Robust approachability
regret minimization with partial monitoring

S. Mannor, V. Perchet, G. Stoltz

Technion (Haifa), ENS (Cachan), ENS & HEC (Paris)

Conference On Learning Theory
06-11-2011

Introduction

- **Online-learning:** Data obtained and treated sequentially (bandits, regrets, repeated games, Markov Decision Processes, etc.), **adversarial** no stochasticity/stationary assumption (individual sequences)

- **Approachability:** equivalent to the **multi-criteria optimization** (generalizes the usual regret – convex optimization). No treatment of the objectives (as sequential optimization, or linear/convex combination of objectives, etc.).

- **Robustness (Partial monitoring):** **Payoffs/data not** (totally) **observed**; noisy feedback (image and/or signals), partial observations (bandit, congested networks, etc.)...

Model

Sequential Framework: at round $t \in \mathbb{N}$

- DM chooses (possibly at random) action $a_t \in \mathcal{A}$, law $x_t \in \Delta(\mathcal{A})$;
- Nature chooses state of world $b_t \in \mathcal{B}$, law $y_t \in \Delta(\mathcal{B})$;
- Reward (payoff) obtained: $r_t = r(a_t, b_t) \in \mathbb{R}^d$

Multi-objectives criteria

- **Target set** $\mathcal{C} \subset \mathbb{R}^d$ (convex set, that is *approachable* if...)
- **Construct an algorithm** such that $\bar{r}_T = \frac{1}{T}\sum_{t=1}^{T} r_t \to_{T\to\infty} \mathcal{C}$

Simplified model

- Expected payoff: $\mathbf{r}_t = r(x_t, y_t) = \mathbb{E}_{x_t,y_t}[r(a_t, b_t)]$. $\bar{\mathbf{r}}_T \to_{T\to\infty} \mathcal{C}$
- Concentration (or Doob) inequalities

Blackwell's condition

Blackwell's condition [Blackwell, 1956]

$\mathcal{C}$ is approachable iff $\forall y \in \Delta(\mathcal{B}),\ \exists x \in \Delta(\mathcal{A})$ s.t. $\mathbf{r}(x, y) \in \mathcal{C}$.

Interpretation: Objectives can be fulfilled simultaneously in an **adversarial, sequential** framework iff it is possible in a **known stochastic** framework ($b_t \sim y_0$ i.i.d. and y_0 known).

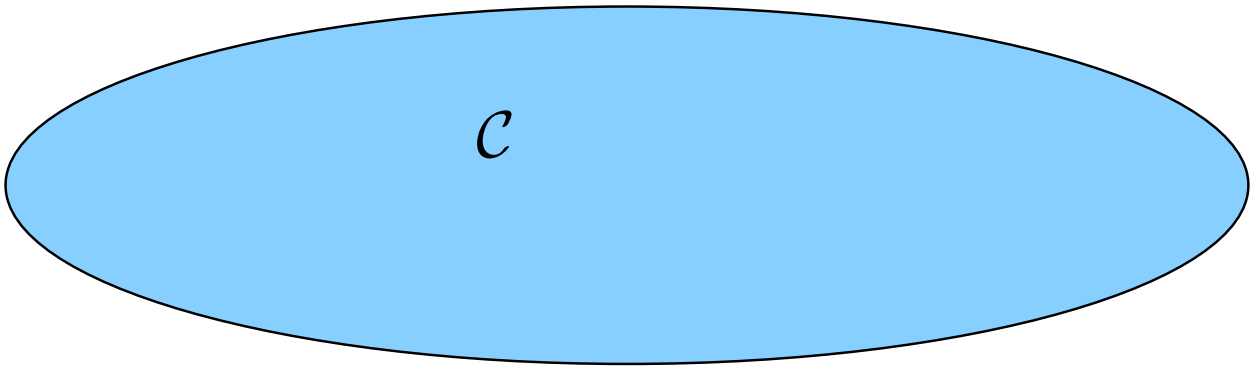
Necessity: condition not satisfied for some y_0;

- **choosing $\mathbf{b_t} \sim \mathbf{y_t} = \mathbf{y_0}$** ensures that $\bar{\mathbf{r}}_T \notin \mathcal{C}$ (in expectation).
- Concentration ineq. : $\bar{r}_T$ does not converge to $\mathcal{C}$ either (a.s.).

Sufficiency: $d\left(\bar{\mathbf{r}}_T, \mathcal{C}\right) \leq \square \frac{1}{\sqrt{T}}$; holds also for $\bar{r}_T$.

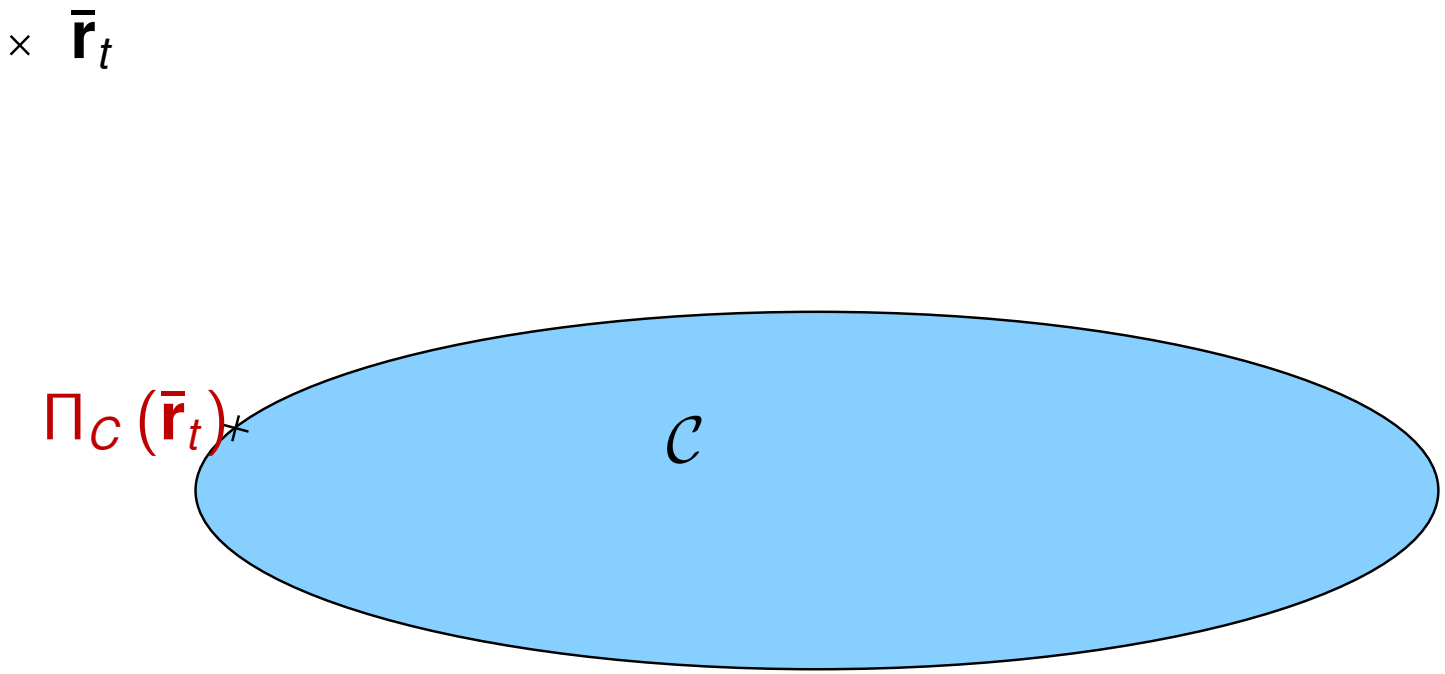
Sufficiency, insights

$\times \quad \bar{\mathbf{r}}_t$



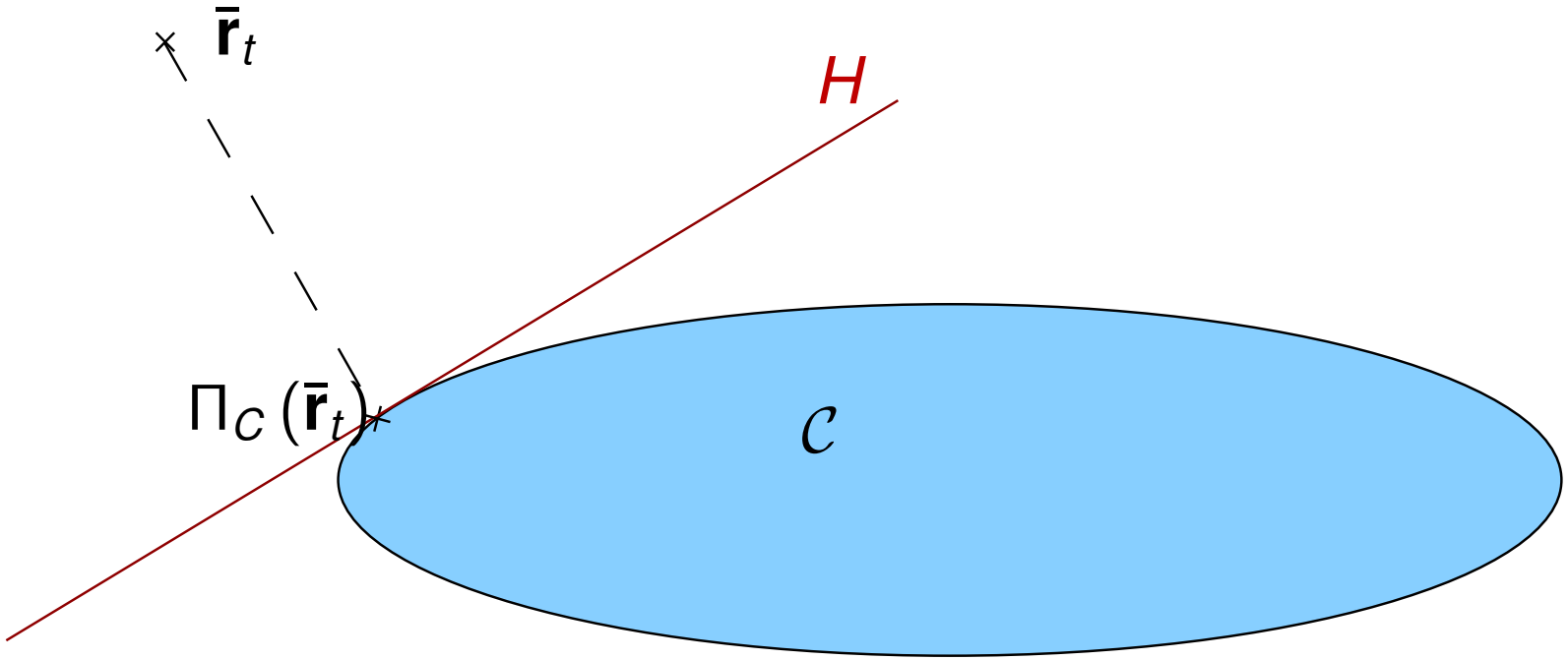
At round t, the average (expected) reward is $\bar{\mathbf{r}}_t$.

