Game Theory Project - Adaptive Heuristics

Théo Delemazure

March 2020

Abstract

In this short report, I will present the work of Hart on its paper *"Adaptive heuristics"*, published in 2005 in *Econometrica*. This paper summarizes the work done by Hart and Mas-Colell, and some other, on Adaptive heuristics between 1999 and 2005, and build a strong link between the dynamic approach of adaptive heuristic, and the static approach of correlated equilibria. The paper largely discusses REGRET MATCHING strategy and more generally GENERALIZED REGRET MATCHING , and how they converge to the set of correlated equilibria and leads to a fully rational outcome, as highlighted in the experiments I conducted. The authors also show that such strategies are really close to the behavior of real agent. The last major result is about the impossibility to have a simple adaptive heuristic which converge to the set of Nash equilibria in the general case.

Introduction

An important notion when considering a strategy in game theory, is its *degree of rationality.* We can informally define the degree of Rationality of a strategy as the complexity of the reasoning and computation for the player using it.

For instance, on the *"low degree of rationality"* side, we have **Evolutionary dynamics**. In such models, a player is represented by a population of individuals, and each individual plays an action dictated by his genotype. The proportion of individuals playing an action gives the probability associated to it in the mixed action of the player. Evolutionary dynamics rely on two main forces to improve the payoff of the player: **Selection**, which represent the *exploitation* of the best action by keeping only the best individuals, and **Mutation**, which represent *exploration* as it generates new actions. There is no rational thinking at all in this model: individual's actions are dictated by their genotype.

On the other side, we have **Learning Dynamics**, which have a high degree of rationality. In a Learning dynamics, each player starts with beliefs about the world, and play an action according to this belief. At each period, the player can update his beliefs with the new information he received. Such strategy is **fully rational**, as it needs beliefs about the world, and computing the best action according to this beliefs.

Adaptive Heuristics lie in between : A strategy is a **heuristic** if it is simple and unsophisticated. Moreover, it is **adaptive** if this simple rule leads the player to choose better action and make better decision. For instance, *Fictitious play* is a well-known adaptive heuristic: It is a heuristic because it follows the simple rule *"play the best action against all previous periods"*, and it is adaptive, as it leads to better decision in the long run.

In **Section 1**, I present notations and important notions, **Section 2** is dedicated to REGRET MATCHING strategy and its variation, which are analyzed with experiments in **Section 3** and compared to real behavior in **Section 4**. Finally, **Section 5** briefly discuss the link,between adaptive heuristics and Nash equilibrium.

1 Preliminaries

1.1 Notations and context

N players are playing a basic N-person game Γ in strategic form. Each player i is associated with a set of action S^i. The set of possible actions profiles is $S = S^1 \times \ldots \times S^N$ and the payoff function of the i^{th} players is a function $u^i : S \to \mathbb{R}$. We denote $m^i = |S^i|$ the number of possible actions of player i. For simplicity, we write m.

The game is played repeatedly over time. We use the *discrete time framework* for now, but all results carry over on the *continuous time framework*. The action of player i at period t is $s^i_t \in S^i$ and the $(N-1)$-tuple of actions of all players except i is denoted $s^{-i} \in S^{-i} = S^1 \times \cdots \times S^{i-1} \times S^{i+1} \times \cdots \times S^N$. Player can also play mixed action $\sigma^i_t \in \Delta(S_i) = \{x \in \mathbb{R}^m_+ | \sum_{k=1}^m x(k) = 1\}$.

Finally, we assume **Perfect monitoring**: At the end of each period, all players observe the entire action profile s_t. This gives the history of play at period t denoted h_t and leads to the notion of *joint distribution of play.*

Definition 1. *The* ***joint distribution of play*** *is the relative frequency of each N-tuple in the history of play. It is a probability distribution $z_T \in \Delta(S)$.*

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

It is different from the products of marginals in general.

1.2 Correlated equilibrium

The most studied equilibrium in Game Theory literature are *Nash equilibrium*, unfortunately, we will see in **Section 5** that adaptive heuristics and Nash equilibrium do not work well together and the most important theorems of this paper are about correlated equilibrium.

Let's assume that before the game start, a referee choose a signal $\overline{s} = (\overline{s}^1, ..., \overline{s}^N) \in S$ with a probability distribution ψ known by all players. The player i receive the signal $\overline{s}^i$. We say that ψ is a correlated equilibria if for each player, it is better to play the action received from the referee than any other action.

