Social Learning with Heterogeneous Preferences

Pedro Brandão Solti

Penn

Oct 11, 2022

Introduction

- **observational learning** has been shown to be a powerful tool for spreading information in society.
- e.g.: Green Revolution in the developing world.
- farmers observe their neighbors using new technologies and techniques and copy their behavior.
- (Krishnan, Patnam 2014) on seeds and fertilizers in Ethiopia.
  - social learning has long-lasting impacts, unlike direct outreach.

Introduction

- evidence that heterogeneity significantly hinders diffusion of technology via social learning.
  - (Beaman, Dillon 2018) social learning is different according across genders, depending on the structure of the networks and seeding strategies.
  - (De Groote et al 2016) information about agricultural practices spread more than info about nutrition.
- **question:** how severe can the impact of heterogeneity (of preferences) on social learning be?
- **answer:** arbitrarily small amounts of heterogeneity can totally breakdown accumulation of information through social learning.

This Presentation

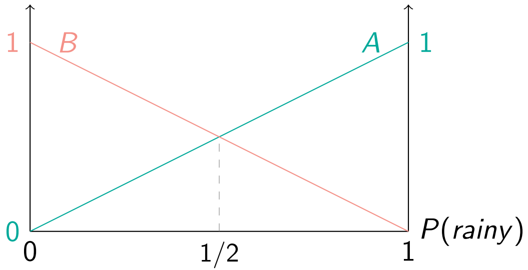
- in this presentation, I'm going to show that heterogeneity of preferences is associated with **loss of information** throughout time.
- an example where heterogeneity of preferences leads agents' actions to reflect only their own private signal. **[anti-herding ex]**
- a **necessary condition** that, if not met, information fails to aggregate well for any sequence of unbounded signal structures. **[th 1]**
- **robustness result**: information aggregates well for any sequence of unbounded signal structures if and only if a certain condition is met. **[ths 2/3]**
  - condition limits the amount of heterogeneity: coarsening of preferences must converge to that of a fictitious agent.
- different **priors** may lead to different asymptotic robustness results. **[th 4]**

Toy Model

- there are $\mathbb{N}$ farmers along an infinite road.
- each farmer decides whether to grow either A(vocados) or B(ananas) for their own consumption.
- rainy seasons benefit Avocados, and dry seasons benefit Bananas.
- before deciding what to grow, each farmer gets a private signal on how rainy the season will be.
- they also observe what their neighbor to the left decided to grow.

Toy Model

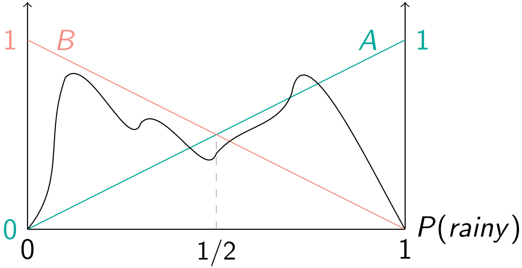
- 1st case (Acemoglu, Dahleh, Lobel 2011): everyone likes A and B equally.
- payoff is 1 if (Avocado, rainy) or (Banana, dry) and 0 otherwise.



.

Toy Model

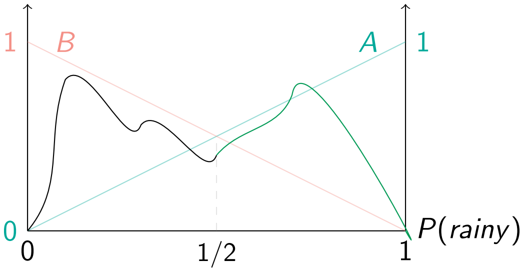
- the distribution of the first player's posteriors is shown below.
- payoff is 1 if (Avocado, rainy) or (Banana, dry) and 0 otherwise.



.

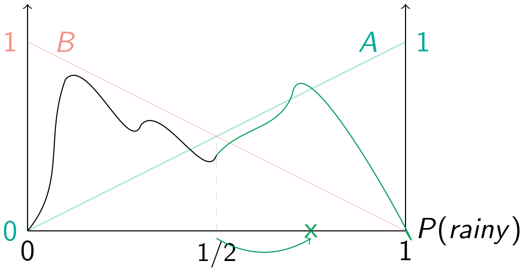
Toy Model

- suppose P2 observes that P1 chose A.
- P2 infers that P1's posterior was in the green part.
- P2 updates his belief to the average posterior in the green distribution.



Toy Model

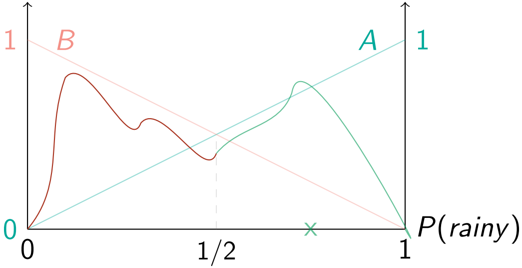
- suppose P2 observes that P1 chose A.
- P2 infers that P1's posterior was in the green part.
- P2 updates his belief to the average posterior in the green distribution.



.

Toy Model

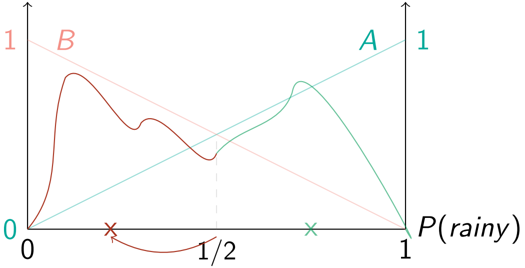
- a similar thing happens if P2 observes P1 choosing B.



.

Toy Model

- a similar thing happens if P2 observes P1 choosing B.



.

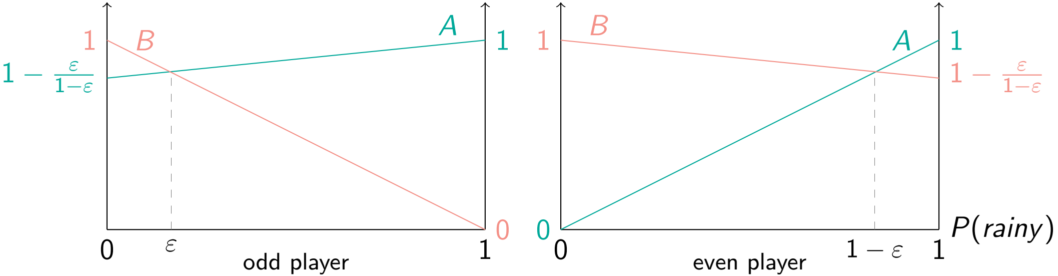
Toy Model

- (Acemoglu Dahlel Lobel 2011): this model has **asymptotic learning**.
  - in the limit, players take the same decision as the perfectly informed agent.

- **improvement principle** : a player can emulate the predecessor.
- by getting more info, she must be weakly better off. this converges to the informed payoff.

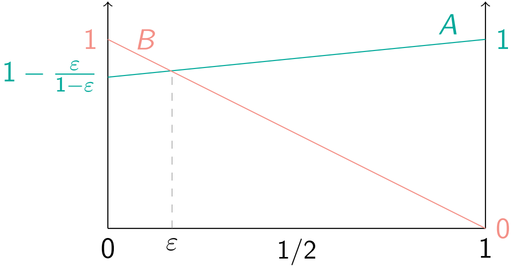
Toy Model - Heterogeneous Preferences

- 2nd case: an agent that doesn't mind bad A is followed by one that doesn't mind bad B, and vice-versa.
- common prior still $1/2$;
- with small prob ϕ, agents get a perfectly informative signal; otherwise, uninformative signal.



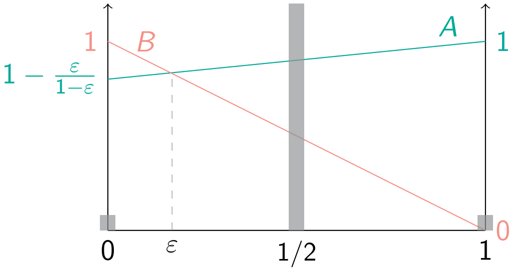
Toy Model - Heterogeneous Preferences

- let's see how the first player behaves.
- P1 only plants Bananas if she learns the season will be dry.



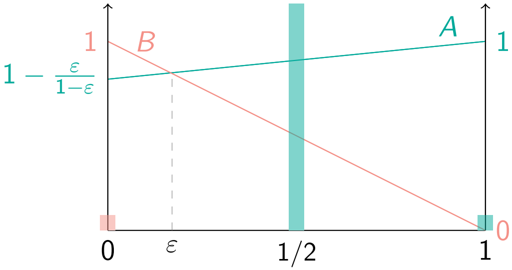
Toy Model - Heterogeneous Preferences

- let's see how the first player behaves.
- P1 only plants Bananas if she learns the season will be dry.



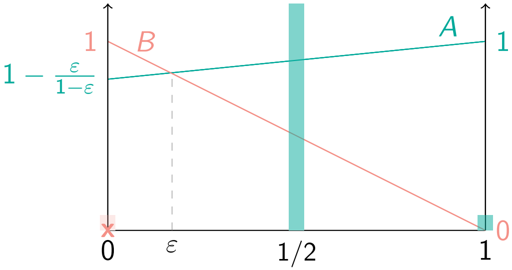
Toy Model - Heterogeneous Preferences

- let's see how the first player behaves.
- P1 only plants Bananas if she learns the season will be dry.



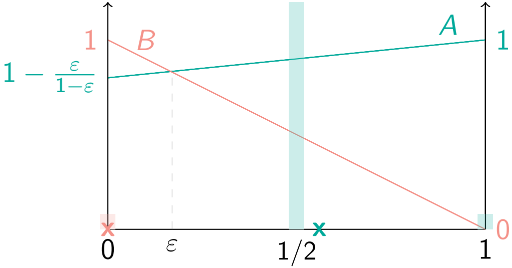
Toy Model - Heterogeneous Preferences

- let's see how the first player behaves.
- P1 only plants Bananas if she learns the season will be dry.



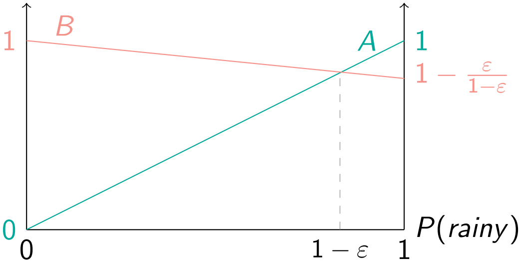
Toy Model - Heterogeneous Preferences

- let's see how the first player behaves.
- P1 only plants Bananas if she learns the season will be dry.



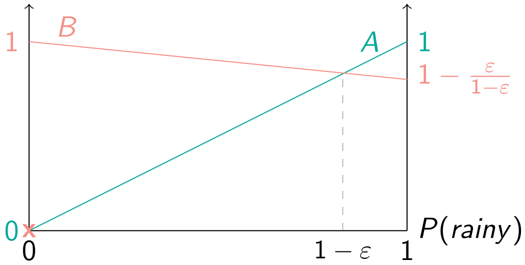
Toy Model - Heterogeneous Preferences

- let's see how the second player behaves.
- if P2 observes P1 planting B, he learns the state to be dry.



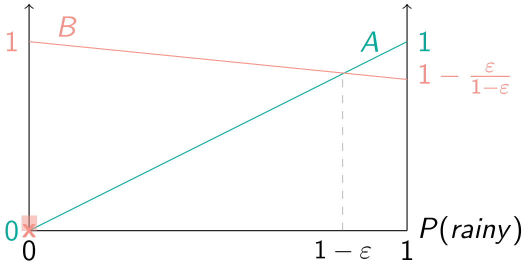
Toy Model - Heterogeneous Preferences

- let's see how the second player behaves.
- if P2 observes P1 planting B, he learns the state to be dry.



