Equilibria for games with asymmetric information: from guesswork to systematic evaluation

Achilleas Anastasopoulos
anastas@umich.edu

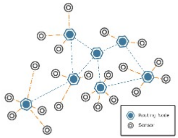
EECS Department
University of Michigan

February 11, 2016

- Joint work with Deepanshu Vasal (PhD student graduating May 2016) and Prof. Vijay Subramanian

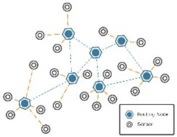
Decentralized decision making in dynamic systems

- Communication networks
- Sensor networks
- Social networks
- Queuing systems
- Energy markets
- Wireless resource sharing
- Repeated online advertisement auctions
- Competing sellers/buyers




Salient features

- Multiple agents (cooperative or strategic)
- Objective: Maximize expected (social or self) reward
- Underlying system state (not perfectly observed)
- Agents make observations (asymmetric information) and take actions partially affecting future state



Classification of problems

	Teams	Games
Symmetric Information	Markov decision processes (MDP) or partially observed MDP (POMDP)	subgame-perfect equilibrium (SPE) Markov-perfect equilibrium (MPE)

Classification of problems

	Teams	Games
Symmetric Information	Markov decision processes (MDP) or partially observed MDP (POMDP)	subgame-perfect equilibrium (SPE) Markov-perfect equilibrium (MPE)
Asymmetric Information	Common information approach [1]	

[1] 2015 IEEE Control Theory Axelby paper award [Nayyar, Mahajan, Teneketzis, 2013]

Classification of problems

	Teams	Games
Symmetric Information	Markov decision processes (MDP) or partially observed MDP (POMDP)	subgame-perfect equilibrium (SPE) Markov-perfect equilibrium (MPE)
Asymmetric Information	Common information approach [1]	Perfect Bayesian (PBE) Sequential eq. (SE) and refinements No methodology! ?

[1] 2015 IEEE Control Theory Axelby paper award [Nayyar, Mahajan, Teneketzis, 2013]

Model

- Discrete-time dynamical system with N strategic agents over finite horizon T
- Player i privately observes her (static[2]) type $X^i \in \mathcal{X}^i$ where

$$P(X) = \prod_{i=1}^{N} Q^i(X^i), \qquad X = (X^1, X^2, \ldots X^N) \in \mathcal{X}$$

- Player i takes action $A_t^i \in \mathcal{A}^i$ which is publicly observed
- Player i's observations: Private: X^i,
  Common: $A_{1:t-1} = (A_1, A_2, \ldots, A_{t-1}) = (A_k^j)_{k \leq t-1}^{j \in \mathcal{N}}$
- Action (randomized) $A_t^i \sim \sigma_t^i(\cdot | X^i, A_{1:t-1})$
- Instantaneous reward $R^i(X, A_t)$
- Player i's objective

$$\max_{\sigma^i} \ \mathbb{E}^{\sigma} \left\{ \sum_{t=1}^{T} R^i(X, A_t) \right\}$$

[2] Generalization to dynamic types straightforward.

Concrete example: A public goods game[3]

- Two players take action to either contribute ($A_t^i = 1$) or not contribute ($A_t^i = 0$) to the production of a public good
- Player i's type (private information) is her cost of contributing: $X^i \in \{L, H\}$, where X^i's are i.i.d. with $P(X^i = H) = q$. (Assume $0 < L < 1 < H < 2$)
- If either player contributes, the public good is produced and the utility enjoyed is 1 for both users (free riding)
- Per-period rewards $(R^1(X^1, A_t), R^2(X^2, A_t))$ are

	contribute($A_t^2 = 1$)	don't contribute($A_t^2 = 0$)
contribute($A_t^1 = 1$)	$(1 - X^1, 1 - X^2)$	$(1 - X^1, 1)$
don't contribute($A_t^1 = 0$)	$(1, 1 - X^2)$	$(0,0)$

- Each player's action $A_t^i \sim \sigma_t^i(\cdot | X^i, A_{1:t-1})$.

[3] Adapted from [Fudenberg and Tirole, 1991, Example 8.3]

Classification of problems

	Teams	Games
Symmetric Information	Markov decision processes (MDP) or partially observed MDP (POMDP)	
Asymmetric Information		

Team with perfect observation of X

- X is observed by everyone
- Single team objective $R(X, A_t) = \sum_{i \in \mathcal{N}} R^i(X, A_t)$

	contribute($A_t^2 = 1$)	don't contribute($A_t^2 = 0$)
contribute($A_t^1 = 1$)	$2 - X^1 - X^2$	$2 - X^1$
don't contribute($A_t^1 = 0$)	$2 - X^2$	0

Team with perfect observation of X

- X is observed by everyone
- Single team objective $R(X, A_t) = \sum_{i \in \mathcal{N}} R^i(X, A_t)$

	contribute($A_t^2 = 1$)	don't contribute($A_t^2 = 0$)
contribute($A_t^1 = 1$)	$2 - X^1 - X^2$	$2 - X^1$
don't contribute($A_t^1 = 0$)	$2 - X^2$	0

- Optimal decisions are myopic (just look at instantaneous reward) and functions of the current system "state" $X = (X^1, X^2)$

$$(A_t^{*1}, A_t^{*2}) = \begin{cases} (1,0) & \text{if } (X^1, X^2) = (L, H) \\ (0,1) & \text{if } (X^1, X^2) = (H, L) \\ (1,0) \text{ or } (0,1) & \text{if } (X^1, X^2) = (L, L) \\ (1,0) \text{ or } (0,1) & \text{if } (X^1, X^2) = (H, H) \end{cases}$$

Team with perfect observation of X

- X is observed by everyone
- Single team objective $R(X, A_t) = \sum_{i \in \mathcal{N}} R^i(X, A_t)$

	contribute($A_t^2 = 1$)	don't contribute($A_t^2 = 0$)
contribute($A_t^1 = 1$)	$2 - X^1 - X^2$	$2 - X^1$
don't contribute($A_t^1 = 0$)	$2 - X^2$	0

- Optimal decisions are myopic (just look at instantaneous reward) and functions of the current system "state" $X = (X^1, X^2)$

$$(A_t^{*1}, A_t^{*2}) = \begin{cases} (1,0) & \text{if } (X^1, X^2) = (L, H) \\ (0,1) & \text{if } (X^1, X^2) = (H, L) \\ (1,0) \text{ or } (0,1) & \text{if } (X^1, X^2) = (L, L) \\ (1,0) \text{ or } (0,1) & \text{if } (X^1, X^2) = (H, H) \end{cases}$$

