' beliefs that a designer can induce through the choice of information structure
  - A designer cannot persuade agents of arbitrary beliefs - there are constraints
- Eliciting the structure of agents' belief distributions and characterize the designer's problem in terms of it
  - Theorem 1 shows that any belief distribution that the designer can induce is a combination of basic communication schemes
  - And Corollary 1 suggests a two-step solution to the designer's problem: first selects an optimal components and second finds a best mixture of them

Contributions

- The results can apply to a variety of solution concepts and equilibrium selection rules
  - e.g. the case that agents only have bounded depths of reasoning, that they can deviate in coalitions, or that they can communicate
  - e.g. the robust mechanism that considers pessimistic equilibrium selection
- Finally we see an application to an investment game under an adversarial equilibrium selection (section 5)

2. Model

Model - 1

- Finite set of players: $N = \{1, 2, \dots, n\}$
- Uncertain state of the world $\theta \in \Theta$, where Θ is a finite set
- Finite action set of each player: A_i, and let $A = \prod_{i \in N} A_i$
- Utility function of each player: $u_i : A \times \Theta \to \mathbb{R}$, and let $u = \prod_{i \in N} u_i$
- θ is distributed according to $\mu_0 \in \Delta\Theta$, which is common knowledge
- Each player would like to maximize her expected payoff
- Refer to $G = (\Theta, \mu_0, N, A, u)$ as the **base game**

Model - 2

- A designer commits to disclosing information to the players about θ
- This is modeled by an **information structure** (S, π)
  - A finite set of signals that player i can receive: S_i, and $S = \prod_{i \in N} S_i$
  - An information map $\pi : \Theta \to \Delta S$
- For any state $\theta \in \Theta$, the message profile drawns from $\pi(s \mid \theta)$ and player i privately observes s_i
- The designer commits to a mechanism **before** she knows the true state θ

Model - 3

- The designer's payoff function: $v : A \times \Theta \to \mathbb{R}$
- The designer would like to maximize her expected payoff
- The combination of a base game and an information structure constitutes a Bayesian game $\mathcal{G} = \langle G, (S, \pi) \rangle$
- **Solution concept**: $\Sigma(\mathcal{G}) \subseteq \{\sigma = (\sigma_i) \mid \sigma_i : S_i \to \Delta A_i \text{ for all } i \in N\}$
  - Σ is a function that maps a Bayesian game to a set of equilibrium strategies
  - An equlibrium concept is arbitrary. e.g. BNE

Model - 4

- The resulting outcomes are represented by

$$O_\Sigma(\mathcal{G}) = \{\gamma \in \Delta(A \times \Theta) \mid \exists \sigma \in \Sigma(\mathcal{G}) \\ \text{s.t. } \gamma(a,\theta) = \sum_s \sigma(a \mid s)\pi(s \mid \theta)\mu_0(\theta)\}$$

- Assumption: O_Σ is non-empty and compact-valued
  - Given G is finite, this holds when Σ is BNE (?)
  - Non-emptiness comes from the fact that $\Sigma(\mathcal{G}) \neq \emptyset$
  - $\Delta(A \times \Theta) = \{p \in \mathbb{R}_+^{|A| \times |\Theta|)} \mid \sum_j p_j = 1\}$, so $O_\Sigma(\mathcal{G})$ is compact $\iff$ $O_\Sigma(\mathcal{G})$ is closed and bounded
  - If $\Sigma(\mathcal{G})$ is compact, then O_Σ is compact since γ is a linear function of σ
  - Boundedness comes from $\Sigma(\mathcal{G}) \subseteq \{\sigma = (\sigma_i) \in \mathbb{R}_+^{\prod_i(|A_i| \times |S_i|)} \mid \sum \sigma_i = 1\}$
  - $\Sigma(\mathcal{G})$ is closed by the closed graph theorem

Model - 5

- For a fixed game G, we just write $O_\Sigma(S,\pi)$ instead of $O_\Sigma(\mathcal{G})$
- When $O_\Sigma(\mathcal{G})$ is not singleton, the designer expects that one of them will happen, which is described by a **selection rule**:

$$g : D \ (\subseteq \Delta(A \times \Theta)) \mapsto g(D) \ (\in D)$$

  - e.g. A pessimistic designer, or one interested in robust information design, expects the worst outcome:

$$g(D) \in \arg\min_{\gamma \in D} \sum_{a,\theta} \gamma(a,\theta) v(a,\theta)$$

    for all compact (so a minimizer exists) $D \subseteq \Delta(A \times \Theta)$
  - Other rules, such as optimistic selection rule and random choise rule, can be considered

Model - 6

- Let $g^{(S,\pi)} = g(O_\Sigma(S,\pi))$
- Now the designer's expected payoff is given by

$$V(S,\pi) \equiv \sum_{a,\theta} g^{(S,\pi)}(a,\theta)v(a,\theta)$$

- And the information design problem is $\sup_{(S,\pi)} V(S,\pi)$

3. Belief Distributions

Belief Hierarchy

- A **belief hierarchy** t_i for player i is an infinite sequence $(t_i^1, t_i^2, \ldots)$, whose components are **coherent** beliefs of all orders
  - $t_i^1 \in \Delta(\Theta)$ is i's first order belief, $t_i^2 \in \Delta(\Theta \times (\Delta(\Theta))^{n-1})$ is i's second order belief, and so on
- A hierarchy t is **coherent** if any belief t_i^k coincides with all beliefs of lower order, $\{t_i^n\}_{n=1}^{k-1}$, on lower order events:

$$\operatorname*{marg}_{X_{k-1}} t_i^k \equiv \sum_{\Theta^{n-1}} t_i^k = t_i^{k-1}, \text{ where } X_{k-1} \equiv \text{supp } t_i^{k-1}$$

for all $k \geq 2$

Belief Hierarchy

- Whereas a player's belief hierarchies are coherent, they may assign positive probability to other players' belief hierarchies that are not coherent

- However Brandenburger and Dekel (1993)[2] showed that we can construct a set of coherent belief hierarchies T_i for every $i \in N$ such that there exists a homeomorphism (bijective and continuous mapping)

$$\beta_i^* : T_i \to \Delta(\Theta \times T_{-i})$$

  for all $i \in N$

  - β_i^* describes i's belief about (θ, t_{-i}) given t_i, and makes coherency common knowledge

- Define $T \equiv \prod_{i \in N} T_i$

[2] Adam Brandenburger and Eddie Dekel. "Hierarchies of beliefs and common knowledge". In: *Journal of Economic Theory* 59.1 (1993), pp. 189–198.

