Evolutionary Games - Winter 2005

Chapter 2

From Psychology to Game Theory

Antonio Cabrales

February 2, 2005

Summary

Introduction: Learning Models *(1/2)*

Main assumption: strategies that have given the higher payoffs are used more frequently.

Two strands:

1) Choose best strategy against some average of past behavior.

   - Fictitious play: Brown (1951), Robinson (1951).
   - Best response dynamics: Cournot models, Matsui (1991).

2) reinforcement model: Choose strategies with probabilities in proportion to average of past payoff.
   Bush and Mosteller (1951), Cross (1973), Roth and Erev (1995).

Camerer and Ho (1999): unification of the two strands of models


Introduction: Learning Models *(2/2)*

The Model

- N-player game, each player i has n_i strategies.
- $s^i_j \in S^i$: strategy j of player i.
- $s^i(t)$: strategy played by player i at time t.
- $s(t)$, $s^{-i}(t)$: strategy profiles at time t.
- $u^i(s^i_j, s^{-i})$: payoff function. $u^i(s^i_j, s^{-i}) > 0$ for all $s \in S$
- $P^i_j(t)$: probability that strategy $s^i_j \in S^i$ is chosenat time t.


Learning by reinforcement *(1/7)*

Roth and Erev (1995).

$I(s_j^i, s^i(t))$: indicator - 1 if j was used by i at time t, 0 otherwise.

$A_j^i(t)$: "propensity" to play strategy j.

$$A_j^i(t) = A_j^j(t-1) + I(s_j^i, s^i(t))u^i(s_j^i, s^{-i}(t))$$

A strategy not played does not increase its "propensity".

Strategies that are played increase their propensity by their payoff.

Strategy j for player i will be played with probability:

$$P_j^i(t+1) = \frac{A_j^i(t)}{\sum_{k=1}^{n_i} A_k^i(t)}.$$

Rewrite this stochastic process, to use stochastic approximation tools.

Learning by reinforcement *(2/7)*

Let for all $i \in N$, $A^i(t) = (A^i_1(t), ..., A^i_{n_i}(t))$, $A(t) = (A^1(t), ..., A^N(t))$.

Let also for all $i \in N$, $P^i(t) = (P^i_1(t), ..., P^i_{n_i}(t))$, $P(t) = (P^1(t), ..., P^N(t))$.

Finally let $\gamma^i(t) = \sum_{k=1}^{n_i} A^i_k(t)$ and $\Pi^i_j(s(t)) = I(s^i_j, s^i(t))u^i(s^i_j, s^{-i}(t))$.

We can then write

$$P^i_j(t+1) = P^i_j(t) + \frac{1}{\gamma^i(t)}\Phi^i_j(P(t), \gamma(t)) \quad (1)$$

$$\gamma^i(t+1) = \gamma^i(t) + \sum_{k=1}^{n_i} \Pi^i_k(s(t)) \quad (2)$$

with

$$\Phi^i_j(P(t), A(t)) = \frac{\Pi^i_j(s(t)) - P^i_j(t)\sum_{k=1}^{n_i}\Pi^i_k(s(t))}{1 + \frac{1}{\gamma^i(t)}\sum_{k=1}^{n_i}\Pi^i_k(s(t))}$$

Learning by reinforcement *(3/7)*

The process defined by 1 is related to the replicator dynamics, defined as:

$$\frac{\partial \widehat{P}^i_j(t)}{\partial t} = \widehat{P}^i_j(t)\left[u^i(s^i_j, \widehat{P}^{-i}(t)) - \sum_{k=1}^{n} \widehat{P}^i_k(t) u^i(s^i_k, \widehat{P}^{-i}(t))\right] \text{ for all } i \in N, j \in S^i \quad (3)$$

$$= f^{RD}(\widehat{P}(t)) \quad (4)$$

Proposition 1 (Ianni 2002) *There exist n', ε' and C such that for $\varepsilon < \varepsilon'$, $n > n'$, the solutions to the stochastic process $P(t)$ defined by 1 and the differential equation system $\widehat{P}(t)$ defined by 3 satisfy*

$$\Pr\left[\sup_{k \in K} \left|\widehat{P}(t_k) - P(t_k)\right| > \varepsilon\right] \leq \frac{C}{\varepsilon^2} \sum_{k=1}^{\overline{K}} \frac{1}{\inf_{i \in N} \gamma^i(t_k)^2}$$

In other words, for t high enough the trajectories of $\widehat{P}(t)$ and $P(t)$ are guaranteed to stay arbitrarily close.

Learning by reinforcement *(4/7)*

The trick is to transform the stochastic process in a stochastic approximation algorithm.