Definition 2. *Let ψ be a probability distribution on S induced by the probability of signals $\overline{s}$, then it is a* ***correlated equilibrium*** *if and only if*

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

A classical example is the *Chicken Game* described in **Table 1**. In this example, if a player receive the signal STAY, he know for sure that the other player received LEAVE, so he can play STAY. However, if a player receive the signal LEAVE, then there is a probability of $\frac{1}{2}$ that the other 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.

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 1: The chicken game (left) and its correlated equilibrium (right)

2 Regret matching and generalization

2.1 Regret matching

The first adaptive heuristic strategy proposed by Hart in this paper is the REGRET MATCHING strategy, introduced in [4], which follow the following rule : *"Switch next period to a different action k with probability which is proportional to the regret for that action"*. I will define it more formally. We denote $U_T := \frac{1}{T}\sum_{t=1}^T u^i(s_t)$ the average payoff up to period T and $V_T(j,k) := \frac{1}{T}\sum_{t=1}^T v^i_t$ such that every occurrence of action j is replaced by action k in the history of play:

$$v^i_t = \begin{cases} v^i_t = u^i(k, s^{-i}_t) & \text{if } s^i_t = j \\ v^i_t = u^i(s^i_t, s^{-i}_t) & \text{otherwise} \end{cases}$$

This give us the internal regret $D_T(j,k) := V_T(j,k) - U_T$, and the non-negative regret $R_T(j,k) := D_T^+(j,k) = max(D_T(j,k), 0)$.

We can now define the mixed action of the i^{th} player with REGRET MATCHING strategy :

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

with c a well chosen constant such that $\sigma_{T+1}(j) > 0$. For instance, we can use $c = \frac{1}{2mM}$ with $M = \max|u^i|$. Moreover, the first action is arbitrary chosen.

Theorem 1 (Hart and Mas-Colell, 2000). *Let each player play* REGRET MATCHING . *Then the joint distribution of play converges to the set of correlated equilibria of the stage game. [4]*

Proof. The full proof of this theorem can be found in [4]. I will only give main elements of the proof, avoiding calculus details.

The proof is a variation of *Blackwell's approachability theorem*. Indeed, what we need to show is that all regrets vanish at the limit, by showing that the distance between all regrets and the non-positive orthant tends to 0. We denote this distance ρ_t :

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

We also denote A_t the regret at period t, such that $D_t = \frac{1}{t}\sum_{i=1}^{T} A_i$.

$$A_t(j,k) = \begin{cases} 0 & \text{if } s_t^i \neq j \\ u^i(k, s^{-i}) - u^i(j, s^{-i}) & \text{if } s_t^i = j \end{cases}$$

First, we want to find a recursive inequality on ρ_t. Since $D_t^- \in \mathbb{R}^-$, we have:

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

We have $|A_t| \leq \sum_{k\neq j}|A_t(j,k)| \leq m(m-1)2|u^i| \leq m(m-1)2M$. We have the same bound on $|D_t|$, and obtain the following recursive inequality:

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

The middle term on the right-hand side does not immediately vanish, so we need to estimate this term. We use the definition of expectation:

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

We take the product with the regret, and knowing that the probability to switch from action j to action k depends on this regret, with some elementary calculus we obtain the following:

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

Where α is define as:

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

We now define an **auxiliary stochastic process** $\hat{s}$. For all history h_t, $(\hat{s}_{t+w})_{w\in\mathbb{N}}$ is such that $\hat{s}_t = s_t$ and $\forall w > 0, \mathbb{P}(\hat{s}_{t+w} = s|\hat{s}_{t+w-1}) = \Pi_{i=1}^{N}\sigma_t^i(\hat{s}_{t+w-1}^i \to s^i)$. While the probability to switch from j to k change each period for s, it remains the same for $\hat{s}$. Indeed, **we always use the mixed action of period** t. We say that the process is stationary. We can then define $\hat{\alpha}$ by replacing s by $\hat{s}$ in α. The idea is to show that (i) α and $\hat{\alpha}$ are close, and that (ii) $\hat{\alpha}$ is small.

Let's first prove (i). By definition of D_t, we have $(t+v)[D_{t+v}(j,k)-D_t(j,k)] = \sum_{w=1}^{v} A_{t+w}(j,k)-vD_t(j,k)$. Moreover, both A_{t+w} and D_t can be bound by $2M = 2\max|u^i|$. That gives us that $D_{t+v}(j,k) - D_t(j,k) = O(\frac{v}{t})$, and therefore $R_{t+v}(j,k) - R_t(j,k) = O(\frac{v}{t})$.

