

Learning: Reinforcement, Fictitious Play and EWA

學習理論：制約、計牌與EWA

Joseph Tao-yi Wang (王道一)
Lecture 9, EE-BGT

Outline: Learning

1. **Learning:** What you do after you see "results" ...
2. **What we know now:** (various learning rules)
 1. Reinforcement
 2. Belief learning
 3. EWA: a hybrid of reinforcement and belief learning
 4. Others: Evolutionary, anticipatory learning, imitation, learning direction theory, rule learning, etc.
3. **Further research:**
 1. Beyond: New direction for research in learning
 2. Application: How can we use these tools?

What you do after you see...

- ▶ Suppose you are playing **Stag Hunt**
- ▶ (B, L) happened last time
- ▶ What would you do now?

- ▶ Change strategy?
- ▶ Stick to it?

	L	R
T	3, 3	0, 1
B	1, 0	1, 1

What you do after you see...

- ▶ A robot (pre-programmed) would stick to it
 - ▶ Evolutionary approach
- ▶ But humans think twice
- ▶ How would you switch?
- ▶ **Reinforcement:**
 - ▶ Choices "reinforced" by previous payoffs
 - ▶ "Very bad" reasoning

	L	R
T	3, 3	0, 1
B	1, 0	1, 1

Reinforcement Learning

- ▶ Update attractions (tendency to play a certain strategy) after given history
- ▶ **Reinforcement:**
 - ▶ Choices "reinforced" by previous payoffs
 - ▶ Allow spillovers to "neighboring strategies" ϵ
- ▶ Example: (**cumulative reinforcement**)

$$\underline{\underline{A^B(t)}} = \varphi \underline{\underline{A^B(t-1)}} + (1 - \epsilon) \cdot \boxed{1}$$

$$\underline{\underline{A^T(t)}} = \varphi \underline{\underline{A^T(t-1)}} + \epsilon \cdot \boxed{1}$$

Reinforcement Learning

- ▶ (More General) **Cumulative Reinforcement:**

$$\underline{\underline{A^B(t)}} = \varphi \underline{\underline{A^B(t-1)}} + (1 - \epsilon) \cdot 1 \cdot [1 - \rho(t-1)]$$

$$\underline{\underline{A^T(t)}} = \varphi \underline{\underline{A^T(t-1)}} + \epsilon \cdot 1 \cdot [1 - \rho(t-1)]$$

- ▶ Alternatively,

- ▶ **Weighted Average Reinforcement**

$$\underline{\underline{A^B(t)}} = \varphi \underline{\underline{A^B(t-1)}} + (1 - \phi) \cdot (1 - \epsilon) \cdot 1$$

$$\underline{\underline{A^T(t)}} = \varphi \underline{\underline{A^T(t-1)}} + (1 - \phi) \cdot \epsilon \cdot 1$$

What "else" could you do...

- ▶ Would you **update your beliefs** about what others do"?

- ▶ Belief learning models

- ▶ **Fictitious Play**

- ▶ Keep track of frequency
- ▶ Ex: rock-paper-scissors

- ▶ **Cournot Best-Response**

- ▶ What you did last time is what you'll do now

	L	R
T	3, 3	0, 1
B	1, 0	1, 1

Weighted Fictitious Play

- ▶ Other weights? Weighted fictitious play
 - ▶ Fictitious play: weigh all history equally ($\rho = 1$)
 - ▶ Cournot: focus only on the last period ($\rho = 0$)
- ▶ Prior:
 - ▶ $P_{t-1}(L) = 3/5, P_{t-1}(R) = 2/5$
- ▶ Posterior:
 - ▶ $P_{t-1}(L) = (3\rho + 1) / (5\rho + 1)$
 - ▶ $P_{t-1}(R) = (2\rho + 0) / (5\rho + 1)$
 - ▶ $\rho =$ decay factor

Weighted Fictitious Play

- ▶ Posterior:

- ▶ $P_{t-1}(L) = (3\rho + 1) / (5\rho + 1)$

- ▶ $P_{t-1}(R) = (2\rho + 0) / (5\rho + 1)$

- ▶ Use this belief to compute payoffs and use them as attractions:

- ▶ $A^T(t) = [3(3\rho + 1) + 0(2\rho + 0)] / (5\rho + 1)$

- ▶ $A^B(t) = [1(3\rho + 1) + 1(2\rho + 0)] / (5\rho + 1)$

- ▶ Note: Actually payoff received play no role

Could you be doing both?