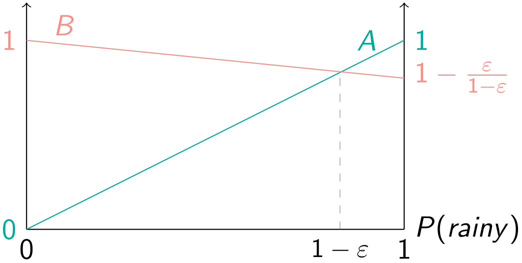
Toy Model - Heterogeneous Preferences

- let's see how the second player behaves.
- if P2 observes P1 planting B, he learns the state to be dry.



Toy Model - Heterogeneous Preferences

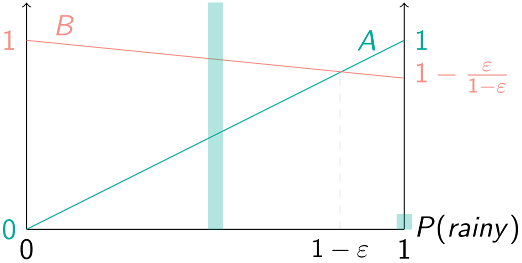
- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

Toy Model - Heterogeneous Preferences

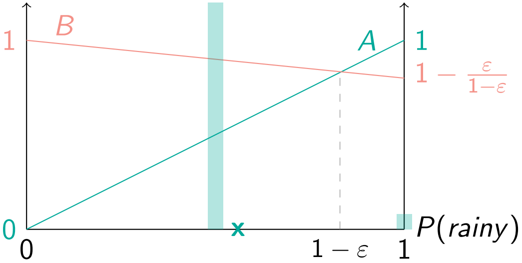
- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

Toy Model - Heterogeneous Preferences

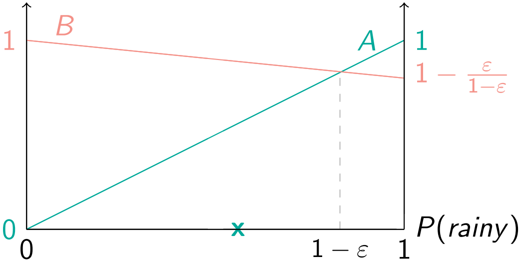
- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

Toy Model - Heterogeneous Preferences

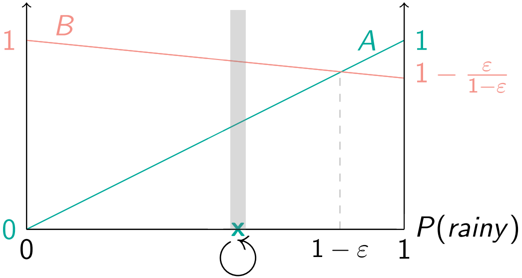
- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

Toy Model - Heterogeneous Preferences

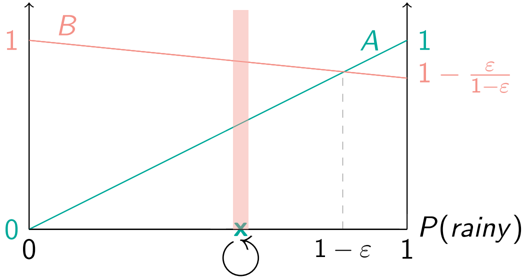
- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

Toy Model - Heterogeneous Preferences

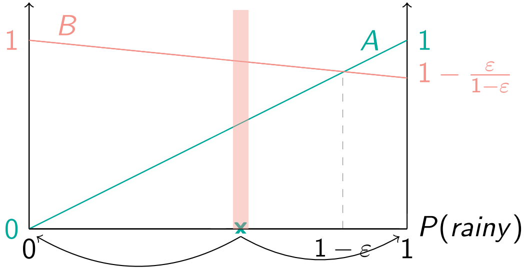
- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

Toy Model - Heterogeneous Preferences

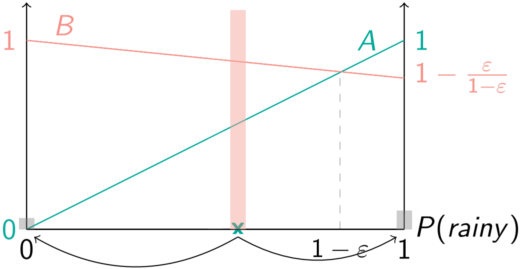
- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

Toy Model - Heterogeneous Preferences

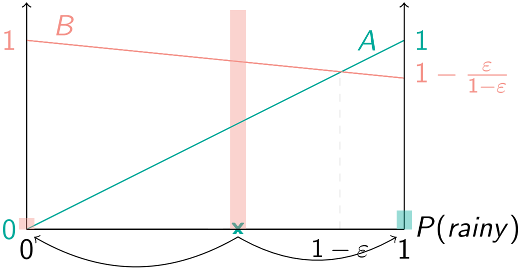
- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

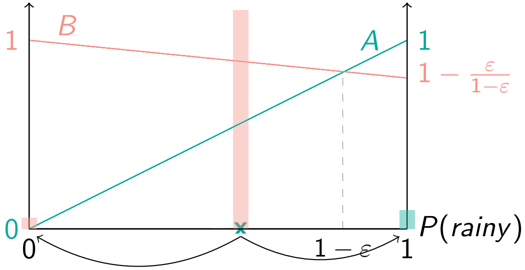
Toy Model - Heterogeneous Preferences

- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



Toy Model - Heterogeneous Preferences

- let's see how the second player behaves.
- if P2 observes P1 planting A, he slightly updates his prior.



- P2 only plants A if he gets a signal that reveals the season to be rainy.

Toy Model - Heterogeneous Preferences

- induction: players will only plant the fruit they "dislike" when they get a private signal saying that it's a good season for it.
- a player's action will be **uninformative** about past players' information.
  - **anti-herding:** player's actions only reflect their private information.
  - opposite of herding (Bikhchandani Hirshleifer Welch 92, Banerjee 92).
- information fails to accumulate over generations.
- in this talk, i'll show that:
  - this failure depends on heterogeneity of preferences;
  - arbitrarily small heterogeneity of preferences can lead to breakdown of info accumulation;
  - discuss extensions, such as optimal network design.

The Model

- discrete time model: $t = 1, 2, 3, \ldots$.
- at each period, a different player i gets to play.
- *(finite)* space of uncertainty Θ.
  - common prior $\mu \in \Delta(\Theta)$.
- player i maximizes $u_i : A_i \times \Theta \to \mathbb{R}$ uniformly bounded by M.
  - $\{u_i\}_i$ is common knowledge.
  - *(A_i is finite).*
  - for all θ, i has a uniformly strict best response at δ_θ.
- information structure:
  - each player i gets a private signal s_i, that induces conditional distribution over posteriors $F^i_\theta \in \Delta(\Delta(\Theta))$.
  - player i observes the action taken by his immediate predecessor.
  - "no herding" suff condition (e.g.: arbitrarily precise signals or gaussian signals).
- model: $\{\Theta, \{A_i, u_i\}_i, \{\{F^i_\theta\}_\theta\}_i\}$.

The Model

- discrete time model: $t = 1, 2, 3, \ldots$.
- at each period, a different player i gets to play.
- finite space of uncertainty Θ.
  - common prior $\mu \in \Delta(\Theta)$.
- player i maximizes $u_i : A_i \times \Theta \to \mathbb{R}$ uniformly bounded by M.
  - $\{u_i\}_i$ is common knowledge.
  - *(A_i is finite)*.
  - for all θ, i has a uniformly strict best response at δ_θ.
- information structure:
  - each player i gets a private signal s_i, that induces conditional distribution over posteriors $F^i_\theta \in \Delta(\Delta(\Theta))$.
  - **player i observes the action taken by his immediate predecessor**.
  - "no herding" suff condition (e.g.: arbitrarily precise signals or gaussian signals).
- model: $\{\Theta, \{A_i, u_i\}_i, \{\{F^i_\theta\}_\theta\}_i\}$.

The Model

- discrete time model: $t = 1, 2, 3, \ldots$.
- at each period, a different player i gets to play.
- finite space of uncertainty Θ.
  - common prior $\mu \in \Delta(\Theta)$.
- player i maximizes $u_i : A_i \times \Theta \to \mathbb{R}$ uniformly bounded by M.
  - $\{u_i\}_i$ is common knowledge.
  - A_i is finite.
  - for all θ, i has a uniformly strict best response at δ_θ.
- information structure:
  - each player i gets a private signal s_i, that induces conditional distribution over posteriors $F^i_\theta \in \Delta(\Delta(\Theta))$.
  - player i observes the action taken by his immediate predecessor.
  - **"no herding" suff condition (e.g: arbitrarily precise signals or gaussian signals)**.
- model: $\{\Theta, \{A_i, u_i\}_i, \{\{F^i_\theta\}_\theta\}_i\}$.

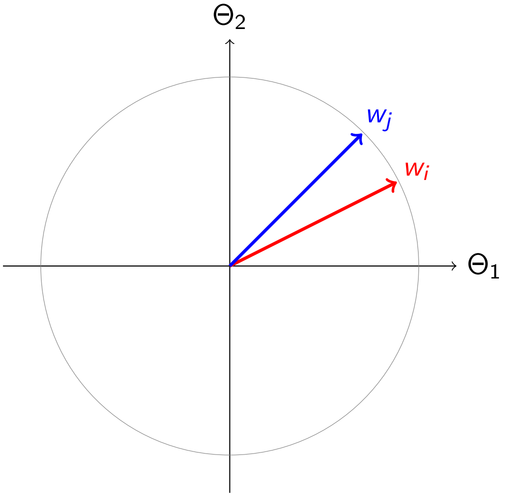
Info Structure Assumptions

- **assumption 1:** "no herding sufficient assumption"
  - arbirarily precise signals (based on Smith Sorensen 2000): $\delta_\theta \in suppF^i_\theta$.
  - gaussian signal structure (as Example 1).

- **assumption 2:** informativeness level of the signal structure bounded away from zero: $I(\mu, \{F^i\}_i) > \varepsilon$ for some ε.