We know that the probability to switch from k to j at period $t+v$ is proportional to $R_{t+v}(j,k)$ for s and proportional to $R_t(j,k)$ for $\hat{s}$, because it is a stationary process. We apply a Lemma (see [4] for details) and we obtain $\mathbb{P}(s_{t+w} = s|h_t) - \mathbb{P}(\hat{s}_{t+w} = s|h_t) = O(\frac{w^2}{t})$. We deduce by definition of α that

$$\alpha_{t,w}(j, s^{-i}) - \hat{\alpha}_{t,w}(j, s^{-i}) = O(\frac{w^2}{t}) \tag{3}$$

Let's now show (ii). We denote Π_t the transition matrix of $\hat{s}$, i.e. $\Pi_t(i,j) = \sigma_t^i(i \to j)$. Since the process is stationary, we have the following: $\mathbb{P}(\hat{s}^i_{t+w} = j|h_t) = \Pi_t^w(i,j)$. Moreover, the players choose their action independently:

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

Noticing that $\sum_{k\in S^i} \sigma_t(k \to j)\Pi^w(s_t^i, k) = \Pi^{w+1}(s_t^i, j)$, we obtain

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

Lemma 1. *Let Π be a $m \times m$ stochastic matrix s.t. $\forall j, \Pi(j,j) > 0$. Then $\Pi^{w+1} - \Pi^w = O(\frac{1}{\sqrt{w}})$.*

Using the Lemma above (see [4] for details), we obtain $\hat{\alpha}_{t,w} = O(\frac{1}{\sqrt{w}})$, which, combined with equation (3), gives us:

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

Thanks to the term$1/c$ in (2), we obtain the same bound on $R_T \cdot \mathbb{E}[A_{t+w}|h_t]$ and it follows from (1):

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

We define a sequence t_n such that $\forall n, t_n = \lfloor n^{\frac{5}{3}} \rfloor$, and $v = t_{n+1} - t_n = O(n^{\frac{2}{3}})$. Then $\forall n, \mathbb{E}[t_{n+1}^2\rho_{t_{n+1}}|h_{t_n}] \le t_n^2\rho_{t_n} + O(n^2)$. Now, we use *Strong Law of Large Numbers for Dependent Random Variable Theorem* (see [4] for details) and obtain $\lim_{n\to\infty} \rho_{t_n} = 0$, which is equivalent to $\forall k, j, \lim_{n\to\infty} R_{t_n}(j,k) = 0$.

Finally, for some $t \in [t_n, t_{n+1}]$, we apply $R_{t+v}(j,k) - R_t(j,k) = O(\frac{v}{t})$ to $t = t_n$ and $v = t - t_n$. That gives us $R_t(j,k) - R_{t_n}(j,k) = O(\frac{n^{2/3}}{n^{5/3}}) = O(\frac{1}{n})$, and finally that **all internal regrets vanish in the limit:**

$$\forall j, k, R_t(j,k) \to 0 \tag{4}$$

The convergence to the set of correlated equilibrium follows. □

2.2 Generalized regret matching

The Regret Matching theorem is actually generalizable to a wide range of strategy. Let f be a function such that (1) f is Lipchitz continuous, and (2) f verify the sign preserving property, i.e. $f(x) > 0$ for $x > 0$ and $f(0) = 0$. Then, the following mixed action corresponds to a Generalized Regret Matching strategy.

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

Theorem 2 (Hart and Mas-Colell, 2001)**.** *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.*

This theorem first appears in [3] and was proved in [1].

Note that we can also have different $f_{k,j}$ for each $k \neq j$ or allow f to depend on the whole vector of regrets $(R^1, ..., R^N)$, or even the vector of non-signed regrets $(D^1, ..., D^N)$. For basic Regret Matching , we just use $f : x \to \frac{x}{c}$ which verify both conditions (1) and (2).

2.3 Other extensions

Many variations of Regret Matching strategy are presented in this paper. I briefly describe two of them in this section.

- Unconditional regret matching : We take **external regret** instead of internal regret, i.e

$$V_T(k) := \frac{1}{T}\sum_{t=1}^{T} u^i(k, s_t^{-i})$$

Theorem 3 (Hart and Mas-Colell, 2000)**.** Unconditioned Regret Matching *is Hannan-consistent, i.e.* $R_T(k) \to 0$ *for all players playing* Unconditioned Regret Matching *strategy. [4]*

- Proxy regret matching : This strategy correspond to the case when **a player don't know that he is playing a game** and only have access to his payoffs. The regret is computed with a proxy :

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

Theorem 4 (Hart and Mas-Colell, 2001)**.** *If players play a* Proxy Regret Matching *, then the joint distribution of play converges to the set of correlated approximate equilibria. [5]*

3 Experiments

I ran some experiments to check the convergence of the Regret Matching and the Generalized Regret Matching strategies to the set of correlated equilibria, and to compare their behavior.