Sufficiency, insights



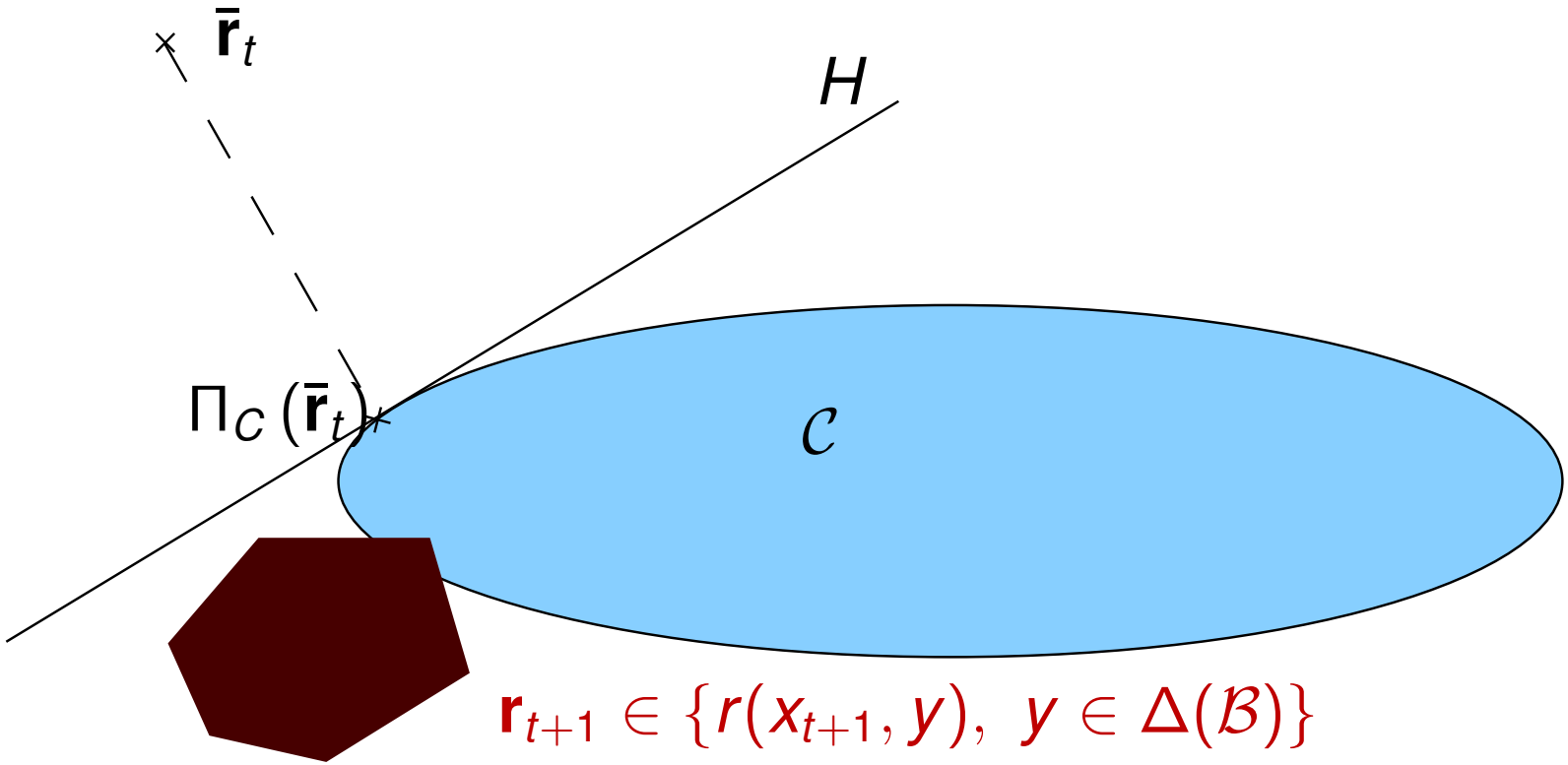
Aim: that $\bar{\mathbf{r}}_{t+1}$ gets closer to $\Pi_C(\bar{\mathbf{r}}_t)$, the projection of $\bar{\mathbf{r}}_t$.

Sufficiency, insights



This is true, at least, if r_{t+1} is on the other side of H.

Sufficiency, insights



Find and play this x_{t+1} (solution of a minmax program, existence via Blackwell's condition).

Uncertainties and partial monitoring

- Uncertainties: payoff $\mathbf{r}_t = r(x_t, y_t) \in \mathbb{R}^d$ no longer observed;
  Only information: it belongs to $\mathbf{m}(x_t, y_t) \subset \mathbb{R}^d$

  with $\mathbf{m}$ bi-linear and $\mathbf{m}(a, b)$ a **subset** of $\mathbb{R}^d$, for every $a \in \mathcal{A}$ and $b \in \mathcal{B}$.

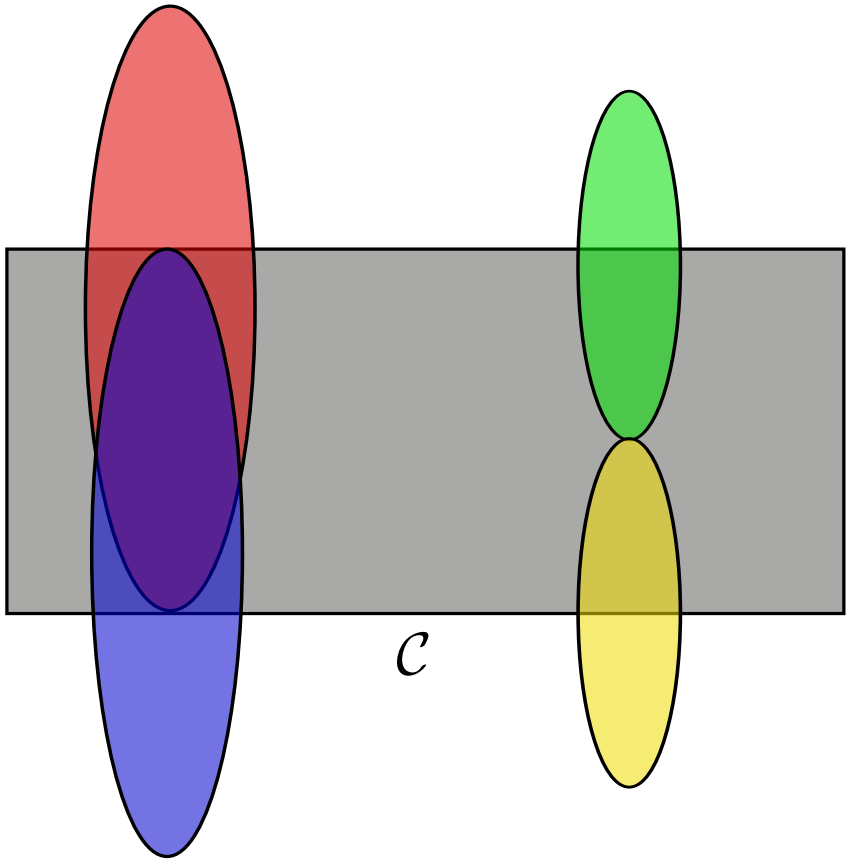
Example (part. monit.): DM receives signal $s_t \sim s(x_t, y_t) \in \Delta(\mathcal{S})$;

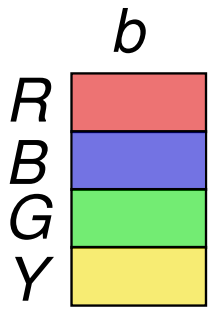
- y and y' are undistinguishable if $s(a, y) = s(a, y')$, for all $a \in \mathcal{A}$;
  or if $\mathbf{s}(y) := \left[s(a, y)\right]_{a \in \mathcal{A}} = \mathbf{s}(y')$:
- $\mathbf{m}(x, y) := \left\{ r(x, y'), \text{ s.t. } \mathbf{s}(y') = \mathbf{s}(y) \right\}$ almost bi-linear