- What about time-varying types, e.g., $Q(X_{t+1}|X_t)$ or $Q(X_{t+1}|X_t, A_t)$? MDP

Team with no observation of X

- X is not observed at all (symmetric information)
- Single team objective $R(X, A_t) = \sum_{i \in \mathcal{N}} R^i(X, A_t)$
- Previous actions are not informative of X
- Same as before with average rewards (w.r.t. prior belief $P(X^i = H) = q$)

	contribute($A_t^2 = 1$)	don't contribute($A_t^2 = 0$)
contribute($A_t^1 = 1$)	$2 -$ (mean total cost)	$2 - (qH + \bar{q}L)$
don't contribute($A_t^1 = 0$)	$2 - (qH + \bar{q}L)$	0

Team with no observation of X

- X is not observed at all (symmetric information)
- Single team objective $R(X, A_t) = \sum_{i\in\mathcal{N}} R^i(X, A_t)$
- Previous actions are not informative of X
- Same as before with average rewards (w.r.t. prior belief $P(X^i = H) = q$)

	contribute$(A_t^2 = 1)$	don't contribute$(A_t^2 = 0)$
contribute$(A_t^1 = 1)$	$2-$(mean total cost)	$2-(qH+\bar{q}L)$
don't contribute$(A_t^1 = 0)$	$2-(qH+\bar{q}L)$	0

- Optimal decisions are constant

$$(A_t^{*1}, A_t^{*2}) = (1,0) \text{ or } (0,1)$$

Team with no observation of X

- X is not observed at all (symmetric information)
- Single team objective $R(X, A_t) = \sum_{i \in \mathcal{N}} R^i(X, A_t)$
- Previous actions are not informative of X
- Same as before with average rewards (w.r.t. prior belief $P(X^i = H) = q$)

	contribute($A_t^2 = 1$)	don't contribute($A_t^2 = 0$)
contribute($A_t^1 = 1$)	$2 -$ (mean total cost)	$2 - (qH + \bar{q}L)$
don't contribute($A_t^1 = 0$)	$2 - (qH + \bar{q}L)$	0

- Optimal decisions are constant

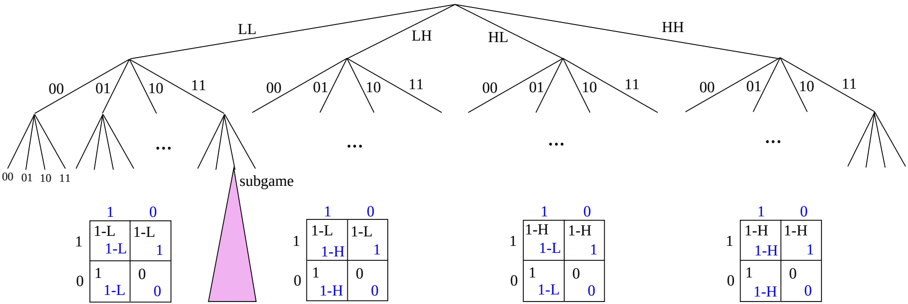
$$(A_t^{*1}, A_t^{*2}) = (1, 0) \text{ or } (0, 1)$$

- What about time-varying types, e.g., $Q(X_{t+1}|X_t)$ or $Q(X_{t+1}|X_t, A_t)$? POMDP

Classification of problems

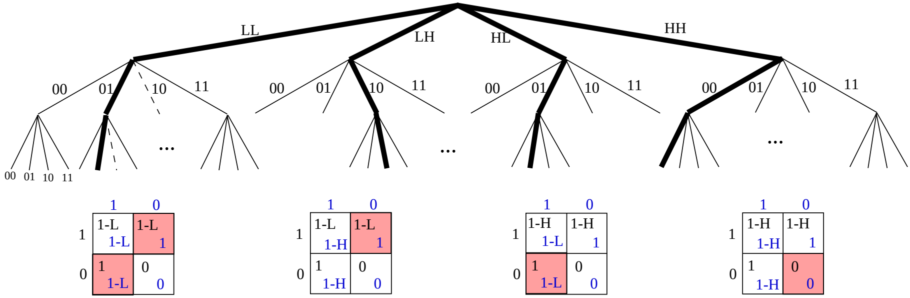
	Teams	Games
Symmetric Information	✓	subgame-perfect equilibrium (SPE) Markov-perfect equilibrium (MPE)
Asymmetric Information		

Game with perfect observation of X



- Players know exactly what branch they are on at each stage of the game
- Sub-game perfect equilibrium (SPE): given any history (path) players "see" a continuation game (sub-game) and do not want to deviate
- Algortihm: Backward induction

Game with perfect observation of X



- Here, at each stage of the game, the continuation game is the same
- SPE strategy profile does not depend on the entire history of actions but only on state X.
- Even with time-varying states, similar algorithm (backward induction) can be used

Classification of problems

	Teams	Games
Symmetric Information	✓	✓
Asymmetric Information	Common information approach	

Decentralized team problem

- Player i's observations: Private: X^i,
  Common: $A_{1:t-1}$
- Action (randomized) $A_t^i \sim \sigma_t^i(\cdot | X^i, A_{1:t-1})$
- Design objective for entire team

$$\max_{\sigma} \; \mathbb{E}^{\sigma}\left\{ \sum_{t=1}^{T} \underbrace{R(X, A_t)}_{\text{e.g., } \sum_{i \in \mathcal{N}} R^i(X, A_t)} \right\}$$

Decentralized team problem

- Player i's observations: Private: X^i,
  Common: $A_{1:t-1}$
- Action (randomized) $A_t^i \sim \sigma_t^i(\cdot | X^i, A_{1:t-1})$
- Design objective for entire team

$$\max_{\sigma} \ \mathbb{E}^{\sigma}\left\{ \sum_{t=1}^{T} \underbrace{R(X, A_t)}_{\text{e.g., } \sum_{i \in \mathcal{N}} R^i(X, A_t)} \right\}$$

- Problems to be addressed[4]
  1. Presence of common $A_{1:t-1}$ and private X^i information for agent i
  2. Decentralized, non-classical information structure (this is **not** a MDP/POMDP-like problem!)
  3. Domain of policies $A_t^i \sim \sigma_t^i(\cdot | X^i, A_{1:t-1})$ increases with time.

