Adaptive Heuristics

Game Theory Presentation

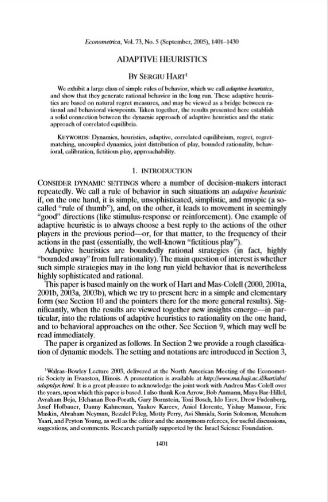
Théo Delemazure

Ecole normale supérieure, PSL University

March 23, 2020

Paper presentation

Title	*Adaptive Heuristics*
Author	*Sergiu Hart* (Former president of the *Game Theory Society*)
Year	*2005*
Journal	*Econometrica*
Citations	279 to this day



Econometrica, Vol. 73, No. 5 (September, 2005), 1401–1430

ADAPTIVE HEURISTICS

BY SERGIU HART[1]

We exhibit a large class of simple rules of behavior, which we call *adaptive heuristics*, and show that they generate rational behavior in the long run. These adaptive heuristics are based on natural regret measures, and may be viewed as a bridge between rational and behavioral viewpoints. Taken together, the results presented here establish a solid connection between the dynamic approach of adaptive heuristics and the static approach of correlated equilibria.

KEYWORDS: Dynamics, heuristics, adaptive, correlated equilibrium, regret, regret-matching, uncoupled dynamics, joint distribution of play, bounded rationality, behavioral, calibration, fictitious play, approachability.

1. INTRODUCTION

CONSIDER DYNAMIC SETTINGS where a number of decision-makers interact repeatedly. We call a rule of behavior in such situations an *adaptive heuristic* if, on the one hand, it is simple, unsophisticated, simplistic, and myopic (a so-called "rule of thumb"), and, on the other, it leads to movement in seemingly "good" directions (like stimulus-response or reinforcement). One example of adaptive heuristic is to always choose a best reply to the actions of the other players in the previous period—or, for that matter, to the frequency of their actions in the past (essentially, the well-known "fictitious play").

Adaptive heuristics are boundedly rational strategies (in fact, highly "bounded away" from full rationality). The main question of interest is whether such simple strategies may in the long run yield behavior that is nevertheless highly sophisticated and rational.

This paper is based mainly on the work of Hart and Mas-Colell (2000, 2001a, 2001b, 2003a, 2003b), which we try to present here in a simple and elementary form (see Section 10 and the pointers there for the more general results). Significantly, when the results are viewed together new insights emerge—in particular, into the relations of adaptive heuristics to rationality on the one hand, and to behavioral approaches on the other. See Section 9, which may well be read immediately.

The paper is organized as follows. In Section 2 we provide a rough classification of dynamic models. The setting and notations are introduced in Section 3,

[1]Walras–Bowley Lecture 2003, delivered at the North American Meeting of the Econometric Society in Evanston, Illinois. A presentation is available at *http://www.ma.huji.ac.il/hart/abs/adaptdyn.html*. It is a great pleasure to acknowledge the joint work with Andreu Mas-Colell over the years, upon which this paper is based. I also thank Ken Arrow, Bob Aumann, Maya Bar-Hillel, Avraham Beja, Elchanan Ben-Porath, Gary Bornstein, Toni Bosch, Ido Erev, Drew Fudenberg, Josef Hofbauer, Danny Kahneman, Yaakov Kareev, Aniol Llorente, Yishay Mansour, Eric Maskin, Abraham Neyman, Bezalel Peleg, Motty Perry, Avi Shmida, Sorin Solomon, Menahem Yaari, and Peyton Young, as well as the editor and the anonymous referees, for useful discussions, suggestions, and comments. Research partially supported by the Israel Science Foundation.

1401

Summary

Summary

Degrees of rationality

We can define the degree of Rationality of a strategy as the complexity of the reasoning and computation for the player using it.

Low rationality $\longrightarrow$ High rationality

Evolutionary Dynamics	*Learning Dynamics*
Player $\Leftrightarrow$ Population of individuals who play an action dictated by their genotype. Two main forces : Selection and Mutation.	Bayesian learning : each player play according to his belief on the world, which is updated every period.

Degrees of rationality

We can define the degree of Rationality of a strategy as the complexity of the reasoning and computation for the player using it.

Low rationality $\longrightarrow$ High rationality

Evolutionary Dynamics	*Learning Dynamics*
Player $\Leftrightarrow$ Population of individuals who play an action dictated by their genotype. Two main forces : Selection and Mutation.	Bayesian learning : each player play according to his belief on the world, which is updated every period.

Degrees of rationality

We can define the degree of Rationality of a strategy as the complexity of the reasoning and computation for the player using it.

Low rationality $\longrightarrow$ High rationality

Evolutionary Dynamics	*Learning Dynamics*
Player $\Leftrightarrow$ Population of individuals who play an action dictated by their genotype. Two main forces : Selection and Mutation.	Bayesian learning : each player play according to his belief on the world, which is updated every period.

What about medium rationality ?

Adaptive	*Heuristics*
The player tends to choose better actions every time.	Define simple, unsophisticated, simplistic rules (*"rules of thumb"*) that the player uses to make his decisions.

What about medium rationality ?

Adaptive

The player tends to choose better actions every time.

