

實驗經濟學一：行為賽局論

Experimental Economics I: Behavioral Game Theory

第九講：學習理論：制約、計牌與 EWA

Lecture 9: Learning: Reinforcement, Fictitious Play, and EWA

授課教師：國立臺灣大學 經濟學系 王道一教授（Joseph Tao-yi Wang）

本課程指定教材: Colin E. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction*. New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003.



【本著作除另有註明外，採取 [創用 CC](#)
[「姓名標示—非商業性—相同方式分享」臺灣 3.0](#)
[版授權釋出](#)】

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)
- $A^B(t) = \phi A^B(t-1) + (1 - \varepsilon) \star 1$
- $A^T(t) = \phi A^T(t-1) + \varepsilon \star 1$

Reinforcement Learning

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

- $A^B(t) = \phi A^B(t-1) + (1 - \varepsilon) * \mathbf{1} * [1 - \rho(t-1)]$

- $A^T(t) = \phi A^T(t-1) + \varepsilon * \mathbf{1} * [1 - \rho(t-1)]$

- Alternatively,

- **Weighted Average Reinforcement:**

- $A^B(t) = \phi A^B(t-1) + \underline{(1 - \phi)} (1 - \varepsilon) * \mathbf{1}$

- $A^T(t) = \phi A^T(t-1) + \underline{(1 - \phi)} \varepsilon * \mathbf{1}$

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**:
 - $P_{t-1}(L) = (3 \rho + 1) / (5 \rho + 1)$
 - $P_{t-1}(R) = (2 \rho + 0) / (5 \rho + 1)$
 - ρ = 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:
- $A^T(t) = [3 (3\rho + 1) + 0 (2\rho + 0)] / (5\rho + 1)$
- $A^B(t) = [1 (3\rho + 1) + 1 (2\rho + 0)] / (5\rho + 1)$ 
- 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

- Add δ : the weight players give to forgone payoffs from unchosen strategies
 - Law of effect vs. Law of simulated effect
- $A^B(t) = [\phi N(t-1) A^B(t-1) + 1] / N(t)$
- $A^T(t) = [\phi N(t-1) A^T(t-1) + 3\delta] / N(t)$ 
- where $N(t) = \phi(1 - \kappa) N(t-1) + 1$
- $N(t)$: Experience weight (weakly increasing)

EWA Special Case: Reinforcement

- $A^B(t) = [\phi N(t-1) A^B(t-1) + \pi(B,L)] / N(t)$
- $A^T(t) = [\phi N(t-1) A^T(t-1) + \pi(T,L) \delta] / N(t)$
where $N(t) = \phi(1 - \kappa) N(t-1) + 1$
- $\delta = 0, N(0) = 1$: Reinforcement!
- $\kappa = 1$: (Simple) cumulative reinforcement
 - $N(t) = 1$ for all t
- $\kappa = 0$: (Weighted) average reinforcement
 - Weights are $\phi / (\phi + 1)$ and $1 / (\phi + 1)$



EWA S.C.: Weighted Fictitious Play

- $A^B(t) = [\phi N(t-1) A^B(t-1) + \pi(B,L)] / N(t)$
- $A^T(t) = [\phi N(t-1) A^T(t-1) + \pi(T,L) \delta] / N(t)$ 
- where $N(t) = \phi(1 - \kappa) N(t-1) + 1$
- $\delta = 1, \kappa = 0$: Weighted Fictitious Play!
 - Good Homework exercise...
 - Hint: $N(t) = 1 + \phi + \dots + \phi^{t-1}$; 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 S.C.: Weighted Fictitious Play

- $A^B(t) = [\phi N(t-1) A^B(t-1) + \pi(B,L)] / N(t)$
- $A^T(t) = [\phi N(t-1) A^T(t-1) + \pi(T,L) \delta] / N(t)$

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

- $\delta = 1, \kappa = 0$: Weighted Fictitious Play!
 - $\phi = 1$: Fictitious Play
 - $\phi = 0$: Cournot best-response 

EWA Cube: Camerer, Wang, Ho (EJ 2008) vs. Wang, Knoepfle, Camerer (JEEA 2009)