[4] All these have been addressed in [Nayyar, Mahajan, Teneketzis, 2013]

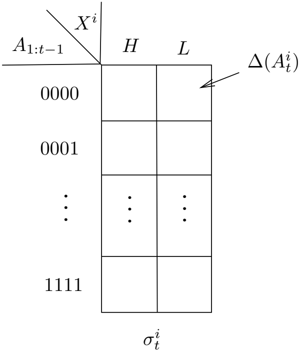
A simple but powerful idea

A policy $\sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be interpreted in two equivalent ways:

A simple but powerful idea

A policy $\sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be interpreted in two equivalent ways:

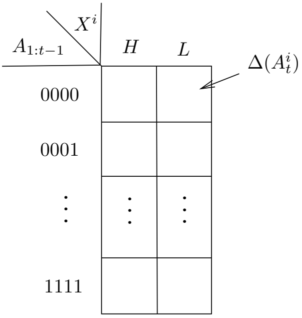
1) A function of $A_{1:t-1}$ and X^i
to $\Delta(\mathcal{A}^i)$



A simple but powerful idea

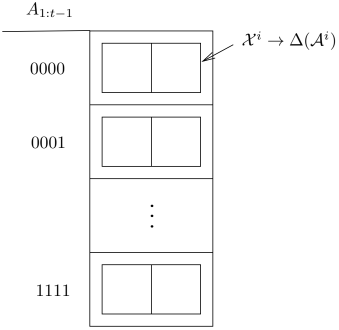
A policy $\sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be interpreted in two equivalent ways:

1) A function of $A_{1:t-1}$ and X^i to $\Delta(\mathcal{A}^i)$



σ_t^i

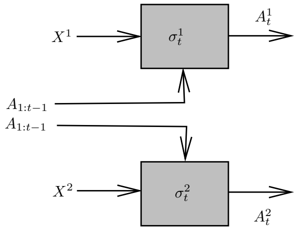
2) A function of $A_{1:t-1}$ to **mappings** from $\mathcal{X}^i$ to $\Delta(\mathcal{A}^i)$



ψ_t^i

A simple but powerful idea

In the first interpretation, the policies to be designed $(\sigma^i)_{i\in\mathcal{N}}$ have inherent **asymmetric** information structure



A simple but powerful idea

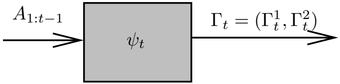
In the second interpretation, each agent's action $A_t^i \sim \sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be thought of as a **two-stage** process

A simple but powerful idea

In the second interpretation, each agent's action $A_t^i \sim \sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be thought of as a **two-stage** process

1. Based on common info $A_{1:t-1}$ select **"prescription"** functions $\Gamma_t^i : \mathcal{X}^i \to \Delta(\mathcal{A}^i)$ through the mapping ψ^i

$$\Gamma_t^i = \psi_t^i[A_{1:t-1}]$$



A simple but powerful idea

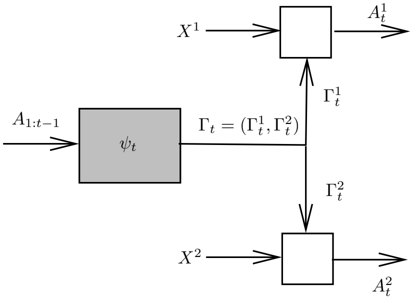
In the second interpretation, each agent's action $A_t^i \sim \sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be thought of as a **two-stage** process

1. Based on common info $A_{1:t-1}$ select **"prescription"** functions $\Gamma_t^i : \mathcal{X}^i \to \Delta(\mathcal{A}^i)$ through the mapping ψ^i

$$\Gamma_t^i = \psi_t^i[A_{1:t-1}]$$

2. The actions A_t^i are determined by "evaluating" Γ_t^i at the private information X^i, i.e.,

$$A_t^i \sim \Gamma_t^i(\cdot|X^i)$$



A simple but powerful idea

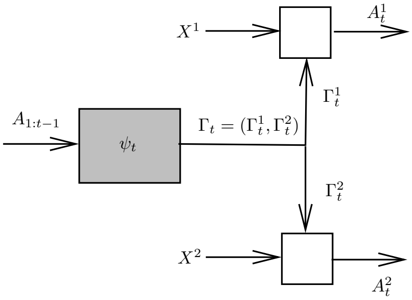
In the second interpretation, each agent's action $A^i_t \sim \sigma^i_t(\cdot|X^i, A_{1:t-1})$ can be thought of as a **two-stage** process

1. Based on common info $A_{1:t-1}$ select **"prescription"** functions $\Gamma^i_t : \mathcal{X}^i \to \Delta(\mathcal{A}^i)$ through the mapping ψ^i

$$\Gamma^i_t = \psi^i_t[A_{1:t-1}]$$

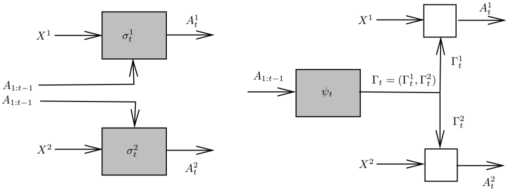
2. The actions A^i_t are determined by "evaluating" Γ^i_t at the private information X^i, i.e.,

$$A^i_t \sim \Gamma^i_t(\cdot|X^i)$$



Overall $\quad A^i_t \sim \Gamma^i_t(\cdot|X^i) = \psi^i_t[A_{1:t-1}](\cdot|X^i) = \sigma^i_t(\cdot|X^i, A_{1:t-1})$

Transformation to a centralized problem



- Generation of A_t^i is a "dumb" evaluation $A_t^i \sim \Gamma_t^i(\cdot|X^i)$ (nothing to be designed here)
- The control problem boils down to selecting prescription functions $\Gamma_t^i = \psi_t^i[A_{1:t-1}]$ through policy $\psi = (\psi_t^i)_{t\in\mathcal{T}}^{i\in\mathcal{N}}$
- The decentralized control problem has been transformed to a **centralized control** problem with a **fictitious common agent** who observes $A_{1:t-1}$ and takes actions Γ_t
- Last issue to address: increasing domain $\mathcal{A}^{t-1}$ of the pre-encoder mappings ψ_t.

