What Preferences Can—and Cannot—Predict in Multi-Agent Online Learning

Omar Abbadi[12], Rida Laraki[1] & Panayotis Mertikopoulos[2]

[1] Moroccan Center for Game Theory, CMSIS, UM6P Rabat
[2] Université Grenoble Alpes, Inria, CNRS

Multi-Agent Online Learning

Learning in games

for each time t and player i:

Choose strategy $x_i(t)$	play
Receive payoff $v_i(x(t))$	depends on strategies of all players
Update strategy	adapt

Multi-Agent Online Learning

Learning in games	
for each time t and player i:	
Choose strategy $x_i(t)$	play
Receive payoff $v_i(x(t))$	depends on strategies of all players
Update strategy	adapt

What's the relation between **player preferences** *and the* **emerging patterns of play***?*

Example 1: **Discoordination**

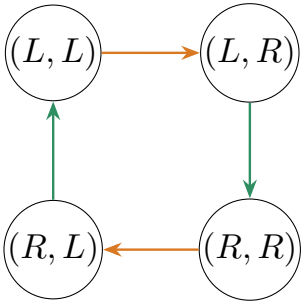
- **Robot 1** and **Robot 2** independently choose between L (left) and R (right).

Example 1: **Discoordination**

- **Robot 1** and **Robot 2** independently choose between L (left) and R (right).
- **Robot 2** wants to pass through, **Robot 1** wants to block it.

Example 1: **Discoordination**

- **Robot 1** and **Robot 2** independently choose between L (left) and R (right).
- **Robot 2** wants to pass through, **Robot 1** wants to block it.

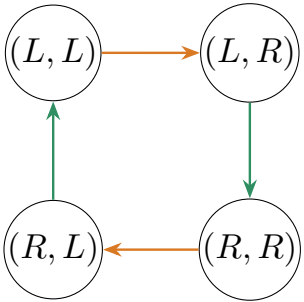


Vertices = outcomes

Edges = preferences

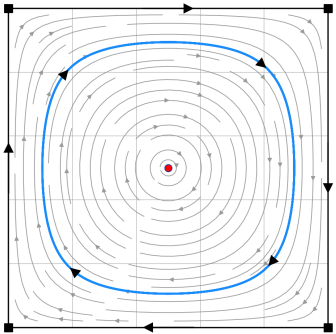
Example 1: **Discoordination**

- **Robot 1** and **Robot 2** independently choose between L (left) and R (right).
- **Robot 2** wants to pass through, **Robot 1** wants to block it.



Vertices = outcomes

Edges = preferences

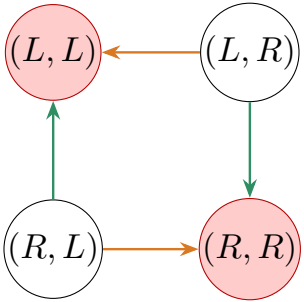


x-**axis** = $\mathbb{P}$(**Robot 1** chooses R)

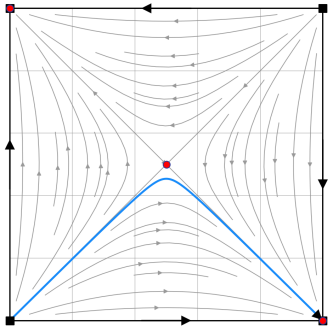
y-**axis** = $\mathbb{P}$(**Robot 2** chooses L)

Example 2: **Coordination**

- **Robot 1** and **Robot 2** want to meet at the same spot.
- **Red vertices** are **preferentially stable**: no edge leaves them.