- Robust approachabilty: choose x_t s.t. $\overline{\mathbf{m}}_T = \sum_{t=1}^T \mathbf{m}(x_t, y_t)/T$
  (and in particular $\overline{\ell}_T$) converges to $\mathcal{C}$:

$$\sup_{z_1 \in \mathbf{m}(x_1, y_1)} .. \sup_{z_T \in \mathbf{m}(x_T, y_T)} d\left(\frac{1}{T} \sum_{t=1}^T z_t, \ \mathcal{C} \right) \to 0.$$

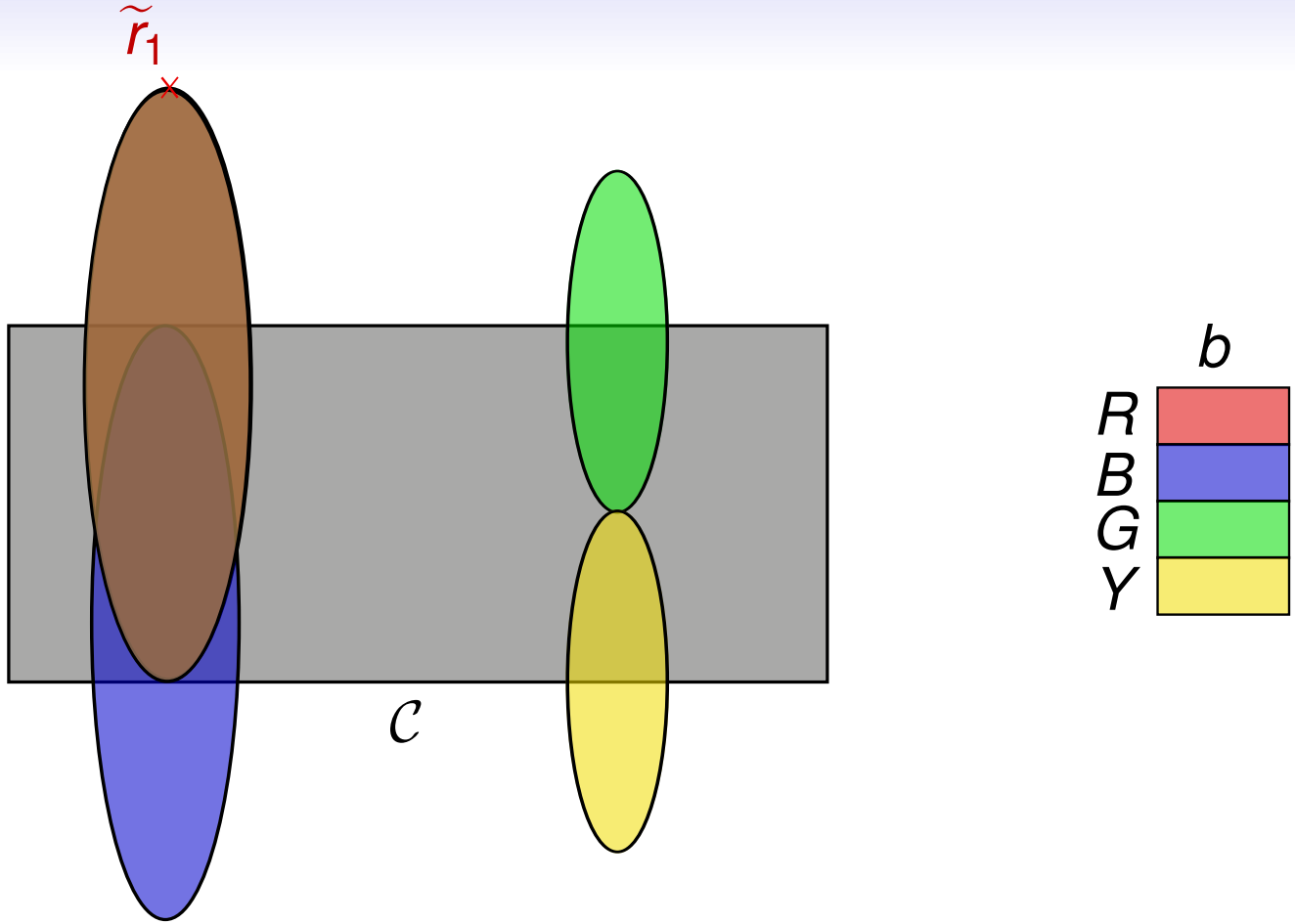
Blackwell strategy to the farthest point





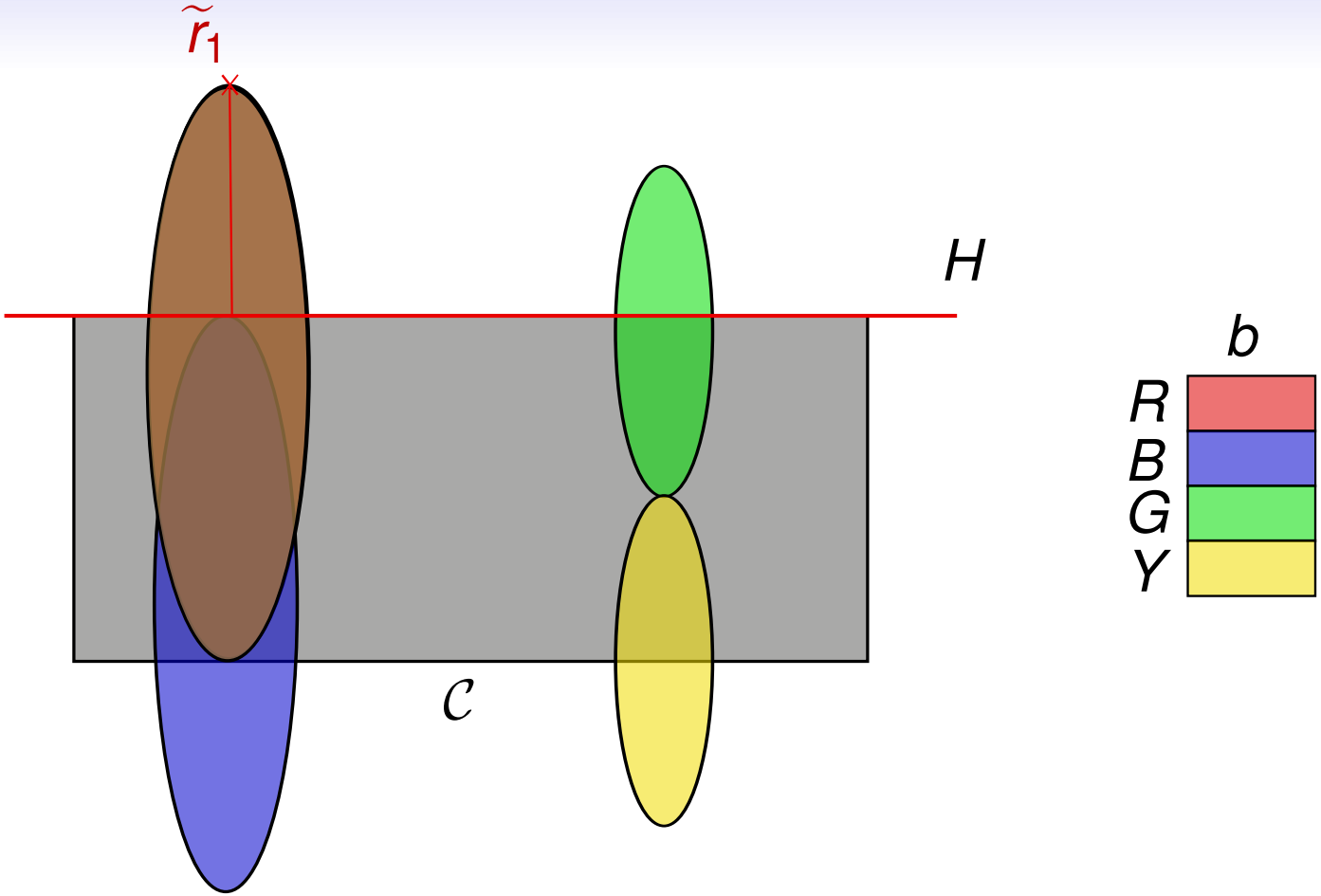
- $\mathcal{A} = \{R, B, G, Y\}$ and $\mathcal{B} = \{b\}$.
- $\mathbf{m}(R, b)$ is the red set, $\mathbf{m}(B, b)$ the blue set, etc...