Introduction of information state

- We would like to summarize $A_{1:t-1}$ in a quantity (state) with time invariant domain

Introduction of information state

- We would like to summarize $A_{1:t-1}$ in a quantity (state) with time invariant domain
- Consider the dynamical system with
  state: (X, A_{t-1})
  observation: A_{t-1}
  action: Γ_t
  reward: $\mathbb{E}\{R(X, A_t)|X, A_{1:t-1}, \Gamma_{1:t}\} = \sum_{a_t} \Gamma_t(a_t|X)R(X, a_t) := \tilde{R}(X, \Gamma_t)$

Introduction of information state

- We would like to summarize $A_{1:t-1}$ in a quantity (state) with time invariant domain
- Consider the dynamical system with
  state: (X, A_{t-1})
  observation: A_{t-1}
  action: Γ_t
  reward: $\mathbb{E}\{R(X, A_t)|X, A_{1:t-1}, \Gamma_{1:t}\} = \sum_{a_t} \Gamma_t(a_t|X)R(X, a_t) := \tilde{R}(X, \Gamma_t)$
- This is a POMDP! Define the posterior belief $\Pi_t \in \Delta(\mathcal{X})$

$$\Pi_t(x) := P(X = x|A_{1:t-1}, \Gamma_{1:t-1}) \qquad \text{for all } x \in \mathcal{X}$$

- Can show that Π_t can be updated using common information

$$\Pi_{t+1} = F(\Pi_t, \Gamma_t, A_t) \qquad \text{(Bayes law)}$$

(*) for this problem it also factors into its marginals

$$\Pi_t(x) = \prod_{i \in \mathcal{N}} \Pi^i_t(x^i) \qquad \text{with} \qquad \Pi^i_{t+1} = F(\Pi^i_t, \Gamma^i_t, A^i_t)$$

Characterization of optimal team policy

- From standard POMDP results, optimal policy is Markovian, i.e.,

$$\Gamma_t = (\Gamma_t^i)_{i\in\mathcal{N}} = \psi_t[A_{1:t-1}] = \theta_t[\Pi_t]$$

$$A_t^i \sim \Gamma_t^i(\cdot|X^i) = \theta_t^i[\Pi_t](\cdot|X^i) = m_t^i(\cdot|X^i, \Pi_t)$$

and can be obtained using backward dynamic programming (DP)

$$\boxed{\theta_t[\pi_t] = \gamma_t^* = \arg\max_{\gamma_t} \mathbb{E}\left\{R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t, A_t))|\pi_t, \gamma_t\right\}}$$

$$\boxed{V_t(\pi_t) = \max_{\gamma_t} \mathbb{E}\left\{R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t, A_t))|\pi_t, \gamma_t\right\}}$$

on the space of beliefs $\pi_t \in \Delta(\mathcal{X})$ over prescriptions $\gamma_t \in \underset{i\in\mathcal{N}}{\times}(\mathcal{X}^i \to \mathcal{A}^i)$

Characterization of optimal team policy

- From standard POMDP results, optimal policy is Markovian, i.e.,

$$\Gamma_t = (\Gamma_t^i)_{i \in \mathcal{N}} = \psi_t[A_{1:t-1}] = \theta_t[\Pi_t]$$

$$A_t^i \sim \Gamma_t^i(\cdot|X^i) = \theta_t^i[\Pi_t](\cdot|X^i) = m_t^i(\cdot|X^i, \Pi_t)$$

and can be obtained using backward dynamic programming (DP)

$$\boxed{\theta_t[\pi_t] = \gamma_t^* = \arg\max_{\gamma_t} \mathbb{E}\left\{R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t, A_t)) | \pi_t, \gamma_t\right\}}$$

$$\boxed{V_t(\pi_t) = \max_{\gamma_t} \mathbb{E}\left\{R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t, A_t)) | \pi_t, \gamma_t\right\}}$$

on the space of beliefs $\pi_t \in \Delta(\mathcal{X})$ over prescriptions $\gamma_t \in \bigtimes_{i \in \mathcal{N}} (\mathcal{X}^i \to \mathcal{A}^i)$

- In the public goods example:
$\pi_t \equiv (\pi_t^1(H), \pi_t^2(H)) \in [0,1]^2$ and
$\gamma_t \equiv (\gamma_t^1(0|H), \gamma_t^1(0|L), \gamma_t^2(0|H), \gamma_t^2(0|L)) \in [0,1]^4$

Summary of team problem

- Introduction of prescription functions was crucial
- We gained:
  - Decentralized non-classical information structure $\Rightarrow$ POMDP
    $\Rightarrow A_t^i \sim \theta_t^i[\Pi_t](\cdot|X^i)$ and θ can be obtained using DP

Summary of team problem

- Introduction of prescription functions was crucial

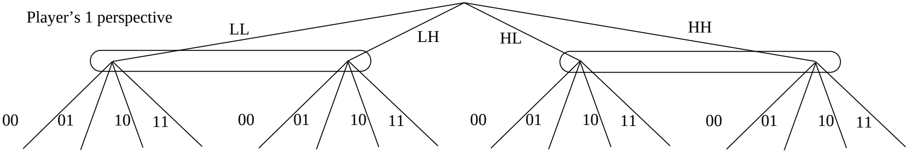
- We gained:
  - Decentralized non-classical information structure $\Rightarrow$ POMDP
  $\Rightarrow A_t^i \sim \theta_t^i[\Pi_t](\cdot|X^i)$ and θ can be obtained using DP

- We gave up:
  - *Fictitious common* agent does not observe X^i.
  - Can only maximize average reward-to-go $\mathbb{E}\{\sum_{t'=t}^{T} R(X, A_{t'})|A_{1:t-1}\}$ **before** seeing private information,
  - This is not a problem in teams since we are interested in maximizing the average reward

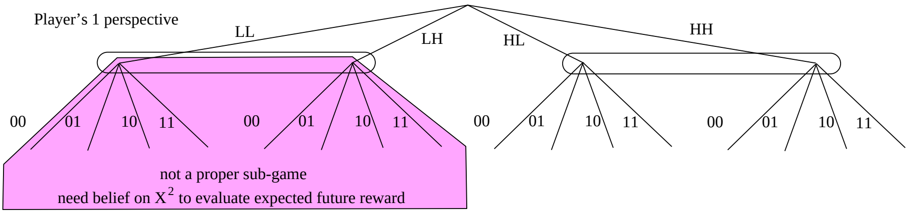
Classification of problems

	Teams	Games
Symmetric Information	✓	✓
Asymmetric Information	✓	Perfect Bayesian (PBE) Sequential eq. (SE) and refinements No methodology! ?