Heuristics

Define simple, unsophisticated, simplistic rules (*"rules of thumb"*) that the player uses to make his decisions.

What about medium rationality ?

Adaptive

The player tends to choose better actions every time.

Heuristics

Define simple, unsophisticated, simplistic rules (*"rules of thumb"*) that the player uses to make his decisions.

Example : Fictitious Play

"At each period, play the action which is the best reply to all the previous actions of the opponent"

More formally, we have :

$$\alpha^i(T+1) = argmax_{\alpha \in A^i} \frac{1}{T} \sum_{t=1}^{T} u^i(\alpha, s_t^{-i})$$

Notations

stage game	Γ
players	$1, ..., N$
N-tuples of actions	$S = S_1 \times ... \times S_N$
payoff function	$u^i : S \to \mathbb{R}$
i's out	$s^{-i} \in S_1 \times S_{i-1} \times S_{i+1} \times ... \times S_N$
i's action	$s^i \in S^i$
Mixed action	$\sigma^i_t \in \Delta(S^i)$

Perfect monitoring assumption : At the end of each period, all players observe the actions taken by everyone s_t.

Notations

stage game	Γ
players	$1, ..., N$
N-tuples of actions	$S = S_1 \times ... \times S_N$
payoff function	$u^i : S \to \mathbb{R}$
i's out	$s^{-i} \in S_1 \times S_{i-1} \times S_{i+1} \times ... \times S_N$
i's action	$s^i \in S^i$
Mixed action	$\sigma_t^i \in \Delta(S^i)$

Perfect monitoring assumption : At the end of each period, all players observe the actions taken by everyone s_t.

Correlated equilibrium

Before the game starts, each player receives a signal θ^i. The distribution of signal $\theta = (\theta^1, ...\theta^n)$ is known to all players, and the signal do not affect the payoffs. $\Rightarrow$ It can affect the outcome

(Trivial example : Battle of the Sexes)

Let ψ be a probability distribution on S induced by the probability of signals θ, then it is a correlated equilibrium iff

$$\forall k \neq j, \sum_{s^{-i} \in S^{-i}} \psi(j, s^{-i}) u^i(j, s^{-i}) \geq \sum_{s^{-i} \in S^{-i}} \psi(j, s^{-i}) u^i(k, s^-i)$$

Correlated equilibrium

Before the game starts, each player receives a signal θ^i. The distribution of signal $\theta = (\theta^1, ...\theta^n)$ is known to all players, and the signal do not affect the payoffs. $\Rightarrow$ It can <u>affect the outcome</u>
(Trivial example : Battle of the Sexes)

Let ψ be a probability distribution on S induced by the probability of signals θ, then it is a correlated equilibrium iff

$$\forall k \neq j, \sum_{s^{-i} \in S^{-i}} \psi(j, s^{-i}) u^i(j, s^{-i}) \geq \sum_{s^{-i} \in S^{-i}} \psi(j, s^{-i}) u^i(k, s^{-}i)$$

Correlated equilibrium

Before the game starts, each player receives a signal θ^i. The distribution of signal $\theta = (\theta^1, ...\theta^n)$ is known to all players, and the signal do not affect the payoffs. $\Rightarrow$ It can affect the outcome
(Trivial example : Battle of the Sexes)

Let ψ be a probability distribution on S induced by the probability of signals θ, then it is a correlated equilibrium iff

$$\forall k \neq j, \sum_{s^{-i} \in S^{-i}} \psi(j, s^{-i}) u^i(j, s^{-i}) \geq \sum_{s^{-i} \in S^{-i}} \psi(j, s^{-i}) u^i(k, s^- i)$$

Correlated equilibrium : The chicken game

Γ	LEAVE	STAY
LEAVE	5,5	3,6
STAY	6,3	0,0

ψ	LEAVE	STAY
LEAVE	1/3	1/3
STAY	1/3	0

TABLE – The chicken game (left) and its correlated equilibrium (right)

- If a player receive STAY, he knows that the other player received LEAVE.
- If a player receives LEAVE, then there is a probability of $\frac{1}{2}$ that the other one received STAY and $\frac{1}{2}$ that he received LEAVE. If he follows the recommendation, the expected payoff is 4, and 3 if he choose STAY instead.

Correlated equilibrium : The chicken game

Γ	LEAVE	STAY
LEAVE	5,5	3,6
STAY	6,3	0,0

ψ	LEAVE	STAY
LEAVE	1/3	1/3
STAY	1/3	0

TABLE – The chicken game (left) and its correlated equilibrium (right)

- If a player receive STAY, he knows that the other player received LEAVE.
- If a player receives LEAVE, then there is a probability of $\frac{1}{2}$ that the other one received STAY and $\frac{1}{2}$ that he received LEAVE. If he follows the recommendation, the expected payoff is 4, and 3 if he choose STAY instead.

Correlated equilibrium : The chicken game

Γ	LEAVE	STAY
LEAVE	5,5	3,6
STAY	6,3	0,0

ψ	LEAVE	STAY
LEAVE	1/3	1/3
STAY	1/3	0

TABLE – The chicken game (left) and its correlated equilibrium (right)

- If a player receive STAY, he knows that the other player received LEAVE.
- If a player receives LEAVE, then there is a probability of $\frac{1}{2}$ that the other one received STAY and $\frac{1}{2}$ that he received LEAVE. If he follows the recommendation, the expected payoff is 4, and 3 if he choose STAY instead.