Belief Hierarchy

- Given (S, π) and μ_0, player i recieves s_i and use Bayes' rule to formulate beliefs $\mu_i(s_i) \in \Delta(\Theta \times S_{-i})$, called first-order belief
- $\mu_i^1(s_i) \equiv \underset{\Theta}{\text{marg}}\, \mu_i(s_i)$: i's first-order belief about the state
- $\mu_i^2(s_i)$: i's second-order belief about the state, derived from i's first-order belief about s_{-i} and $j's$ $(\neq i)$ first-order beliefs about the state
- and so on...

Belief Hierarchy Distribution

- Every $s_i \in S_i$ induces a belief hierarchy $h_i(s_i) \in T_i$
- And so every $s \in S$ induces a profile of belief hierarchies $h(s) \equiv (h_i(s_i))_{i \in N}$

Definition (Definition 1)

An information structure (S, π) induces a distribution $\tau \in \Delta T$ over profiles of belief hierarchies, called a belief(-hierarchy) distribution, if

$$\tau(t) = \sum_{\theta} \pi(\{s \mid h(s) = t\} \mid \theta)\mu_0(\theta) \tag{3}$$

for all $t \in T$

Belief Hierarchy: Example

- The information structure given by the following table induces

$$\tau = \frac{3}{4} t_{1/3} + \frac{1}{4} t_1$$

when $\mu_0 \equiv \mu_0(\theta = 1) = \frac{1}{2}$, where t_μ is a hierarchy profile in which $\mu \equiv \mu(\theta = 1)$ is commonly believed

$\pi(\cdot\|0)$	s_1	s_2
s_1	1	0
s_2	0	0

$\pi(\cdot\|1)$	s_1	s_2
s_1	$\frac{1}{2}$	0
s_2	0	$\frac{1}{2}$

TABLE 1: A (Public) Information Structure

Public/Private Belief Distribution

- We categorize belief distributions into public and private

Definition (Definition 2)

A belief distribution τ is **public** if

- $t_i^1 = t_j^1$ for all $i, j \in N$
- $(\operatorname{marg}_{T_{-i}} \beta_i^*(\cdot \mid t_i))(t'_{-i}) = \mathbb{1}\{t'_{-i} = t_{-i}\}$ for all $i \in N$

for all $t \in \operatorname{supp} \tau$ and $|\operatorname{supp} \tau| \geq 2$. Otherwise τ is **private**

- we categorize the $|\operatorname{supp} \tau| = 1$ degenerate case as private

Belief Manipulation: Example

- Consider the following ‘guessing the state’ game

Example

- Players: $N = \{1, 2\}$
- States: $\Theta = \{0, 1\}$
- Actions: $A_i = \{0, 1\}$ for $i = 1, 2$
- Utilities: $u_i(a_i, \theta) = -(a_i - \theta)^2$ for $i = 1, 2$
- Designer's payoff: $v = u_1 - u_2$
- $\mu_0(\theta = 0), \mu_0(\theta = 1) > 0$

Here the designer could obtain her maximal payoff of 1, if she could

- somehow reveal the state perfectly to player 1
- persuade player 2 that the opposite state has realized

Is it possible? - If they have a common prior, no.

Belief Manipulation: Common Priors

- Aumann (1976)[3] showed that Bayesian agents cannot agree to disagree if they have a common prior
- Say $p \in \Delta(\Theta \times T)$ is a **common prior** if

$$p(\theta, t) = \beta_i^*(\theta, t_{-i} \mid t_i) \times (\operatorname*{marg}_{T_i} p)(t_i) \tag{4}$$

  for all θ, t and i

- That is, all players i obtain their belief map β_i^* by Bayesian updating of the same distribution p

[3] Robert J Aumann. "Agreeing to disagree". In: *The annals of statistics* (1976), pp. 1236–1239.

Belief Manipulation: Common Priors

- Denote by Δ^f the probability measures with finite support
- Define

$$\mathcal{C} \equiv \{\tau \in \Delta^f(T) \mid \exists \text{a common prior } p \text{ s.t. } \tau = \underset{T}{\operatorname{marg}}\, p\}$$

  to be the space of **consistent** (belief-hierarchy) distributions
- In a consistent distribution, all players' beliefs arise from a common prior that draws every t with the same probability as τ
- Let p_τ be the unique distribution p in the above equation (uniquness comes from Mertens and Zamir (1985, Proposition 4.5)[4])

[4] Jean-François Mertens and Shmuel Zamir. "Formulation of Bayesian analysis for games with incomplete information". In: *International Journal of Game Theory* 14.1 (1985), pp. 1–29.