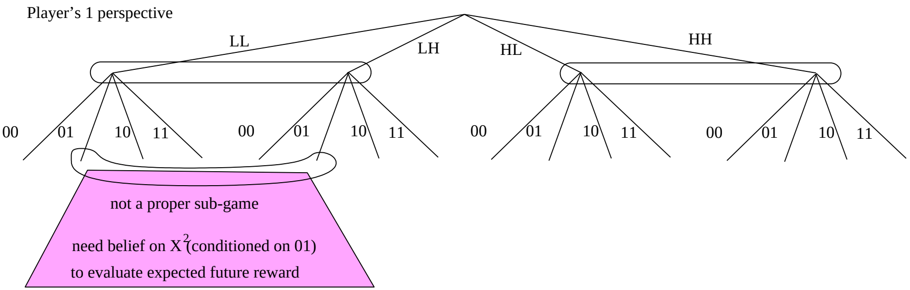
Perfect Bayesian equilibria (PBE)



Perfect Bayesian equilibria (PBE)



Perfect Bayesian equilibria (PBE)



- SPE is not appropriate equilibrium concept!
- Perfect Bayesian equilibrium (PBE)

Perfect Bayesian equilibria (PBE)

- A PBE is an assessment (σ^*, μ^*) of strategy profiles σ^* and beliefs μ^* satisfying (a) sequential rationality and (b) consistency

(a) For every $t \in \mathcal{T}$, agent $i \in \mathcal{N}$, information set $(A_{1:t-1}, X^i)$, and unilateral deviation σ^i

$$\mathbb{E}^{\mu^*,\sigma^{*i}\sigma^{*-i}}\{\sum_{t'=t}^{T} R^i(X, A_{t'})|A_{1:t-1}, X^i\} \geq \mathbb{E}^{\mu^*,\sigma^{i}\sigma^{*-i}}\{\sum_{t'=t}^{T} R^i(X, A_{t'})|A_{1:t-1}, X^i\}$$

(b) Beliefs μ^* should be updated by Bayes law (whenever possible) given σ^* and satisfy further consistency conditions [Fudenberg and Tirole, 1991, ch. 8]

- Due to the circular dependence of μ^* and σ^* finding PBE is a large fixed-point problem (no time decomposition)

Ideas from teams: structured equilibrium strategies σ^*

- Useful idea from teams:
  Instead of considering equilibria with general strategies $\sigma^* = (\sigma_t^{*i})_{t \in \mathcal{T}}^{i \in \mathcal{N}}$ of the form

  $$A_t^i \sim \sigma_t^{*i}(\cdot | X^i, A_{1:t-1})$$

  consider equilibria with **structured** strategies $\theta = (\theta_t^i)_{t \in \mathcal{T}}^{i \in \mathcal{N}}$ of the form

  $$A_t^i \sim \Gamma_t^i(\cdot | X^i) = \theta_t^i[\Pi_t](\cdot | X^i) = m_t^i(\cdot | X^i, \Pi_t)$$

  where

$$\Pi_{t+1} = F(\Pi_t, \Gamma_t, A_t) = F(\Pi_t, \theta_t[\Pi_t], A_t) = F_t^{\theta}(A_{1:t}) \qquad \text{(essentially Bayes law)}$$

- $\sigma^* \Leftrightarrow \theta$
- Note: although equilibrium strategies are structured, unilateral deviations may be anything

Parenthesis: are structured strategies restrictive?

Lemma

For any given strategy profile $\sigma = (\sigma^i)_{i \in \mathcal{N}}$*, there exists a structured strategy profile* $\theta \leftrightarrow m = (m^i)_{i \in \mathcal{N}}$ *with the players receiving the same average rewards for both* σ *and* m*.*

Parenthesis: are structured strategies restrictive?

Lemma

For any given strategy profile $\sigma = (\sigma^i)_{i \in \mathcal{N}}$, there exists a structured strategy profile $\theta \leftrightarrow m = (m^i)_{i \in \mathcal{N}}$ with the players receiving the same average rewards for both σ and m.

- Bottom line: Structured strategy profiles m are a sufficiently rich class so that we can concentrate on equilibria within this class.
- **Caveat:** Each m^i depends on the entire $\sigma = (\sigma^i)_{i \in \mathcal{N}}$, so unilateral deviations in σ^i result in multilateral deviations in m

Ideas from teams: beliefs μ^*

- Recall that in PBE, μ^* is a set of beliefs on unobserved types X^{-i} for each agent i and for each private history (information set) $(A_{1:t-1}, X^i)$
- Consider beliefs that are:
  (a) only functions of the common history $A_{1:t-1}$ and
  (b) are generated from a common belief in product form

$$\mu_t^*[A_{1:t-1}](X) = \prod_{j \in \mathcal{N}} \mu_t^{*j}[A_{1:t-1}](X^j)$$

- So, for each agent i and for each history $(A_{1:t-1}, X^i)$ belief on X^{-i} is

$$\prod_{j \in \mathcal{N} \setminus \{i\}} \mu_t^{*j}[A_{1:t-1}](X^j)$$

- In addition, given strategies $\sigma^* \Leftrightarrow \theta$, these beliefs are updated as

$$\underbrace{\mu_{t+1}^{*i}[A_{1:t}]}_{\Pi_{t+1}^i} = F(\underbrace{\mu_t^{*i}[A_{1:t-1}]}_{\Pi_t^i}, \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}, A_t^i)$$

Ideas from teams: beliefs μ^*

- Recall that in PBE, μ^* is a set of beliefs on unobserved types X^{-i} for each agent i and for each private history (information set) $(A_{1:t-1}, X^i)$
- Consider beliefs that are:
  (a) only functions of the common history $A_{1:t-1}$ and
  (b) are generated from a common belief in product form

$$\mu_t^*[A_{1:t-1}](X) = \prod_{j \in \mathcal{N}} \mu_t^{*j}[A_{1:t-1}](X^j)$$

- So, for each agent i and for each history $(A_{1:t-1}, X^i)$ belief on X^{-i} is

$$\prod_{j \in \mathcal{N} \setminus \{i\}} \mu_t^{*j}[A_{1:t-1}](X^j)$$

- In addition, given strategies $\sigma^* \Leftrightarrow \theta$, these beliefs are updated as

$$\underbrace{\mu_{t+1}^{*i}[A_{1:t}]}_{\Pi_{t+1}^i} = F(\underbrace{\mu_t^{*i}[A_{1:t-1}]}_{\Pi_t^i}, \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}, A_t^i)$$

- Bottom line: all "consistency" conditions are satisfied automatically.