- ▶ Reinforcement does not update beliefs
 - ▶ But people DO update!
- ▶ Fictitious play doesn't react to actual payoffs
 - ▶ But people DO respond
- ▶ **EWA: a hybrid of two**
 - ▶ Camerer and Ho (Econometrica, 1999)

	L	R
T	3, 3	0, 1
B	1, 0	1, 1

Experience-Weighted Attraction

- ▶ $N(t)$: Experience Weight (weakly increasing)
$$N(t) = \varphi(1 - \kappa)N(t - 1) + 1, N(t) \leq \frac{1}{1 - \varphi(1 - \kappa)}$$

- ▶ **Attraction** (for chosen action B)

$$A^B(t) = [\varphi N(t - 1)A^B(t - 1) + \mathbf{1}]/N(t)$$

- ▶ For unchosen action T , add δ :

- ▶ Weight players give to forgone payoffs from unchosen strategies

- ▶ Law of effect vs. Law of simulated effect

$$A^T(t) = [\varphi N(t - 1)A^T(t - 1) + \mathbf{3}\underline{\underline{\delta}}]/N(t)$$

EWA Special Case: Reinforcement

- ▶ $A^B(t) = [\varphi N(t-1)A^B(t-1) + \pi(B, L)]/N(t)$
- ▶ $A^T(t) = [\varphi N(t-1)A^T(t-1) + \pi(T, L)\underline{\delta}]/N(t)$
where $N(t) = \varphi(1 - \kappa)N(t-1) + 1$
- ▶ Becomes **Reinforcement** if $\delta = 0, N(0) = 1$
- ▶ (Simple) cumulative reinforcement: $\kappa = 1$
 - ▶ $N(t) = 1$ for all t
- ▶ (Weighted) average reinforcement: $\kappa = 0$
 - ▶ Weights are $\frac{\phi}{\phi+1}$ and $\frac{1}{\phi+1}$

EWA Special Case: W. Fictitious Play

- ▶ $A^B(t) = [\varphi N(t-1)A^B(t-1) + \pi(B, L)]/N(t)$
- ▶ $A^T(t) = [\varphi N(t-1)A^T(t-1) + \pi(T, L)\underline{\underline{\delta}}]/N(t)$

where $N(t) = \varphi(1 - \kappa)N(t-1) + 1$

- ▶ **Weighted Fictitious Play** if $\delta = 1, \kappa = 0$

- ▶ Good Homework exercise...

- ▶ Hint: Since $N(t) = 1 + \varphi + \varphi^2 + \dots + \varphi^{t-1}$

- ▶ Posterior is

$$P_t(L) = \frac{I(L, h(t)) + (\varphi + \dots + \varphi^{t-1}) \cdot P_{t-1}(L)}{1 + \varphi + \dots + \varphi^{t-1}}$$

EWA Special Case: W. Fictitious Play

- ▶ $A^B(t) = [\varphi N(t-1)A^B(t-1) + \pi(B, L)]/N(t)$
- ▶ $A^T(t) = [\varphi N(t-1)A^T(t-1) + \pi(T, L)\underline{\delta}]/N(t)$
where $N(t) = \varphi(1 - \kappa)N(t-1) + 1$
- ▶ **Weighted Fictitious Play** if $\delta = 1, \kappa = 0$
 - ▶ Fictitious Play: $\varphi = 1$
 - ▶ Cournot best-response: $\varphi = 0$

EWA Cube: Camerer, Wang, Ho (EJ 08) vs. Wang, Knoepfle, Camerer (JEEA 09)