Belief Manipulation: Bayes plausible

- A distribution $\tau \in \Delta^f(T)$ is **Bayes plausible** if the expected first-order belief of at least one player equals the prior:

$$\forall \theta \in \Theta, \sum_{t_i} (\operatorname*{marg}_{\Theta} \beta_i^*(\cdot \mid t_i))(\theta)\tau_i(t_i) = \mu_0(\theta)$$

for some $i \in N$

Characterization of Beliefs

Proposition (Proposition 1)

There exists (S, π) that induces $\tau \in \Delta^f(T)$, if and only if, τ is consistent and Bayes plausible

- This proposition shows the limitation of the designer's belief manipulation
- It does not matter which player i satisfies Bayes plausibility, because by consistency, if it is true for one player, then it will hold for all

Characterization of Beliefs

- Before seeing the proof, let's go back to the example
- To reach the upper bound of 1 of the designer's payoff, it would have to be that

$$\beta_1^*(\theta = 1, t_2 \mid t_1) = 1 \wedge \beta_2^*(\theta = 0, t_1 \mid t_2) = 1$$

for some (t_1, t_2), which violates (4)

The designer's problem: Another approach

- Consistency is the main difference and technical challenge compared to the single-agent case
- One approach to implementing it is to design the individual distributions of players' beliefs and then couple them in a consistent way (Ely (2017) approach)
- Another approach: interpreting the designer's information design problem as a choice of a distribution over her own beliefs about the state, and then viewing players' first-order belief distributions as a special garbling of the designer's information

The designer's problem: Another approach

Proposition (Proposition 2)

If τ is consistent and Bayes plausible, then there exists $v : \text{supp}\ \tau \to \Delta\Theta$ such that:

$$\sum_t \tau(t)v(t) = \mu_0 \tag{6}$$

and

$$\sum_{t_{-i}} \tau(t_{-i} \mid t_i)v(t_i, t_{-i}) = \operatorname*{marg}_{\Theta} \beta_i^*(\cdot \mid t_i),\ \forall i, t_i \tag{7}$$

- $v(t)$ represents the designer's beliefs about the state given t
- τ can be interpreted as the distribution of the designer's beliefs

The designer's problem: Another approach

- (6): τ is consistent and Bayes plausible, then on average the designer's beliefs must be equal to the prior
- (7): τ each player's first-order belief, $\underset{\Theta}{\text{marg}}\ \beta_i^*(\cdot \mid t_i)$ must be derived from the designer's beliefs
  - The (LHS) of (7) is a mean-preserving transformation of the designer's beliefs

The designer's problem: Another approach

Proposition (Proposition 2 (Cont.))

Conversely, if $\xi \in \Delta^f(\Delta\Theta)^n$ and $v : \text{supp } \xi \to \Delta\Theta$ satisfies

$$\sum_{\boldsymbol{\mu}=(\mu_i)} \xi(\boldsymbol{\mu})v(\boldsymbol{\mu}) = \mu_0 \tag{8}$$

(here μ_i is i's first-order belief <u>only for θ</u>, corresponding to previous μ_i^1) and

$$\sum_{\boldsymbol{\mu}_{-i}} \xi(\boldsymbol{\mu}_{-i} \mid \mu_i)v(\boldsymbol{\mu}_{-i}, \mu_i) = \mu_i, \ \forall i, \mu_i, \tag{9}$$

then there exist a consistent and Bayes plausible τ such that
$\text{supp } \tau_i \simeq \text{supp } \xi_i$ (a bijection $\phi : \text{supp } \xi_i \to \text{supp } \tau_i$ exists) and
$\mu_i = \underset{\Theta}{\text{marg}}\ \beta_i^*(\cdot \mid \phi_i(\mu_i))$ for all i, and $\tau(t) = \xi(\phi^{-1}(t))$ for all t

- A method constructing a consistent, Bayes plausible belief distributions, s.t. hierarchies identified with their first-order beliefs

The designer's problem: Another approach

- ξ is a distribution over players' first order beliefs $\boldsymbol{\mu} = (\mu_i)$
- $v(\boldsymbol{\mu}) \in \Delta\Theta$ is the designer's belief about the state given $\boldsymbol{\mu}$
- In the latter part of the proposition, different first-order beliefs generate diffrent hierarchies (?)
- Hence we can find a unique $t \equiv \phi(\boldsymbol{\mu})$ induced by $\boldsymbol{\mu}$. So we can write the common prior p by ξ and v:

$$p(\theta, \boldsymbol{\mu}) \equiv p(\theta, t) = \xi(\boldsymbol{\mu})(v(\boldsymbol{\mu}))(\theta)$$

  for all $\theta, \boldsymbol{\mu}$, which satisfies Bayes plausibility and consistency
- [Proof of proposition 2]

Outcomes from Belief Distributions

- Given a consistent distribution τ , denote a solution concept in the space of beliefs as

$$\Sigma^B(\tau) \subseteq \{\sigma = (\sigma_i) \mid \sigma_i : \text{supp } \tau_i \to \Delta A_i, \forall i \in N\}$$

- This set describes players' behavior in the Bayesian game $\langle G, p_\tau \rangle$, where p_τ is the unique common prior

- The corresponding outcome distributions are

$$\begin{aligned} O_{\Sigma^B}(\tau) \equiv \{\gamma \in \Delta(A \times \Theta) \mid \exists \sigma \in \Sigma^B(\tau) \\ \text{s.t. } \gamma(a, \theta) = \sum_t \sigma(a \mid t) p_\tau(t, \theta), \forall a, \theta\} \end{aligned}$$

Outcomes from Belief Distributions

- We concretely define $\Sigma^B(\tau)$ such that for all consistent τ,

$$O_{\Sigma^B}(\tau) = \cup_{(S,\pi) \text{ induces } \tau} O_\Sigma(S,\pi)$$

  holds

- For example when Σ is BNE, then Liu (2015)[5]'s (beliefpreserving) correlated equilibrium

[5] Qingmin Liu. "Correlation and common priors in games with incomplete information". In: *Journal of Economic Theory* 157 (2015), pp. 49–75.