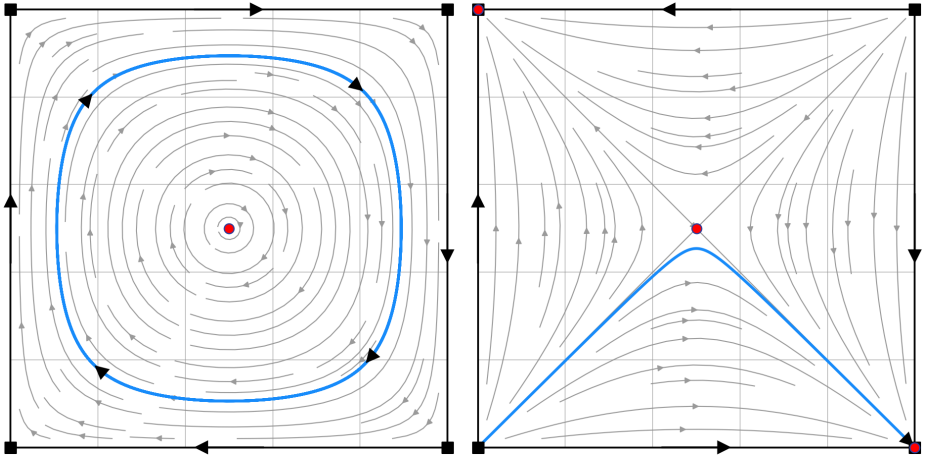


Preference graph



Dynamics

Are Preferences Enough?



Can we predict **patterns of play** *that* **emerge in the long-run** *from* **preferences alone***?*

Main Results

- **Dynamic stability** *implies* **preferential stability**.

Main Results

- **Dynamic stability** *implies* **preferential stability**.
- *Surprisingly*, the converse is **false** in general.

Main Results

- **Dynamic stability** *implies* **preferential stability**.
- *Surprisingly*, the converse is **false** in general.
- We recover **dynamic stability** via a **cardinal strengthening** of **preferential stability**.

Finite Games

- **Player set** $\mathcal{N} = \{1, \ldots, N\}$
- For each player i, **finite action set** $\mathcal{A}_i$
- For each player i, **payoff**

$$u_i : \mathcal{A} := \prod_{j \in \mathcal{N}} \mathcal{A}_j \to \mathbb{R}$$

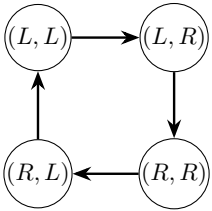
Preference Graph

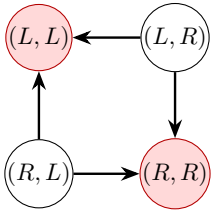
Preference graph

- **Vertices**: pure profiles $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathcal{A}$
- **Edges**: $\alpha \to \beta$ if they differ only in one player i's action and $u_i(\beta) \geq u_i(\alpha)$

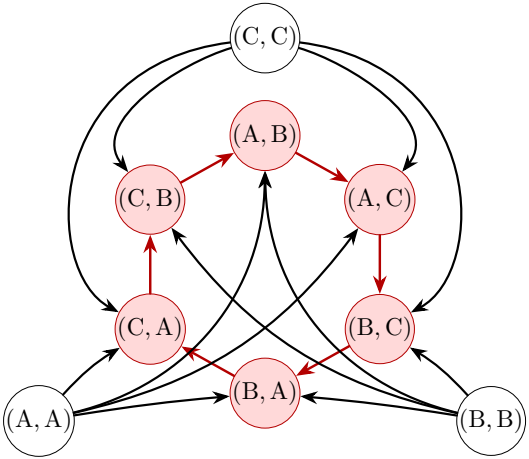
Thus, the graph records **better replies**.

A subset of pure profiles $\mathcal{H} \subseteq \mathcal{A}$ is **closed under better replies** if **no preference leaves** $\mathcal{H}$.





