INTRODUCTION

• Framework

<table>
<thead>
<tr>
<th>Evolutionist game theory</th>
<th>vs</th>
<th>Classical game theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bounded rationality</td>
<td>⇐</td>
<td>Strong rationality</td>
</tr>
<tr>
<td>Sequential time</td>
<td>⇐</td>
<td>Virtual time</td>
</tr>
<tr>
<td>Dynamic process</td>
<td>⇐</td>
<td>Equilibrium notions</td>
</tr>
</tbody>
</table>

-no precise definition of learning

• Motivation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Equilibrium state as asymptotic state of a dynamic process</td>
<td>⇐</td>
<td>Equilibrium state as coordinated players’s beliefs</td>
</tr>
</tbody>
</table>

-no problem of equilibrium selection

-no problem-solving aim (ex: collective optimization)
- theoretical aim
- empirical aim
    (experimental game theory)
LEARNING PRINCIPLES

• **Satisfaction principle**
  (common with classical game theory)
  
  basic game: normal form (matrix) → Utilities
  + extensive form (tree) (rewards)
  repetition (stationary structure)

• **Interaction principle**
  interaction neighbourhood of each player (network)
  stochastic matching interaction through action

• **Information principle**
  information neighbourhood of each player
  selected information on other's actions (free information)
  own's actions

• **Evaluation principle**
  Indices of past information
  other's past frequency of actions
  own's past performance of action
  Expectations (bounded memory)
  (stationarity assumptions)

• **Decision principle**
  Adaptation of actions
  Exploration-exploitation (implicit) dilemma
  → behavioural randomness
LEARNING PROCESS

rational learning
(belief based learning)

other’s determiners unknown
own’s determiners known
observation of others’ actions
quasi perfect instrumental rationality
Ex: fictitious play (+- perturbed)

belief revision
bounded cognitive rationality
Hybrid model, mimetisme.

rational learning

behavioural learning
(reinforcement learning)

other’s determiner unknown
own’s determiner unknown
observation of own’s felt utilities
bounded instrumental rationality
CPR model
(isomorphic to replicator)
asymptotically

evolutionary process
isomorphism

reinforcement rule
( aspiration levels)
CPR LEARNING RULE
(cumulative proportional reinforcement)

• Evaluation rule
At each period $t$, player $i$ computes an index $G_t^i(a)$ for action $a$, equal to the cumulative payoffs obtained in the past

$$G_t^i(a) = \sum_{\tau \leq t} \delta_t^i(a) u_t^i + G_0^i(a)$$

• Decision rule
At each period $t$, player $i$ chooses randomly an action with a probability proportional to its present index

$$p_t^i = G_t^i(a) / \sum_a G_t^i(a)$$

Good (implicit) exploration-exploitation trade-off (no optimal !!!!)
exploration at beginning
exploitation at end, but exploration never stops.
CPR LEARNING PROCESS

→ Stochastic process :
  non linear Polya urn
  – each ball colour corresponds to a combination of players’ actions
    state = frequency of \((a, b)\)
  – if some combination is played, the modeller introduces one corresponding ball
  – the played actions are given by a stationary « transition urn function »
    \[ \text{Action} = p_i(t) \times q_j(t) \]
    \[ = r(t) = f(x(t)) \]

Evolution law :
\[
x(t+1) - x(t) = \frac{1}{t} (-x(t) + \varepsilon(t))
\]
with \(\varepsilon(t) = 1\) with probability \(r(t)\)
\(\varepsilon(t) = 0\) with probability \(1-r(t)\)

Th : any action is chosen an infinite number of time

\(\$\) Index with discounted utility does not work average (low speed)
RESULTS FOR NORMAL-FORM GAMES
(2 players, generic)

- **Decision against (random) nature**
  Th: The process converges almost surely towards the expected utility maximizing actions
  ⇨ Positive feedback (high utility actions are played more and more often)

- **Two-player game [LTW,00]**
  Th: The process converges
  - with positive probability toward any strict Nash equilibrium
  - with zero probability toward some (characterized) mixed Nash equilibrium.
  ⇨ No result inbetween

- **Ex: 2x2 games**
  no pure Nash equilibrium (Ex: matching pennies)
  one pure Nash equilibrium (Ex: prisoner’s dilemma) convergence with positive prob. toward it.
  two pure Nash equilibrium (Ex: battle of sexes)
  one mixed Nash equilibrium convergence with positive prob. toward pure convergence with zero prob. toward mixed
RESULTS FOR EXTENSIVE-FORM GAMES (generic)

• **s-CPR learning rule**
  - Extensive-form game is transformed in normal-form game (through strategies) and CPR rule is applied to it.
  - No specific result

• **a-CPR learning rule**
  - CPR rule is applied at each node of the game tree:
    - when some trajectory is followed at date t, its global payoff is affected to all moves in the trajectory (for each player)
    - the trajectory is chosen by selecting a move at each node according to the local indices.

  Th: The a-CPR process converges almost surely towards the subgame perfect equilibrium path (general result)

• **Ex:**

  \[
  \begin{array}{ccc}
  1 & C & 2 \\
  S & S & C \\
  (3,3) & (1,1) & (4,2)
  \end{array}
  \quad
  \begin{array}{ccc}
  1 & S & 2 \\
  S & (3,3)^N & (3,3) \\
  \uparrow & \downarrow & (1,1) \rightarrow (4,2)^N
  \end{array}
  \]

  Actions and strategies coincide, but the rules differ (s-CPR less sharp)

  ❧ application to incomplete information game
CONCLUSION

• Specific framework (/ AI)
  – Framework restricted to games
    repeated actions  
natural rewards  
  ⇨ weak link with stochastic decision theory
    (no mimic of backward induct.)
  – Correlated learning of several players
    non stationarity of player’s environment
    observation of conjoint results
    ⇨ other player ≠ nature (no transition function)

• Different topics (/ AI)
  – Given learning models
    computationally simple
    introspectively reasonable
    ⇨ positive point of view (tested models)
  – Asymptotic rather than transitory properties
    long term efficiency
    convergence speed
    computation costs
    regularities