4. Optimal Solutions

Some Assumptions

- **Assumption 1 (Linear Selection)**: g is linear
  - An assumption that requires the selection criterion to be independent of the subsets of outcomes to which it is applied
  - The best, worst and random selection satisfies linearlity
- **Assumption 2 (Invariant Solution)**: For all consistent τ, τ', if $\sigma \in \Sigma^B(\tau)$, then there exists $\sigma' \in \Sigma^B(\tau')$ such that $\sigma(t) = \sigma(t')$ for all $t \in \text{supp } \tau \cap \text{supp } \tau'$
  - This assumption says that play at a profile of belief hierarchies t under Σ^B is independent of the ambient distribution from which t is drawn

Proposition (Proposition 3)

If Σ^B is invariant, then O_{Σ^B} is linear

Minimality

Lemma (Lemma 1)

C is convex

- A belief distribution $\tau \in C$ is minimal if there is no $\tau' \in C$ such that supp $\tau' \subsetneq$ supp τ
- Let C^M be the set of all minimal distributions in C
- $\tau \in C$ is an **extreme point** of the convex set C if $\tau', \tau'' \in C$, $\alpha \in [0, 1]$ and $\tau = \alpha\tau' + (1 - \alpha)\tau''$, then $\alpha = 0$ or 1
- Let $\mathcal{E}$ be the set of all extreme points in C

Lemma (Lemma 2)

$C^M = \mathcal{E}$

Representation of Belief Distributions

Proposition (Proposition 6)

For any $\tau \in C$, there exist unique $L \in \mathbb{N}$, $\{e_\ell\}_{\ell=1}^{L} \subseteq C^M$ and $\{\alpha_\ell\}_{\ell=1}^{L} \subseteq \mathbb{R}_+$ such that

- $\sum_{\ell=1}^{L} \alpha_\ell = 1$
- $\tau = \sum_{\ell=1}^{L} \alpha_\ell e_\ell$

- For this proposition, it is necessary that $\tau \in \Delta^f(T)$: finite support

Representation Theorem

- Given any consistent distribution τ and the selected outcome $g^{\tau} \equiv g(O_{\Sigma^B}(\tau))$, the designer's ex ante expected payoff is given by

$$w(\tau) \equiv \sum_{\theta, a} g^{\tau}(a, \theta) v(a, \theta)$$

Lemma (Lemma 3)

w is linear in τ over C^M

Representation Theorem

Theorem (Theorem 1)

The designer's maximization problem can be represented as

$$\sup_{(S,\pi)} V(S,\pi) = \sup_{\lambda \in \Delta^f(C^M)} \sum_{e \in C^M} w(e)\lambda(e)$$

$$\text{s.t.} \sum_{e \in C^M} \operatorname*{marg}_{\Theta} p_e \lambda(e) = \mu_0$$

- [Proof of theorem 1]

Representation Theorem

- Theorem 1 states that the designer maximizes her expected payoff as if she were optimally randomizing over minimal consistent distributions, subject to posterior beliefs averaging to μ_0 across those distributions
- Every minimal distribution e induces a Bayesian game and leads to an outcome for which the designer receives expected payoff $w(e)$
- Every minimal distribution has a distribution over states, $\operatorname{marg}_{\Theta} p_e \lambda(e) = \mu_0$, and the further that is from μ_0 , the more costly it is for designers
  - A kind of budget constraints

Within-Between Maximizations

Corollary (Corollary 1)

For any $\mu \in \Delta\Theta$, let

$$w^*(\mu) \equiv \sup_{e \in C^M \text{ s.t. } \operatorname{marg}_{\Theta} p_e = \mu} w(e)$$

Then the designer's maximization problem can be represented as

$$\sup_{(S,\pi)} V(S,\pi) = \sup_{\lambda \in \Delta^f(\Delta\Theta)} \sum_{\mu \in \text{supp } \lambda} w^*(e)\lambda(e)$$
$$\text{s.t.} \sum_{\mu \in \text{supp } \lambda} \mu\lambda(\mu) = \mu_0$$

- w^* is ‘maximization within’ and $\sup_{\lambda \in \Delta^f \Delta\Theta} \sum_{\text{supp } \lambda} w^*(e)\lambda(e)$ is ‘maximization between’

5. Application

Application: Adversarial Selection in Investment Game

- Consider the investment game with a pessimistic selection rule:

Example (Investment Game)

- Players: $N = \{1, 2\}$
- States: $\Theta = \{2, -1\}$
- Designer's payoff: $v(I, I) > v(I, N) = v(N, I) > v(N, N)$, indep of θ
- Note that under complete information, playing I is dominant strategy when $\theta = 2$, and playing N is dominant strategy when $\theta = -1$

(u_1, u_2)	I	N
I	θ, θ	$\theta - 1, 0$
N	$0, \theta - 1$	$0, 0$

Table 2: Investment Game

Application: Adversarial Selection in Investment Game

Example (Investment Game (Cont.))

- Let $\mu_0 \equiv \mu(\theta = 2)$
- Solution concept: BNE
- min selection rule (robust information design):

$$g(O_\Sigma(S,\pi)) \in \underset{\gamma \in O_\Sigma(S,\pi)}{\arg\min} \sum_{a,\theta} \gamma(a,\theta) v(a,\theta)$$

(u_1, u_2)	I	N
I	θ, θ	$\theta - 1, 0$
N	$0, \theta - 1$	$0, 0$

TABLE 2: Investment Game

BCE Program

- The current method available for Bayes-Nash information design (as in Bergemann and Morris (2016)[6]), based on BCE outputs a suboptimal information structure in this problem

- BCE program:

$$\max_{\pi} \sum_{\theta, a} v(a, \theta)\pi(a \mid \theta)\mu_0(\theta)$$

$$\text{s.t. } \pi \text{ is a Bayes Correlated Equilibrium}$$

- For any $\mu_0 \geq \frac{1}{3}$, this outputs 'no information' (like $\pi^*((I, I) \mid \theta) = 1$ for all θ) as solution

- The above program implicitly adopts the 'max' selection rule

[6] Dirk Bergemann and Stephen Morris. "Bayes correlated equilibrium and the comparison of information structures in games". In: *Theoretical Economics* 11.2 (2016), pp. 487–522.

Signal Space Restriction

- Under the 'min' rule, it is known that relevant mechanisms can have infinite message spaces (Maskin (1999)[7])
- In this paper, we have restricted attention to finite message spaces, as this guarantees existence of BNE for all belief-hierarchy distributions
- Here we restrict attention to at most $m = 4$ messages in the maximization within
- This will be without loss for some priors and constraining for others, which is true for any finite m

[7] Eric Maskin. "Nash equilibrium and welfare optimality". In: *The Review of Economic Studies* 66.1 (1999), pp. 23–38.

