WHY LEARNING IN GAMES?

[---]

Initial Response LEARNING Equilibrium (?)

Equilibrium concepts do not explain how people reach equilibrium beliefs or equilibrium play.

Two alternative interpretations for equilibrium play:

1. **Traditional explanation**: players have perfect mental models of others, and know the theory so they play equilibrium immediately.

2. **Adaptive explanation**: people learn, adapt and evolve toward equilibrium

Three contributions of learning:

1. **Learning can explain how people evolve toward equilibrium play.** Example of Cooper, Garvin & Kagel (RAND97,EJ97) on signaling games.

2. If subjects do not reach equilibrium, **learning can explain how individuals were making decisions and therefore why did not reach equilibrium**. Mixed strategy equilibrium and reinforcement learning by Erev & Roth (AER98).

3. If there are many equilibria **learning can shed light on equilibrium selection**. Example of continental divide game (Van Huyck, Battalio & Cook (JEBO97)) and Ho, Camerer & Chong (02)'s EWA.

LEARNING MODELS

Learning models are adaptive behavioral rules that describe how individuals use both information and rational abilities in order to make decisions from one period to the next.

There have been **two different approaches** in the learning models literature: **reinforcement learning** and **belief-based learning** (fictitious play, cournot) models. They have different informational and rationality assumptions.

<u>REINFORCEMENT LEARNING</u>: successful past actions will be used more often in the future.

- Closer to animal behavior. No rationality required.
- Information required:

  -Own past actions and the associated realized payoffs

  -Structure of the game no needed. Game theoretical setting is not different from a decision making setting.

<u>BELIEF-BASED LEARNING MODELS</u>: assumes a simple model of opponent's play. Individuals build beliefs about opponent's future behavior based on opponent past behavior and best respond to them.

- Individuals are rational, able to build beliefs, compute expected payoffs and best respond accordingly.
- Information required:

  -Opponent's past actions.

  -Structure of the game, all of it is needed.

Experienced-Weighted Attraction (EWA), Camerer & Ho(EMT1999) is a general model that includes both reinforcement and belief-based learning as special cases. The EWA merges both information requirements.

Information Requirements Summary Table

Learning Rules' Information Requirements for Updating (for Forming Initial beliefs)

	Own payoff function	Own realized payoffs*	Own decisions*	Others' payoff functions	Others' realized payoffs^	Others' decisions*^
Reinforcement learning	No (No)	Yes (No)	Yes (No)	No (No)	No (No)	No (No)
Beliefs-based learning+	Yes (Yes)	No (No)	No (No)	No (Yes)	No (No)	Yes (No)
EWA	Yes (Yes)	Yes (No)	Yes (No)	No (Yes)	No (No)	Yes (No)

*at least the most recent value, and also previous values if the subject does not store history
+for a subject without a dominant strategy
^or the relevant summary

Description of a learning model: initial values and two types of rules

- **Initial values of state variables:** reinforcements, beliefs, expected payoffs or in general attractions associated with each of the available strategies.
- **Decision rule:** how these attractions determine the probability of taking each action. It can be deterministic or stochastic. Noise can be added to the learning rules => stochastic.
- **Updating rule:** how the state variable(s) are updated from period *t* to *t+1* with the information obtained at *t*.

LEARNING MODELS I: EWA by Camerer & Ho(EMT1999)

EWA model for individual i , j is for strategies and t for time.

<u>**Two variables**</u>:

- **Attractions (positive numerical values)** $A_i^j(t)$ associated with each strategy j=1, 2, …, N.
- **Experience-measures** $N(t)$ is interpreted as the number of observation-equivalent of past experience.

<u>**Updating rule**</u>**:** how attractions and experience-measure are updated from t to $t+1$

$$N(t) = \rho \cdot N(t-1) + 1$$

$$A_i^j(t) = \frac{N(t-1)\cdot\phi\cdot A_i^j(t-1) + [\delta + (1-\delta)\cdot I(s_i^j, s_i(t))]\cdot \pi_i(s_i^j, s_{-i}(t))}{N(t)}$$

ρ discount factor for past experience-measure
ϕ discount factor for past attraction
δ relevance of *law of simulated effect*

Closer look to how attractions are updated. $I(s_i^j, s_i(t))$ is an indicator function that takes value 1 when the strategy chosen is strategy j.