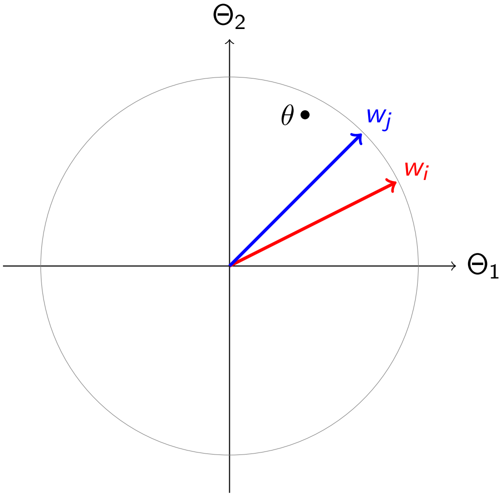
Example 2: Gaussian World

- $\Theta = \mathbb{R}^2 = \Theta_1 \times \Theta_2$
- $u_i(a, \theta) = -(a - w_i \cdot \theta)^2$
  - $A_i = \mathbb{R}$ (actions are scalars)
  - $||w_i|| = 1$
  - w_i determines what dimension i cares about.
- $\mu \sim N(0, I)$
- $s_i = \theta + \varepsilon_i$, where $\varepsilon_i \sim N(0, \Sigma)$



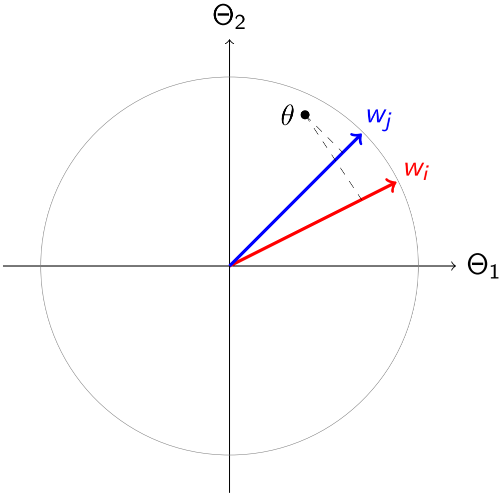
Example 2: Gaussian World

- $\Theta = \mathbb{R}^2 = \Theta_1 \times \Theta_2$
- $u_i(a, \theta) = -(a - w_i \cdot \theta)^2$
  - $A_i = \mathbb{R}$ (actions are scalars)
  - $||w_i|| = 1$
  - w_i determines what dimension i cares about.
- $\mu \sim N(0, I)$
- $s_i = \theta + \varepsilon_i$, where $\varepsilon_i \sim N(0, \Sigma)$



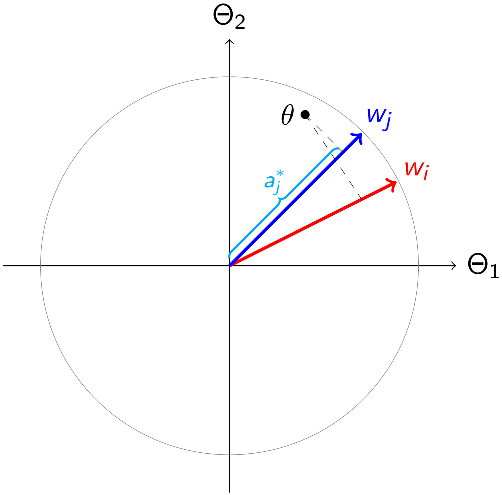
Example 2: Gaussian World

- $\Theta = \mathbb{R}^2 = \Theta_1 \times \Theta_2$
- $u_i(a, \theta) = -(a - w_i \cdot \theta)^2$
  - $A_i = \mathbb{R}$ (actions are scalars)
  - $||w_i|| = 1$
  - w_i determines what dimension i cares about.
- $\mu \sim N(0, I)$
- $s_i = \theta + \varepsilon_i$, where $\varepsilon_i \sim N(0, \Sigma)$



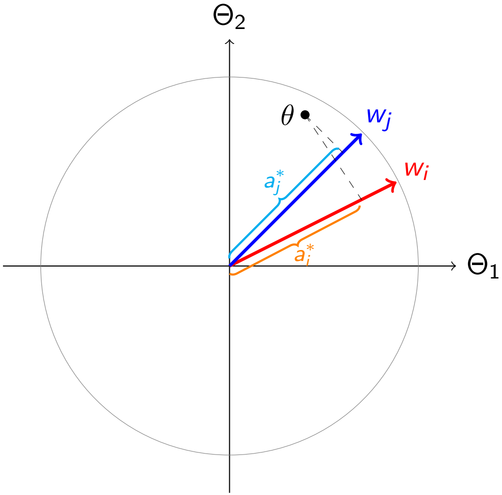
Example 2: Gaussian World

- $\Theta = \mathbb{R}^2 = \Theta_1 \times \Theta_2$
- $u_i(a, \theta) = -(a - w_i \cdot \theta)^2$
  - $A_i = \mathbb{R}$ (actions are scalars)
  - $||w_i|| = 1$
  - w_i determines what dimension i cares about.
- $\mu \sim N(0, I)$
- $s_i = \theta + \varepsilon_i$, where $\varepsilon_i \sim N(0, \Sigma)$



Example 2: Gaussian World

- $\Theta = \mathbb{R}^2 = \Theta_1 \times \Theta_2$
- $u_i(a, \theta) = -(a - w_i \cdot \theta)^2$
  - $A_i = \mathbb{R}$ (actions are scalars)
  - $||w_i|| = 1$
  - w_i determines what dimension i cares about.
- $\mu \sim N(0, I)$
- $s_i = \theta + \varepsilon_i$, where $\varepsilon_i \sim N(0, \Sigma)$



Benchmark: Asymptotic Learning

Definition 1

The model features **asymptotic learning** if, in equilibrium σ^*, agents get arbitrarily close to the best payoff possible for the true state of the world. That is,

$$|u_i(\sigma_i^*, \theta) - \max_a u_i(a, \theta)| \to_p 0.$$

The Gaussian Example

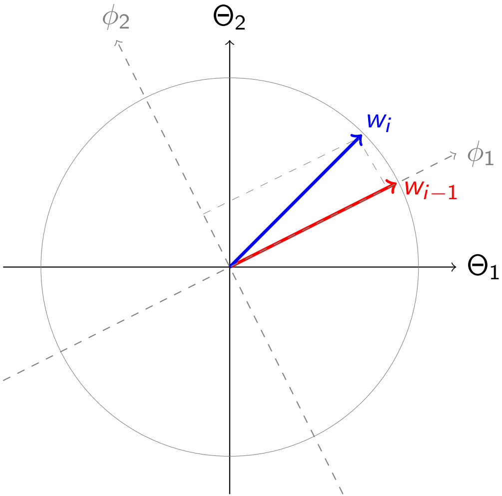
- let's see if asymptotic learning happens under the Gaussian example.

Proposition 1

In the Gaussian World model, asymptotic learning happens if and only if the preference vectors converge to being on the same direction, that is, if

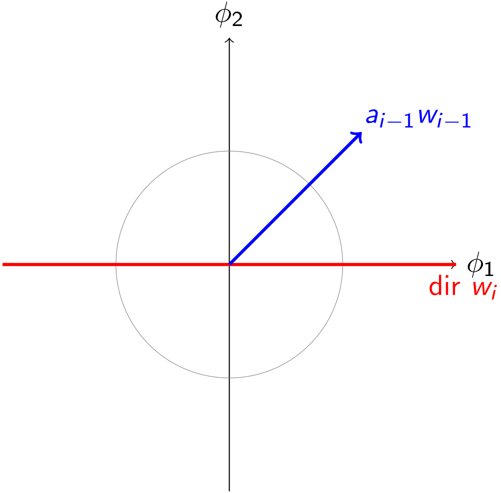
$$\lim_{i\to\infty} |w_i \cdot w_{i+1}| = 1.$$

The Gaussian Example



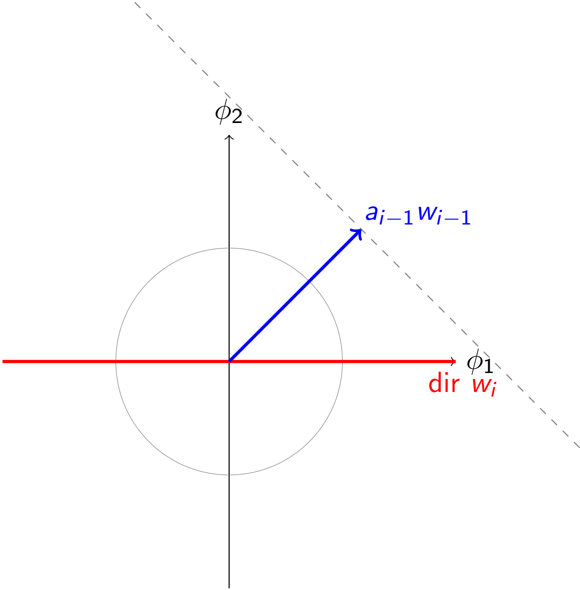
The Gaussian Example

- suppose predecessor is perfectly informed.
- by observing the predecessor's actions, the posterior will have support only on the dimension orthogonal to the predecessor's preferences w_{i-1}.



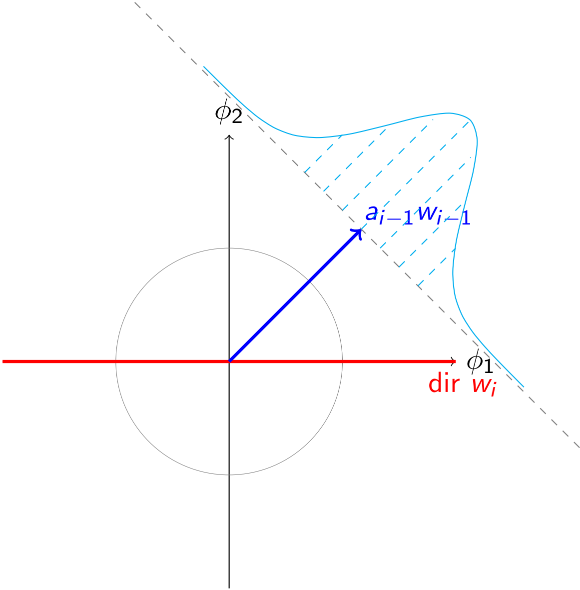
The Gaussian Example

- suppose predecessor is perfectly informed.
- by observing the predecessor's actions, the posterior will have support only on the dimension orthogonal to the predecessor's preferences w_{i-1}.



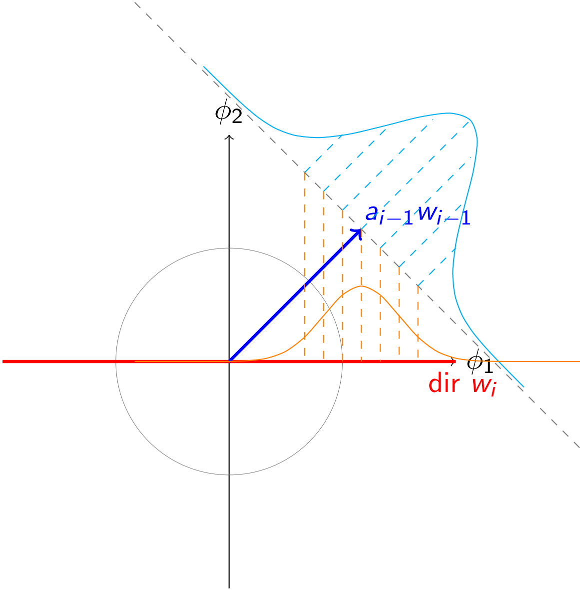
The Gaussian Example

- suppose predecessor is perfectly informed.
- by observing the predecessor's actions, the posterior will have support only on the dimension orthogonal to the predecessor's preferences w_{i-1}.



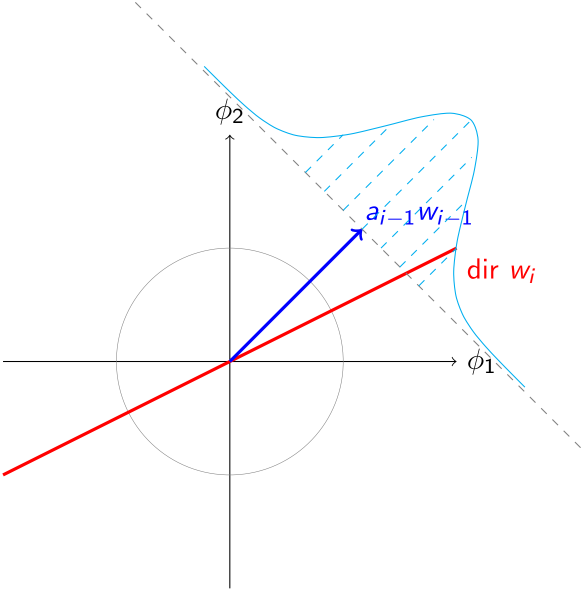
The Gaussian Example

- but what matters for the successor is the projection of that interim belief onto the direction w_i.