Worst-Equilibrium Characterization

- First we can see that playing I is the unique rationalizable strategy when $\mu_0 > \frac{2}{3}$, irrespective of his belief about the other player's action, since

$$\begin{aligned} \mathrm{E}_\theta[u_1((I,N),\theta)] &= \mu_0 \times 1 + (1-\mu_0) \times (-2) \\ &= 3\mu_0 - 2 \\ &\geq 0 = \mathrm{E}_\theta[u_1((N,N),\theta)] \end{aligned}$$

- Even if $\mu \leq \frac{2}{3}$, I is also the unique rationalizable strategy when she believes $\theta = 2$ with high probability

(u_1, u_2)	I	N
I	θ, θ	$\theta - 1, 0$
N	$0, \theta - 1$	$0, 0$

TABLE 2: Investment Game

Worst-Equilibrium Characterization

- Let ρ_i^k be the set of hierarchies defined as:

$$\rho_i^1 \equiv \left\{ t_i \mid \beta_i^*(\{\theta = 2\} \times T_j \mid t_i) > \frac{2}{3} \right\}$$
$$\rho_i^k \equiv \left\{ t_i \mid \beta_i^*(\{\theta = 2\} \times T_j \mid t_i) + \frac{1}{3}\beta_i^*(\Theta \times \rho_j^{k-1} \mid t_i) > \frac{2}{3} \right\}$$

- Let $\rho_i = \cup_{k \geq 1} \rho_i^k$. The unique optimal action for a player with belief in ρ_i is I
- Hence in all BNEs, i's equilibrium strategy must choose I when her hierarchy is in ρ_i

Worst-Equilibrium Characterization

Proof of the fact in the previous slide.

- If $t_i \in \rho_i^1$, then

$$\begin{aligned}\mathrm{E}_{\theta,a_j}[u_1((I,a_j),\theta)] &\geq E_\theta[u_1((I,N),\theta)] \\ &= \beta_i^*(\{\theta=2\} \mid t_i) \times 1 + (1-\beta_i^*(\{\theta=2\} \mid t_i)) \times (-2) \\ &= 3\beta_i^*(\{\theta=2\} \mid t_i) - 2 \\ &\geq \mathrm{E}_\theta[u_1((N,N),\theta)] = 0,\end{aligned}$$

  which implies I is the unique rationalizable if $t_i^1 \in \rho_i^1$

- If $t_i \in \rho_i^k$ for $k \geq 2$, then

$$\begin{aligned}&E_{\theta,a_j}[u_1((I,a_j),\theta)] = 3\beta_i^*(\{\theta=2\} \mid t_i) - 2 + \beta_i^*(\Theta \times \rho_j^{k-1} \mid t_i) \times 1 \\ &\geq E_{\theta,a_j}[u_1((N,a_j),\theta)] \geq 0\end{aligned}$$

  which implies I is the unique rationalizable if $t_i \in \rho_i^k$

Worst-Equilibrium Characterization

- Given a belief distribution $\tau \in \Delta(T_1 \times T_2)$, the worst equilibrium for the designer is such that all players play I only when their beliefs belong to ρ_i and play N otherwise:

$$\sigma_i^{\text{MIN}}(I \mid t_i) = \begin{cases} 1 & \text{if } t_i \in \rho_i \\ 0 & \text{otherwise} \end{cases}$$

- This is the worst equilibrium since, when both I and N are rationalizable, N is always played
- Given the structure of the worst equilibrium, it is never optimal to induce a belief hierarchy at which a player has multiple rationalizable actions

Worst-Equilibrium Characterization

Proposition (Proposition 4)

If τ^* is an optimal belief-hierarchy distribution and player i plays N with positive probability at $t_i \in \text{supp } \tau_i^*$ in any BNE, then $\beta_i^*(\theta = 2 \mid t_i) = 0$

Corollary (Corollary 2)

All optimal belief-hierarchy distributions generate a unique BNE

Worst-Equilibrium Characterization

- Equilibrium multiplicity occurs when hierarchies have moderate beliefs
- Instead of inducing moderate beliefs that do not contribute to investment under adversarial selection, the designer should instead create very pessimistic hierarchies for which N is uniquely rationalizable, allowing her to put more weight on optimistic hierarchies, for which I is uniquely rationalizable
- The designer optimally fosters investment in the worst equilibrium by redistributing the mass from hierarchies with multiple rationalizable actions to hierarchies with a uniquely rationalizable action while preserving Bayes plausibility

Worst-Equilibrium Characterization

Proposition (Proposition 5)

Suppose τ^* is an optimal belief-hierarchy distribution. Let $m = \sum_{i=1,2} |\text{supp } \tau_i^*|$. Then, for all i and $t_i \in \text{supp } \tau_i^*$, $\sigma_i^{\text{MIN}}(I \mid t_i) = 1$ if and only if $t_i \in \cup_{k=1}^{m} \rho_i^k$

Solution to Maximization Problem

- Let $\mu \equiv \mu_0(\theta = 2)$. Given consistent and Bayse plausible τ, consider an information design (S, π) where $S_i = \{t_i', t_i''\}$ and π satisfies

$$\tau(t) = \sum_{\theta} \pi(t \mid \theta)\mu_0(\theta)$$

(then $\tau(t) = (\underset{T}{\text{marg}}\ p)(t)$: consistency satisfied)

- We would like to find the range of μ in which there exists π that induce players to play (I, I) always

- Let $\pi_t^\theta \equiv \pi(t \mid \theta)$. Again by consistency, each player's first-order belief will be

$$\beta_i^*(\theta, t_j \mid t_i) = \frac{p(\theta, t)}{p(t_i)} = \frac{\pi_t^\theta \mu_0(\theta)}{\sum_\theta \sum_{t_j} \pi_t^\theta \mu_0(\theta)}$$