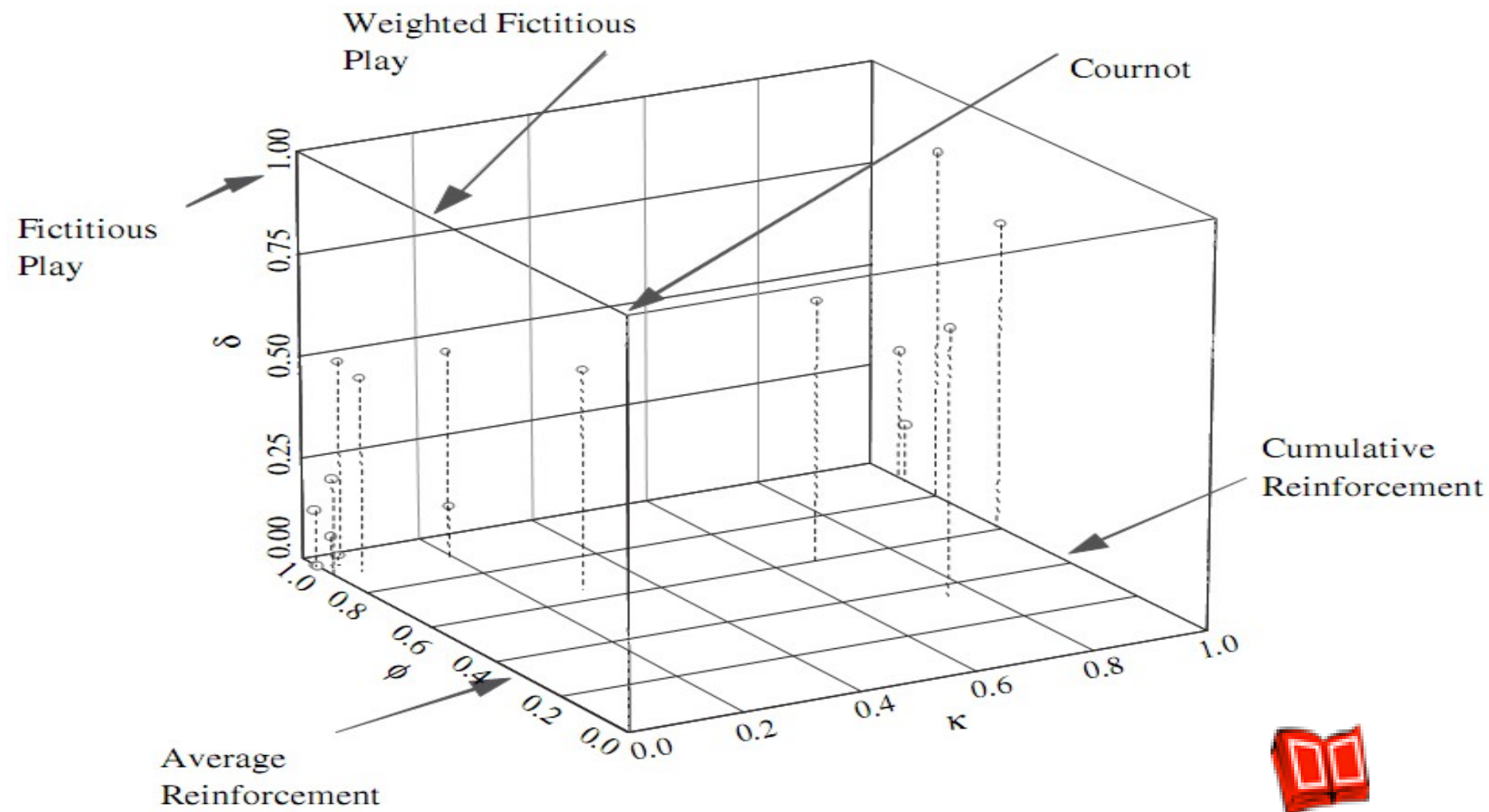
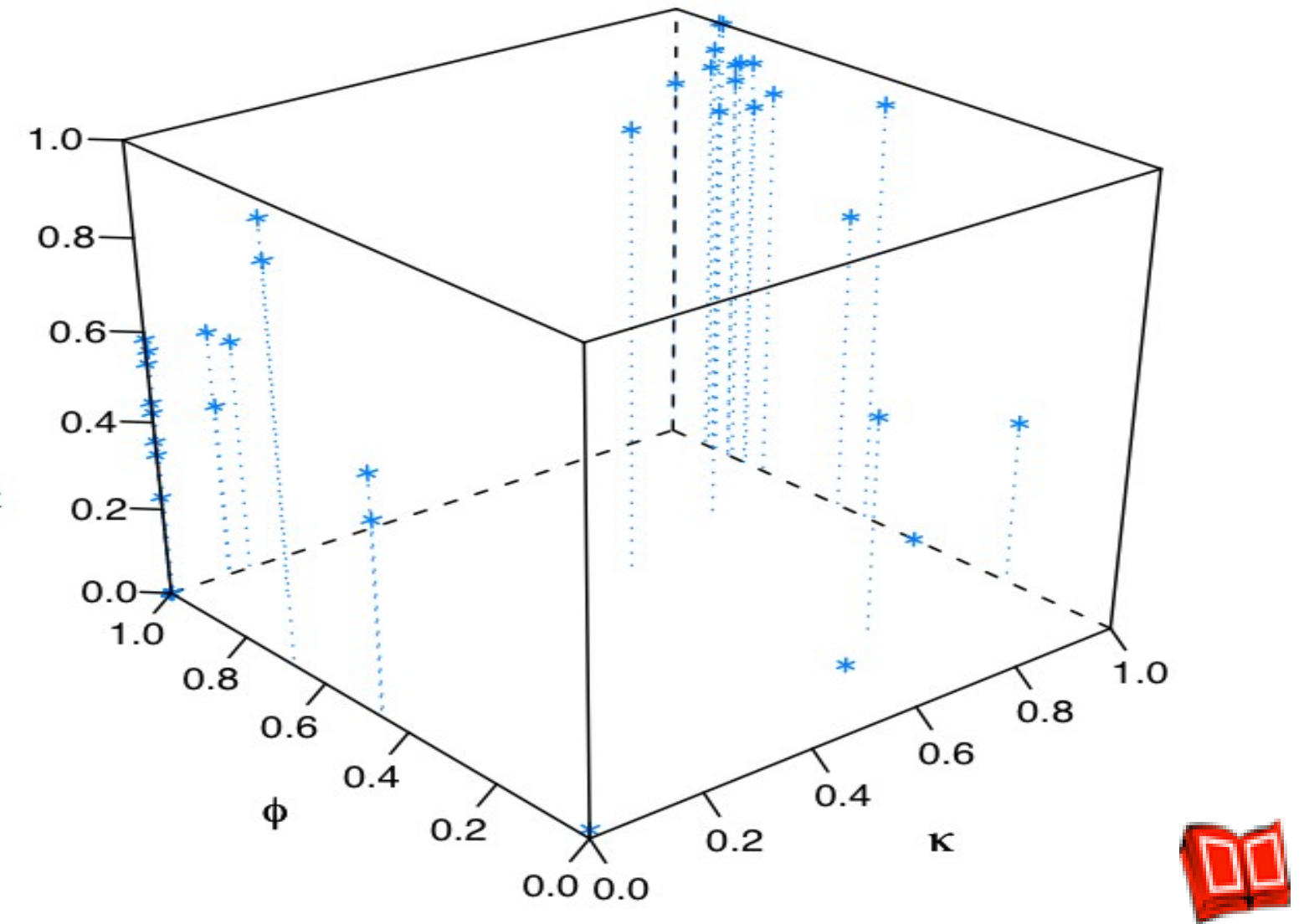