Example: **Shapley's Game**



	A	B	C
A	$(0,0)$	$(2,1)$	$(1,2)$
B	$(1,2)$	$(0,0)$	$(2,1)$
C	$(2,1)$	$(1,2)$	$(0,0)$

2-player 3-action game

The **red hexagon** is **closed under better replies**

Strategy Space

- Player i's mixed strategy:

$$x_i \in \mathcal{X}_i := \Delta(\mathcal{A}_i), \qquad x_{i\alpha_i} = \mathbb{P}(i \text{ plays } \alpha_i)$$

- Mixed profile:

$$x = (x_1, \ldots, x_N) \in \mathcal{X} := \prod_{i \in \mathcal{N}} \mathcal{X}_i$$

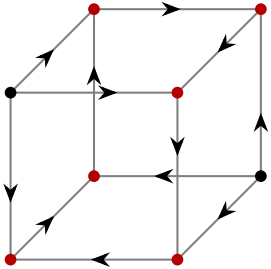
- Probability of α under x:

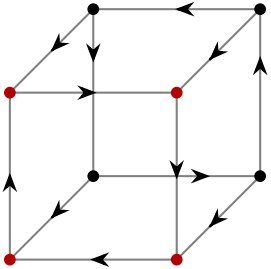
$$x_\alpha = \prod_{i \in \mathcal{N}} x_{i\alpha_i}$$

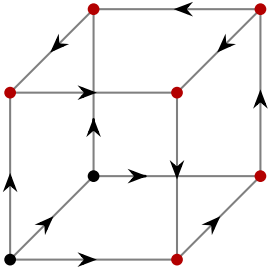
- Expected payoff:

$$u_i(x) = \sum_{\alpha \in \mathcal{A}} x_\alpha u_i(\alpha)$$

Spans & Skeletons



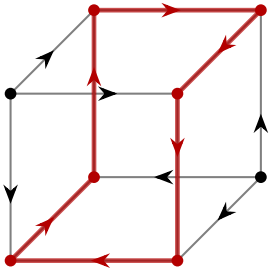


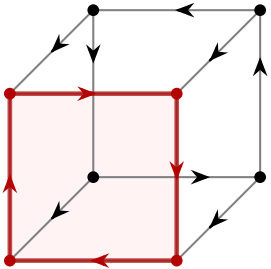


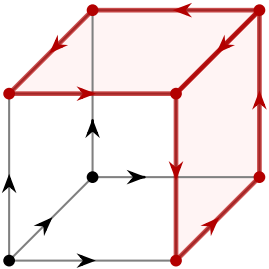
Spans & Skeletons

For $\mathcal{H} \subseteq \mathcal{A}$, its **span** is the set of mixed profiles whose support is in $\mathcal{H}$:

$$\text{span}(\mathcal{H}) := \{x \in \mathcal{X} : x_\alpha = 0 \text{ for every } \alpha \notin \mathcal{H}\}.$$



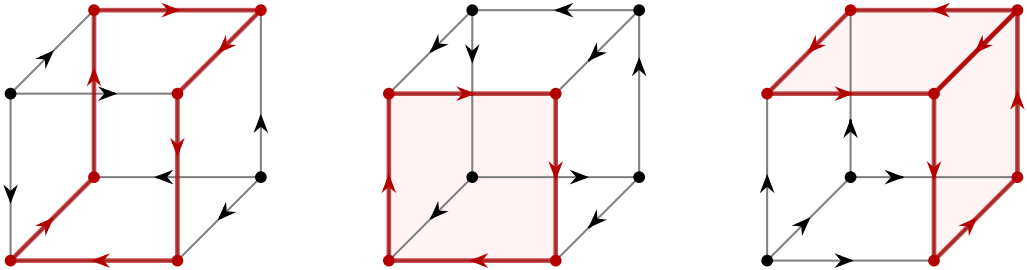




Spans & Skeletons

For $\mathcal{H} \subseteq \mathcal{A}$, its **span** is the set of mixed profiles whose support is in $\mathcal{H}$:

$$\operatorname{span}(\mathcal{H}) := \{x \in \mathcal{X} : x_\alpha = 0 \text{ for every } \alpha \notin \mathcal{H}\}.$$



For $S \subseteq \mathcal{X}$, its **skeleton** is the set of *pure profiles contained in* S.