Solution to Maximization Problem

- e.g.

$$\beta_1^*(\theta = 2, t_2' \mid t_1') = \frac{\pi^2_{t_1',t_2'}\mu}{(\pi^2_{t_1',t_2'} + \pi^2_{t_1',t_2''})\mu + (\pi^{-1}_{t_1',t_2'} + \pi^{-1}_{t_1',t_2''})(1-\mu)}$$

- Then ρ_1^1 is

$$\begin{aligned}\rho_1^1 &= \left\{ t_1 \mid \beta_1^*(\theta = 2 \mid t_1) > \frac{2}{3} \right\} \\ &= \left\{ t_1 \mid \frac{(\pi^2_{t_1,t_2'} + \pi^2_{t_1,t_2''})\mu}{(\pi^2_{t_1,t_2'} + \pi^2_{t_1,t_2''})\mu + (\pi^{-1}_{t_1,t_2'} + \pi^{-1}_{t_1,t_2''})(1-\mu)} > \frac{2}{3} \right\}\end{aligned}$$

Solution to Maximization Problem

- Then ρ_1^2 is

$$\rho_1^2 = \left\{ t_1 \mid \beta_1^*(\theta = 2 \mid t_1) + \frac{1}{3}\beta_1^*(\Theta \times \rho_2^1 \mid t_1) > \frac{2}{3} \right\}$$

$$= \begin{cases} \rho_1^1 & \text{if } \rho_2^1 = \emptyset \\ X & \text{if } \rho_2^1 = \{t_2'\} \\ \left\{ t_1 \mid \frac{(\pi^2_{t_1,t_2'}+\pi^2_{t_1,t_2''})\mu}{(\pi^2_{t_1,t_2'}+\pi^2_{t_1,t_2''})\mu+(\pi^{-1}_{t_1,t_2'}+\pi^{-1}_{t_1,t_2''})(1-\mu)} > \frac{1}{3} \right\} & \text{if } \rho_2^1 = \{t_2', t_2''\} \end{cases}$$

where

$$X = \left\{ t_1 \mid \frac{(\pi^2_{t_1,t_2'} + \pi^2_{t_1,t_2''})\mu}{(\pi^2_{t_1,t_2'} + \pi^2_{t_1,t_2''})\mu + (\pi^{-1}_{t_1,t_2'} + \pi^{-1}_{t_1,t_2''})(1-\mu)} + \frac{1}{3}\frac{\pi^2_{t_1,t_2'}\mu + \pi^{-1}_{t_1,t_2'}(1-\mu)}{(\pi^2_{t_1,t_2'} + \pi^2_{t_1,t_2''})\mu + (\pi^{-1}_{t_1,t_2'} + \pi^{-1}_{t_1,t_2''})(1-\mu)} > \frac{2}{3} \right\}$$

Solution to Maximization Problem

- Let A be the probability that (t_1', t_2') occurs:

$$A \equiv \pi^2_{t_1,t_2'}\mu + \pi^{-1}_{t_1,t_2'}(1-\mu)$$

- Then the designer's belief $\mu_A \equiv v(\theta = 2 \mid t_1', t_2')$ is

$$\mu_A \equiv \frac{\pi^2_{t_1,t_2'}\mu}{\pi^2_{t_1,t_2'}\mu + \pi^{-1}_{t_1,t_2'}(1-\mu)}$$

- Similary we define B, C, D be the probability that $(t_1', t_2''), (t_1'', t_2'), (t_1'', t_2'')$

Solution to Maximization Problem

- Then ρ_1^1, ρ_1^2 can be written as

$$\rho_1^1 = \left\{ t_1 \mid \frac{A\mu_A + B\mu_B}{A+B} > \frac{2}{3} \right\}$$

$$\rho_1^2 = \begin{cases} \rho_1^1 & \text{if } \rho_2^1 = \emptyset \\ \{t_1 \mid \frac{A\mu_A+B\mu_B}{A+B} + \frac{1}{3}\frac{A}{A+B} > \frac{2}{3}\} & \text{if } \rho_2^1 = \{t_2'\} \\ \{t_1 \mid \frac{A\mu_A+B\mu_B}{A+B} > \frac{1}{3}\} & \text{if } \rho_2^1 = \{t_2', t_2''\} \end{cases}$$

e_μ^*	t_2'	t_2''
t_1'	A	B
t_1''	C	D

Solution to Maximization Problem

- If $\mu > \frac{2}{3}$, then $\rho_1^1 = \{t_1', t_1''\}, \rho_2^1 = \{t_2', t_2''\}$, no information signal can always induce players to play (I, I)
- How about when $\mu \leq \frac{2}{3}$?
  - If $\rho_1^1 = \rho_2^1 = \emptyset$, then all $\rho_i^k = \emptyset$, which contradicts Proposition 5
  - By Proposition 5, $\rho_1^4 = T_1, \rho_2^4 = T_2$ is necessary

Solution to Maximization Problem

$$
\begin{cases}
\dfrac{A}{A+B}\mu_A + \dfrac{B}{A+B}\mu_B > \dfrac{2}{3} & (17)\\
\left(\dfrac{A}{A+C}\mu_A + \dfrac{C}{A+C}\mu_C\right) + \dfrac{1}{3}\dfrac{A}{A+C} > \dfrac{2}{3} & (18)\\
\left(\dfrac{C}{C+D}\mu_C + \dfrac{D}{C+D}\mu_D\right) + \dfrac{1}{3}\dfrac{C}{C+D} > \dfrac{2}{3} & (19)\\
\left(\dfrac{B}{B+D}\mu_B + \dfrac{D}{B+D}\mu_D\right) + \dfrac{1}{3} > \dfrac{2}{3} & (20)\\
A\mu_A + B\mu_B + C\mu_C + D\mu_D = \mu & (21)
\end{cases}
$$