Joint distribution of play

Definition

The joint distribution of play is the relative frequency of each N-tuple in the history of play. It is a probability distribution z_T.

$$\forall s \in S, z_T(s) := \frac{1}{T}|\{1 \leq t \leq T : s_t = s\}|$$

It is different from the products of marginals in general.

Summary

Regret Matching

Switch next period to a different action k with a probability that is proportional to the regret for that action. [5]

- j the action played by i^{th} player at T.
- $U_T := \frac{1}{T}\sum_{t=1}^{T} u^i(s_t)$ the average payoff up to T^{th} period.
- $V_T(j,k) := \frac{1}{T}\sum_{t=1}^{T} v_t^i$ where $v_t^i = u^i(k, s_t^{-i})$ whenever $s_t^i = j$.
- $D_T(j,k) := V_T(j,k) - U_T$ the internal regret associated to action k.
- $R_T(j,k) := [D_T(j,k)]_+$ the non-negative regret.
- The mixed action when j was the last action played is :

$$\sigma_{T+1}(k) = \begin{cases} cR_T(j,k) & \text{if } k \neq j \\ 1 - c\sum_{k\neq j} R_T(j,k) & \text{if } k = j \end{cases}$$

Regret Matching

Switch next period to a different action k with a probability that is proportional to the regret for that action. [5]

- j the action played by i^{th} player at T.
- $U_T := \frac{1}{T}\sum_{t=1}^{T} u^i(s_t)$ the average payoff up to T^{th} period.
- $V_T(j,k) := \frac{1}{T}\sum_{t=1}^{T} v_t^i$ where $v_t^i = u^i(k, s_t^{-i})$ whenever $s_t^i = j$.
- $D_T(j,k) := V_T(j,k) - U_T$ the internal regret associated to action k.
- $R_T(j,k) := [D_T(j,k)]_+$ the non-negative regret.
- The mixed action when j was the last action played is :

$$\sigma_{T+1}(k) = \begin{cases} cR_T(j,k) & \text{if } k \neq j \\ 1 - c\sum_{k\neq j} R_T(j,k) & \text{if } k = j \end{cases}$$

Regret Matching

Switch next period to a different action k with a probability that is proportional to the regret for that action. [5]

- j the action played by i^{th} player at T.
- $U_T := \frac{1}{T}\sum_{t=1}^{T} u^i(s_t)$ the average payoff up to T^{th} period.
- $V_T(j,k) := \frac{1}{T}\sum_{t=1}^{T} v_t^i$ where $v_t^i = u^i(k, s_t^{-i})$ whenever $s_t^i = j$.
- $D_T(j,k) := V_T(j,k) - U_T$ the internal regret associated to action k.
- $R_T(j,k) := [D_T(j,k)]_+$ the non-negative regret.
- The mixed action when j was the last action played is :

$$\sigma_{T+1}(k) = \begin{cases} cR_T(j,k) & \text{if } k \neq j \\ 1 - c\sum_{k \neq j} R_T(j,k) & \text{if } k = j \end{cases}$$

Regret Matching

> Switch next period to a different action k with a probability that is proportional to the regret for that action. [5]

- j the action played by i^{th} player at T.
- $U_T := \frac{1}{T}\sum_{t=1}^{T} u^i(s_t)$ the average payoff up to T^{th} period.
- $V_T(j,k) := \frac{1}{T}\sum_{t=1}^{T} v_t^i$ where $v_t^i = u^i(k, s_t^{-i})$ whenever $s_t^i = j$.
- $D_T(j,k) := V_T(j,k) - U_T$ the internal regret associated to action k.
- $R_T(j,k) := [D_T(j,k)]_+$ the non-negative regret.
- The mixed action when j was the last action played is :

$$\sigma_{T+1}(k) = \begin{cases} cR_T(j,k) & \text{if } k \neq j \\ 1 - c\sum_{k \neq j} R_T(j,k) & \text{if } k = j \end{cases}$$

Regret Matching

Switch next period to a different action k with a probability that is proportional to the regret for that action. [5]

- j the action played by i^{th} player at T.
- $U_T := \frac{1}{T}\sum_{t=1}^{T} u^i(s_t)$ the average payoff up to T^{th} period.
- $V_T(j,k) := \frac{1}{T}\sum_{t=1}^{T} v_t^i$ where $v_t^i = u^i(k, s_t^{-i})$ whenever $s_t^i = j$.
- $D_T(j,k) := V_T(j,k) - U_T$ the internal regret associated to action k.
- $R_T(j,k) := [D_T(j,k)]_+$ the non-negative regret.
- The mixed action when j was the last action played is :

$$\sigma_{T+1}(k) = \begin{cases} cR_T(j,k) & \text{if } k \neq j \\ 1 - c\sum_{k\neq j} R_T(j,k) & \text{if } k = j \end{cases}$$

Regret Matching

Switch next period to a different action k with a probability that is proportional to the regret for that action. [5]

- j the action played by i^{th} player at T.
- $U_T := \frac{1}{T}\sum_{t=1}^{T} u^i(s_t)$ the average payoff up to T^{th} period.
- $V_T(j,k) := \frac{1}{T}\sum_{t=1}^{T} v_t^i$ where $v_t^i = u^i(k, s_t^{-i})$ whenever $s_t^i = j$.
- $D_T(j,k) := V_T(j,k) - U_T$ the internal regret associated to action k.
- $R_T(j,k) := [D_T(j,k)]_+$ the non-negative regret.
- The mixed action when j was the last action played is :

$$\sigma_{T+1}(k) = \begin{cases} cR_T(j,k) & \text{if } k \neq j \\ 1 - c\sum_{k\neq j} R_T(j,k) & \text{if } k = j \end{cases}$$