Exponential Weights Dynamics

At time t, each action α_i has a **score** $y_{i\alpha_i}(t)$:

$$u_i\big(\alpha_i, x_{-i}(t)\big) \longrightarrow \dot{y}_{i\alpha_i}(t) = u_i\big(\alpha_i, x_{-i}(t)\big) \longrightarrow x_{i\alpha_i}(t) = \frac{\exp\big(y_{i\alpha_i}(t)\big)}{\sum_{\beta_i} \exp\big(y_{i\beta_i}(t)\big)}$$

"observe payoff" — "follow payoff gradient" — "update strategy"

higher payoff $\Longrightarrow$ **larger score** $\Longrightarrow$ **played more often**

In the paper, we consider the more general *Follow The Regularized Leader* dynamics.

Exponential Weights Dynamics

At time t, each action α_i has a **score** $y_{i\alpha_i}(t)$:

$$u_i\big(\alpha_i, x_{-i}(t)\big) \longrightarrow \dot{y}_{i\alpha_i}(t) = u_i\big(\alpha_i, x_{-i}(t)\big) \longrightarrow x_{i\alpha_i}(t) = \frac{\exp\big(y_{i\alpha_i}(t)\big)}{\sum_{\beta_i} \exp\big(y_{i\beta_i}(t)\big)}$$

"observe payoff" — "follow payoff gradient" — "update strategy"

higher payoff $\Longrightarrow$ **larger score** $\Longrightarrow$ **played more often**

Equivalently,

$$\dot{x}_{i\alpha_i} = x_{i\alpha_i}\Big(u_i(\alpha_i, x_{-i}) - u_i(x)\Big)$$

This is the well-known **replicator dynamics**.

In the paper, we consider the more general *Follow The Regularized Leader* dynamics.

Dynamic Stability & Attraction

Definitions

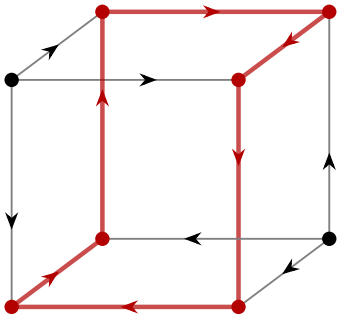
For a subset of mixed profiles $S \subseteq \mathcal{X}$,

- S is **stable** if: $x(0)$ near $S \Longrightarrow x(t)$ remains near S for all $t \geq 0$.
- S is **attracting** if: $x(0)$ near $S \Longrightarrow x(t)$ converges to S.
- S is **asymptotically stable** if it is both stable and attracting.

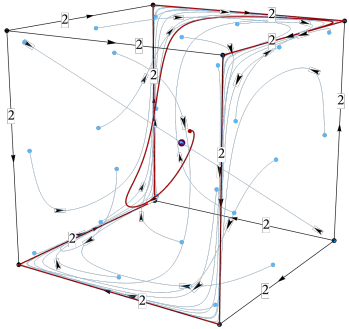
Dynamic Stability ⇒ Preferential Stability

Theorem 1

If S is **asymptotically stable**, *its skeleton is* **closed under better replies**.



Preferentially stable

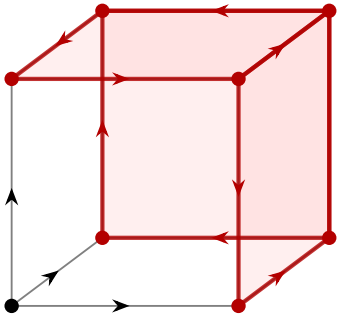


Dynamically stable

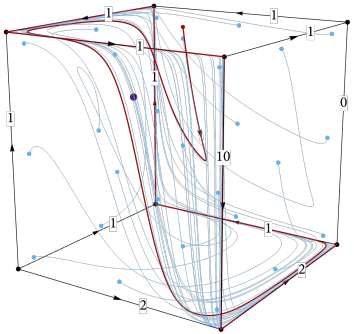
Preferential Stability $\not\Rightarrow$ Dynamic Stability

Theorem 2

If $\mathcal{H}$ is **closed under better replies**, *its span* **need not be asymptotically stable**.