- The above equations can be solved when $\mu > \frac{1}{2}$

Solution to Maximization Problem

$e^*_{1/2}$	$t'_2 : \mu'_2 = 3/5$	$t''_2 : \mu''_2 = 1/3$
$t'_1 : \mu'_1 = 2/3$	$1/8$	$1/8$
$t''_1 : \mu''_1 = 4/9$	$1/2$	$1/4$

- $\mu \downarrow \frac{1}{2}$ case solution

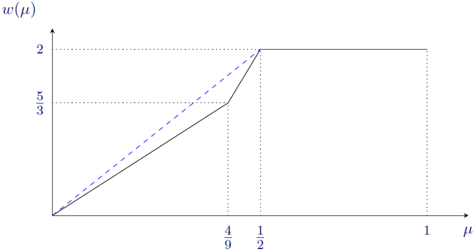
Solution to Maximization Problem

- When $\frac{4}{9} < \mu \leq \frac{1}{2}$, playing (I, I) only when $(\{t_1', t_1''\}, t_2')$ can be induced
- When $\mu < \frac{4}{9}$, the designer can only ensure I will be played at two hierarchies

$e^*_{1/2}$	$t_2' : \mu_2' = 3/5$	$t_2'' : \mu_2'' = 1/3$
$t_1' : \mu_1' = 2/3$	$1/8$	$1/8$
$t_1'' : \mu_1'' = 4/9$	$1/2$	$1/4$

Solution to Maximization Problem

- The solid line shows the designer's maximum-within value for $\{t_1', t_1''\} \times \{t_2', t_2''\}$
- For $\mu \leq \frac{1}{2}$, the designer can increase the profit by the between maximization using the public randomization device
- The designer uses a public signal and sends both players to $e^*_{1/2}$ (both always play I) with probability 2μ, and to e^*_0 (both always play N) with probability $1 - 2\mu$. The designer's value is then $w^*(\mu) = 4\mu$



Solution to Maximization Problem

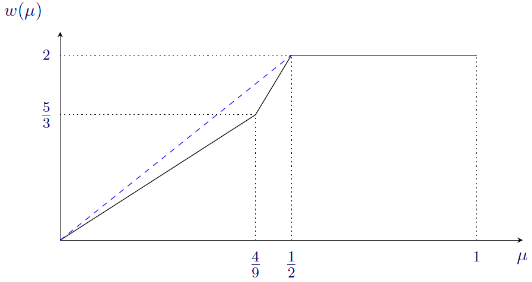


FIGURE 1: Value of maximization within (solid) and between (dashed)

6. Conclusion

Conclusion

- Extending Kamenica and Gentzkow (2011) model into a multi-agent Bayesian game setting
- Formulating the belief-based approach to the problem, and decompose it into maximization within and between basic communication schemes C^M
- Accommodating various equilibrium selection rules and solution concepts, which can be used to analyze diverse topics, such as robustness, bounded rationality, collusion, or communication
- Future extensions:
  - Generalize this robust information design to a class of games with strategic complementarities, for which the results from the previous investment game provide the fundamental logic
  - Examining the implications of heterogenous prior distributions among the agents

7. Proofs

Proof of Proposition 2

Proof of Proposition 2.

τ: consistent and Bayes plausible $\implies \exists v$ part:

Fix $\tau \in C$: Bayes plausible. By consistency, $\exists$common prior $p \in \Delta(\Theta \times T)$ s.t.

$$\forall \theta, t, p(\theta, t) = \beta_i^*(\theta, t_{-i} \mid t_i) \times (\operatorname*{marg}_{T_i} p)(t_i)$$

and $\operatorname*{marg}_{T} p = \tau$. Define $v(t) \equiv p(\cdot \mid t) = \frac{p(\cdot, t)}{(\operatorname*{marg}_{T} p)(t)}$. Then we have for all θ,

$$\begin{aligned}\sum_t \tau(t) v(t) &= \sum_t (\operatorname*{marg}_{T} p)(t) p(\cdot \mid t) = \sum_t p(\cdot, t) \\ &= \beta_i^*(\cdot, t_{-i} \mid t_i)(\operatorname*{marg}_{T_i} p)(t_i) \\ &= \mu_0\end{aligned}$$

by the Bayes plausiblity.

Proof of Proposition 2

Proof of Proposition 2 (Cont.)

Moreover for all t_{-i},

$$\begin{aligned}\sum_{t_{-i}} \tau(t_{-i} \mid t_i) v(t) &= \sum_{t_{-i}} \frac{p(\cdot \mid t)(\operatorname*{marg}_{T} p)(t)}{(\operatorname*{marg}_{T_i} p)(t_i)} \\ &= \sum_{t_{-i}} p(\cdot, t_{-i} \mid t_i) \\ &= p(\cdot \mid t_i) = \operatorname*{marg}_{\Theta} \beta_i^*(\cdot \mid t_i)\end{aligned}$$

Proof of Proposition 2

Proof of Proposition 2 (Cont.)

$\xi, v \implies \exists\tau$: consistent and Bayes plausible part:

Pick $\xi \in \Delta^f(\Delta\Theta)^n$ and $v : \text{supp } \xi \to \Delta\Theta$ that satisfies (8), (9).

Define $g(\theta, \boldsymbol{\mu}) \equiv \xi(\boldsymbol{\mu})v(\boldsymbol{\mu})(\theta)$ for each $\theta, \boldsymbol{\mu}$.

We can think of

$$g(\boldsymbol{\mu} \mid \theta) \equiv \frac{\xi(\boldsymbol{\mu})v(\boldsymbol{\mu})(\theta)}{\mu_0(\theta)}$$

as the information map of the information structure $(S, \pi) = (\text{supp } \xi, g)$.

This is actually an information structure since

- supp ξ is finite
- $\sum_{\boldsymbol{\mu}} \xi(\boldsymbol{\mu})v(\boldsymbol{\mu})(\theta) = \mu_0(\theta)$ by (8), which implies $\sum_{\boldsymbol{\mu}} g(\boldsymbol{\mu} \mid \theta) = 1$ for all θ

Proof of Proposition 2

Proof of Proposition 2 (Cont.)