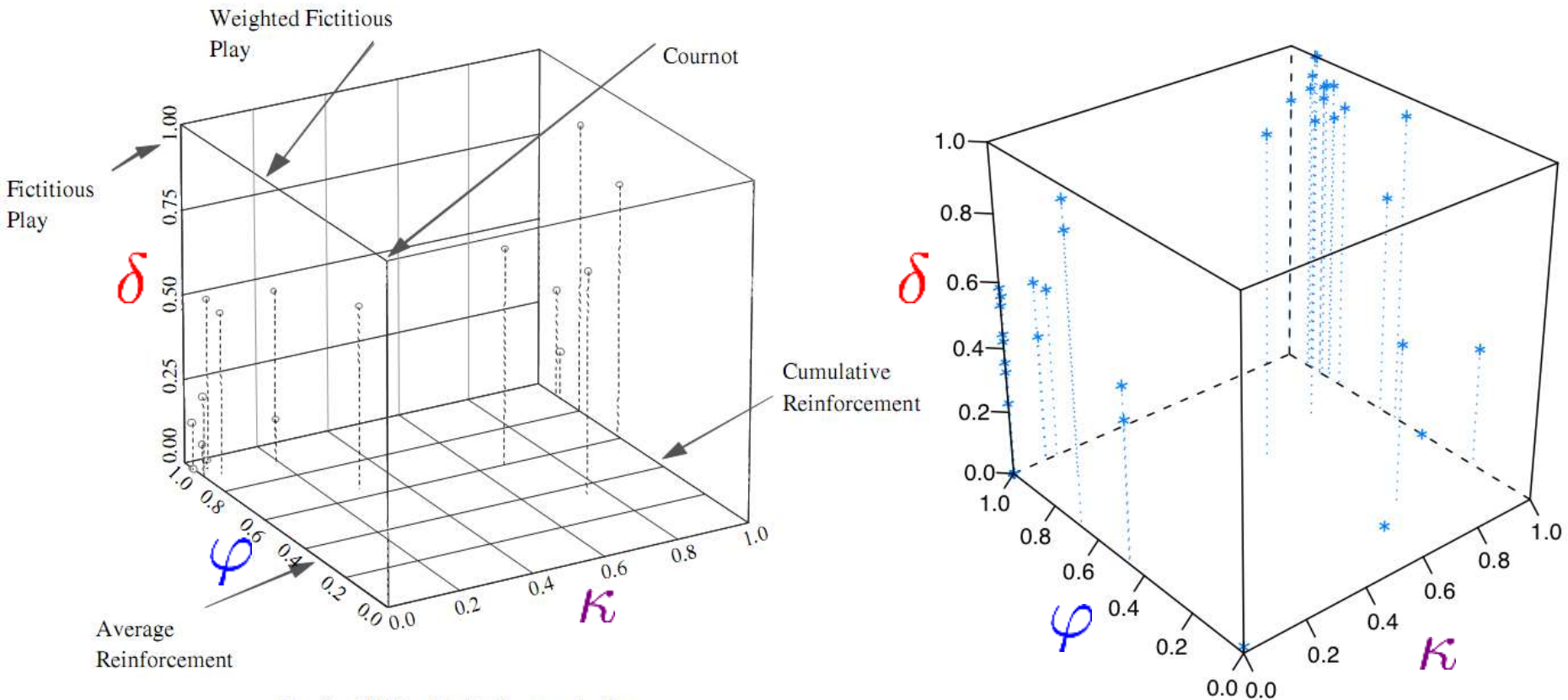


Fig. 2. EWA's Model Parametric Space

- ▶ δ : Attraction weight on forgone payoffs
- ▶ φ : Decay of previous attractions; κ : Growth rate of attractions

Interpretation of EWA Parameters

- ▶ δ : Attraction weight on forgone payoffs
 - ▶ Diff. between received vs. opportunity gains
- ▶ κ : The rate attractions grow
 - ▶ Cumulative vs. Average
- ▶ $N(t)$: The strength of initial attractions
 - ▶ (in units of "experience-equivalence")
- ▶ φ : Weight in $N(t)$
 - ▶ Decay of previous attractions

EWA Cube: Camerer, Wang, Ho (EJ 08) vs. Wang, Knoepfle, Camerer (JEEA 09)