- **If strategy *j* is taken at *t* then** $I(s_i^j, s_i(t)) = 1$ **: the *law of actual effect*, the strategy chosen is reinforced by the payoff.**

$$A_i^j(t) = \frac{N(t-1)\cdot\phi\cdot A_i^j(t-1) + \pi_i(s_i^j, s_{-i}(t))}{N(t)}$$

- **If strategy *j* is not taken at *t* then** $I(s_i^j, s_i(t)) = 0$ **:the *law of simulated effect:*** even if an action is not taken individuals internalize the payoff that it would have yielded in the case it was chosen. Parameter δ measures how much of this effect is included in the learning process.

$$A_i^j(t) = \frac{N(t-1)\cdot\phi\cdot A_i^j(t-1) + \delta\cdot\pi_i(s_i^j, s_{-i}(t))}{N(t)}$$

<u>Decision rule</u>: how attractions determine the probability of taking one action. Different options: logit, probit and power decision rules. Most popular ones: logit and power decision rules. $x_i^j(t)$: probability of taking strategy *j*

- **Logit decision rule**

$$x_i^j(t) = \frac{\exp(\lambda\cdot A_i^j(t))}{\sum_{k=1}^{N}\exp(\lambda\cdot A_i^k(t))}$$

- **Power decision rule**

$$x_i^j(t) = \frac{A_i^j(t)^\lambda}{\sum_{k=1}^{N}(A_i^k(t))^\lambda}$$

In both cases λ measures the sensitivity of the decision rule to attractions. In the logit specification $\frac{1}{\lambda}$ can be interpreted as noise.

LEARNING MODELS II: DERIVATION OF OTHER LEARNING MODELS

REINFORCEMENT LEARNING: Erev & Roth (GEB 95, AER98)

Two-parameter Reinforcement learning (s1, ϕ)

-**s1**: it is the **initial strength**, in the interpretation is very similar to experience-measure, it measures the sensitivity of attractions to payoffs.

- ϕ : they call it ***recency effect***, how much individuals discount past attractions.

-Only one variable, no experience measure ($N(0)$=1 and ρ =0) just attractions (reinforcements or propensities).

-No law of simulated effect (δ =0): only chosen actions get reinforcement

-Decision rule: power decision rule (stochastic)

-Initial values of reinforcements are equal for different players

Initial Values:

$$A_i^j(0) = \frac{s1 \cdot averageabs\ olutepayoff}{numberofst\ rategies}$$

Updating rule:

$$A_i^j(t) = \phi \cdot A_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t))$$

Decision rule:

$$x_i^j(t) = \frac{A_i^j(t)}{\sum_{k=1}^{N}(A_i^k(t))}$$

Three-parameter reinforcement learning: s1, ϕ , ε

-There is not an exact equivalence with EWA.

-They add an ***experimentation effect*** which can be related to *simulated law of effect*. Non-chosen actions get equal reinforcement.

-Same decision rule and initial values.

Updating rule:

- If strategy *j* was chosen:

$$A_i^j(t) = \phi \cdot A_i^j(t-1) + (1-\varepsilon) \cdot \pi_i(s_i^j, s_{-i}(t))$$

- If strategy *j* was not chosen:

$$A_i^j(t) = \phi \cdot A_i^j(t-1) + \frac{\varepsilon}{N_i - 1} \cdot \pi_i(s_i^j, s_{-i}(t))$$

BELIEF-BASED LEARNING MODEL:

<u>Weighted fictitious play:</u> Fudenberg & Levine(98)

-Weights on opponent's possible strategy combinations $w^k_{-i}(t)$

-Update rule for weights: add 1 to the observed strategy combination

$$w^k_{-i}(t) = \rho \cdot w^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))$$

-It is a weighted average: $\rho = 1$ fictitious play and $\rho = 0$ Cournot

-Build beliefs about opponent's future action based on past actions. Belief associated with strategy combination *k*.

$$B^k_{-i}(t) = \frac{w^k_{-i}(t)}{\sum_{h=1}^{N-i} (w^h_i(t))}$$

-Calculate expected payoffs with these beliefs and choose the strategy that gives the highest payoff (deterministic choice rule) or a logistic rule where the λ measures the sensitivity to expected payoffs.

<u>Belief-based learning models are special cases of EWA.</u>

Attractions are expected payoffs calculated with beliefs built using initial experience-measures for each of the strategy combinations as weights.

-Initial weights are initial experience-measures for each of the strategy combinations.

$$w^k_{-i}(0) = N^k_{-i}(0)$$

-Beliefs are built using the initial weights or initial experience-measures. Beliefs can be written in terms of beliefs last period.

$$B_{-i}^{k}(0) = \frac{N_{-i}^{k}(0)}{\sum_{h=1}^{N-i}(N_{i}^{h}(0))}$$

-Calculate expected payoffs with these beliefs and write current expected payoffs in terms of past expected payoffs. This is the EWA model in which $\phi = \rho$ and δ=1 (law of simulated effect is in full charge).

$$E_{i}^{j}(t) = \frac{N(t-1)\cdot\rho\cdot E_{i}^{j}(t-1) + \pi_{i}(s_{i}^{j}, s_{-i}(t))}{N(t)}$$

- $\phi = \rho$ refers to the weight, $\rho = 1$ fictitious play and $\rho = 0$ Cournot.