Regret Matching

Switch next period to a different action k with a probability that is proportional to the regret for that action. [5]

- j the action played by i^{th} player at T.
- $U_T := \frac{1}{T}\sum_{t=1}^{T} u^i(s_t)$ the average payoff up to T^{th} period.
- $V_T(j,k) := \frac{1}{T}\sum_{t=1}^{T} v_t^i$ where $v_t^i = u^i(k, s_t^{-i})$ whenever $s_t^i = j$.
- $D_T(j,k) := V_T(j,k) - U_T$ the internal regret associated to action k.
- $R_T(j,k) := [D_T(j,k)]_+$ the non-negative regret.
- The mixed action when j was the last action played is :

$$\sigma_{T+1}(k) = \begin{cases} cR_T(j,k) & \text{if } k \neq j \\ 1 - c\sum_{k\neq j} R_T(j,k) & \text{if } k = j \end{cases}$$

Regret Matching Theorem

Theorem (Hart and Mas-Colell, 2000 [5])

Let each player play regret matching. Then the joint distribution of play converges to the set of correlated equilibria of the stage game.

Regret Matching Theorem : Proof

Theorem (Hart and Mas-Colell, 2000 [5])

Let each player play regret matching. Then the joint distribution of play converges to the set of correlated equilibria of the stage game.

$\Rightarrow$ The proof is a variation of *Blackwell's approachability theorem*.

We denote ρ_t the distance between internal regrets and the negative orthant :

$$\rho_t = [dist(D_t, \mathbb{R}^-)]^2 = \left\| D_t - D_t^- \right\|^2 = \left\| D_t^+ \right\|^2 = \sum_{k \neq j} D_t^+(k,j)^2$$

We also denote A_t the regret of period t : $D_t = \frac{1}{t} \sum_{i=1}^{T} A_i$.

Regret Matching Theorem : Proof

Step 1/8 : Recursion equation

- Since $D_t^- \in \mathbb{R}^-$, $\rho_{t+v} \leq \left\|D_{t+v} - D_t^-\right\|^2$ and

$$\rho_{t+v} \leq \left\| \frac{1}{t+v}(tD_t + \sum_{w=1}^{v} A_{t+w}) - D_t^- \right\|^2$$

- We have $|A_t| \leq \sum_{k \neq j} |A_t(j,k)| \leq m(m-1)2|u^i| \leq m(m-1)2M$. This gives us a bound on $|D_t|$.
- With some calculus we obtain :

$$\rho_{t+v} \leq \frac{t^2}{(t+v)^2}\rho_t + \frac{2t}{(t+v)^2}\sum_{w=1}^{v} A_{t+w} \cdot R_t + \frac{v^2}{(t+v)^2}C$$

with $C = m(m-1)16M^2$.

Regret Matching Theorem : Proof

Step 1/8 : Recursion equation

- This gives us our basic recursion equation :

$$\mathbb{E}[(t+v)^2\rho_{t+v}|h_t] \leq t^2\rho_t + 2t\sum_{w=1}^{v} R_t \cdot \mathbb{E}[A_{t+w}|h_t] + O(v^2) \qquad (1)$$

- The middle term on right-hand side does not immediately vanish, so we need to estimate this term.

Regret Matching Theorem : Proof

Step 2/8 : Rewriting the middle term

- Using the definition of expectation

$$\mathbb{E}[A_{t+w}(j,k)|h_t] = \sum_{s^{-i}} \mathbb{P}((j,s^{-i})_{t+w}|h_t)[u^i(k,s^{-i}) - u^i(j,s^{-i})]$$

- Using the definition of the mixed action σ_t and doing some simple calculus, we obtain :

$$R_t \cdot \mathbb{E}[A_{t+w}|h_t] = \frac{1}{c}\sum_{s^{-i}}\sum_{j\in S^i} u^i(j,s^{-i})\alpha_{t,w}(j,s^{-i}) \tag{2}$$

where $\alpha_{t,w} = \sum_{k\in S^i} \sigma_t(k\to j)\mathbb{P}((k,s^{-i})_{t+w}|h_t) - \mathbb{P}((j,s^{-i})_{t+w}|h_t)$

Regret Matching Theorem : Proof

Step 3/8 : An auxiliary stochastic process

- For all history h_t, define an auxiliary stochastic process $(\hat{s}_{t+w})$ s.t.
  - Initial value : $\hat{s}_t = s_t$
  - Transitions : $\mathbb{P}(\hat{s}_{t+w} = s | \hat{s}_{t+w-1}) = \Pi_{i=1}^{N} \sigma_t^i(\hat{s}^i_{t+w-1} \to s^i)$

  This process is stationary, because it only uses probability of period t.
- We can define

$$\hat{\alpha}_{t,w} = \sum_{k \in S^i} \sigma_T(k \to j) \mathbb{P}((k, \hat{s}^{-i})_{t+w} | h_T) - \mathbb{P}((j, \hat{s}^{-i})_{t+w} | h_T)$$

- Now, we want to show that $\hat{\alpha}$ is close to α and that $\hat{\alpha}$ remains small.