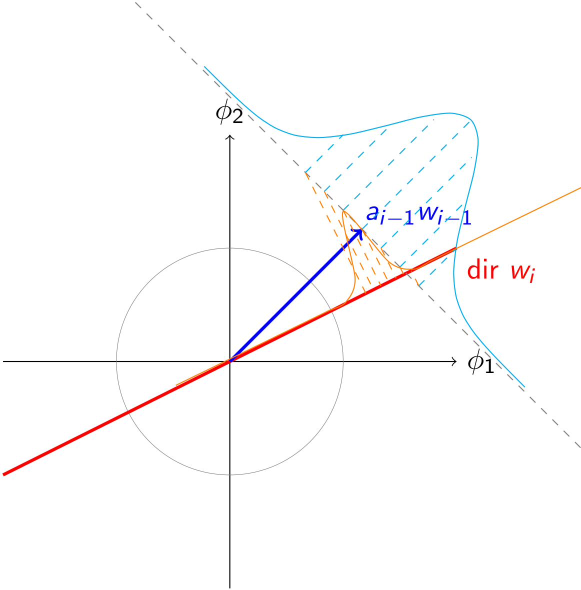
The Gaussian Example

- as the directions converge, those projections give rise to a posterior with lower variance.

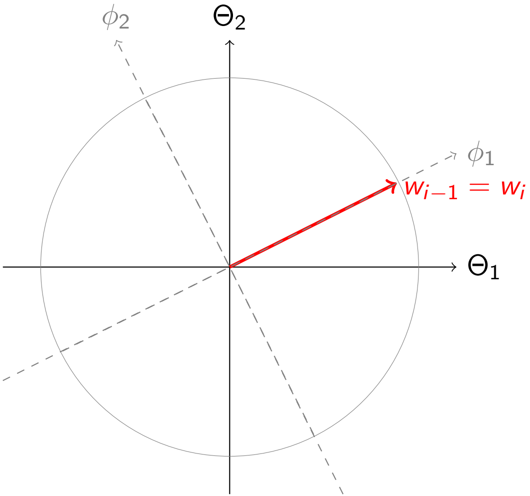


The Gaussian Example

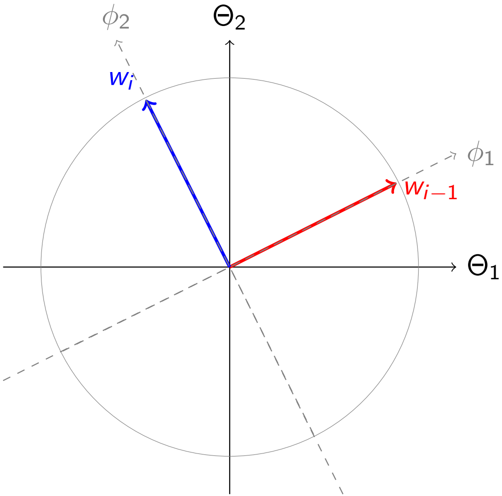
- as the directions converge, those projections give rise to a posterior with lower variance.



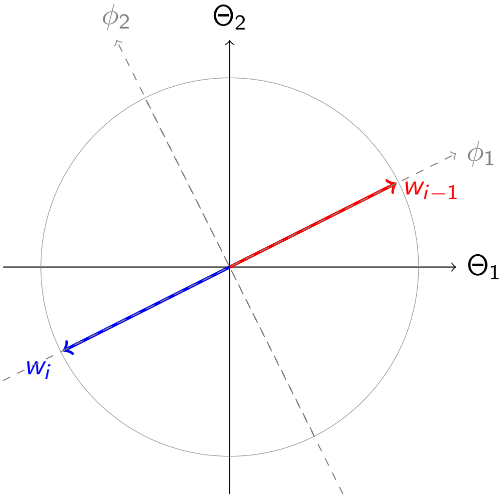
The Gaussian Example



The Gaussian Example



The Gaussian Example



Informational Content of Behavior at Certainty

- the Gaussian world suggests that players must care about the same "aspects" of the world, at least at certainty.

Definition 2

A **best-response correspondence** for player i is a mapping $BR_i : \Delta(\Theta) \rightrightarrows A_i$ such that

$$BR_i(\tilde{\mu}) = \arg\max_{a \in A_i} \mathbb{E}_{\tilde{\mu}}[u_i(a, \theta)].$$

Definition 3

Player i's **informational content of behavior at certainty** is a partition X_i of Θ such that :

$$BR_i(\delta_\theta) = BR_i(\delta_{\theta'}) \iff \exists x \in X_i \text{ such that } \theta, \theta' \in x$$

Informational Content of Behavior at Certainty

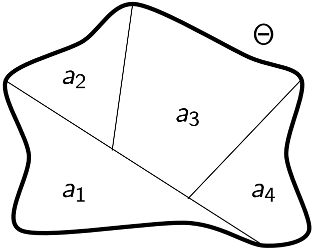


Figure: illustration of an informational content of behavior at certainty

Informational Content of Behavior at Certainty

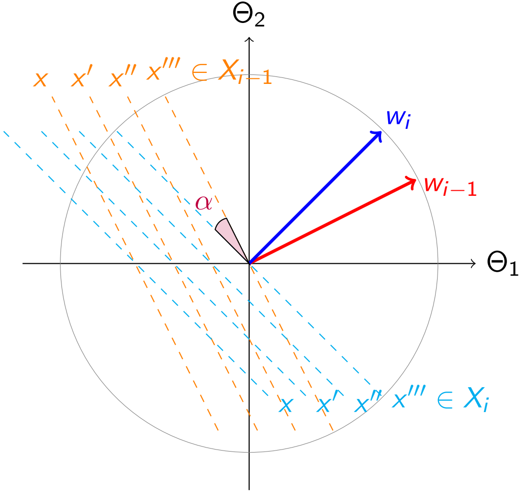


Figure: illustration of an informational content of behavior at certainty for the gaussian world. $\alpha \to 0$ iff $VI(X_i, x_{i-1}) \to 0$.

Theorem 1 - A Necessary Condition

Theorem 1

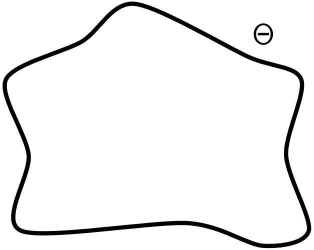
The existence of a partition X of Θ such that the sequence of informational contents at certainty converges to X, $\{X_i\} \rightarrow X$, is a necessary condition for the model to have asymptotic learning.

▸▸ The Metric of Convergence (Variation of Information)

Theorem 1 - Sketch of the Proof

- perfectly informed predecessor as the benchmark.
- "discrepancy sets"
  - states that wouldn't be maximized by learning X_{i-1}.

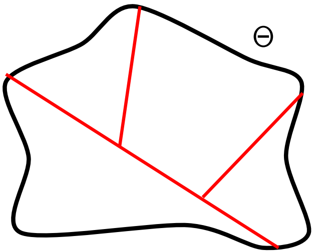
$$SD_i = \{\theta : BR_i(X_{i-1}(\theta)) \neq BR_i(\delta_\theta)\}$$



Theorem 1 - Sketch of the Proof

- perfectly informed predecessor as the benchmark.
- "discrepancy sets"
  - states that wouldn't be maximized by learning X_{i-1}.

    $$SD_i = \{\theta : BR_i(X_{i-1}(\theta)) \neq BR_i(\delta_\theta)\}$$

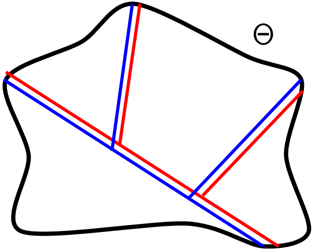


red: X_{i-1}

Theorem 1 - Sketch of the Proof

- perfectly informed predecessor as the benchmark.
- "discrepancy sets"
  - states that wouldn't be maximized by learning X_{i-1}.

$$SD_i = \{\theta : BR_i(X_{i-1}(\theta)) \neq BR_i(\delta_\theta)\}$$

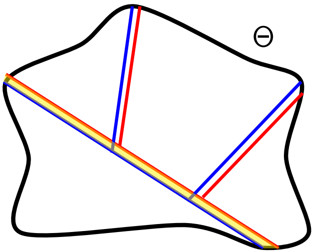


red: X_{i-1}
blue: X_i

Theorem 1 - Sketch of the Proof

- perfectly informed predecessor as the benchmark.
- "discrepancy sets"
  - states that wouldn't be maximized by learning X_{i-1}.

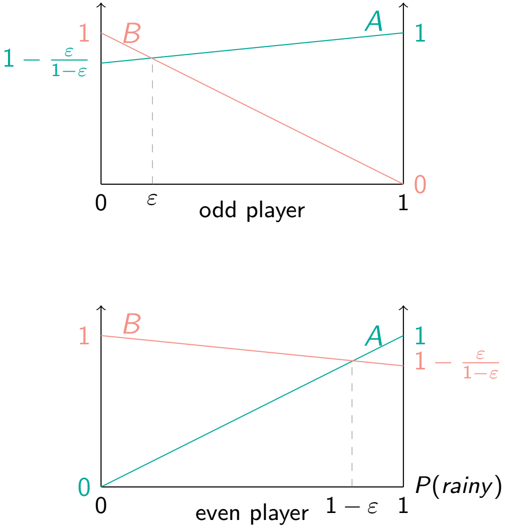
$$SD_i = \{\theta : BR_i(X_{i-1}(\theta)) \neq BR_i(\delta_\theta)\}$$



red: X_{i-1}
blue: X_i

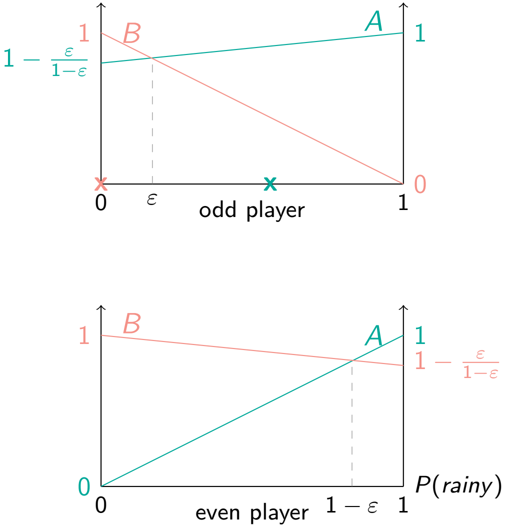
Necessary but not Sufficient

- similar behavior at certainty is necessary, but is is sufficient?
- no. counterexample: anti-herding model.
- informational content at certainty is the same for all players.
  - $X_i = \{\{\text{dry}\}, \{\text{rainy}\}\}$
- as we established, there is no asymptotic learning.



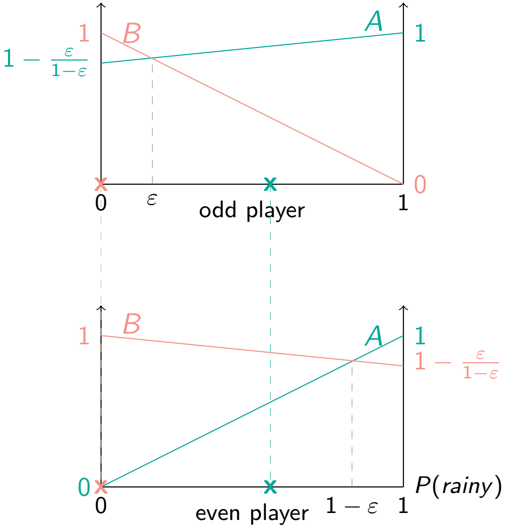
Necessary but not Sufficient

- similar behavior at certainty is necessary, but is is sufficient?
- no. counterexample: anti-herding model.
- informational content at certainty is the same for all players.
  - $X_i = \{\{\text{dry}\}, \{\text{rainy}\}\}$
- as we established, there is no asymptotic learning.