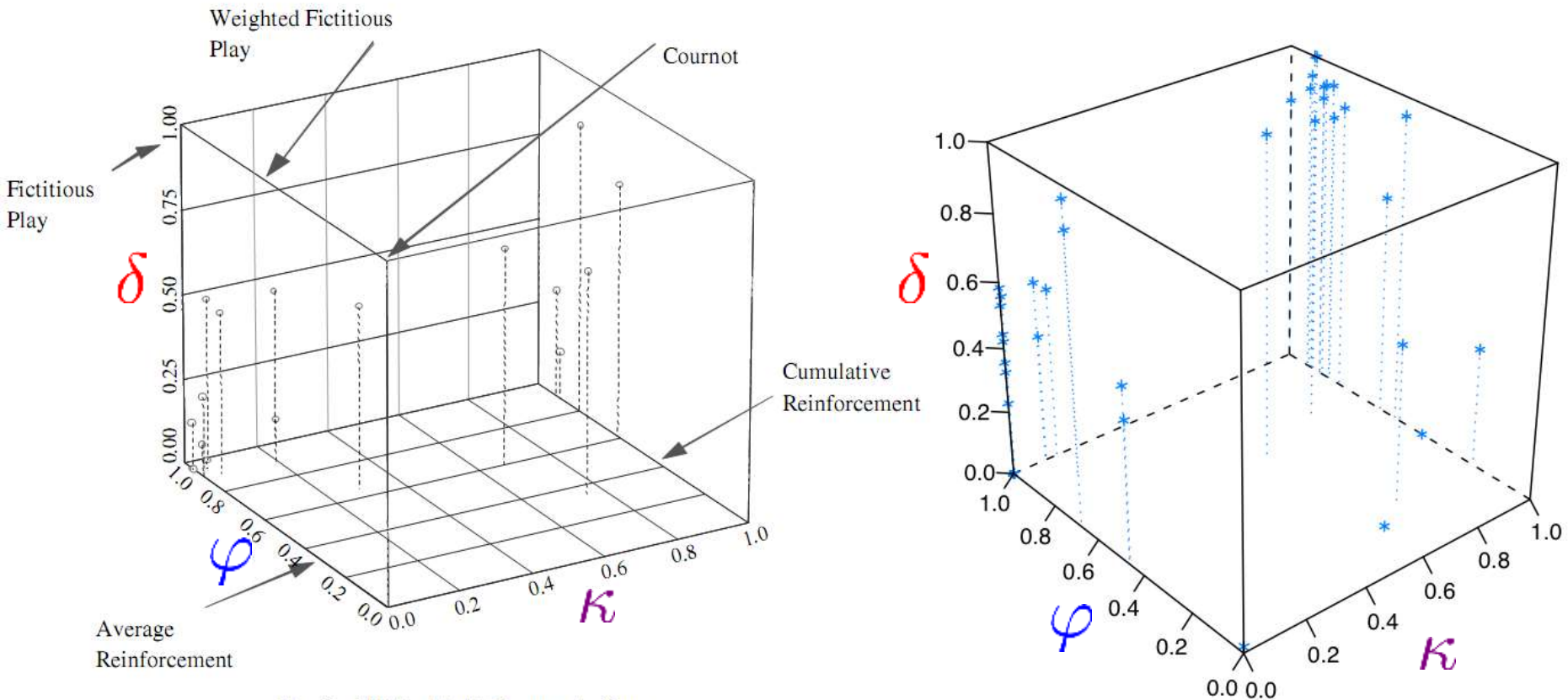


Fig. 2. EWA's Model Parametric Space

- ▶ δ : Attraction weight on forgone payoffs
- ▶ φ : Decay of previous attractions; κ : Growth rate of attractions

Prediction Power of EWA

- ▶ EWA generally improves accuracy in about 35 games (except for mixed ones)
 - ▶ See Camerer and Ho (book chapter, 1999)
 - ▶ "Long version" of the Econometrica paper?
- ▶ BGT, Ch. 6 provides two examples:
 - ▶ Continental Divide
 - ▶ p-Beauty Contest

Prediction Power of EWA

- ▶ **Overfitting**: Too many parameters?
- ▶ Can be tested:
 - ▶ LR test: Restricted fit vs. Unrestricted
- ▶ Better **Out-of-sample** Prediction Power:
 - ▶ Estimate parameters and predict “new data”
 - ▶ Not prone to overfitting (because of new data)
- ▶ 1-parameter **self-tuned EWA** works too:
 - ▶ This EWA-Lite does as good as reinforcement or fictitious play, even on data with new games

Other Learning Rules

- ▶ Other rules include:
- ▶ **Anticipatory learning (Sophistication):**
 - ▶ Sophisticated players are aware that others are learning – BR to Cournot, etc. (level-k)
 - ▶ **Soph. EWA:** Camerer, Ho, Chong (JET 2002)
- ▶ **Imitation:** Imitate average or "best" player
- ▶ **Learning direction theory:** Move toward BR
- ▶ **Rule learning:** Learn which "rule" to use
 - ▶ Stahl (GEB 2000)

Further Research

- ▶ Here is where we stand.
- ▶ Are there new direction for research in learning?
 - ▶ How does **information acquisition** help us study how people learn?
 - ▶ Learning direction theory and imitation are still **loose ends** to be filled

Holy Grail: How do people "actually" learn?

Further Research

- ▶ How can we use these tools?
- ▶ Econometric Properties of learning rules:
 - ▶ Salmon (Econometrica 2001): Simulate data via certain learning rules and estimate them
 - ▶ Identification is bad for mixed strategy equilibrium and games with few strategies
 - ▶ EWA estimation does well on δ ; others okay only for 1000 periods (but not 30 periods)
- ▶ Can use this to test designs

Conclusion

- ▶ **Learning:** How people react to past history
- ▶ Reinforcement
- ▶ Belief Learning
 - ▶ Fictitious play, Cournot, etc.
- ▶ **EWA:** a Hybrid model
 - ▶ Performs better even out-of-sample
- ▶ **Design tests:** simulate and estimate