Regret Matching Theorem : Proof

Step 4/8 : α and $\hat{\alpha}$ are close

- By definition, we have :

$$(t+v)[D_{t+v}(j,k) - D_t(j,k)] = \sum_{w=1}^{v} A_{t+w}(j,k) - vD_t(j,k)$$

- Using that $|A_{t+w}(j,k)| \leq 2M$ and $|D_{t+w}(j,k)| \leq 2M$:

$$R_{t+v}(j,k) - R_t(j,k) = O(\frac{v}{t}) \tag{3}$$

- As R_{t+v} are used for s_{t+v} and R_t for $\hat{s}_{t+v}$, we go from this 1-transition inequality to a w-transitions inequality :

$$\mathbb{P}(s_{t+w} = s|h_t) - \mathbb{P}(\hat{s}_{t+w} = s|h_t) = O(\frac{w^2}{t})$$

- Finally, we obtain that α and $\hat{\alpha}$ are close :

$$\alpha_{t,w}(j, s^{-i}) - \hat{\alpha}_{t,w}(j, s^{-i}) = O(\frac{w^2}{t}) \tag{4}$$

Regret Matching Theorem : Proof

Step 5/8 : Dominate $\hat{\alpha}$

- Denoting Π_t the transition matrix such that $\Pi_t(i,j) = \sigma_t^i(i \to j)$, we have that

$$\mathbb{P}(\hat{s}^i_{t+w} = j | h_t) = \Pi_t^w(i,j)$$

- By independence, we also have :

$$\begin{aligned} \mathbb{P}(\hat{s}_{t+w} = (j, s^{-i}) | h_t) &= \mathbb{P}(\hat{s}^{-i}_{t+w} = s^{-i} | h_t)\mathbb{P}(\hat{s}^i_{t+w} = j | h_t) \\ &= \mathbb{P}(\hat{s}^{-i}_{t+w} = s^{-i} | h_t)\Pi_t^w(i,j) \end{aligned}$$

- We inject that into the definition of $\hat{\alpha}$ and we obtain :

$$\hat{\alpha}_{t,w} = \mathbb{P}(\hat{s}^{-i}_{t+w} = s^{-i} | h_t)[\Pi_t^{w+1} - \Pi_t^w](k,j)$$

Regret Matching Theorem : Proof

Step 6/8 : Dominate α

Lemma

Let Π be a $m \times m$ stochastic matrix with all of its diagonal entries > 0. Then $\Pi^{w+1} - \Pi^w = O(\frac{1}{\sqrt{w}})$.

- In the definition of σ_t, we chose a constant c such that $\sigma_t(j \to j) > 0$ in any cases.
- We can then apply the lemma, and we obtain that $\hat{\alpha}$ tends to 0 with w :

$$\hat{\alpha}_{t,w} = O(\frac{1}{\sqrt{w}}) \tag{5}$$

- Using equations (4) and (5), we obtain that

$$\alpha_{t,w} = O(\frac{w^2}{t} + \frac{1}{\sqrt{w}}) \tag{6}$$

Regret Matching Theorem : Proof

Step 7/8 : Dominating the middle term of (1)

- Equations (6) and (2) give us that

$$R_t \cdot \mathbb{E}[A_{t+w}|h_t] = O(\frac{w^2}{t} + \frac{1}{\sqrt{w}})$$

- It follow from (1) that

$$\mathbb{E}[(t+v)^2\rho_{t+v}|h_t] = t^2\rho_T + O(v^3 + t\sqrt{v})$$

- $\forall n, t_n = \lfloor n^{\frac{5}{3}} \rfloor$, and $v = t_{n+1} - t_n = O(n^{\frac{2}{3}})$:

$$\mathbb{E}[t_{n+1}^2\rho_{t_{n+1}}|h_{t_n}] \leq t_n^2\rho_{t_n} + O(n^2)$$

Regret Matching Theorem : Proof

Step 8/8 : Show the convergence

- We use *Strong Law of Large Numbers for Dependent Random Variable Theorem* and obtain $\lim_{n\to\infty} \rho_{t_n} = 0$.
- $\rho_{t_n} = \sum_{k\neq j}[R_{t_n}(j,k)]^2$, so $\forall k, j, \lim_{n\to\infty} R_{t_n}(j,k) = 0$.
- For $t_n \leq t \leq t_{n+1}$, we have from (3) that

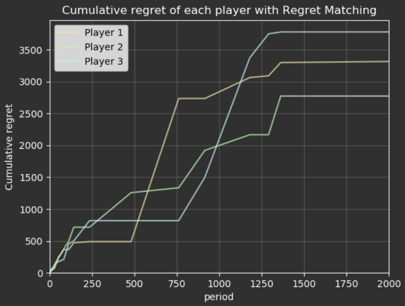
$$R_t(j,k) - R_{t_n}(j,k) = O(\frac{n^{2/3}}{n^{5/3}}) = O(\frac{1}{n})$$

- We obtain what we wanted : All internal regrets vanish at the limit

$$\forall j, k, R_t(j,k) \to 0 \tag{7}$$

- The convergence to the set of correlated equilibrium follows.