Summary so far

- We have motivated the use of structured (equilibrium) strategies $\sigma^* \Leftrightarrow \theta$

$$A_t^i \sim \sigma_t^{*i}(\cdot|A_{1:t-1}, X^i) = \underbrace{\theta_t^i[\overbrace{\mu_t^*[A_{1:t-1}]}^{\Pi_t}]}_{\Gamma_t^i}(\cdot|X^i)$$

- We have restricted attention to a class of beliefs μ^* that remain independent and updated as

$$\underbrace{\mu_{t+1}^{*i}[A_{1:t}]}_{\Pi_{t+1}^i} = F(\underbrace{\mu_t^{*i}[A_{1:t-1}]}_{\Pi_t^i}, \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}, A_t^i)$$

- PBE equilibrium $(\sigma^*, \mu^*) \equiv (\theta, \mu^*)$ even in this restricted class is still the solution of a large fixed point equation. Circularity between θ and μ^* still present

Summary so far

- We have motivated the use of structured (equilibrium) strategies $\sigma^* \Leftrightarrow \theta$

$$A_t^i \sim \sigma_t^{*i}(\cdot|A_{1:t-1}, X^i) = \underbrace{\theta_t^i[\overbrace{\mu_t^*[A_{1:t-1}]}^{\Pi_t}]}_{\Gamma_t^i}(\cdot|X^i)$$

- We have restricted attention to a class of beliefs μ^* that remain independent and updated as

$$\underbrace{\mu_{t+1}^{*i}[A_{1:t}]}_{\Pi_{t+1}^i} = F(\underbrace{\mu_t^{*i}[A_{1:t-1}]}_{\Pi_t^i}, \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}, A_t^i)$$

- PBE equilibrium $(\sigma^*, \mu^*) \equiv (\theta, \mu^*)$ even in this restricted class is still the solution of a large fixed point equation. Circularity between θ and μ^* still present
- How can we find θ with a simple algorithm?
- Beliefs and policies are decomposed by considering the policies for all possible beliefs π; not just for μ^*

First erroneous attempt

- Recall DP equation from team problem
- For each $t = T, T-1, \ldots, 1$ and for every $\pi_t \in \Delta(\mathcal{X})$ solve the following maximization problem

$$\theta_t[\pi_t] = \gamma_t^* = \arg\max_{\gamma_t^i \gamma_t^{-i}} \mathbb{E}^{\pi_t, \gamma_t^i \gamma_t^{-i}} \left\{ R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t^i \gamma_t^{-i}, A_t)) \right\}$$

- What is the logical extension in games?

First erroneous attempt

- Recall DP equation from team problem
- For each $t = T, T-1, \ldots, 1$ and for every $\pi_t \in \Delta(\mathcal{X})$ solve the following maximization problem

$$\theta_t[\pi_t] = \gamma_t^* = \arg\max_{\gamma_t^i \gamma_t^{-i}} \mathbb{E}^{\pi_t, \gamma_t^i \gamma_t^{-i}} \left\{ R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t^i \gamma_t^{-i}, A_t)) \right\}$$

- What is the logical extension in games?
- Transform it into a best-response type equation (fix γ_t^{*-i} and maximize over γ_t^i)

for all $i \in \mathcal{N}$

$$\gamma_t^{*i} \in \arg\max_{\gamma_t^i} \mathbb{E}^{\pi_t, \gamma_t^i \gamma_t^{*-i}} \left\{ R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, A_t)) \right\}$$

First erroneous attempt: what is the catch?

$$\text{for all } i \in \mathcal{N}$$

$$\gamma_t^{*i} \in \arg\max_{\gamma_t^i} \mathbb{E}^{\pi_t, \gamma_t^i \gamma_t^{*-i}} \left\{ R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, A_t)) \right\}$$

- Why erroneous?

First erroneous attempt: what is the catch?

for all $i \in \mathcal{N}$

$$\gamma_t^{*i} \in \arg\max_{\gamma_t^i} \mathbb{E}^{\pi_t, \gamma_t^i \gamma_t^{*-i}} \left\{ R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, A_t)) \right\}$$

- Why erroneous?
- **Explanation:** reward-to-go is not conditioned on the entire history $(A_{1:t-1}, X^i)$ for user i but only on part of it $A_{1:t-1} \leftrightarrow \Pi_t$.
  This was OK in teams but is not sufficient to prove sequential rationality in games!

$$\mathbb{E}^{\mu^*, \sigma^{*i} \sigma^{*-i}} \{ \sum_{t'=t}^{T} R^i(X, A_{t'}) | A_{1:t-1}, X^i \} \geq \mathbb{E}^{\mu^*, \tilde{\sigma}^i \sigma^{*-i}} \{ \sum_{t'=t}^{T} R^i(X, A_{t'}) | A_{1:t-1}, X^i \}$$

Special case[5]

- Consider dynamical systems for which belief update is prescription-independent, i.e., $\Pi_{t+1} = F(\Pi_t, A_t)$
- In that case the backward process decomposes and conditioning on X^i is irrelevant
- A strong statement can be made for this special case: "For every PBE there exists a structured PBE that corresponds to a SPE of an equivalent symmetric-information game"

[5] [Nayyar, Gupta, Langbort, Başar, 2014], [Gupta, Nayyar, Langbort, Başar, 2014]

Second erroneous attempt

Condition on X^i in the backward induction step to be consistent with sequential rationality condition

- For each $t = T, T-1, \ldots, 1$ and for every $\pi_t \in \Delta(\mathcal{X})$ solve the following one-step fixed-point equation

$$\text{for all } i \in \mathcal{N} \text{ and for all } x^i \in \mathcal{X}^i$$

$$\gamma_t^{*i} \in \arg\max_{\gamma_t^i} \mathbb{E}^{\pi_t, \gamma_t^i(\cdot|x^i)\gamma_t^{*-i}} \left\{ R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, A_t), x^i) | x^i \right\}$$

- Note in this case reward-to-go is $V_t^i(\pi_t, x^i)$

Second erroneous attempt: explanation

$$\mathbb{E}\{\cdot|\cdot\} = \sum_{a_t, x^{-i}} \gamma_t^i(a_t^i|x^i)\gamma_t^{*-i}(a_t^{-i}|x^{-i})\pi^{-i}(x^{-i})\times$$

$$\left(R^i(x^i x^{-i}, a_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i\gamma_t^{*-i}, a_t), x^i)\right)$$