Blackwell strategy to the farthest point



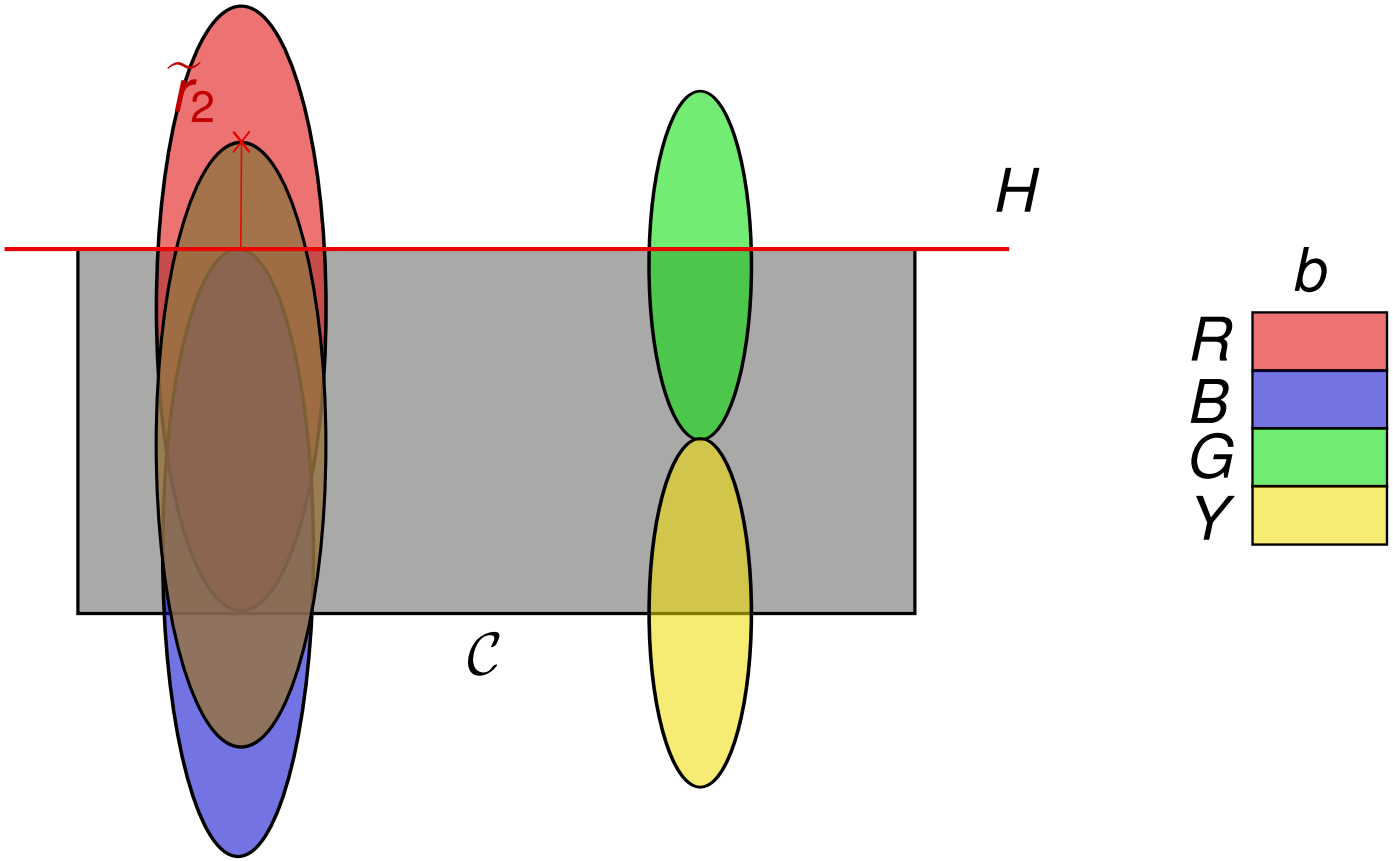
- Assume that R played at first stage; $\overline{\mathbf{m}}_1$ in brown
- $\widetilde{r}_1$ is the farthest point in $\overline{\mathbf{m}}_1$ to $\mathcal{C}$

Blackwell strategy to the farthest point



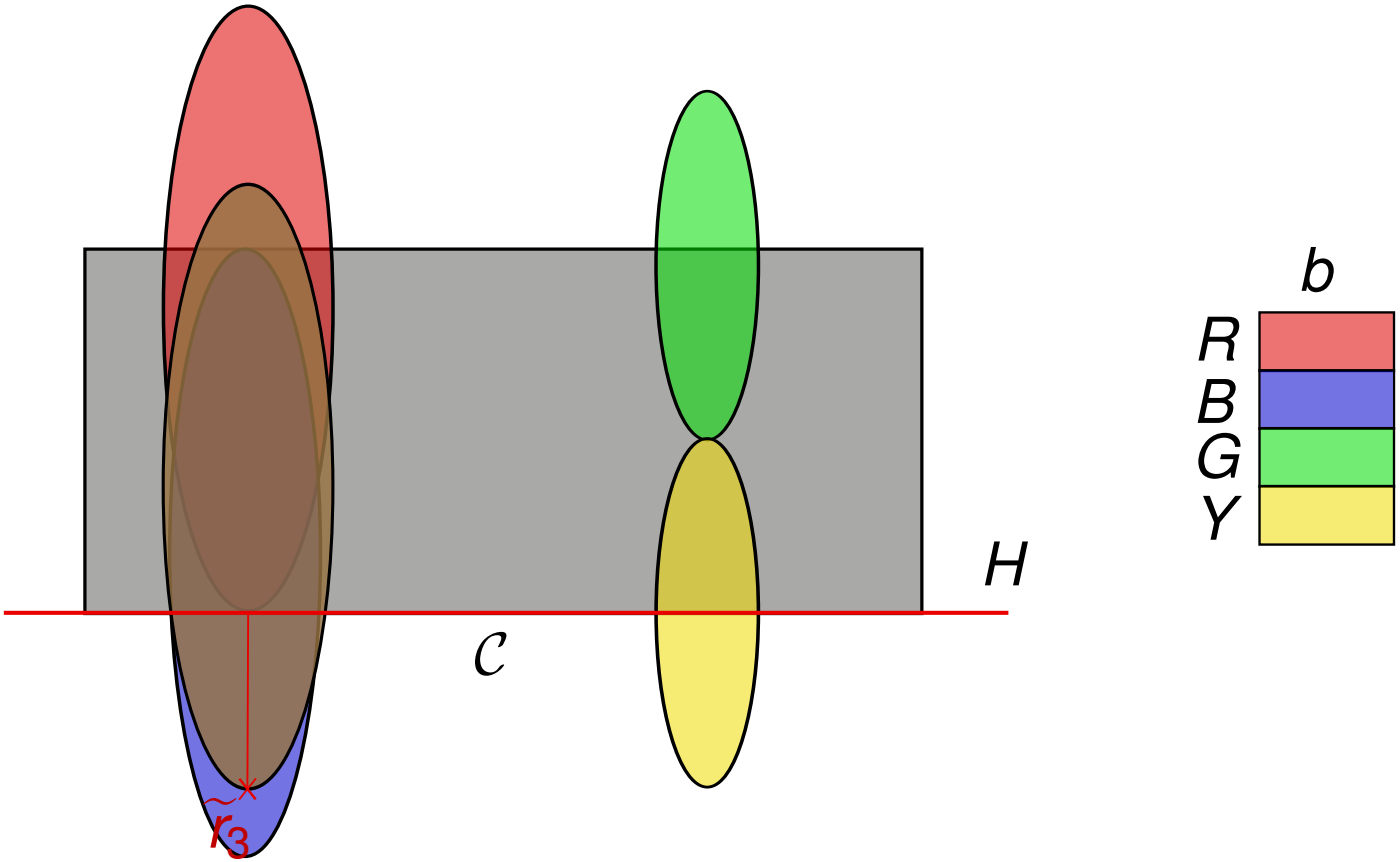
- The blue set is on the other side of the hyperplane;
- Blackwell strategy recommend to play B at stage 2.

Blackwell strategy to the farthest point



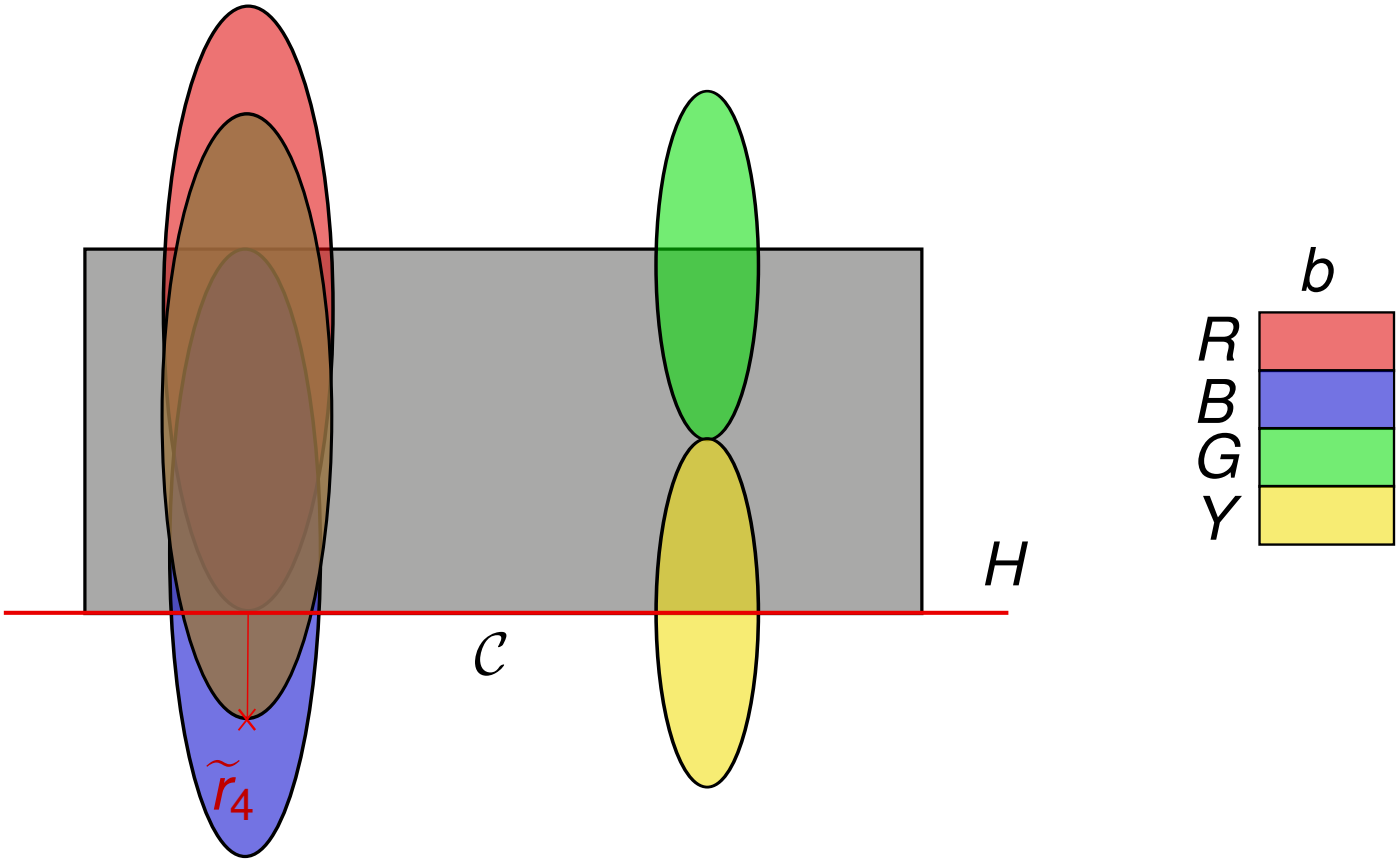
- The blue set is again on the other side of the hyperplane;
- Blackwell strategy recommend to play B at stage 3.