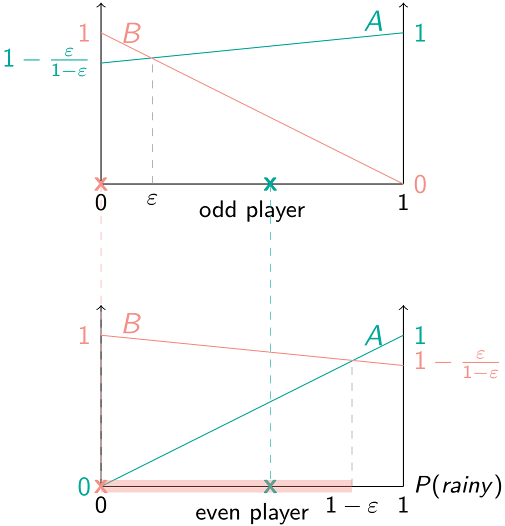
Necessary but not Sufficient

- similar behavior at certainty is necessary, but is is sufficient?
- no. counterexample: anti-herding model.
- informational content at certainty is the same for all players.
  - $X_i = \{\{\text{dry}\}, \{\text{rainy}\}\}$
- as we established, there is no asymptotic learning.



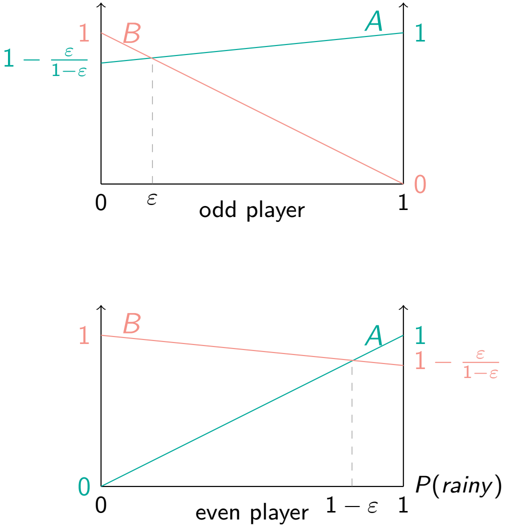
Necessary but not Sufficient

- similar behavior at certainty is necessary, but is is sufficient?
- no. counterexample: anti-herding model.
- informational content at certainty is the same for all players.
  - $X_i = \{\{\text{dry}\}, \{\text{rainy}\}\}$
- as we established, there is no asymptotic learning.



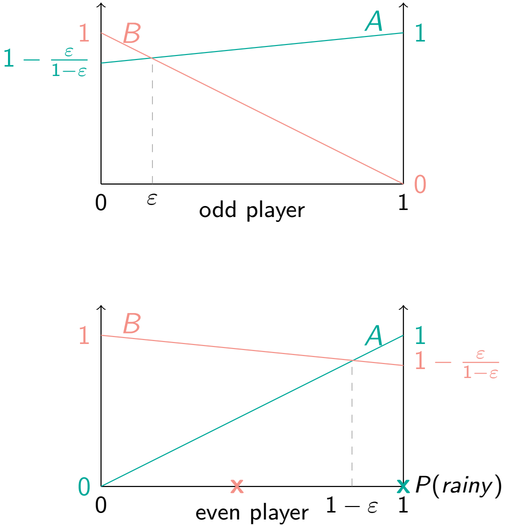
Necessary but not Sufficient

- similar behavior at certainty is necessary, but is is sufficient?
- no. counterexample: anti-herding model.
- informational content at certainty is the same for all players.
  - $X_i = \{\{\text{dry}\}, \{\text{rainy}\}\}$
- as we established, there is no asymptotic learning.



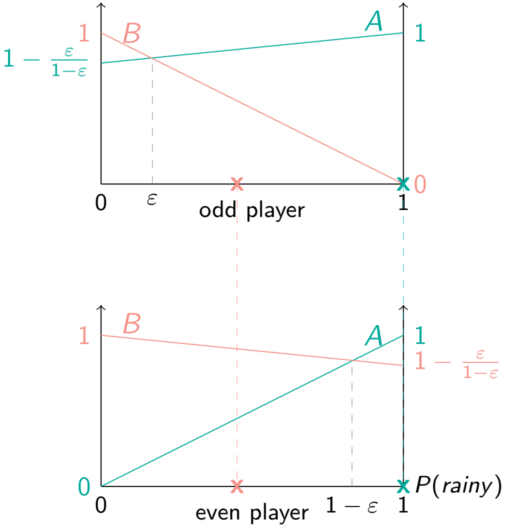
Necessary but not Sufficient

- similar behavior at certainty is necessary, but is is sufficient?
- no. counterexample: anti-herding model.
- informational content at certainty is the same for all players.
  - $X_i = \{\{\text{dry}\}, \{\text{rainy}\}\}$
- as we established, there is no asymptotic learning.



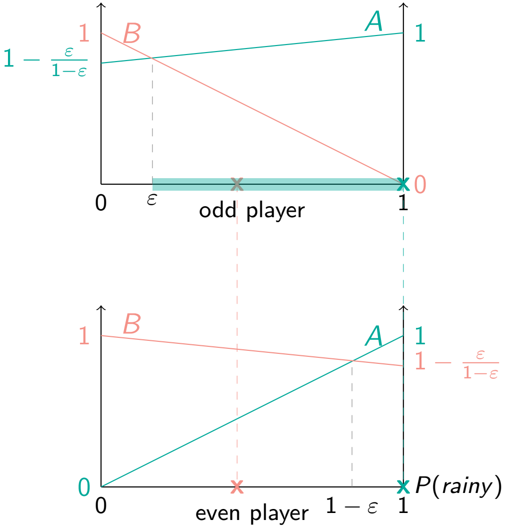
Necessary but not Sufficient

- similar behavior at certainty is necessary, but is is sufficient?
- no. counterexample: anti-herding model.
- informational content at certainty is the same for all players.
  - $X_i = \{\{\text{dry}\}, \{\text{rainy}\}\}$
- as we established, there is no asymptotic learning.



Necessary but not Sufficient

- similar behavior at certainty is necessary, but is is sufficient?
- no. counterexample: anti-herding model.
- informational content at certainty is the same for all players.
  - $X_i = \{\{\text{dry}\}, \{\text{rainy}\}\}$
- as we established, there is no asymptotic learning.



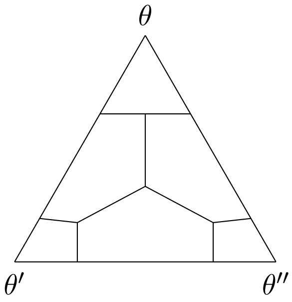
Theorem 2

- intuition: the driving force in the anti herding example is that precise beliefs get contaminated by beliefs close to the prior, losing the strength of the accumulation over time.
- Theorem 2 shows a sufficient condition to avoid that contamination for every signal structure.

Definition 4

A **best response coarsening for player** i is a partition C_i of $\Delta(\Theta)$ such that for every $c \in C_i$, there exists a subset $\tilde{A} \subseteq A_i$ for which $c_i = \cup_{a \in \mathcal{P}^{\tilde{A}}} BR_i^{-1}(a)$.
A best response coarsening is **convex** if all its elements are convex sets.

Theorem 2

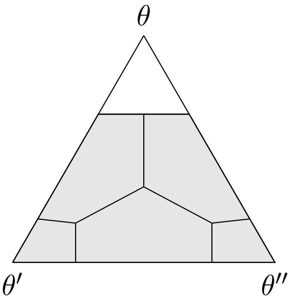


Definition 4

A **best response coarsening for player** i *is a partition* C_i *of* $\Delta(\Theta)$ *such that for every* $c \in C_i$*, there exists a subset* $\tilde{A} \subseteq A_i$ *for which* $c_i = \cup_{a \in \mathcal{P}\tilde{A}} BR_i^{-1}(a)$*.*
A best response coarsening is **convex** *if all its elements are convex sets.*

- suppose we have a player with this best response partition.

Theorem 2

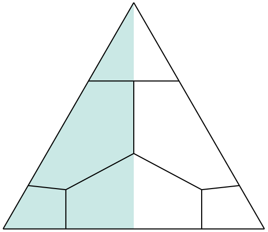


Definition 4

A **best response coarsening for player i** *is a partition C_i of $\Delta(\Theta)$ such that for every $c \in C_i$, there exists a subset $\tilde{A} \subseteq A_i$ for which $c_i = \cup_{a \in \mathcal{P}\tilde{A}} BR_i^{-1}(a)$.*
A best response coarsening is **convex** *if all its elements are convex sets.*

- suppose we have a player with this best response partition.
- the partition {white triangle, gray diamond} is a convex coarsening of this player's best response partition.

Theorem 2

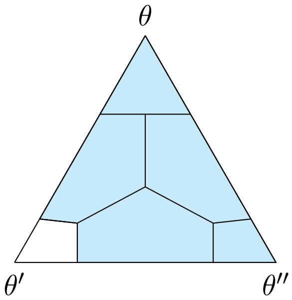


Definition 4

A **best response coarsening for player** i *is a partition* C_i *of* $\Delta(\Theta)$ *such that for every* $c \in C_i$*, there exists a subset* $\tilde{A} \subseteq A_i$ *for which* $c_i = \cup_{a \in \mathcal{P}\tilde{A}} BR_i^{-1}(a)$*.*
A best response coarsening is **convex** *if all its elements are convex sets.*

- suppose we have a player with this best response partition.
- the partition {white triangle, gray diamond} is a convex coarsening of this player's best response partition.

Theorem 2



Definition 4

A **best response coarsening for player** i *is a partition* C_i *of* $\Delta(\Theta)$ *such that for every* $c \in C_i$*, there exists a subset* $\tilde{A} \subseteq A_i$ *for which* $c_i = \cup_{a \in \mathcal{P}^{\tilde{A}}} BR_i^{-1}(a)$*.*
A best response coarsening is **convex** *if all its elements are convex sets.*

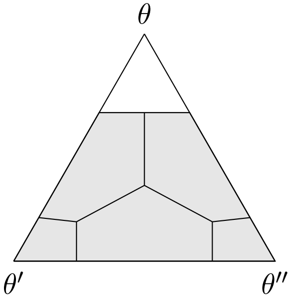
- suppose we have a player with this best response partition.
- the partition {white diamond, blue area} is a coarsening of this player's best response partition, but is not convex.

Theorem 2 - Separation

Definition 5

An event $e \subseteq E \subseteq \Theta$ is **separated** at E if there exists a sequence of best response coarsenings of $\Delta(E)$, $\{C_i\}$, that converges to a convex partition C of E such that:

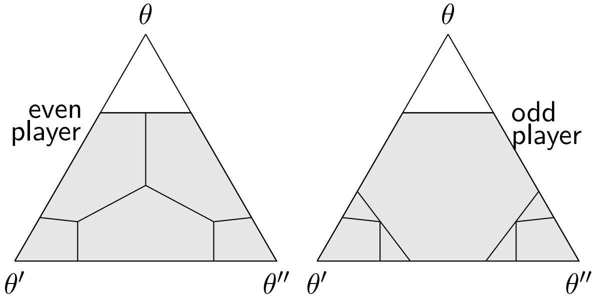
- there exists $c \in C$ for which $\delta_e \subseteq c$, and
- $\delta_{e^c} \cap c = \emptyset$.