- This is an unusual fixed point equation: dependence on $\gamma_t^i(\cdot|x^i)$ but also on the entire $\gamma_t^i(\cdot|\cdot)$ (inside the belief update)

Second erroneous attempt: explanation

$$\mathbb{E}\{\cdot|\cdot\} = \sum_{a_t, x^{-i}} \gamma_t^i(a_t^i|x^i)\gamma_t^{*-i}(a_t^{-i}|x^{-i})\pi^{-i}(x^{-i})\times$$

$$\left(R^i(x^i x^{-i}, a_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i\gamma_t^{*-i}, a_t), x^i)\right)$$

- This is an unusual fixed point equation: dependence on $\gamma_t^i(\cdot|x^i)$ but also on the entire $\gamma_t^i(\cdot|\cdot)$ (inside the belief update)

- Unfortunately this results in FP solution θ with $\gamma_t^* = \theta_t[\pi_t, x]$ so resulting policy is of the form

$$A_t^i \sim \Gamma_t^{*i}(\cdot|X^i) = \theta_t^i[\Pi_t, X](\cdot|X^i)$$

which is **not implementable** (requires unknown private information X^{-i} for the strategy of i).

Third erroneous attempt

Condition on X^i in the backward induction step to be consistent with sequential rationality **and** optimize only over some part of the prescription

- For each $t = T, T-1, \ldots, 1$ and for every $\pi_t \in \Delta(\mathcal{X})$ solve the following one-step fixed-point equation

for all $i \in \mathcal{N}$ and for all $x^i \in \mathcal{X}^i$

$$\gamma_t^{*i}(\cdot|x^i) \in \arg\max_{\gamma_t^i(\cdot|x^i)} \mathbb{E}^{\pi_t, \gamma_t^i(\cdot|x^i)\gamma_t^{*-i}} \left\{ R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i(\cdot|x^i)\gamma_t^{*i}(\cdot|\cdot)\gamma_t^{*-i}, A_t), x^i)|x^i \right\}$$

Third erroneous attempt

Condition on X^i in the backward induction step to be consistent with sequential rationality **and** optimize only over some part of the prescription

- For each $t = T, T-1, \ldots, 1$ and for every $\pi_t \in \Delta(\mathcal{X})$ solve the following one-step fixed-point equation

for all $i \in \mathcal{N}$ and for all $x^i \in \mathcal{X}^i$

$$\gamma_t^{*i}(\cdot|x^i) \in \arg\max_{\gamma_t^i(\cdot|x^i)} \mathbb{E}^{\pi_t, \gamma_t^i(\cdot|x^i)\gamma_t^{*-i}} \left\{ R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i(\cdot|x^i)\gamma_t^{*i}(\cdot|\cdot)\gamma_t^{*-i}, A_t), x^i) | x^i \right\}$$

- This results in FP solution θ with $\gamma_t^* = \theta_t[\pi_t]$ for all $\pi_t \in \Delta(\mathcal{X})$
- Unfortunately, does not work in the proof: something more fundamental is going on...

An algorithm for PBE evaluation: backward recursion

- For each $t = T, T-1, \ldots, 1$ and for every $\pi_t \in \Delta(\mathcal{X})$ solve the following one-step fixed-point equation

for all $i \in \mathcal{N}$ and for all $x^i \in \mathcal{X}^i$

$$\gamma_t^{*i}(\cdot|x^i) \in \arg\max_{\gamma_t^i(\cdot|x^i)} \mathbb{E}^{\pi_t, \gamma_t^i(\cdot|x^i)\gamma_t^{*-i}} \left\{ R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \boxed{\gamma_t^{*i}\gamma_t^{*-i}}, A_t), x^i)|x^i \right\}$$

- This results in FP solution θ with $\gamma_t^* = \theta_t[\pi_t]$ for all $\pi_t \in \Delta(\mathcal{X})$
- This is **not a best-response** type function: γ_t^{*i} present on left/right hand side
- **Intuition:** Find $\gamma_t^i(\cdot|x^i)$ that is optimal under unperturbed belief update! Remember the core concept in PBE...

An algorithm for PBE evaluation: forward recursion

- From backard recursion we have obtained $\theta = (\theta_t^i)_{t\in\mathcal{T}}^{i\in\mathcal{N}}$.
- For each $t = 1, 2, \ldots, T$ and for every $i \in \mathcal{N}$, $A_{1:t}$, and X^i

$$\sigma_t^{*i}(A_t^i|A_{1:t-1}, X^i) := \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}(A_t^i|X^i)$$

$$\underbrace{\mu_{t+1}^*[A_{1:t}]}_{\Pi_{t+1}} := F(\underbrace{\mu_t^*[A_{1:t-1}]}_{\Pi_t}, \underbrace{\theta_t[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t}, A_t)$$

- In fact we can obtain a family of PBEs for any type distribution $\prod_{i\in\mathcal{N}} Q^i(X^i)$ with appropriate initialization of μ_1^*

Main Result

Theorem

(σ^*, μ^*) generated by the backward/forward algorithm (whenever it exists) is a PBE, i.e. for all $i, t, A_{1:t-1}, X^i, \sigma^i$,

$$\mathbb{E}^{\sigma^{*i}_{t:T}\sigma^{*-i}_{t:T}\mu^*_t}\left\{\sum_{n=t}^{T} R^i(X, A_n)\Big|A_{1:t-1}X^i\right\}$$
$$\geq \mathbb{E}^{\sigma^{i}_{t:T}\sigma^{*-i}_{t:T}\mu^*_t}\left\{\sum_{n=t}^{T} R^i(X, A_n)\Big|A_{1:t-1}X^i\right\}$$

and μ^* satisfies the consistency conditions.