Preferentially stable



Dynamically unstable

Can one guarantee **dynamic stability** *by adding* **cardinal payoff information**, *rather than only* **ordinal preference information**?

Payoff Flux & Leaklessness

The **payoff flux** at $\beta \in \mathcal{A}$ toward $\alpha \in \mathcal{A}$ is

$$\Phi(\beta, \alpha) := \sum_{i \in \mathcal{N}} \big(u_i(\alpha_i, \beta_{-i}) - u_i(\beta)\big)$$

$\Phi(\beta, \alpha)$ is the **aggregate of all players' unilateral gains** from β toward α.

Payoff Flux & Leaklessness

The **payoff flux** at $\beta \in \mathcal{A}$ toward $\alpha \in \mathcal{A}$ is

$$\Phi(\beta, \alpha) := \sum_{i \in \mathcal{N}} \big(u_i(\alpha_i, \beta_{-i}) - u_i(\beta)\big)$$

$\Phi(\beta, \alpha)$ is the **aggregate of all players' unilateral gains** from β toward α.

Definition

Given $\mathcal{H} \subseteq \mathcal{A}$:

- $\mathcal{H}$ is **leakless** if for every $\beta \in \mathcal{H}$ and $\alpha \notin \mathcal{H}$,

$$\Phi(\beta, \alpha) \leq 0.$$

- $\mathcal{H}$ is **strictly leakless** if the inequality is strict for every $\beta \in \mathcal{H}$ and $\alpha \notin \mathcal{H}$.

Basic Properties

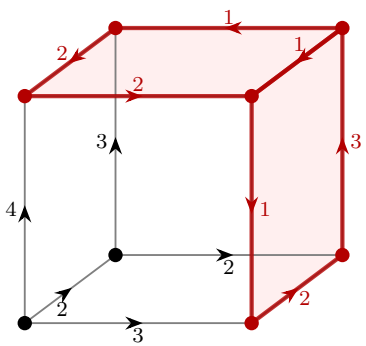
- **Strict leaklessness** *implies* **closedness under better replies**.
- For a single pure profile α

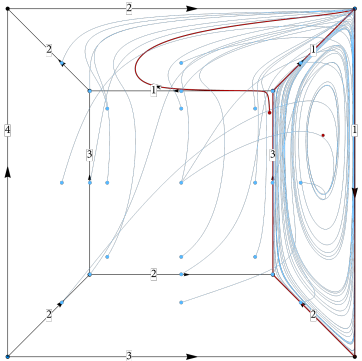
$$\alpha \text{ is } \textbf{(strictly) leakless} \iff \alpha \text{ is a } \textbf{(strict) Nash equilibrium}$$

Restoring Dynamic Stability

Theorem 3

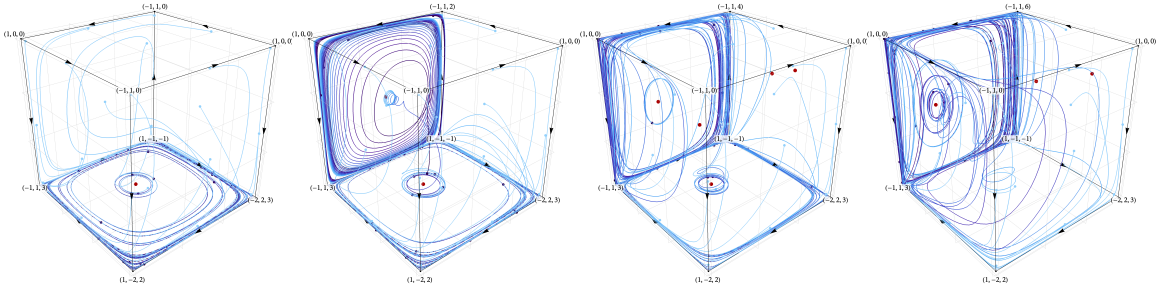
- *If* $\mathcal{H}$ *is* **strictly leakless**, *then* $\mathrm{span}(\mathcal{H})$ *is* **asymptotically stable**.
- *If* $\mathcal{H}$ *is* **leakless** *and* **closed under better replies**, *then* $\mathrm{span}(\mathcal{H})$ *is* **asymptotically stable**.