- in this example, if the sequence of best response partitions converges to the one illustrated on the right, then $\{\theta\}$ is separated.

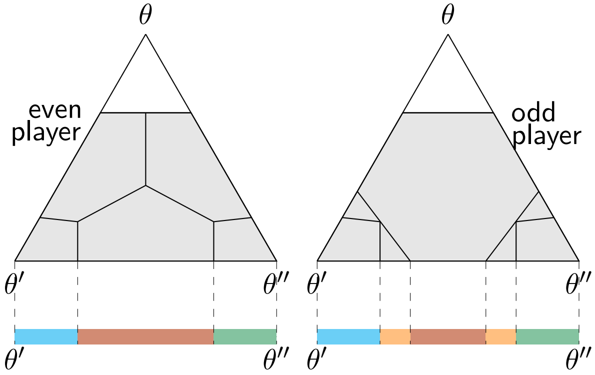
Theorem 2 - Separation

- theorem 2 will show that if $\{\theta\}$ can be separated, then, in the limit, players will be able to learn if the event is $\{\theta\}$ or $\{\theta', \theta''\}$.
- but then we can use the same criterium to try to distinguish between θ' and θ'.
- in the second round, we can merge the orange and the pink areas, and separate θ' and θ''.



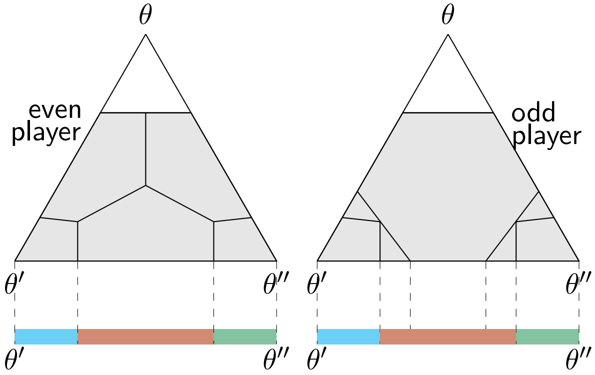
Theorem 2 - Separation

- theorem 2 will show that if $\{\theta\}$ can be separated, then, in the limit, players will be able to learn if the event is $\{\theta\}$ or $\{\theta', \theta''\}$.
- but then we can use the same criterium to try to distinguish between θ' and θ'.
- in the second round, we can merge the orange and the pink areas, and separate θ' and θ''.



Theorem 2 - Separation

- theorem 2 will show that if $\{\theta\}$ can be separated, then, in the limit, players will be able to learn if the event is $\{\theta\}$ or $\{\theta', \theta''\}$.
- but then we can use the same criterium to try to distinguish between θ' and θ'.
- in the second round, we can merge the orange and the pink areas, and separate θ' and θ''.



Theorem 2 - Iteration Process

- consider the following iteration algorithm:
  - set $t = 0$ and $S_0 = \{\Delta(\Theta)\}$;
  - consider each element $E \in S_t$. find the finest partition of E such that all of its elements can be separated at E. call this partition P_E
  - set $S_{t+1} = \times_{E \in S_t} P_E$.
  - if $S_{t+1} = \Theta$ or $S_{t+1} = S_t$, stop. otherwise, advance one step on t and return to step 2.

Theorem 2

- may X be a partition of Θ such that $VI(X_i, X) \to 0$.

Theorem 2

If the sequence of preferences leads the iteration algorithm to end at X, then the model will feature asymptotic learning for any sequence of unbounded signal structures.

- **corollary:** homogeneous preferences implies asymptotic learning.

Theorem 2- Outline of the Proof

- **goal:** recover some version of the improvement principle.
  - improvement principle: with homogeneous preferences, the successor can always imitate the predecessor, so his payoff has to be weakly higher.

- **challenge:** payoffs are changing, so how to compare?

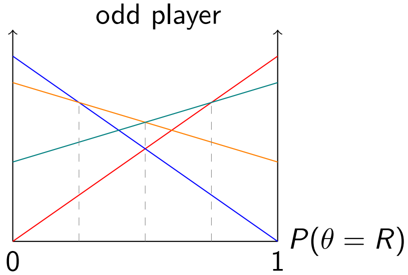
- **solution:** construct a sequence of players with homogeneous preferences ("limit player") that:
  1. have a payoff somehow related to the sequence of players;
  2. observes a signal of the action of player i;
  3. the sequence of payoffs converges to the full information payoff.

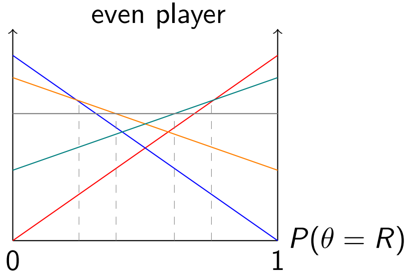
- show that the limit player's payoff is higher by observing i than by observing $i-1$.

- argue that, if the limit player is doing as well as full info by observing a signal of the actions of the actual players, then the actual players must also be doing as well as full info.

Theorem 2- Outline of the Proof

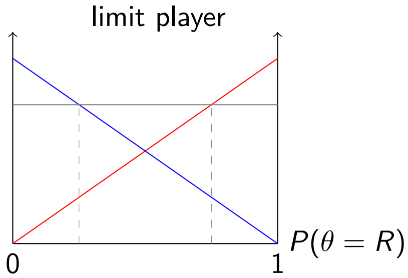
- i will give a graphic outline of the proof for a simple, two states case.
- suppose all players have a common convex coarsening of their best response partition.



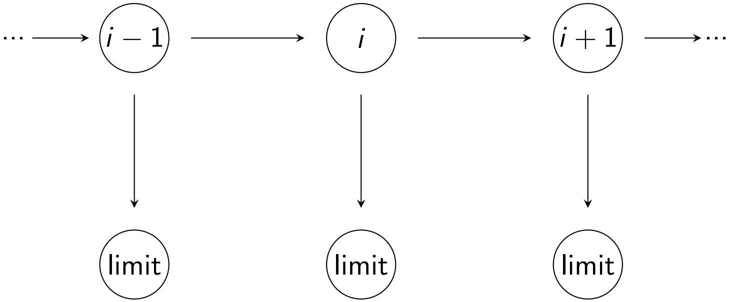


Theorem 2- Outline of the Proof

- i will give a graphic outline of the proof for a simple, two states case.
- suppose all players have a common convex coarsening of their best response partition.
- we can define a 'limit player' whose best response partition is exactly that convex coarsening.
  - (that limit player is not unique, but it doesn't matter)

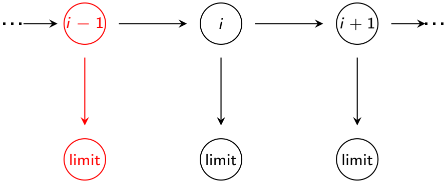


Theorem 2- Outline of the Proof

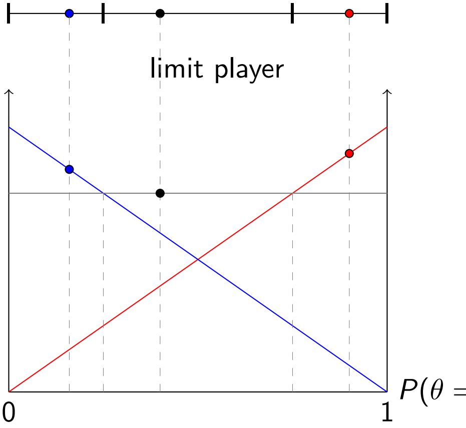


- I will consider a version of the game in which, at each period, a new limit player shows up and observes only the (censored) action of the player from that period.
- **censoring:** if two actions a and a' have their inverse best response regions merged, then the limit player won't be able to distinguish between them.
  - in our example, the limit player can't distinguish between the intermediate actions.

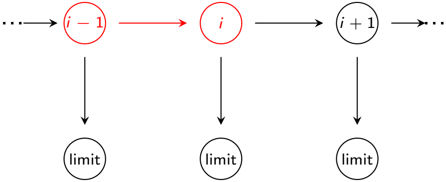
Theorem 2- Outline of the Proof

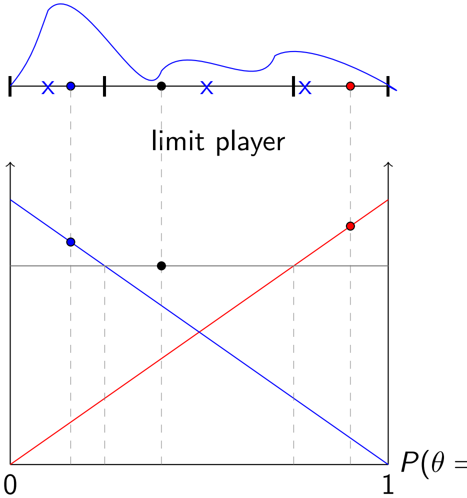


- let's start by looking at the situation when the limit player observes $i-1$'s censored action.
- given the structure of the equilibrium, the limit players know the distribution of posteriors of player $i-1$;
- the dots represent the average posterior conditioning on falling on each one of those regions.



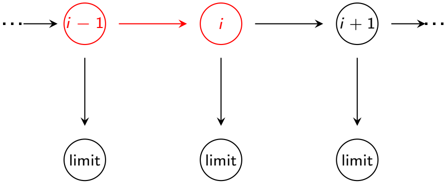
Theorem 2- Outline of the Proof



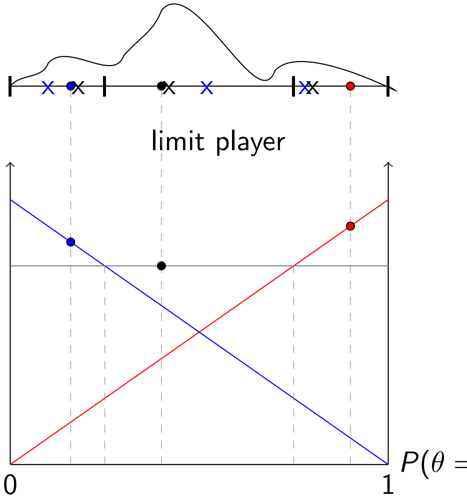


- now let's look at i observing $i-1$.
- suppose i observes $i-1$ took the blue action;
- on top of that, she will get a private signal that will lead to a distribution of posteriors conditional on $i-1$ picking blue.

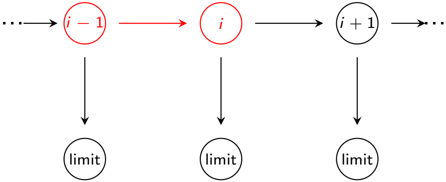
Theorem 2- Outline of the Proof



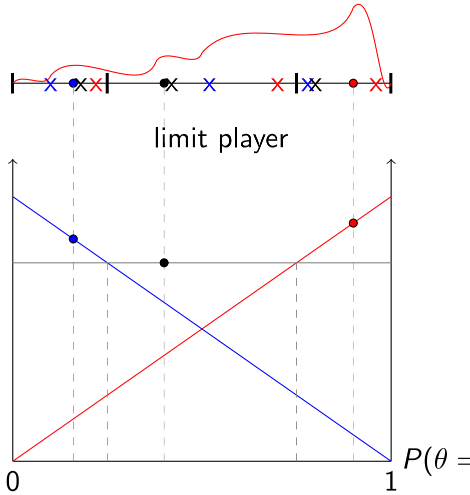
- now let's look at i observing $i-1$.
- we can do that for the average of all of $i-1$'s actions in the censored black region...