Blackwell strategy to the farthest point



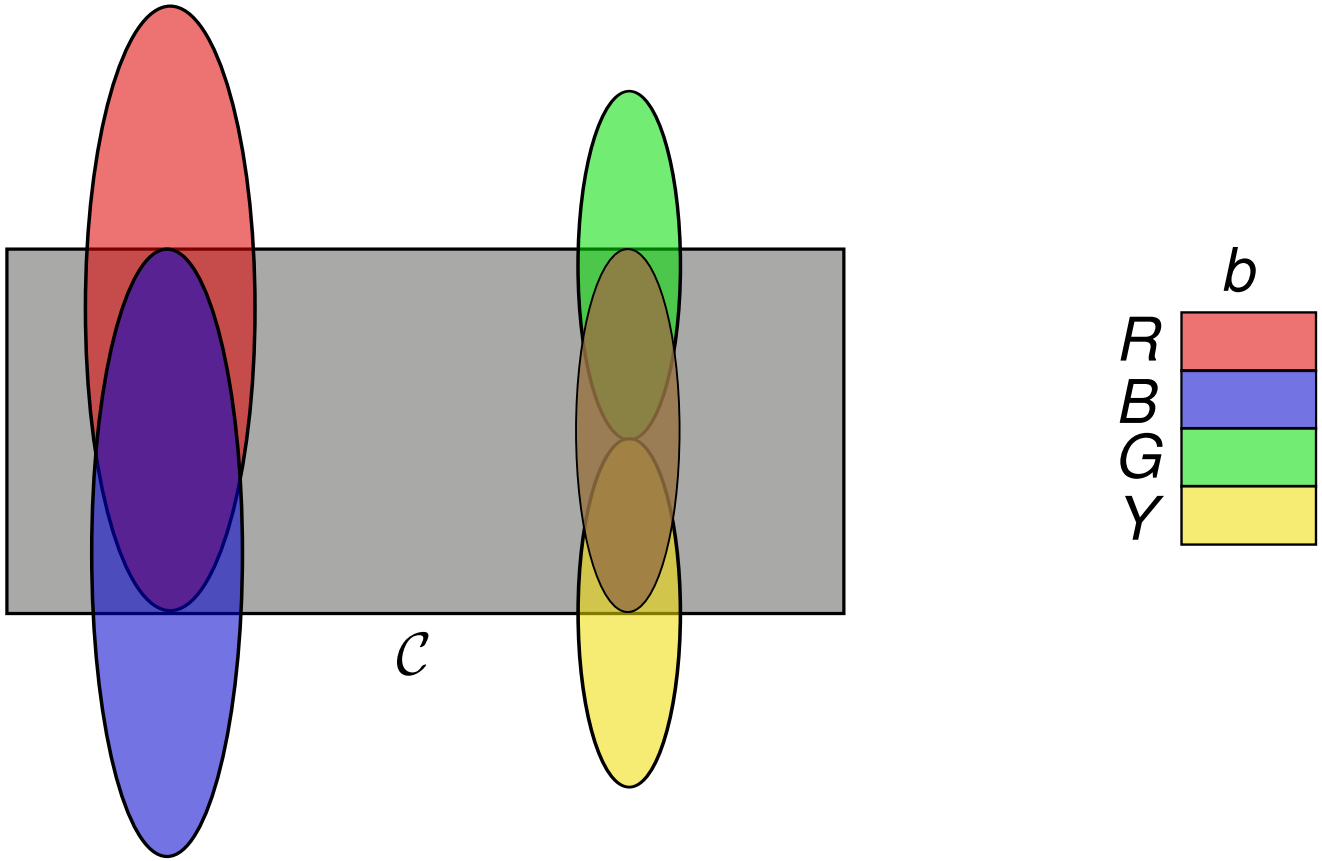
- At this stage, the red set is on the other side of the hyperplane;
- Blackwell strategy recommend to play R at stage 4.

Blackwell strategy to the farthest point



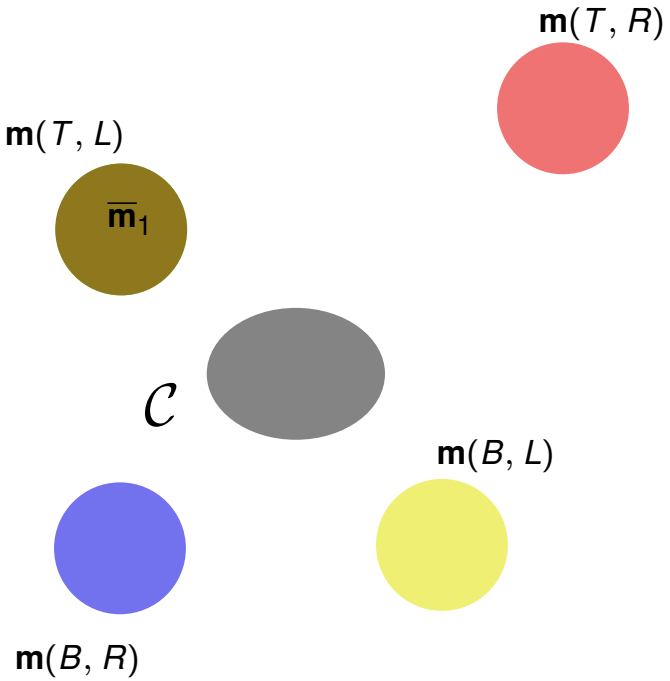
- The algorithm can oscillate indefinitely between R and B;
- $\overline{\mathbf{m}}_T$ does not converge to $\mathcal{C}$.... But on the other hand...

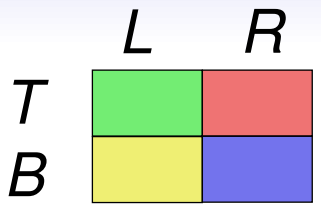
Blackwell strategy to the farthest point



- with G and Y chosen i.i.d. with proba. $1/2$, $\overline{\mathbf{m}}_T$ converges to $\mathcal{C}$;
- Must not focus on payoffs but on sequence of actions.

Main insights of the result



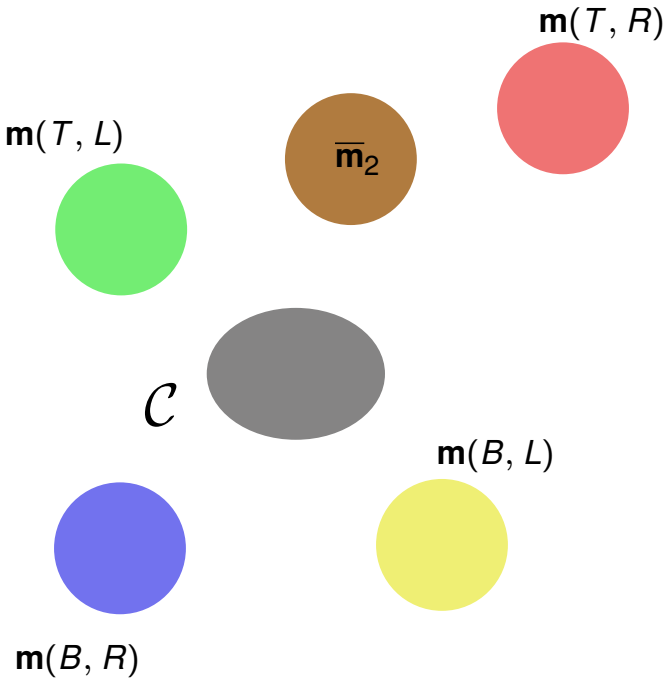


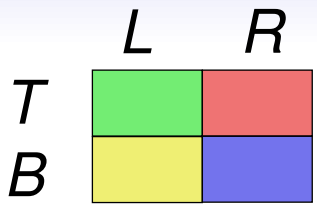
Sequence played:

(T, L)

- $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the convex $\mathcal{C}$ is approachable.
- At the first stage, T and L were chosen; $\overline{\mathbf{m}}_1 = \mathbf{m}(T, L)$ in brown.