Sketch of the proof

- Independence of types and specific DP equation are crucial in proving the result
- Modified comparison principle (backward induction)

Sketch of the proof

- Independence of types and specific DP equation are crucial in proving the result
- Modified comparison principle (backward induction)

- Specific DP guarantees that unperturbed reward-to-go (LHS) at time t is the obtained value function $V_t^i = R^i + V_{t+1}^i$
- Specific DP guarantees that unilateral deviations with fixed belief update reduce V_t^i
- Induction step reduces V_{t+1}^i to (perturbed) reward-to-go at time $t+1$
- Independence of types guarantees that resulting expression is exactly the (perturbed) reward-to-go at time t (RHS)

Comments on the new per-stage FP equation

- This is not a best-response type of FP equation (due to presence of γ^{*i} on both the LHS and RHS of equation)
- Standard tools for existence of solution (e.g., Brouwer, Kakutani) do not apply (problem with continuity of $V(\cdot)$ functions)

Comments on the new per-stage FP equation

- This is not a best-response type of FP equation (due to presence of γ^{*i} on both the LHS and RHS of equation)
- Standard tools for existence of solution (e.g., Brouwer, Kakutani) do not apply (problem with continuity of $V(\cdot)$ functions)

- Existence can be shown for a special case[6] where $R^i(X, A_t)$ does not depend on its own type X^i
- In that case prescriptions $\Gamma_t^i(\cdot|X^i) = \Gamma_t^i(\cdot)$ do not depend on private type X^i and FP equation reduces to best response.
  No signaling!
  Essentially reduces to the model $\Pi_{t+1} = F(\Pi_t, A_t)$

[6] [Ouyang, Tavafoghi, Teneketzis, 2015]

Current/Future work

- Model generalizations:
  - Types are independent controlled Markov processes (controlled by **all** actions) $P(X_t|X_{1:t-1}, A_{1:t-1}) = \prod_{i\in\mathcal{N}} Q^i(X_t^i|X_{t-1}^i, A_{t-1})$[7]
  - Dependence types with "strategic independence"[8]
  - Types are observed through a noisy channel (even by same user) $Q(Y_t^i|X_t^i)$. Example: "informational cascades" literature
  - Infinite horizon and continuous action spaces
- Existence results: prove existence for the simplest non-trivial class of problems. Core issue: the per-stage FP equation is not a best response
- Dynamic mechanism design (indirect mechanisms with message space smaller than type space)

[7] [Vasal, Subramanian, A, 2015b]
[8] [Battigalli, 1996]

Thank you!

Extra: FP equations

- First attempt

$$\left.\begin{array}{ll} \tilde{\gamma}^1 & = f_1(\gamma^2, \pi) \\ \tilde{\gamma}^2 & = f_2(\gamma^1, \pi) \end{array}\right\} \Rightarrow \tilde{\gamma} = f(\gamma, \pi) \Rightarrow \gamma^* = \theta(\pi)$$

- Second attempt

$$\left.\begin{array}{ll} \tilde{\gamma}^1 & = f_{1H}(\gamma^2, \pi) \\ \tilde{\gamma}^1 & = f_{1L}(\gamma^2, \pi) \\ \tilde{\gamma}^2 & = f_{2H}(\gamma^1, \pi) \\ \tilde{\gamma}^2 & = f_{2L}(\gamma^1, \pi) \end{array}\right\} \Rightarrow \left\{\begin{array}{ll} \tilde{\gamma} & = f_{LL}(\gamma, \pi) \\ \tilde{\gamma} & = f_{LH}(\gamma, \pi) \\ \tilde{\gamma} & = f_{HL}(\gamma, \pi) \\ \tilde{\gamma} & = f_{HH}(\gamma, \pi) \end{array}\right\} \Rightarrow \gamma^* = \theta(\pi, x)$$

Extra: FP equations

- Third attempt

$$\left.\begin{array}{ll}\tilde{\gamma}^1_H & = f_{1H}(\gamma^1_L,\gamma^2,\pi)\\ \tilde{\gamma}^1_L & = f_{1L}(\gamma^1_H,\gamma^2,\pi)\\ \tilde{\gamma}^2_H & = f_{2H}(\gamma^2_L,\gamma^1,\pi)\\ \tilde{\gamma}^2_L & = f_{2L}(\gamma^2_H,\gamma^1,\pi)\end{array}\right\} \Rightarrow \left\{\begin{array}{ll}\tilde{\gamma}^1 & = f_1(\gamma^1,\gamma^2,\pi)\\ \tilde{\gamma}^2 & = f_2(\gamma^1,\gamma^2,\pi)\end{array}\right\} \Rightarrow \tilde{\gamma} = f(\gamma,\pi)$$

$$\Rightarrow \gamma^* = \theta(\pi)$$

- Correct

$$\left.\begin{array}{ll}\tilde{\gamma}^1_H & = f_{1H}(\gamma^1,\gamma^2,\pi)\\ \tilde{\gamma}^1_L & = f_{1L}(\gamma^1,\gamma^2,\pi)\\ \tilde{\gamma}^2_H & = f_{2H}(\gamma^2,\gamma^1,\pi)\\ \tilde{\gamma}^2_L & = f_{2L}(\gamma^2,\gamma^1,\pi)\end{array}\right\} \Rightarrow \left\{\begin{array}{ll}\tilde{\gamma}^1 & = f_1(\gamma^1,\gamma^2,\pi)\\ \tilde{\gamma}^2 & = f_2(\gamma^1,\gamma^2,\pi)\end{array}\right\} \Rightarrow \tilde{\gamma} = f(\gamma,\pi)$$

$$\Rightarrow \gamma^* = \theta(\pi)$$

Battigalli, P. (1996).
Strategic independence and perfect Bayesian equilibria.
Journal of Economic Theory, 70(1):201–234.

Fudenberg, D. and Tirole, J. (1991).
Game Theory.
MIT Press, Cambridge, MA.

Gupta, A., Nayyar, A., Langbort, C., and Başar, T. (2014).
Common information based markov perfect equilibria for linear-gaussian games with asymmetric information.
SIAM Journal on Control and Optimization, 52(5):3228–3260.

Nayyar, A., Gupta, A., Langbort, C., and Başar, T. (2014).
Common information based markov perfect equilibria for stochastic games with asymmetric information: Finite games.
IEEE Trans. Automatic Control, 59(3):555–570.

Nayyar, A., Mahajan, A., and Teneketzis, D. (2013).
Decentralized stochastic control with partial history sharing: A common information approach.
Automatic Control, IEEE Transactions on, 58(7):1644–1658.

Ouyang, Y., Tavafoghi, H., and Teneketzis, D. (2015).
Dynamic oligopoly games with private markovian dynamics.
Available at www-personal.umich.edu/~tavaf/Oligopolygames.pdf.

Vasal, D., Subramanian, V., and Anastasopoulos, A. (2015a).
A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information.
Technical report.

Vasal, D., Subramanian, V., and Anastasopoulos, A. (2015b).
A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information.
In *American Control Conference*.
(Accepted for publication/presentation).