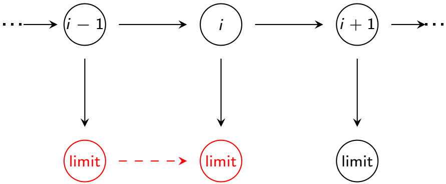
Theorem 2- Outline of the Proof



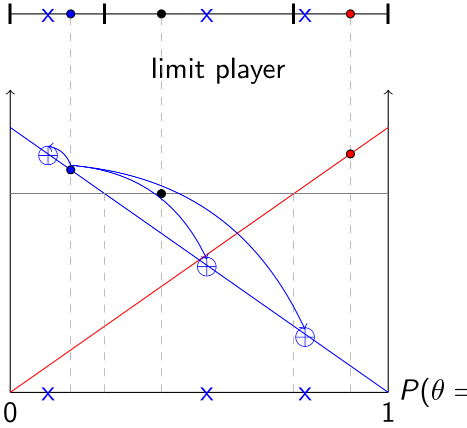
- now let's look at i observing $i-1$.
- ... and the red action.



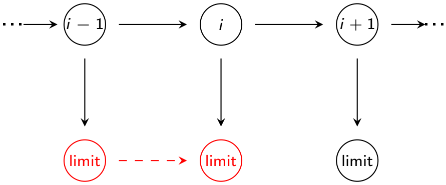
Theorem 2- Outline of the Proof



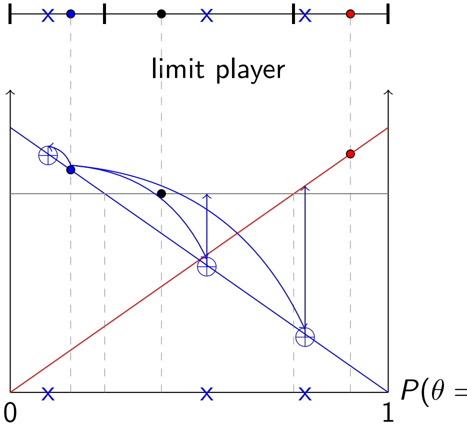
- let's try a thought experiment:
- first, we spread the posterior after observing blue from $i-1$ to match the possible posteriors if he was to observe i.
- if we at first don't allow him to change actions, the expected payoff must not change.
- and then we see what would be the gain in payoff if we allowed the limit player to change his actions.



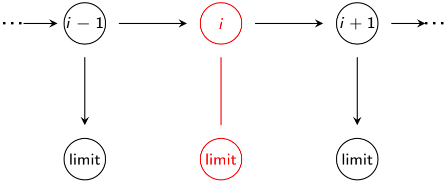
Theorem 2- Outline of the Proof



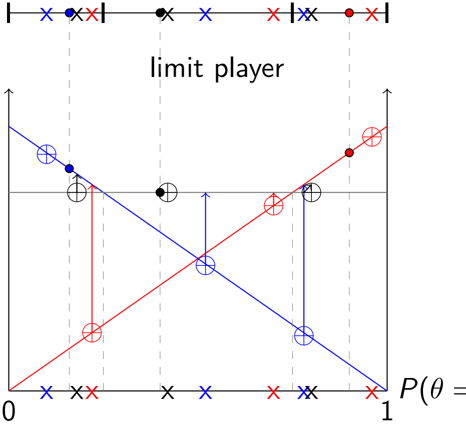
- let's try a thought experiment:
- first, we spread the posterior after observing blue from $i-1$ to match the possible posteriors if he was to observe i.
- if we at first don't allow him to change actions, the expected payoff must not change.
- and then we see what would be the gain in payoff if we allowed the limit player to change his actions.



Theorem 2- Outline of the Proof



- we can do that for all actions of $i-1$.
- the new, improved payoffs are exactly the payoffs the limit player gets when he observes i.
- **takeaway:** when the coarsenings converge, there is an "improvement principle" for the limit player: it is better to observe people that come later in line.



Theorem 2- Outline of the Proof

- this improvement is continuous in the joint distribution of actions and states of the player being observed.
- the payoff of the limit player becomes an increasing, bounded sequence. it converges. because of the continuity argument, it must converge to a stable point. the only stable point is payoff of full information.
- if the limit player converges to the payoff of full information, it is because the actions of players must become arbitrarily informative about the state of the world.
- this gives us asymptotic learning.

Theorem 3 - (or second part of theorem 2)

- Theorem 2 gives us a sufficient condition for asymptotic learning for all sequence of unbounded signal structures.
  - is it necessary?
  - yes, it is.

Theorem 3

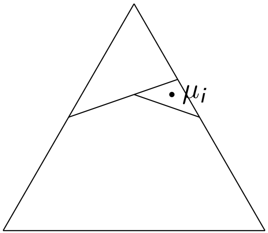
Take any $\varepsilon > 0$. Suppose the algorithm does not converge to X, and suppose it does not feature sparse heterogeneity.
Then, there exists a sequence of unbounded signal structures for which there are infinitely many agents with an expected payoff in equilibrium that is ε-away from the expected payoff they would get if they only had access to their own signal.

- **corollary:** convergence of the algorithm to X is necessary for there to be asymptotic learning for all signal structures.

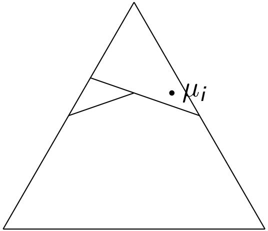
sparse heterogeneity

Theorem 3 - Sketch of the Proof

- for this talk, I'll show how to construct such a sequence of signal structures for a specific example.
- suppose there are three states of the world.
- players' preferences are described in the triangle.
- common prior μ.



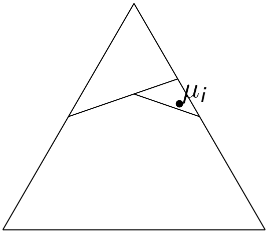
odd players



even players

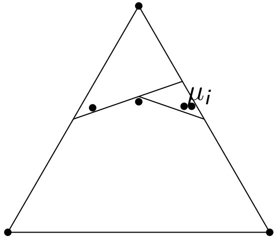
Theorem 3 - Sketch of the Proof

- let's first look at P1.
- suppose his signal structure induces the posteriors represented in the triangle.
- large probability of uninformative signal.
- if P2 observes "bottom", he can't tell which one of the three possible signals P1 received and average them out.



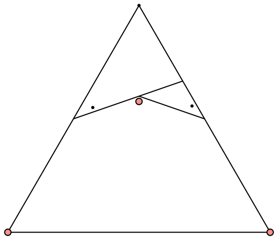
Theorem 3 - Sketch of the Proof

- let's first look at P1.
- suppose his signal structure induces the posteriors represented in the triangle.
- large probability of uninformative signal.
- if P2 observes "bottom", he can't tell which one of the three possible signals P1 received and average them out.



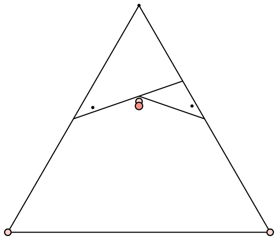
Theorem 3 - Sketch of the Proof

- let's first look at P1
- suppose his signal structure induces the posteriors represented in the triangle.
- large probability of uninformative signal.
- if P2 observes **"bottom"** , he can't tell which one of the three possible signals P1 received and average them out.



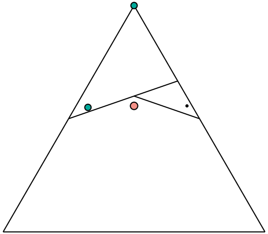
Theorem 3 - Sketch of the Proof

- let's first look at P1
- suppose his signal structure induces the posteriors represented in the triangle.
- large probability of uninformative signal.
- if P2 observes **"bottom"** , he can't tell which one of the three possible signals P1 received and average them out.



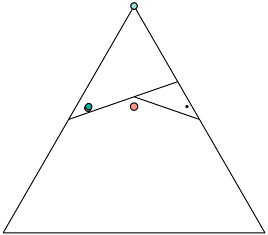
Theorem 3 - Sketch of the Proof

- let's first look at P1.
- suppose his signal structure induces the posteriors represented in the triangle.
- large probability of uninformative signal.
- if P2 observes **"left"** , he can't tell which one of the two possible signals P1 received and average them out.



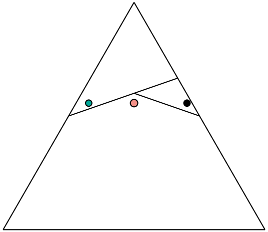
Theorem 3 - Sketch of the Proof

- let's first look at P1.
- suppose his signal structure induces the posteriors represented in the triangle.
- large probability of uninformative signal.
- if P2 observes **"left"** , he can't tell which one of the two possible signals P1 received and average them out.



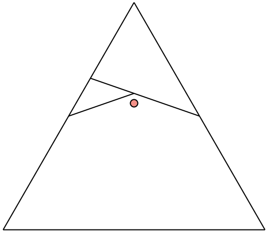
Theorem 3 - Sketch of the Proof

- let's first look at P1.
- suppose his signal structure induces the posteriors represented in the triangle.
- large probability of uninformative signal.
- if P2 observes **"right"**, he learns what was P1's signal.



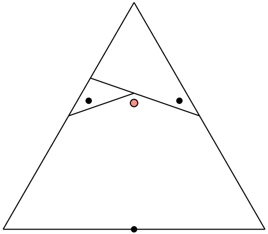
Theorem 3 - Sketch of the Proof

- now let's see what is the distribution of posteriors if P2 observes **"bottom"** .
- the posterior represented already merge all points in the same belief basin.



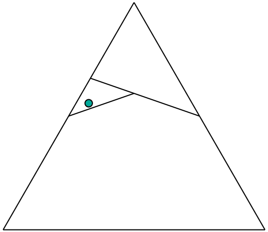
Theorem 3 - Sketch of the Proof

- now let's see what is the distribution of posteriors if P2 observes **"bottom"** .
- the posterior represented already merge all points in the same belief basin.



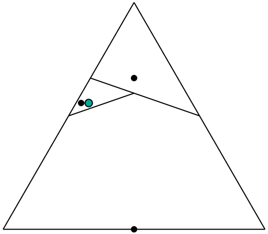
Theorem 3 - Sketch of the Proof

- now let's see what is the distribution of posteriors if P2 observes **"left"** .
- the posterior represented already merge all points in the same belief basin.



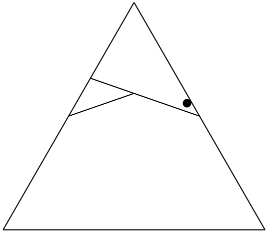
Theorem 3 - Sketch of the Proof

- now let's see what is the distribution of posteriors if P2 observes **"left"** .
- the posterior represented already merge all points in the same belief basin.



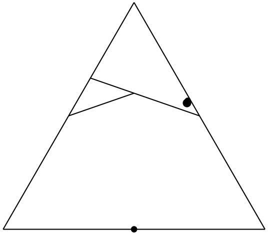
Theorem 3 - Sketch of the Proof

- now let's see what is the distribution of posteriors if P2 observes **"bottom"** .
- the posterior represented already merge all points in the same belief basin.



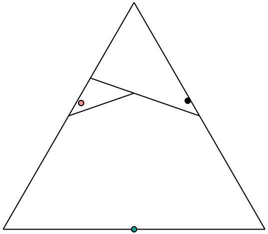
Theorem 3 - Sketch of the Proof

- now let's see what is the distribution of posteriors if P2 observes **"bottom"** .
- the posterior represented already merge all points in the same belief basin.