Is this the end of the road?

- Fully characterizing **asymptotically stable** **spans**...
- Understanding possible **limit sets**...



Thank you! Come see our poster (HALL A Poster #4503) :-)

What Preferences Can—and Cannot—Predict in Multi-Agent Online Learning

Omar Abbadi [1 2] Rida Laraki [1] Panayotis Mertikopoulos [2]

[1]Moroccan Center for Game Theory, CMSIS, UM6P [2]Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG

ICML Thirty-second International Conference on Machine Learning

Motivation

The long-run behavior of multi-agent regularized learning is often **setwise**: convergence to *attractors* instead of equilibria.

These stable outcomes are fundamentally linked to the structure of the *underlying game*, and in particular the *players' preferences*.

To what extent do player preferences constrain their long-run behavior?

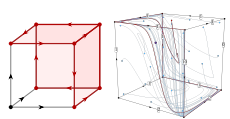
Finite Games

- *Players:* $i \in \mathcal{N} := \{1, \dots, N\}$
- *Pure strategies:* $\alpha_i \in \mathcal{A}_i := \{1, \dots, A_i\}$, $\alpha \in \mathcal{A} := \prod_i \mathcal{A}_i$, $\alpha_{-i} = (\alpha_j)_{j \neq i}$
- *Mixed strategies:* $x_i \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$, $x \in \mathcal{X} := \prod_i \mathcal{X}_i$, $x_\alpha := \prod_i x_{i\alpha_i}$
- *Payoffs:* $u_i(x) = \mathbb{E}_{\alpha_i \sim x}[u_i(\alpha)] = \sum_\alpha x_\alpha u_i(\alpha) \in \mathbb{R}$
- *Payoff vectors:* $v_i(x) := \nabla_i u_i(x) = (u_i(\alpha_i, x_{-i}))_{\alpha_i \in \mathcal{A}_i} \in \mathbb{R}^{\mathcal{A}_i}$
- *Nash equilibrium:* $u_i(x^*) \geq u_i(x_i, x^*_{-i})$, $\forall i, \forall x_i \neq x_i^*$.
  Strict Nash := inequalities are strict ⇒ Pure.

Regularized Learning

*« Players play a **regularized best response** to their cumulative payoff vector »*

- *Follow The Regularized Leader (FTRL)*
$$y_i(t) = y_i(0) + \int_0^t v_i(x(s))ds$$
$$x_i(t) = Q_i(y_i(t))$$
- *Regularized best response*
$$Q_i(y_i) := \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - h_i(x_i)\}$$
  - ↠ *Entropic regularizer / Exponential Weights:*
  $$h_i(x_i) = \sum_\alpha x_{i\alpha_i} \log x_{i\alpha_i} \quad Q_i(y_i) = \exp(y_i) / \sum_\alpha \exp(y_{i\alpha_i})$$
  - ↠ *Euclidean regularizer / Projected Gradient:*
  $$h_i(x_i) = \frac{1}{2}\|x_i\|_{l^2}^2 \quad Q_i(y_i) = \text{proj}_{\mathcal{X}}(y_i)$$



When preferences and behavior do not align.

When preferences and behavior do align.

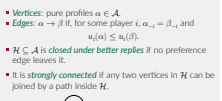
Main Takeaways

1. Every set of outcomes which is *dynamically stable* is also *preferentially stable*.
2. The converse is **false**: despite all players individually *preferring a set of outcomes* they might collectively *diverge from it*.
3. To remedy this, we introduce a cardinal, *payoff-based* condition which guarantees *dynamical stability*.

Preference Graph

« Vertices are outcomes, edges encode individual preferences »

- *Vertices:* pure profiles $\alpha \in \mathcal{A}$.
- *Edges:* $\alpha \to \beta$ if, for some player i, $\alpha_{-i} = \beta_{-i}$ and $u_i(\alpha) \leq u_i(\beta)$.
- $\mathcal{H} \subseteq \mathcal{A}$ is *closed under better replies* if no preference edge leaves it.
- It is *strongly connected* if any two vertices in $\mathcal{H}$ can be joined by a path inside $\mathcal{H}$.