Regret Matching : Experiment



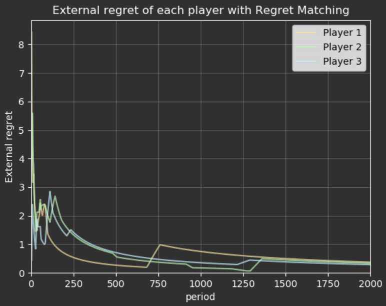


FIGURE – Evolution of cumulative regret (left) and internal regret (right) with Regret matching

$$\textsc{Cumulative Regret}(i,T) = \sum_{t=1}^{T} \left(\max_{k \in S^i} u^i(k, s_t^{-i}) - u^i(s_t^i, s_t^{-i})\right) \quad (8)$$

Real-life behaviours

- No regret $\Rightarrow$ Play the same action : *"Never change a winner team"*
- Some regret $\Rightarrow$ Probability to switch : *"If you invested in A instead of B, you will have gained 20% more by now. So switch to A now !*
- People tend to have too much inertia $\Rightarrow$ probability to not switch > 0 : *"Status quo bias"*.

Real-life behaviours

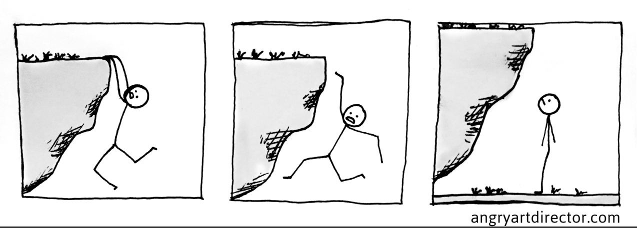
- No regret $\Rightarrow$ Play the same action : *"Never change a winner team"*
- Some regret $\Rightarrow$ Probability to switch : *"If you invested in A instead of B, you will have gained 20% more by now. So switch to A now !*
- People tend to have too much inertia $\Rightarrow$ probability to not switch > 0 : *"Status quo bias"*.

Real-life behaviours

- No regret $\Rightarrow$ Play the same action : *"Never change a winner team"*
- Some regret $\Rightarrow$ Probability to switch : *"If you invested in A instead of B, you will have gained 20% more by now. So switch to A now !*
- People tend to have too much inertia $\Rightarrow$ probability to not switch > 0 : *"Status quo bias"*.

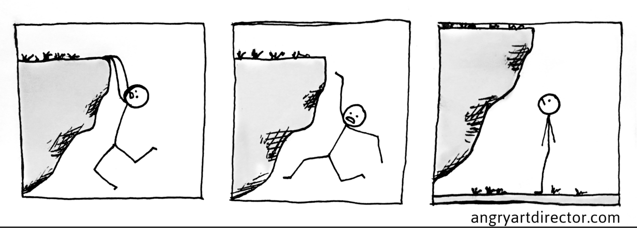
Real-life behaviours

- No regret $\Rightarrow$ Play the same action : *"Never change a winner team"*
- Some regret $\Rightarrow$ Probability to switch : *"If you invested in A instead of B, you will have gained 20% more by now. So switch to A now !*
- People tend to have too much inertia $\Rightarrow$ probability to not switch > 0 : *"Status quo bias"*.



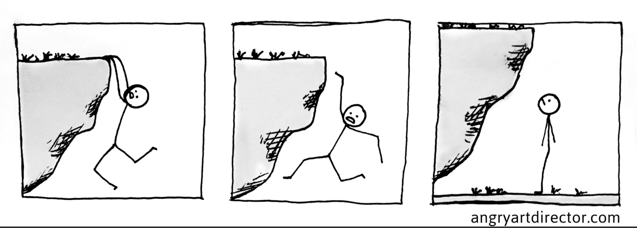
Real-life behaviours

- No regret $\Rightarrow$ Play the same action : *"Never change a winner team"*
- Some regret $\Rightarrow$ Probability to switch : *"If you invested in A instead of B, you will have gained 20% more by now. So switch to A now !*
- People tend to have too much inertia $\Rightarrow$ probability to not switch > 0 : *"Status quo bias"*.



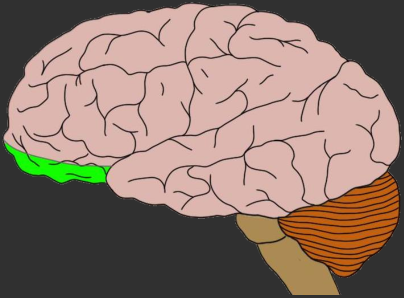
Real-life behaviours

- No regret $\Rightarrow$ Play the same action : *"Never change a winner team"*
- Some regret $\Rightarrow$ Probability to switch : *"If you invested in A instead of B, you will have gained 20% more by now. So switch to A now !*
- People tend to have too much inertia $\Rightarrow$ probability to not switch > 0 : *"Status quo bias"*.



In our brain

Regret influences choices in some aspect and is experienced in the orbitofrontal cortex (Camille et al. 2004) [2, 7]



Generalized Regret Matching

Choose a function f verifying the two conditions below and use the following mixed action.

1. f is Lipchitz continous.
2. Sign preserving property : $f(x) > 0$ for $x > 0$ and $f(0) = 0$

$$\sigma_{T+1}(k) = \begin{cases} f(R_T(k)) & \text{if } k \neq j \\ 1 - \sum_{k \neq j} f(R_T(k)) & \text{if } k = j \end{cases}$$

We can also have different $f_{k,j}$ for each $k \neq j$ or allow f to depend on the whole vector of regrets.

Generalized Regret Matching

Choose a function f verifying the two conditions below and use the following mixed action.

1. f is Lipchitz continous.
2. Sign preserving property : $f(x) > 0$ for $x > 0$ and $f(0) = 0$

$$\sigma_{T+1}(k) = \begin{cases} f(R_T(k)) & \text{if } k \neq j \\ 1 - \sum_{k \neq j} f(R_T(k)) & \text{if } k = j \end{cases}$$

We can also have different $f_{k,j}$ for each $k \neq j$ or allow f to depend on the whole vector of regrets.