There are 3 players with 5 possible actions for each player, and the payoffs of the game are chosen randomly in $[0, 10]$. In the first simulation, all players use Regret Matching strategy, and in the second one, they all play the same Generalized Regret Matching strategy, with a function f on the whole vector of regrets :

$$\forall k \neq j, f_k(R) = C \times \frac{R(k)}{\sum_{k' \neq j} R(k')}$$

Where $C \in]0, 1[$ is a constant corresponding to the probability of switching. Indeed, the player switch to another action with probability C (except if he has no regrets), and he switch to action k proportionally to the regret associated to k. If $\forall k \neq j, R(k) = 0$, then the player has no regret. Consequently, $\forall k \neq j, f_k(R) = 0$ and the player doesn't switch. In this experiment, I used $C = 0.2$.

Figure 1 show the evolution of players' cumulative regret when they all use Regret Matching strategies.

$$\text{Cumulative Regret}(i, T) = \sum_{t=1}^{T}(\max_{k \in S^i} u^i(k, s_t^{-i}) - u^i(s_t^i, s_t^{-i})) \tag{5}$$

It highlights the high inertia of this strategy. Indeed, players rarely change the action they play, which can lead to high regrets for some candidate during a lot of periods. However, as we can see in **Figure 2**, the external regret of all players ultimately converges to 0.

Some Generalized Regret Matching strategy converges more quickly, as one can see on **Figure 4**. We can see on **Figure 3** that players change really often their action, until they reach a correlated equilibrium and their cumulative regret stops growing.

Finally, if we compare the two strategies, the second one seems **more efficient**, but the first one is **closer to real-life behavior of agents**, as explained on **Section 4**.

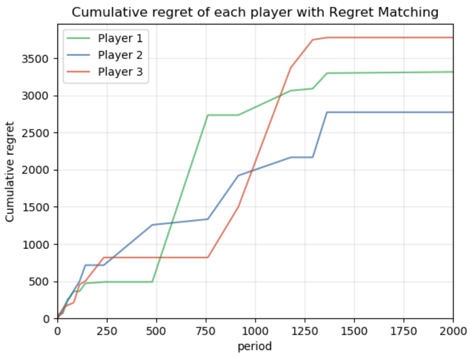

Figure 1: Evolution of players' cumulative regret with Regret Matching strategy.

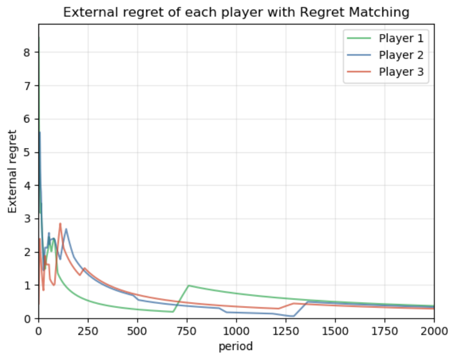

Figure 2: Evolution of players' external regret with Regret Matching strategy.

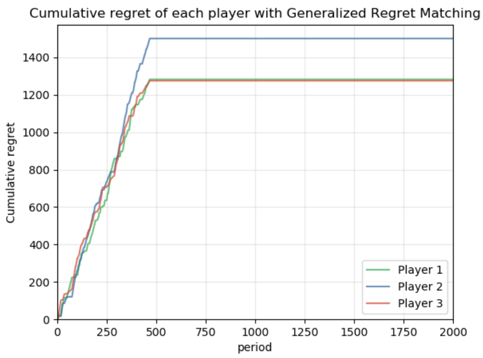

Figure 3: Evolution of players' cumulative regret with Generalized Regret Matching strategy.

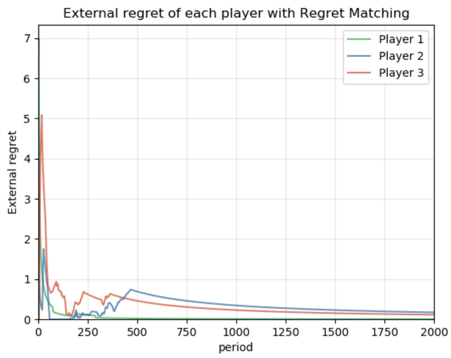

Figure 4: Evolution of players' external regret with Generalized Regret Matching strategy.