Fig. 2. EWA's Model Parametric Space



δ : attraction weight on forgone payoffs

ϕ : decay of previous attractions

κ : growth rate of attractions

Interpretation of EWA Parameters

- δ : Decay of previous attractions
- κ : The rate attractions grow
- $N(t)$: The strength of initial attractions (in units of “experience-equivalence”)
 - ϕ : Weight in $N(t)$

EWA Cube: Camerer, Wang, Ho (EJ 2008) vs. Wang, Knoepfle, Camerer (JEEA 2009)

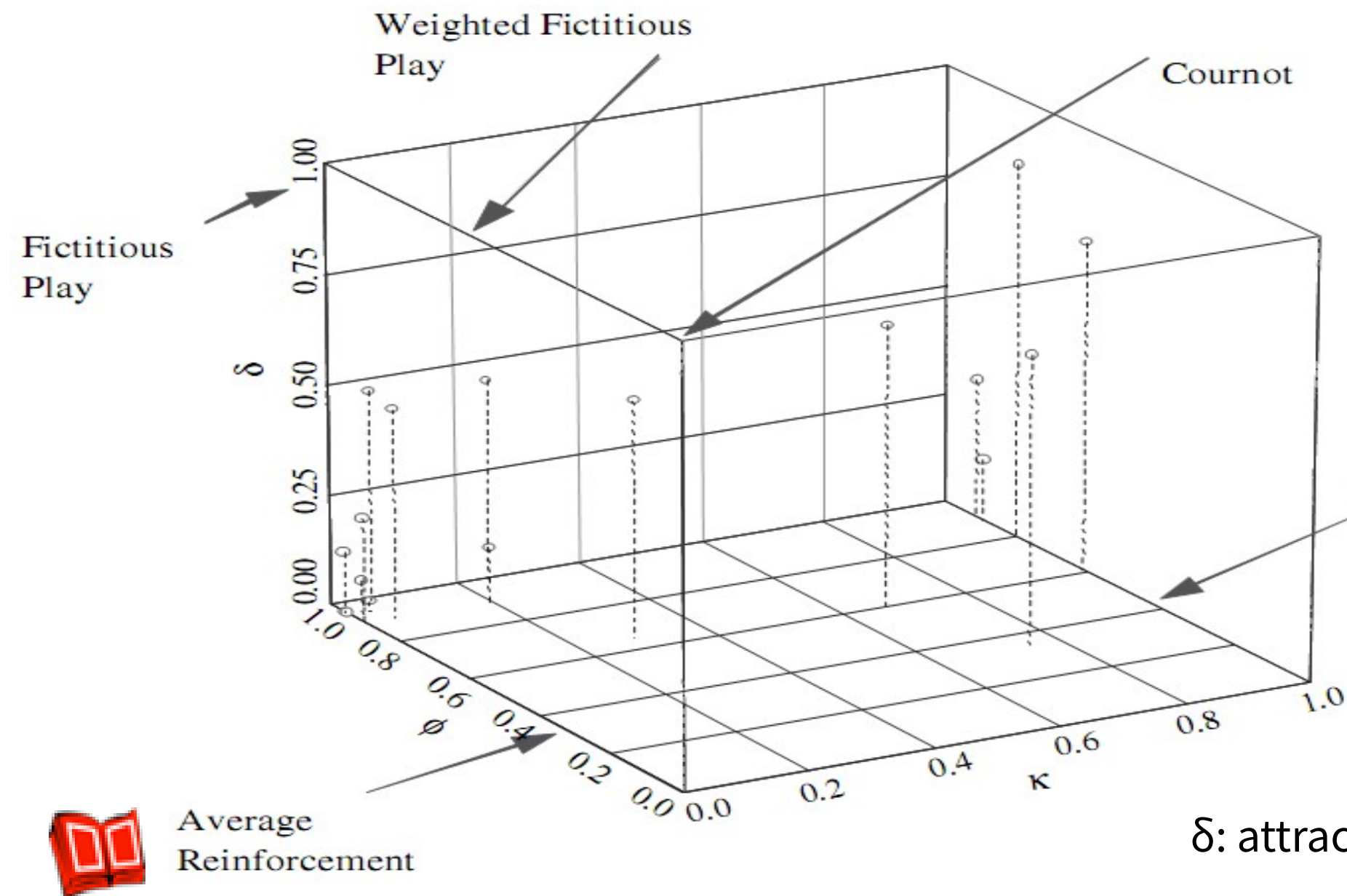
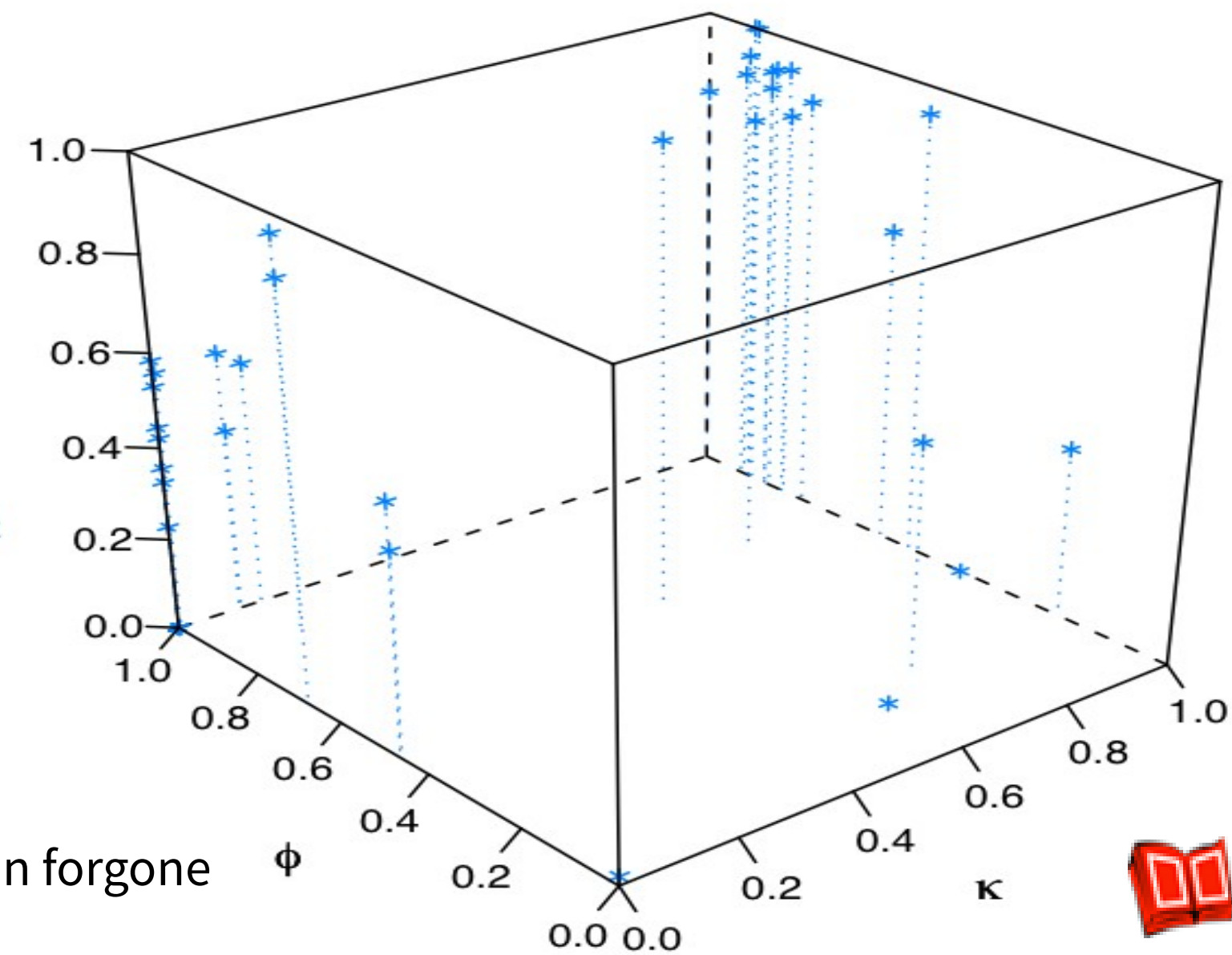