Generalized Regret Matching Theorem

Theorem (Hart and Mas-Colell, 2001 [1, 3])

Let each player play a generalized regret matching strategy. Then the joint distribution of play converges to the set of correlated equilibria of the stage game.

Generalized Regret Matching : Experiments

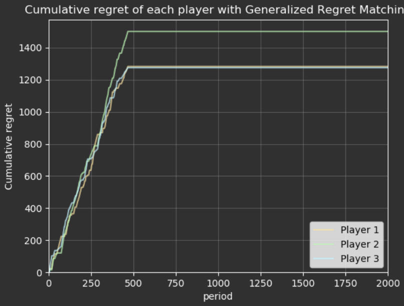


FIGURE – Evolution of cumulative regrets (left) and internal regrets (right) with Regret matching

We use the followingvector of function :

$$f_k : (R(1), ..., R(n)) \to C_0 \times \frac{R(k)}{\sum_i R(i)}$$

Other regret-based strategies

- Unconditional regret matching : We take $V_T(k) := \frac{1}{T}\sum_{t=1}^{T} u^i(k, s_t^{-i})$.
  Theorem (Hart and Mas-Colell, 2000 [5]) : $R_T(k) \to 0$ for all players.
- Proxy regret matching : Used when players don't know that they are playing a game and only have access to their payoffs.

$$\hat{R}_{T+1}(k) := \left[\frac{1}{n_k}\sum_{s_t^i = k} u^i(s_t) - \frac{1}{n_j}\sum_{s_t^i = j} u^i(s_t)\right]_+$$

  Theorem (Hart and Mas-Colell, 2001 [4]) : convergence to correlated approximate equilibria.
- Continuous Time : All the results carry over on the continuous time framework (Hart and Mas-Colell, 2003 [6]).

Other regret-based strategies

- Unconditional regret matching : We take $V_T(k) := \frac{1}{T}\sum_{t=1}^{T} u^i(k, s_t^{-i})$.
  Theorem (Hart and Mas-Colell, 2000 [5]) : $R_T(k) \to 0$ for all players.
- Proxy regret matching : Used when players don't know that they are playing a game and only have access to their payoffs.

$$\hat{R}_{T+1}(k) := \left[\frac{1}{n_k}\sum_{s_t^i = k} u^i(s_t) - \frac{1}{n_j}\sum_{s_t^i = j} u^i(s_t)\right]_+$$

  Theorem (Hart and Mas-Colell, 2001 [4]) : convergence to correlated approximate equilibria.
- Continuous Time : All the results carry over on the continuous time framework (Hart and Mas-Colell, 2003 [6]).

Other regret-based strategies

- Unconditional regret matching : We take $V_T(k) := \frac{1}{T}\sum_{t=1}^{T} u^i(k, s_t^{-i})$.
  Theorem (Hart and Mas-Colell, 2000 [5]) : $R_T(k) \to 0$ for all players.
- Proxy regret matching : Used when players don't know that they are playing a game and only have access to their payoffs.

$$\hat{R}_{T+1}(k) := \left[\frac{1}{n_k}\sum_{s_t^i=k} u^i(s_t) - \frac{1}{n_j}\sum_{s_t^i=j} u^i(s_t)\right]_+$$

  Theorem (Hart and Mas-Colell, 2001 [4]) : convergence to correlated approximate equilibria.
- Continuous Time : All the results carry over on the continuous time framework (Hart and Mas-Colell, 2003 [6]).

Summary

Introduction

$\Rightarrow$ Is there adaptive heuristics that lead to the set of *Nash Equilibria* ?

2-person zero-sum games
2-person potential games
dominance-solvable games
supermodular games

$\rightarrow$ Just use *fictitious play*
(Hart and Mas-Colell, 2003)

$\Rightarrow$ What about general games ?

Dynamic systems definition

Definition (Dynamic system)

A dynamic system in continuous time has the general form

$$\dot{x}(t) = F(x(t); \Gamma) \tag{9}$$

where x is called the *stated variable*

We assume Uncoupledness, i.e. each player's strategy only depends on his own payoff :

$$\forall i, \dot{x}^i(t) = F^i(x(t); u^i)$$

with $x = (x^1, ..., x^N)$ and $F = (F^1, ..., F^N)$

Uncoupled Dynamics Theorem

- We are only studying games with a unique Nash equilibrium, denoted $\overline{x}(\Gamma)$
- A dynamic is said to be *Nash-convergent* on $\mathcal{U}$ if $\forall \Gamma \in \mathcal{U}$, the unique Nash equilibrium is
  1. A rest-point of the dynamic, i.e. $F(\overline{x}(\Gamma); \Gamma) = 0$
  2. A stable point for the dynamic, i.e. $\lim_{t \to \infty} x(t) = \overline{x}(\Gamma)$ for every solution of (9).

Theorem (Hart and Mas-Colell, 2003 [6])

There exist no uncoupled dynamics that guarantee Nash convergence.

Corollary

There exist no uncoupled dynamics that guarantee convergence to the convex hull of the set of Nash equilibria.

Uncoupled Dynamics Theorem

- We are only studying games with a unique Nash equilibrium, denoted $\overline{x}(\Gamma)$
- A dynamic is said to be *Nash-convergent* on $\mathcal{U}$ if $\forall \Gamma \in \mathcal{U}$, the unique Nash equilibrium is
  1. A rest-point of the dynamic, i.e. $F(\overline{x}(\Gamma); \Gamma) = 0$
  2. A stable point for the dynamic, i.e. $\lim_{t \to \infty} x(t) = \overline{x}(\Gamma)$ for every solution of (9).