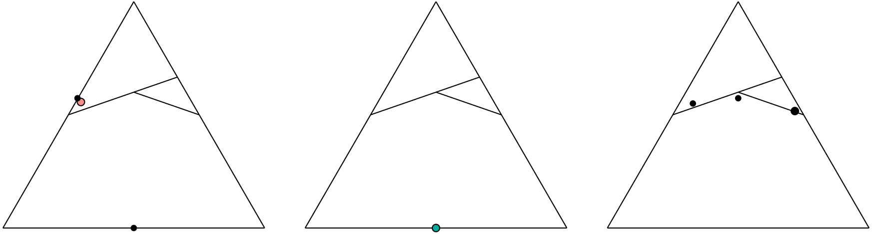
Theorem 3 - Sketch of the Proof

- by averaging all the beliefs in the same basin, the final distribution of interim beliefs for P3 is this one.



Theorem 3 - Sketch of the Proof

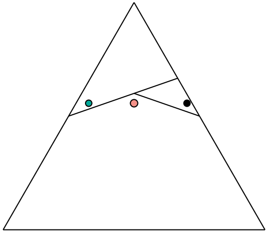
- suppose the signals for P3, conditional on each action observed is given by the graphs below.



P3 observed left P3 observed bottom P3 observed right

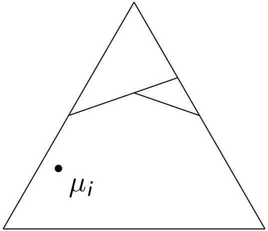
Theorem 3 - Sketch of the Proof

- by averaging out, P4 will have the same distribution of interim beliefs as P2.
- we can construct cycles.
- the interim distribution of infinitely many players will be very close to just being their prior.



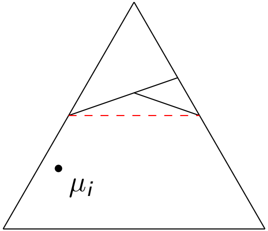
Role of the Prior

- convexity plays a role because beliefs can be "led" to the non-convex part, and then entirely leave the belief space.
- nevertheless, if the prior lies outside of the "beak", the interim belief can never be at the beak.
- the prior must lie in the convex hull of the interim beliefs.



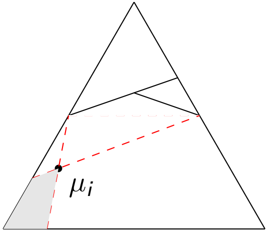
Role of the Prior

- convexity plays a role because beliefs can be “led” to the non-convex part, and then entirely leave the belief space.
- nevertheless, if the prior lies outside of the “beak”, the interim belief can never be at the beak.
- the prior must lie in the convex hull of the interim beliefs.



Role of the Prior

- convexity plays a role because beliefs can be "led" to the non-convex part, and then entirely leave the belief space.
- nevertheless, if the prior lies outside of the "beak", the interim belief can never be at the beak.
- the prior must lie in the convex hull of the interim beliefs.



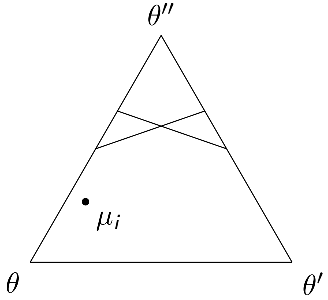
Role of the Prior

Theorem 4 (one detail missing in the proof)

If there exists a hyperplane H that divides the belief space into two half-spaces H^+ and H^- with the following features:

- *the prior $\mu \in H^+$, and*
- *all belief basins that do not contain the prior are themselves contained in H^-,*

then asymptotically, agents will learn the event $\{\theta \in \Theta : \delta_\theta \in H^-\}$.



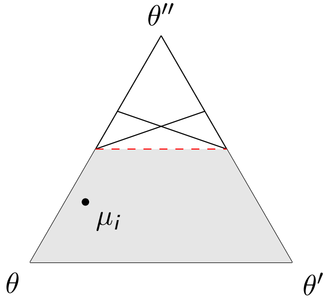
Role of the Prior

Theorem 4 (one detail missing in the proof)

If there exists a hyperplane H that divides the belief space into two half-spaces H^+ and H^- with the following features:

- *the prior $\mu \in H^+$, and*
- *all belief basins that do not contain the prior are themselves contained in H^-,*

then asymptotically, agents will learn the event $\{\theta \in \Theta : \delta_\theta \in H^-\}$.



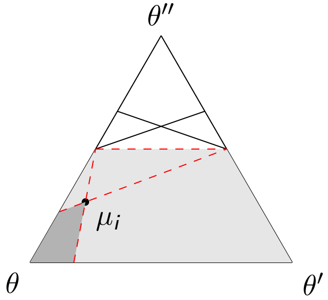
Role of the Prior

Theorem 4 (one detail missing in the proof)

If there exists a hyperplane H that divides the belief space into two half-spaces H^+ and H^- with the following features:

- *the prior $\mu \in H^+$, and*
- *all belief basins that do not contain the prior are themselves contained in H^-,*

then asymptotically, agents will learn the event $\{\theta \in \Theta : \delta_\theta \in H^-\}$.



Extension: Observing More than Predecessor

- suppose that player i observes not just $i-1$, but rather a subset of past players $N(i) \subseteq \{1, ...i-1\}$.
- furthermore, suppose that we have the **expanding observations** and **no long-run sparsity** assumption.
  - expanding observations: excludes the case that everyone only observes the first K players for some finite K.
  - no long-run sparsity: you cannot get disjoined networks if you delete the first K nodes, for any K.
- qualitatively, the results go through, but we now have to look at the **join** of the observed players' belief basins.

formal definitions

Extension: Observing More than Predecessor

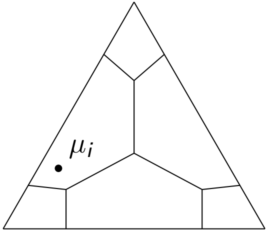
- Thm 1 can be rewritten as:
  Assume expanding observations and no long-run sparsity. The existence of a partition X of Θ such that the sequence of joins of the informational contents at certainty of the observed players converges to X,

$$\wedge_{j \in N(i)} X_j \to X,$$

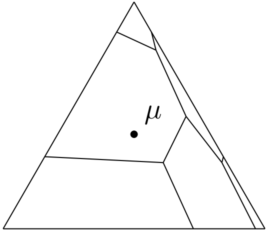
  is a necessary condition for the model to have asymptotic learning.
- the separation algorithm necessary for Thms 2 and 3 looks at the join of the belief basins instead of the belief basins themselves.

Extension: Heterogeneous Priors

- what if, instead of heterogeneous preferences agents have heterogeneous priors?
- we can turn the heterogeneous priors model into a heterogeneous preferences model and use our results.
- in this model, agents know what prior the other ones have and try to infer the signal they received.
- we can replace an agent with a different prior for one with the same prior as everyone else, and changed preferences.
  - this agent acts the same way upon receiving the same information as the original.



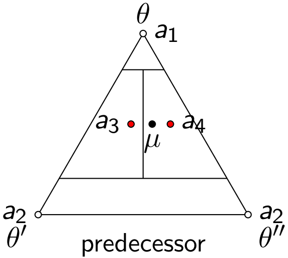
original player

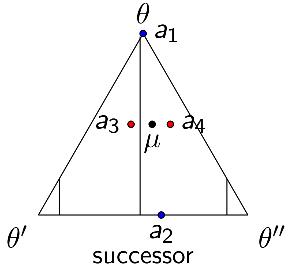


altered player

Extension: Break of Monotonicity

- **Proposition:** with heterogeneous preferences, player i does weakly better if player $i-1$ is Blackwell-better informed.
- corollary: it is better to come later in line.
- this relationship is no longer true with heterogeneous preferences.
  - giving more information to the predecessor may hurt the successor.





Extension: Optimal Network Design

- suppose a central planner can organize who observes whom.
- in general, it is not tractable to talk about optimal queuing without imposing more structure to the problem.
- the 2-dimensional Gaussian example has enough structure for us to be able to say something.
- **Proposition:** in the 2-dimensional Gaussian example, any queue that is not ordered by the angle of v_i is dominated.
  - a queue q is dominated by another queue q' if, in equilibrium, all agents get an expected payoff weakly lower under q than under q'.

Wrapping Up

- when we include heterogeneous preferences to a social learning model, the dynamics get richer in novel ways.
- it is necessary that heterogeneity of **informational content of preferences at certainty vanish** for asymptotic learning to be possible.
- if the preferences can be **'iteratatively coarsened up'** in the limit by that of a representative player then asymptotic learning happens.
  otherwise, there will be information structures for which asymptotic learning won't happen.
- the **geometric position of the prior** is relevant for learning certain events.
- extensions:
  - what if the player observes not only its immediate predecessor?
  - heterogeneous priors.
  - break of monotonicity.
  - optimal queue design (gaussian world).

Other Applications

- many different applications for different subfields of economics:
  - development
    - propagation of technologies and information in developing countries.
  - io/marketing
    - advertisement campaigns and word-of-mouth
  - political economy
    - spread of misinformation in social networks
  - macro/finance
    - spread of information through financial networks

thank you very much!

Variation of Information

Definition 6

Take two informational contents of behavior at certainty X, $Y \in \mathcal{P}^{\Theta}$. The **Variation of Information** $VI : (\mathcal{P}^{\Theta})^2 \to \mathbb{R}$ is given by:

$$VI(X, Y) = H(X) + H(Y) - 2I(X, Y), \text{ where}$$

- $H : \mathcal{P}^{\Theta} \to \mathbb{R}$ is the **entropy** function:

$$H(X) = \int_{x \in X} \mu(x) \log \mu(x) dx,$$

- $I : (\mathcal{P}^{\Theta})^2 \to \mathbb{R}$ is the **mutual information** function:

$$I(X, Y) = \int_{x \in X} \int_{y \in Y} \mu(x, y) \log \left(\frac{\mu(x, y)}{\mu(x)\mu(y)} \right) dxdy.$$

Variation of Information

$$VI(X, Y) = H(X) + H(Y) - 2I(X, Y)$$

- entropy measures the amount of uncertainty that the random object induced by partition X has.
  - deterministic ROs have the lowest possible entropy and uniform ROs the highest possible.
- mutual information measures how much you learn about a random object X by observing random object Y.
  - independent ROs have the lowest possible mutual information, and equal ROs the highest possible (and equal to the entropy).
- variation of information is a metric (Meila 2007) that measures how different the uncertainty of two random objects are.

theorem 1

Sparse Heterogeneity

Definition 7

A model contains **sparse heterogeneity** of preferences if

$$\lim_{i\to\infty} \frac{|\{j \leq i : VI_{unif}(BR_j, BR_{j+1}) \neq 0\}|}{i} = 0$$

- example:

$$u_i = \begin{cases} \tilde{u} \text{ if } i = 2^n \text{ for some natural } n, \\ u \text{ otherwise.} \end{cases}$$

Proposition 2

If a model has sparse heterogeneity and $\lim_{i\to\infty} VI(X_i, X_{i+1}) \to 0$*, then it has asymptotic learning*

theorem 3

Expanding Observations

Definition 8

A network contains **expanding observations** if

$$\lim_{i \to \infty} \mathbb{1}\left(\max_{b \in B(i)} b < K\right) = 0$$

Definition 9

A network contains **long-run sparsity** if, for every $K > 0$, the network obtained by deleting the first K nodes is connected.

▸▸ Extension: networks