Fig. 2. EWA's Model Parametric Space

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:
 - 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?

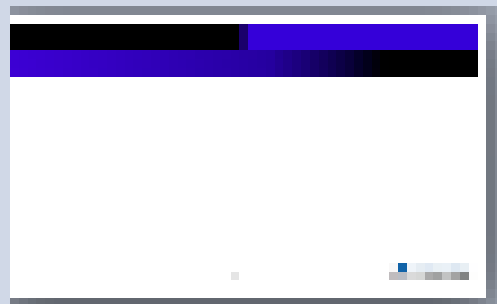
Further research

- 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 are okay if 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

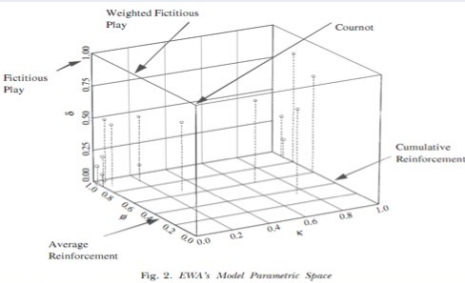
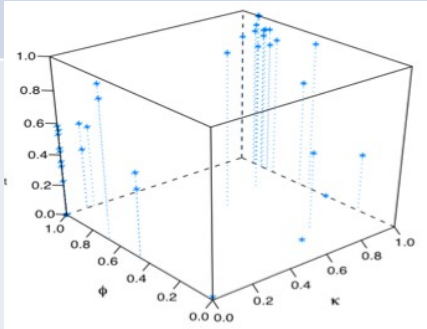
版權聲明