Let τ be a belief distribution induced by $(\operatorname{supp}\xi, g)$. Then $\exists\phi_i : \operatorname{supp}\xi_i \to \operatorname{supp}\tau_i$, bijection, such that $\tau(t) = \xi(\phi^{-1}(t))$ for all t. Define $p(\theta, t) \equiv g(\theta, \phi^{-1}(t))$. For any $t_i \in \operatorname{supp}\tau_i$, define $\mu_i = \phi^{-1}(t_i)$. Then we have for all θ,

$$
\begin{aligned}
(\operatorname*{marg}_{\Theta} \beta_i^*(\cdot \mid t_i))(\theta) &= \sum_{t_{-i}} \frac{p(\theta, t)}{(\operatorname*{marg}_{T_i} p)(t_i)} \frac{\sum_{t_{-i}} p(\theta, t)}{\sum_\theta \sum_{t_{-i}} p(\theta, t)} \\
&= \frac{\sum_{\boldsymbol{\mu}_{-i}} g(\theta, \boldsymbol{\mu})}{\sum_\theta \sum_{\boldsymbol{\mu}_{-i}} g(\theta, \boldsymbol{\mu})} \\
&= \frac{\sum_{\boldsymbol{\mu}_{-i}} \xi(\boldsymbol{\mu}) v(\boldsymbol{\mu})(\theta)}{\sum_\theta \sum_{\boldsymbol{\mu}_{-i}} \xi(\boldsymbol{\mu}) v(\boldsymbol{\mu})(\theta)}
\end{aligned}
$$

Proof of Proposition 2

Proof of Proposition 2 (Cont.)

Here

$$\sum_{\boldsymbol{\mu}_{-i}} \xi(\boldsymbol{\mu})v(\boldsymbol{\mu})(\theta) = \xi(\mu_i)\sum_{\boldsymbol{\mu}_{-i}} \xi(\boldsymbol{\mu}_{-i} \mid \mu_i)v(\boldsymbol{\mu})(\theta) = \xi(\mu_i)\mu_i(\theta)$$

by (9). Therefore

$$(\operatorname*{marg}_{\Theta} \beta_i^*(\cdot \mid t_i))(\theta) = \frac{\xi(\mu_i)\mu_i(\theta)}{\sum_\theta \xi(\mu_i)\mu_i(\theta)} = \mu_i(\theta).$$

Proof of Proposition 2

Proof of Proposition 2 (Cont.)

Note that

$$\sum_{t_i} (\operatorname{marg}_{\Theta} \beta_i^*(\cdot \mid t_i))(\theta)\tau_i(t_i) = \sum_{\mu_i} \mu_i(\theta)\xi(\mu_i)(\theta)$$
$$= \sum_{\boldsymbol{\mu}} \xi(\boldsymbol{\mu})v(\boldsymbol{\mu})(\theta) = \mu_0(\theta)$$

for all i, θ by (8), so τ is Bayes plausibility. Consistency comes from

$$\sum_{\theta} p(\theta, t) = \sum_{\theta} g(\theta, \phi^{-1}(t))$$
$$= \xi(\phi^{-1}(t)) \sum_{\theta} v(\phi^{-1}(t))(\theta) = \xi(\phi^{-1}(t)) = \tau(t)$$

for all t. [Back] $\square$

Proof of Theorem 1

Proof of Theorem 1.

Fix a prior $\mu_0 \in \Delta(\Theta)$ and take any information structure (S, π). By Proposition 1, there exists $\tau \in C$ such that

$$\begin{aligned}
\operatorname*{marg}_{\Theta} p_\tau &= \sum_{t_i} (\operatorname*{marg}_{\Theta} \beta_i^*(\cdot \mid t_i)) \times (\operatorname*{marg}_{T_i} p)(t_i) \\
&= \sum_{t_i} (\operatorname*{marg}_{\Theta} \beta_i^*(\cdot \mid t_i)) \times \tau_i(t_i) \\
&= \mu_0.
\end{aligned}$$

By definition of σ^B, we have $V(S, \pi) \leq w(\tau)$ and so

$$\sup_{(S,\pi)} V(S, \pi) \leq \sup_{\tau \in C \text{ s.t. } \operatorname*{marg}_{\Theta} p_\tau = \mu_0} w(\tau).$$

Proof of Theorem 1

Proof of Theorem 1 (Cont.)

Conversely, again by Proposition 1, for $\tau \in C$ such that $\underset{\Theta}{\text{marg}}\ \pi_\tau$, there exists (S, π) that induces τ such that $V(S, \pi) = w(\tau)$. Therefore

$$\sup_{(S,\pi)} V(S, \pi) \geq \sup_{\tau \in C \text{ s.t. } \underset{\Theta}{\text{marg}}\ p_\tau = \mu_0} w(\tau).$$

So we conclude

$$\sup_{(S,\pi)} V(S, \pi) = \sup_{\tau \in C \text{ s.t. } \underset{\Theta}{\text{marg}}\ p_\tau = \mu_0} w(\tau).$$

Proof of Theorem 1

Proof of Theorem 1 (Cont.)

By Proposition 6, there exists a unique $\lambda \in \Delta^f(C^M)$ such that $\tau = \sum_{e \in \text{supp } \lambda} \lambda(e) e$. Since p and $\text{marg}_{\cdot}$ is linear,

$$\text{marg}_{\Theta}\, p_\tau = \text{marg}_{\Theta}\, p_{\sum_{e \in \text{supp } \lambda} \lambda(e) e} = \sum_{e \in \text{supp } \lambda} \lambda(e) p_e.$$

By Lemma 3, we have

$$\sup_{(S,\pi)} V(S,\pi) = \sup_{\lambda \in \Delta^f(C^M)} \sum_{e \in C^M} w(e)\lambda(e)$$
$$\text{s.t.} \sum_{e \in C^M} \text{marg}_{\Theta}\, p_e \lambda(e) = \mu_0$$

[Back] □