4 Behavioral aspects

The Regret Matching strategy and its generalizations have many common points with how real agents make decisions.

First of all, if an agent is experiencing no regret, he will not change his behavior, as expressed in the saying *"Never change a winning team"*. Moreover, it is not absurd that an agent experiencing regret will change is behavior with a probability proportional to this regret. This corresponds to a reasoning of the sort : *"If you invested in A instead of B, you will have gained* 20% *more by now. So switch to A now !* However, people tend to have too much inertia and wait a lot of time before changing their behavior. This is known as the *"status quo bias"* and have been studied a lot [9].

Moreover, research in neurosciences seems to show that we experience regret in our orbitofrontal cortex and that it affects the decision we make [10, 2].

5 Uncoupled dynamics

We have shown that simple adaptive heuristics are sufficient to obtain the convergence to **the set of correlated equilibria**. The question now, is to know if there are adaptive heuristics leading to **the set of Nash equilibria**.

For particular games, like 2-person zero-sum games, 2-person potential games, dominance solvable games or supermodular games, we can use *Fictitious play* and converge to the set of Nash equilibria [8, 6]. What about general games?

We will see that in the general case, there is not convergence to the set of Nash equilibria. To prove that, we will use the *continuous time framework* and introduce the notion of *Dynamic systems.*

5.1 Dynamic systems

Definition 3 (Dynamic system)**.** *A dynamic system in continuous time for a game* Γ *has the general form*

$$\dot{x}(t) = F(x(t); \Gamma) \tag{6}$$

where $x = (x_1, ..., x_N)$ *is called the* stated variable

We assume the **Uncoupledness** property, which state that the strategy of every player is based on (1) the history of play, represented by the state variable x, and (2) his own payoff at each period u^i, but not the payoff of every other player u^k for $k \neq i$. If we denote $F = (F_1, ..., F_N)$, the dynamic system can be written:

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

Hart considers first the simplest case : Games Γ with **a unique Nash equilibria**, which is 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 (6).

5.2 Uncoupled dynamics theorem

Unfortunately, Hart and Mas-Colell proved the following theorem in [7], which state that uncoupledness and Nash-convergence **are not compatible**, even for simple games. Therefore, for any set of game $\mathcal{U}$, there is no uncoupled dynamics with convergence to **the convex hull of the set of Nash equilibria**, which is the unique Nash equilibria, in the set of game considered above.

Theorem 5 (Hart and Mas-Colell, 2003)**.** *There exist no uncoupled dynamics that guarantee Nash convergence. [7]*

Conclusion

This paper discusses the how simple adaptive heuristics can lead to very rational outcomes, by converging to the set of correlated equilibria. We can summarize the results as follows:

1. *There is a simple adaptive heuristic that lead to the set of correlated equilibria* : Regret Matching Theorem

2. *There is a large class of adaptive heuristics that lead to the set of correlated equilibria* : Generalized Regret Matching Theorem

3. *There is no adaptive heuristics that always lead to the set of Nash equilibria, or its convex hull* : Uncoupled Dynamics Theorem

In short, adaptive heuristics seems to be the nice bridge between *behavioral* and *relational* approaches.

Discussion

At the end of the paper, Hart highlights that many questions on adaptive heuristics remains, and suggest some directions of research: 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? What happen if we 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

[1] Amotz Cahn. General procedures leading to correlated equilibria. *International Journal of Game Theory*, 33(1):21–40, 2004.

[2] Nathalie Camille, Giorgio Coricelli, Jerome Sallet, Pascale Pradat-Diehl, Jean-René Duhamel, and Angela Sirigu. The involvement of the orbitofrontal cortex in the experience of regret. *Science*, 304(5674):1167–1170, 2004.

[3] Sergiu Hart and Andreu Mas-Colell. A general class of adaptive strategies.

[4] Sergiu Hart and Andreu Mas-Colell. A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica*, 68(5):1127–1150, September 2000.

[5] Sergiu Hart and Andreu Mas-Colell. A reinforcement procedure leading to correlated equilibrium. In *Economics Essays*, pages 181–200. Springer, 2001.

[6] Sergiu Hart and Andreu Mas-Colell. Regret-based continuous-time dynamics. *Games and Economic Behavior*, 45(2):375–394, 2003.

[7] Sergiu Hart and Andreu Mas-Colell. Uncoupled dynamics do not lead to nash equilibrium. *American Economic Review*, 93(5):1830–1836, 2003.

[8] Josef Hofbauer and William H Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.

[9] William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1):7–59, 1988.

[10] Jonathan D Wallis. Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.*, 30:31–56, 2007.