頁碼	作品	版權標示	來源 / 作者									
1-23			國立臺灣大學 經濟學系 王道一 教授									
3, 4, 7, 10	<table><tr><td></td><td>L</td><td>R</td></tr><tr><td>T</td><td>3, 3</td><td>0, 1</td></tr><tr><td>B</td><td>1, 0</td><td>1, 1</td></tr></table>		L	R	T	3, 3	0, 1	B	1, 0	1, 1		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.266. 依據著作權法第 46 、 52 、 65 條合理使用
	L	R										
T	3, 3	0, 1										
B	1, 0	1, 1										
5	$A^B(t) = \phi A^B(t-1) + (1 - \varepsilon) * 1$ $A^T(t) = \phi A^T(t-1) + \varepsilon * 1$		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.268. 依據著作權法第 46 、 52 、 65 條合理使用									
6	$A^B(t) = \phi A^B(t-1) + (1 - \varepsilon) * 1 * [1 - \rho(t-1)]$ $\dots A^T(t) = \phi A^T(t-1) + \underline{(1 - \phi)} \varepsilon * 1$		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.268. 依據著作權法第 46 、 52 、 65 條合理使用									

版權聲明

頁碼	作品	版權標示	來源 / 作者
8-9	$Pt-1(L) = (3 \rho + 1) / (5 \rho + 1)$ $Pt-1(R) = (2 \rho + 0) / (5 \rho + 1)$		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.269. 依據著作權法第 46 、 52 、 65 條合理使用
9	$A^T(t) = [3 (3\rho + 1) + 0 (2\rho + 0)] / (5\rho + 1)$ $-A^B(t) = [1 (3\rho + 1) + 1 (2\rho + 0)] / (5\rho + 1)$		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.269. 依據著作權法第 46 、 52 、 65 條合理使用
11	$A^B(t) = [\phi N(t-1) A^B(t-1) + 1] / N(t)$...-N(t) where $N(t) = \phi(1 - \kappa) N(t-1) + 1$		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.269. 依據著作權法第 46 、 52 、 65 條合理使用
12-14	$A^B(t) = [\phi N(t-1) A^B(t-1) + \pi(B,L)] / N(t)$ $A^T(t) = [\phi N(t-1) A^T(t-1) + \pi(T,L) \delta] / N(t)$...Weights are $\phi / (\phi + 1)$ and $1 / (\phi + 1)$		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.267. 依據著作權法第 46 、 52 、 65 條合理使用

版權聲明

頁碼	作品	版權標示	來源 / 作者
13	$P_t(L) = \frac{I(L, h(t)) + (\varphi + \cdots \varphi^{t-1}) \cdot P_{t-1}(L)}{1 + \varphi + \cdots \varphi^{t-1}}$		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.283. 依據著作權法第 46 、 52 、 65 條合理使用
14	$\phi = 1$: Fictitious Play $\phi = 0$: Cournot best-response		Colin E. Camerer, <i>Behavioral Game Theory: Experiments in Strategic Interaction</i> . New York: Russell Sage Foundation; New Jersey: Princeton UP, 2003. pp.283. 依據著作權法第 46 、 52 、 65 條合理使用
15&17			John Nash, “Two-Person Cooperative Games,” <i>Econometrica</i> , Vol.21, No.1, (Jan., 1953), pp136-137. 依據著作權法第 46 、 52 、 65 條合理使用
15&17			Alvin Roth and Michael W.K.Malouf, “Game-Theoretical Models and the Role of Information in Bargaining,” <i>Psychological Review</i> , Vol. 86, No.6, pp586. 依據著作權法第 46 、 52 、 65 條合理使用

版權聲明

頁碼	作品	版權標示	來源 / 作者
22	Simulate data via certain learning rules and estimate them Identification...estimation does well on δ , others are okay if 1000 periods (but not 30 periods)		T.C.Salmon, “An Evaluation of Econometric Models of Adaptive Learning,” <i>Econometrica</i> , Vol.69, No.6, (2001), pp.1597-1628. 依據著作權法第 46 、 52 、 65 條合理使用