	A	B	C
A	(0, 0)	(2, 1)	(1, 2)
B	(1, 2)	(0, 0)	(2, 1)
C	(2, 1)	(1, 2)	(0, 0)

Shapley's game: The red hexagon is both closed under better replies and strongly connected.

↠ *Weakly acyclic games*: From every vertex, there exists a path to a (pure) Nash equilibrium.

Dynamic Stability

- S is *asymptotically stable* if
$$x(0) \text{ sufficiently near } S \implies \begin{cases} x(t) \text{ converges to } S \\ x(t) \text{ remains near } S \end{cases}$$
- S is an *attractor* if asymptotically stable *and* invariant.

Skeletons, Spans & Subgames

- *Skeleton* of $S \subseteq \mathcal{X}$ is $\text{skl}(S) := S \cap \mathcal{A}$, i.e. set of pure profiles it contains.
- *Span* of $\mathcal{H} \subseteq \mathcal{A}$ is
$$\text{span}(\mathcal{H}) := \{x \in \mathcal{X} \mid \forall \alpha \notin \mathcal{H},\ x_\alpha = 0\}.$$
It is the union of faces for which all vertices are in $\mathcal{H}$.
- *Subgame* is $\mathcal{B} \subseteq \mathcal{A}$ such that
$$\mathcal{B} = \prod_i \mathcal{B}_i, \quad \mathcal{B}_i \subseteq \mathcal{A}_i.$$
Its span is a *face*: $\text{span}(\mathcal{B}) = \prod_i \Delta(\mathcal{B}_i)$.

Figure 2. Three 2 × 2 × 2 games; red region is the span of the set of red vertices. Middle one is a subgame / face.

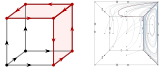
Payoff Flux & Leaklessness

*« The **aggregate unilateral gain** from one pure profile toward another »*

- *Payoff flux* at $\beta \in \mathcal{A}$ toward $\alpha \in \mathcal{A}$:
$$\Phi(\beta, \alpha) = \sum_i (u_i(\alpha_i, \beta_{-i}) - u_i(\beta)).$$

A set $\mathcal{H} \subseteq \mathcal{A}$ is

- *Leakless* if, for all $\beta \in \mathcal{H}$ and $\alpha \notin \mathcal{H}$,
$$\Phi(\beta, \alpha) \leq 0.$$
- *Strictly leakless* if those inequalities are strict.



Formal Statements

Dynamic Stability Implies Preferential Stability

1. If S is *asymptotically stable*, then its *skeleton* is *closed under better replies*.
2. If S is an *attractor* and $\mathcal{H} \subseteq S \cap \mathcal{A}$ is *strongly connected*, then $\text{span}(\mathcal{H}) \subseteq S$.
   - ↠ **Consequence 1:** If the whole preference graph is *strongly connected*, the dynamics admit *no proper attractor*.
   - ↠ **Consequence 2:** for *weakly acyclic games* with no ties, the (minimal) *attractors* are exactly the *strict Nash equilibria*.

Preferences Are *Not* Enough in General

1. A *subgame* is *closed under better replies* *if and only if* its *span* is *asymptotically stable*.
2. This equivalence *breaks down* for *general pure profile sets*: there exists a 2 × 2 × 2 game which admits a *closed under better replies* set whose *span* is *not stable*.

Restoring Dynamic Stability

1. If $\mathcal{H} \subseteq \mathcal{A}$ is *strictly leakless*, then $\text{span}(\mathcal{H})$ is an *attractor*.
2. If $\mathcal{H} \subseteq \mathcal{A}$ is *leakless* *and* *closed under better replies*, then $\text{span}(\mathcal{H})$ is an *attractor* for the *exponential weights dynamics*.