Theorem (Hart and Mas-Colell, 2003 [6])

There exist no uncoupled dynamics that guarantee Nash convergence.

Corollary

There exist no uncoupled dynamics that guarantee convergence to the convex hull of the set of Nash equilibria.

Uncoupled Dynamics Theorem

- We are only studying games with a unique Nash equilibrium, denoted $\overline{x}(\Gamma)$
- A dynamic is said to be *Nash-convergent* on $\mathcal{U}$ if $\forall \Gamma \in \mathcal{U}$, the unique Nash equilibrium is
  1. A rest-point of the dynamic, i.e. $F(\overline{x}(\Gamma);\Gamma) = 0$
  2. A stable point for the dynamic, i.e. $\lim_{t\to\infty} x(t) = \overline{x}(\Gamma)$ for every solution of (9).

Theorem (Hart and Mas-Colell, 2003 [6])

There exist no uncoupled dynamics that guarantee Nash convergence.

Corollary

There exist no uncoupled dynamics that guarantee convergence to the convex hull of the set of Nash equilibria.

Summary

1. *There are simple adaptive heuristics that lead to the set of correlated equilibria* (Regret Matching)
2. *There is a large class of adaptive heuristics that lead to the set of correlated equilibria* (Generalized Regret Matching)
3. *There is no adaptive heuristics that always lead to the set of Nash equilibria, or its convex hull.* (Uncoupled Dynamics Theorem)

$\Rightarrow$ In short, adaptive heuristics seem to be the natural bridge between behavioral and relational approaches.

Summary

1. *There are simple adaptive heuristics that lead to the set of correlated equilibria* (Regret Matching)
2. *There is a large class of adaptive heuristics that lead to the set of correlated equilibria* (Generalized Regret Matching)
3. *There is no adaptive heuristics that always lead to the set of Nash equilibria, or its convex hull.* (Uncoupled Dynamics Theorem)

$\Rightarrow$ In short, adaptive heuristics seem to be the natural bridge between behavioral and relational approaches.

Summary

1. *There are simple adaptive heuristics that lead to the set of correlated equilibria* (Regret Matching)
2. *There is a large class of adaptive heuristics that lead to the set of correlated equilibria* (Generalized Regret Matching)
3. *There is no adaptive heuristics that always lead to the set of Nash equilibria, or its convex hull.* (Uncoupled Dynamics Theorem)

$\Rightarrow$ In short, adaptive heuristics seem to be the natural bridge between behavioral and relational approaches.

Summary

1. *There are simple adaptive heuristics that lead to the set of correlated equilibria* (Regret Matching)
2. *There is a large class of adaptive heuristics that lead to the set of correlated equilibria* (Generalized Regret Matching)
3. *There is no adaptive heuristics that always lead to the set of Nash equilibria, or its convex hull.* (Uncoupled Dynamics Theorem)

$\Rightarrow$ In short, adaptive heuristics seem to be the natural bridge between behavioral and relational approaches.

Direction of research suggested

- Do all correlated equilibria are obtained from adaptive heuristics, or can we define a smaller subset ?
- We know how these strategies behave in the limit, but how do they behave along the way ?
- We can use alternative notions of regret, using for instance time-averaging or discounting.
- Adaptive heuristics must be tested in practice : How much do they fit real behaviors ?

Direction of research suggested

- Do all correlated equilibria are obtained from adaptive heuristics, or can we define a smaller subset ?
- We know how these strategies behave in the limit, but how do they behave along the way ?
- We can use alternative notions of regret, using for instance time-averaging or discounting.
- Adaptive heuristics must be tested in practice : How much do they fit real behaviors ?

Direction of research suggested

- Do all correlated equilibria are obtained from adaptive heuristics, or can we define a smaller subset ?
- We know how these strategies behave in the limit, but how do they behave along the way ?
- We can use alternative notions of regret, using for instance time-averaging or discounting.
- Adaptive heuristics must be tested in practice : How much do they fit real behaviors ?

Direction of research suggested

- Do all correlated equilibria are obtained from adaptive heuristics, or can we define a smaller subset ?
- We know how these strategies behave in the limit, but how do they behave along the way ?
- We can use alternative notions of regret, using for instance time-averaging or discounting.
- Adaptive heuristics must be tested in practice : How much do they fit real behaviors ?

References I

- CAHN, A. General procedures leading to correlated equilibria. *International Journal of Game Theory* **33**, 21-40 (2004).
- CAMILLE, N. *et al.* The involvement of the orbitofrontal cortex in the experience of regret. *Science* **304**, 1167-1170 (2004).
- HART, S. & MAS-COLELL, A. A general class of adaptive strategies.
- HART, S. & MAS-COLELL, A. in *Economics Essays* 181-200 (Springer, 2001).
- HART, S. & MAS-COLELL, A. A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica* **68**, 1127-1150. https://ideas.repec.org/a/ecm/emetrp/v68y2000i5p1127-1150.html (sept. 2000).
- HART, S. & MAS-COLELL, A. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review* **93**, 1830-1836 (2003).

References II

WALLIS, J. D. Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* **30,** 31-56 (2007).

Thanks for your attention !

github.com/TheoDlmz/AdaptativeHeuristics