Main insights of the result



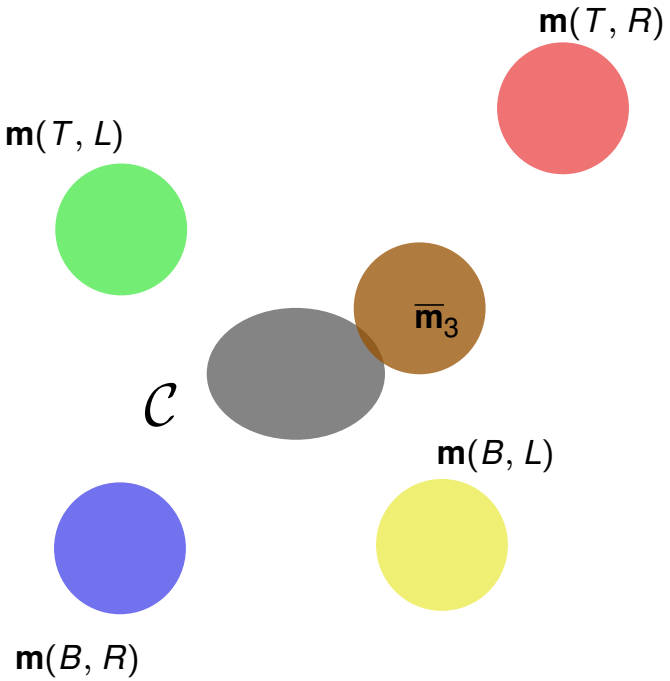


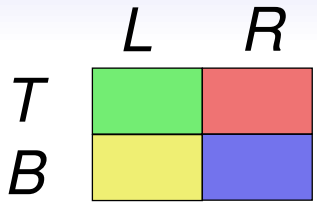
Sequence played:

$(T, L); (T, R)$

- $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the convex $\mathcal{C}$ is approachable.
- Second stage: (T, R) is played; $\overline{\mathbf{m}}_2 = \frac{1}{2}\mathbf{m}(T, L) + \frac{1}{2}\mathbf{m}(T, R)$.

Main insights of the result



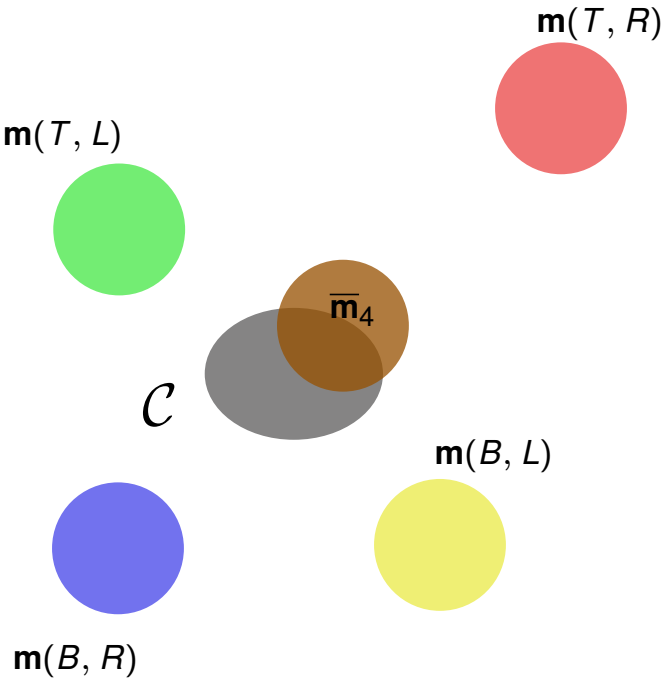


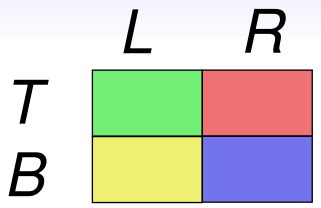
Sequence played:

$(T, L); (T, R); (B, L)$

- $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the convex $\mathcal{C}$ is approachable.
- (B, L) is played; $\overline{\mathbf{m}}_3 = \frac{1}{3}\mathbf{m}(T, L) + \frac{1}{3}\mathbf{m}(T, R) + \frac{1}{3}\mathbf{m}(B, L)$.

Main insights of the result



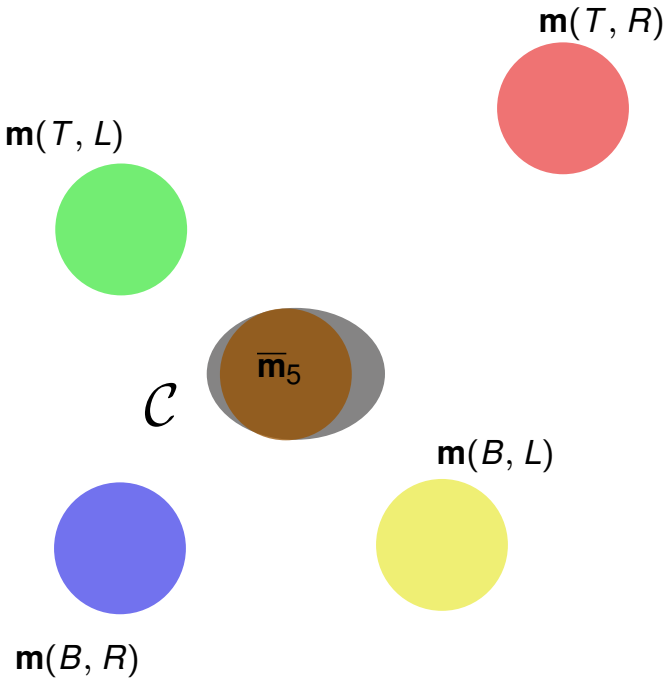


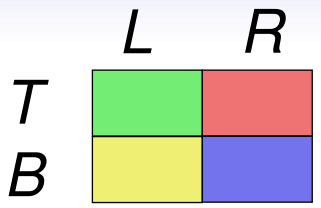
Sequence played:

$(T, L); (T, R); (B, L); (B, R)$

- $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the convex $\mathcal{C}$ is approachable.
- $\overline{\mathbf{m}}_4 = \frac{1}{4}\mathbf{m}(T, L) + \frac{1}{4}\mathbf{m}(T, R) + \frac{1}{4}\mathbf{m}(B, L) + \frac{1}{4}\mathbf{m}(B, R)$.

Main insights of the result



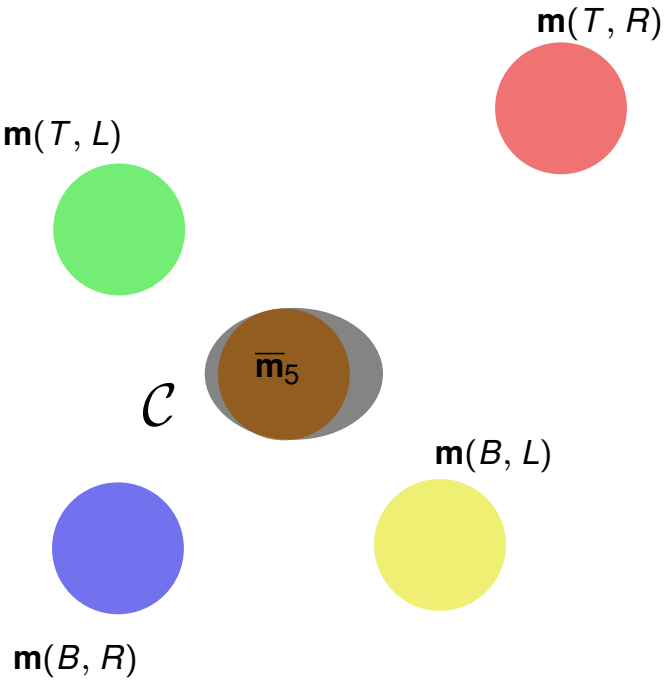


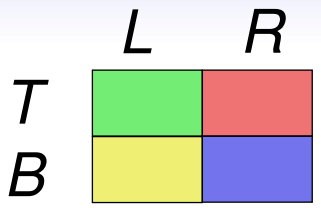
Sequence played:

$(T, L); (T, R); (B, L); (B, R); (B, R)\ldots$

- $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the convex $\mathcal{C}$ is approachable.
- $\overline{\mathbf{m}}_5 = \frac{1}{5}\mathbf{m}(T, L) + \frac{1}{5}\mathbf{m}(T, R) + \frac{1}{5}\mathbf{m}(B, L) + \frac{2}{5}\mathbf{m}(B, R)$.

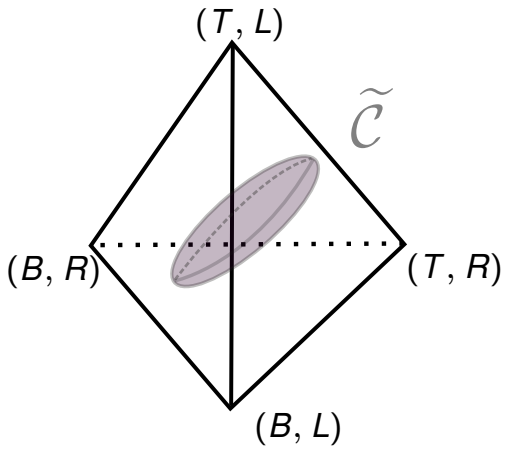
Main insights of the result





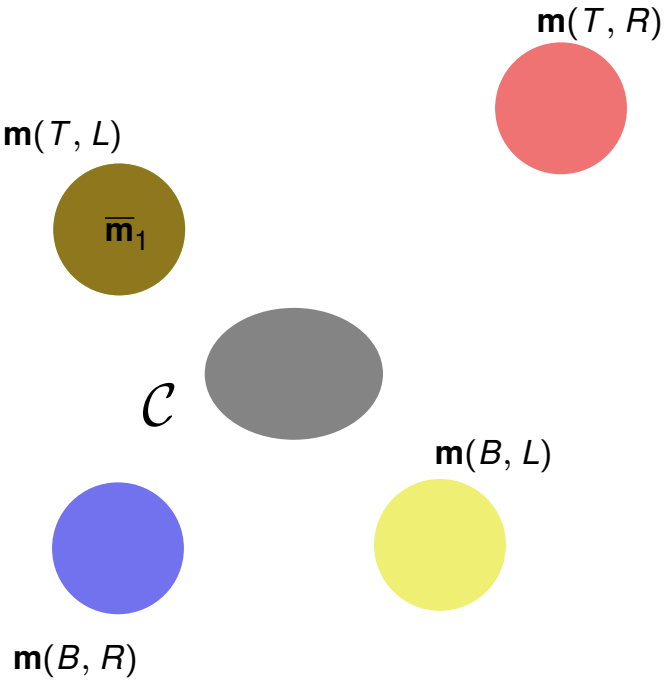
Sequence played:

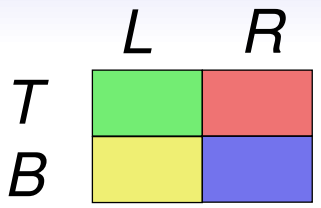
$(T, L); (T, R); (B, L); (B, R); (B, R)$



- $\frac{1}{5}\mathbf{m}(T, L) + \frac{1}{5}\mathbf{m}(T, R) + \frac{1}{5}\mathbf{m}(B, L) + \frac{2}{5}\mathbf{m}(B, R) \subset \mathcal{C}$
- $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5}) \in \widetilde{\mathcal{C}} = \{q \in \Delta(\mathcal{A} \times \mathcal{B}) \; ; \; \mathbb{E}_q [\mathbf{m}(a, b)] \subset \mathcal{C}\}$.

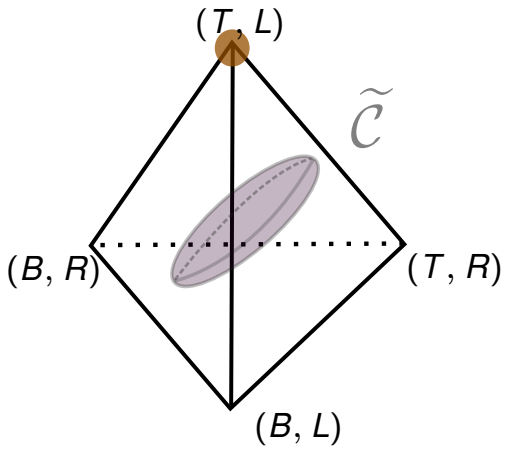
Main insights of the result





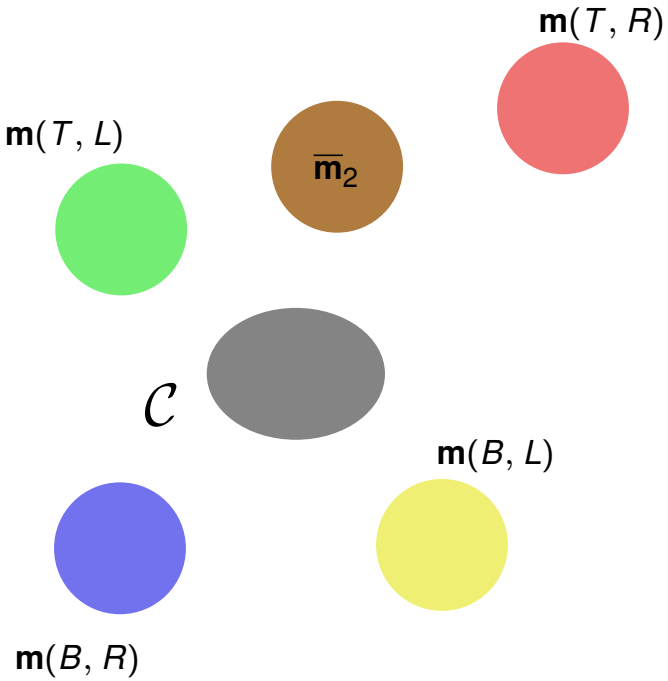
Sequence played:

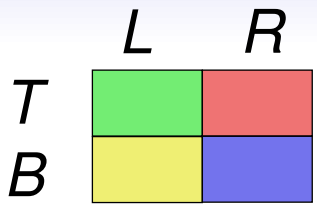
(T, L)



- Approach $\mathcal{C} \subset \mathbb{R}^d$ is equivalent to approach $\widetilde{\mathcal{C}} \subset \Delta(\mathcal{A} \times \mathcal{B})$;
- *Abstract payoff* of first stage: $q_1 = \delta_{(T,L)}$.

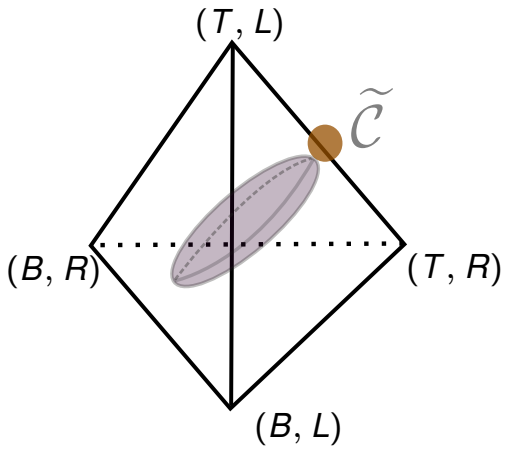
Main insights of the result





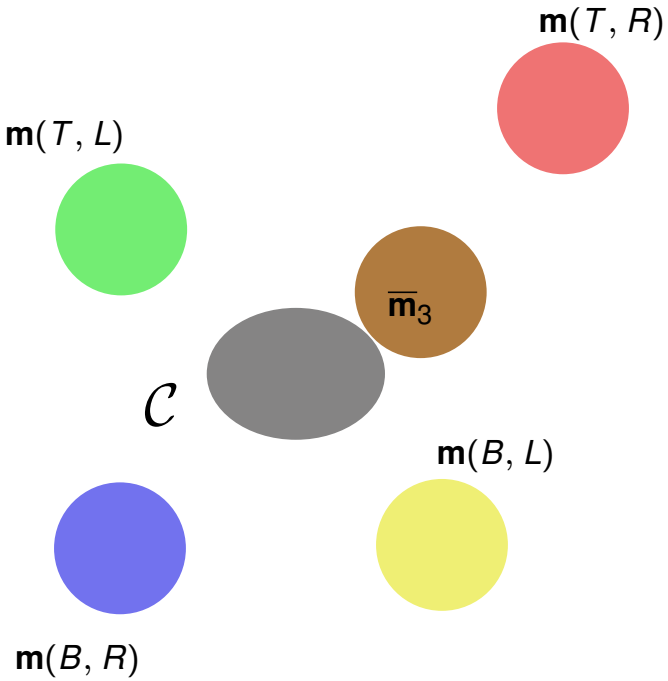
Sequence played:

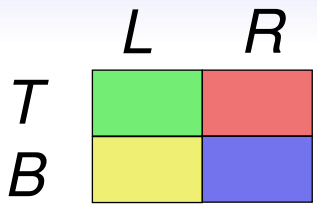
$(T, L); (T, R)$



- *Abstract payoff* of second stage: $q_2 = \delta_{(T,R)}$;
- Empirical distribution of actions: $\overline{q}_2 = \frac{1}{2}\delta_{(T,L)} + \frac{1}{2}\delta_{(T,R)}$.

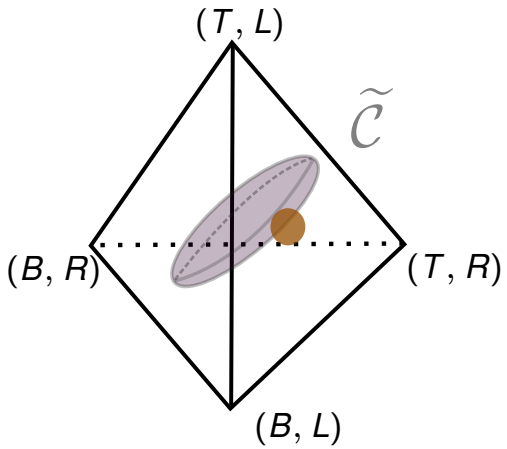
Main insights of the result





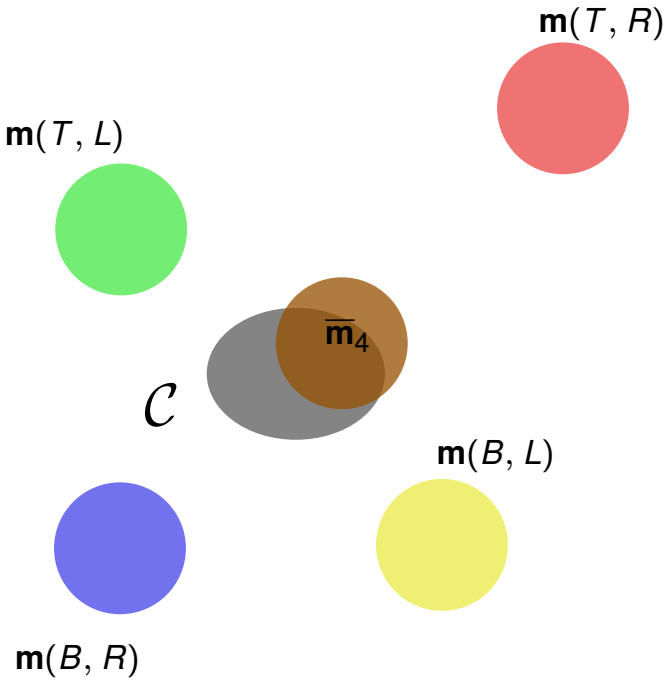
Sequence played:

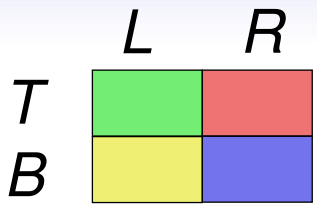
$(T, L); (T, R); (B, L)$



- *Abstract payoff* of third stage: $q_3 = \delta_{(B,L)}$;
- Empirical distribution of actions: $\overline{q}_3 = \frac{1}{3}\delta_{(T,L)} + \frac{1}{3}\delta_{(T,R)} + \frac{1}{3}\delta_{(B,L)}$.

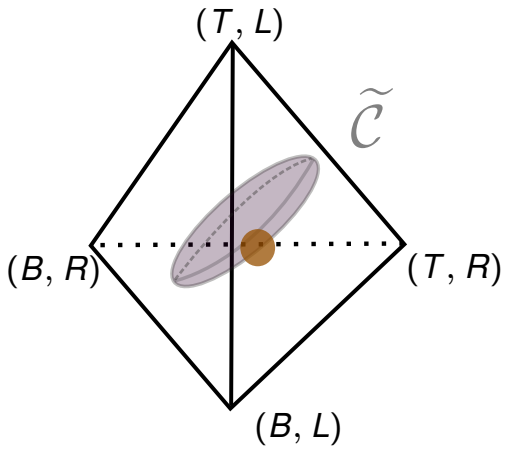
Main insights of the result





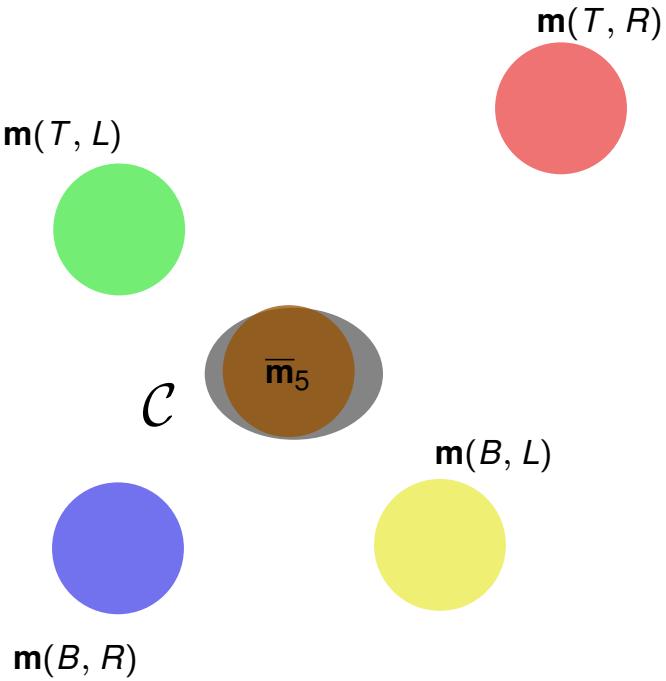
Sequence played:

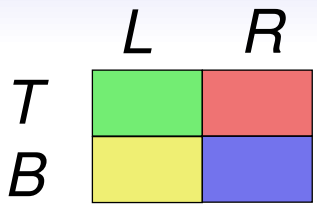
$(T, L); (T, R); (B, L); (B, R)$



- *Abstract payoff* of fourth stage: $q_4 = \delta_{(B,R)}$;
- $\overline{q}_4 = \frac{1}{4}\delta_{(T,L)} + \frac{1}{4}\delta_{(T,R)} + \frac{1}{4}\delta_{(B,L)} + \frac{1}{4}\delta_{(B,R)}$.

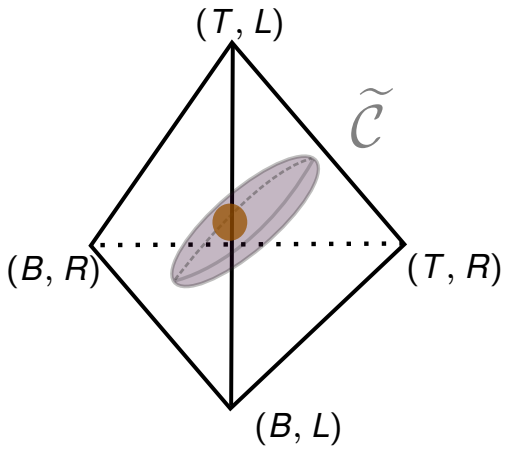
Main insights of the result





Sequence played:

$(T, L); (T, R); (B, L); (B, R); (B, R)$



- $\overline{q}_5 = \frac{1}{5}\delta_{(T,L)} + \frac{1}{5}\delta_{(T,R)} + \frac{1}{5}\delta_{(B,L)} + \frac{2}{5}\delta_{(B,R)}$.
- $\overline{q}_5 \in \widetilde{\mathcal{C}} = \{q \in \Delta(\mathcal{A} \times \mathcal{B}) \; ; \; \mathbb{E}_q[\mathbf{m}(a, b)] \subset \mathcal{C}\}$.

Reduction to full monitoring

	Uncertainties		With full monit.
Actions:	$\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$		$\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$
Payoff:	$\mathbf{m}(x,y) \subset \mathbb{R}^d$		$x \otimes y \in \Delta(\mathcal{A} \times \mathcal{B})$
Target:	$\mathcal{C} \subset \mathbb{R}^d$		$\widetilde{\mathcal{C}} = \{q \in \Delta(\mathcal{A} \times \mathcal{B}) \; ; \; \mathbb{E}_q[\mathbf{m}(a,b)] \subset \mathcal{C}\}$
Result:	$\mathcal{C}$ robust-appr.	iff	$\widetilde{\mathcal{C}}$ approachable.

[Blackwell]'s condition: $\widetilde{\mathcal{C}}$ is approachable iff

$$\forall y \in \Delta(\mathcal{B}),\ \exists x \in \Delta(\mathcal{A}) \text{ s.t. } x \otimes y \in \widetilde{\mathcal{C}}.$$

and $x \otimes y \in \widetilde{\mathcal{C}}$ iff $\mathbb{E}_{x \otimes y}[\mathbf{m}(a,b)] = \mathbf{m}(x,y) \subset \mathcal{C}$.

Condition for robust approachability [M.,P.,S.]

$\mathcal{C}$ is robust-appr. iff $\forall y \in \Delta(\mathcal{B}),\ \exists x \in \Delta(\mathcal{A})$ s.t. $\mathbf{m}(x,y) \subset \mathcal{C}$.

[M.,P.,S.] Rates of convergence

$\mathcal{C}$ is robust approach. iff $\forall y \in \Delta(\mathcal{B})$, $\exists x \in \Delta(\mathcal{A})$ s.t. $\mathbf{m}(x, y) \subset \mathcal{C}$.

$$\sup_{z \in \overline{\mathbf{m}}_T} \inf_{c \in \mathcal{C}} \|z - c\| \leq \square \frac{1}{\sqrt{T}}$$

$\mathcal{C}$ is approachable with partial monitoring iff

$$\forall y \in \Delta(\mathcal{B}), \exists x \in \Delta(\mathcal{A}) \text{ s.t. } \{r(x, y'); \ \mathbf{s}(y') = \mathbf{s}(y)\} \subset \mathcal{C}.$$

$$\inf_{c \in \mathcal{C}} \|\bar{\mathbf{r}}_T - c\| \leq \diamond \frac{1}{T^{\frac{1}{5}}}$$

- Rates independent of actions and signal sets, dimension of losses.
- $T^{-1/5}$: exploration (and policy constant by blocks) to estimate $\mathbf{s}(y)$.

Regret with partial monitoring

- Same model: actions sets $\mathcal{A}$ and $\mathcal{B}$; **real payoffs** $r(a,b) \in \mathbb{R}$;
- External regret: difference between the payoffs DM got and what he would have got **for sure** if he had always played the same $x \in \Delta(\mathcal{A})$

$$R_T^{\text{ext}} = \max_{x \in \Delta(\mathcal{A})} \rho(x, \overline{y}_T) - \frac{1}{T}\sum_{t=1}^{T} r(a_t, b_t), \text{ with } \overline{y}_T = \frac{1}{T}\sum_{t=1}^{T} \delta_{b_t}$$

and $\rho(x,y) = \min\{r(x,y'); \ \mathbf{s}(y') = \mathbf{s}(y)\}$.

Reduction to approachability: ([Blackwell]'s argument in full monit.)

$$\underline{r}(a,b) = \Big[r(a,b); \delta_b\Big] \text{ and } \mathcal{C} = \left\{(z,y); \max_{x \in \Delta(\mathcal{A})} \rho(x,y) \leq z\right\}$$

If $\overline{\underline{r}}_T = \Big[\overline{r}_T; \overline{y}_T\Big] \in \mathcal{C}$ then $R_T^{\text{ext}} = \max_{x \in \Delta(\mathcal{A})} \rho(x, \overline{y}_T) - \overline{r}_T \leq 0$, so

[Lugosi, Mannor, Stoltz] There exists a strategy such that $R_T^{\text{ext}} \leq \diamond \frac{1}{T^{1/5}}$.