These are dynamical systems of the form:

$$x(t) = x(t-1) + \frac{1}{g(t)}\left(F(x(t)) + \varepsilon(t)\right)$$

If $g(t)$ is such that $\sum_{t=1}^{\infty} g(t) = \infty$, and $\sum_{t=1}^{\infty} g(t)^2$ is bounded,

then $x(t)$ converges to the solution of $\frac{\partial x(t)}{\partial t} = F(x(t))$

Proof. (Incomplete.) We can write

$$\Phi_j^i(P(t), A(t)) = \Pi_j^i(s(t)) - P_j^i(t) \sum_{k=1}^{n_i} \Pi_k^i(s(t)) + \delta_j^i(P(t), \gamma(t))$$

with

$$\delta_j^i(P(t), \gamma(t)) = -\frac{1}{\gamma^i(t)} \left[\Pi_j^i(s(t)) - P_j^i(t) \sum_{k=1}^{n_i} \Pi_k^i(s(t)) \right] \frac{\sum_{k=1}^{n_i} \Pi_k^i(s(t))}{1 + \frac{1}{\gamma^i(t)} \sum_{k=1}^{n_i} \Pi_k^i(s(t))}$$

Let $\bar{u} = \max_{i \in N} \max_{s \in S} u^i(s)$. Then one can easily check

$$\frac{\sum_{k=1}^{n_i} \Pi_k^i(s(t))}{1 + \frac{1}{\gamma^i(t)} \sum_{k=1}^{n_i} \Pi_k^i(s(t))} \leq \bar{u}$$

$$\Pi_j^i(s(t)) - P_j^i(t) \sum_{k=1}^{n_i} \Pi_k^i(s(t)) \leq \Pi_j^i(s(t)) \leq \bar{u}$$

So that

$$\left| \delta_j^i(P(t), \gamma(t)) \right| \leq \frac{1}{\gamma^i(t)} \bar{u}^2$$

Now notice that

$$E\left[\Pi_j^i(s(t)) - P_j^i(t)\sum_{k=1}^{n_i}\Pi_k^i(s(t))|P(t),\gamma(t)\right]$$
$$= u^i(s_j^i, P^{-i}(t)) - \sum_{k=1}^{n_i} P_k^i(t)u^i(s_k^i, P^{-i}(t))$$

and then

$$\Phi_j^i(P(t),A(t)) = f^{RD}(\widehat{P}(t)) + \eta_j^i(P(t),\gamma(t)) + \delta_j^i(P(t),\gamma(t))$$

where

$$\eta_j^i(P(t),\gamma(t))$$
$$= \Pi_j^i(s(t)) - P_j^i(t)\sum_{k=1}^{n_i}\Pi_k^i(s(t)) - E\left[\Pi_j^i(s(t)) - P_j^i(t)\sum_{k=1}^{n_i}\Pi_k^i(s(t))|P(t),\gamma(t)\right]$$

Learning by reinforcement *(7/7)*

From these it is obvious that $E\left[\eta_j^i(P(t),\gamma(t))|P(t),\gamma(t)\right] = 0$ with bounded variance. Also as shown above $\delta_j^i(P(t),\gamma(t))$ goes to zero as t goes to infinity

The result follows from straightforward application of the derivations above and Chebyshev's inequality. ■

Remark 2 *Proposition 1 implies that if* $P(0)$ *is close to a strict equilibrium* P^*, $\lim_{t\to\infty} P(t) = P^*$.

Remark 3 *Posch (1997) using similar techniques shows that in* 2×2 *game*

(a) *The stochastic learning process converges almost surely to a stationary point or a cycling path of the replicator dynamics.*

(b) *If the game has a strict equilibrium, the dynamics converge almost surely to a strict equilibrium.*

Other learning models with low rationality *(1/7)*

An alternative approach, by Cross (1973), is to update the probabilities like:

$$P_j^i(t+1) = \begin{cases} u^i(s_j^i, s^{-i}(t)) + (1 - u^i(s^i(t), s^{-i}(t)))P_j^i(t) \text{ if } s_j^i = s^i(t) \\ (1 - u^i(s^i(t), s^{-i}(t)))P_j^i(t) \text{ if } s_j^i \neq s^i(t) \end{cases}$$

It is easy to check that in this case:

$$E\left[P_j^i(t+1) - P_j^i(t)|s(t), P(t)\right] = P_j^i(t)\left[u^i(s_j^i, P^{-i}(t)) - \sum_{k=1}^{n} P_k^i(t)u^i(s_k^i, P^{-i}(t))\right]$$

So the expected rate of change also follows the replicator equation.

Remark, nevertheless, that now the difference

$$P_j^i(t+1) - P_j^i(t) - E\left[P_j^i(t+1) - P_j^i(t)|s(t), P(t)\right]$$

is not a term whose variance vanishes over time (as in Roth-Erev).

Other learning models with low rationality *(2/7)*

Let

$$P_j^{i\theta}(t+1) = \begin{cases} \theta u^i(s_j^i, s^{-i}(t)) + (1 - \theta u^i(s^i(t), s^{-i}(t)))P_j^{i\theta}(t) \text{ if } s_j^i = s^i(t) \\ (1 - \theta u^i(s^i(t), s^{-i}(t)))P_j^{i\theta}(t) \text{ if } s_j^i \neq s^i(t) \end{cases} \tag{5}$$

and the differential equation

$$\frac{\partial \widehat{P}_j^i(t)}{\partial t} = \widehat{P}_j^i(t) \left[u^i(s_j^i, \widehat{P}^{-i}(t)) - \sum_{k=1}^{n} \widehat{P}_k^i(t) u^i(s_k^i, \widehat{P}^{-i}(t)) \right] \text{ for all } i \in N, j \in S^i \tag{6}$$

This presumes that $0 < u^i(s^i(t), s^{-i}(t)) < 1$ for all $s \in S$. Then:

Proposition 4 (Börgers and Sarin 1997) *Consider solutions to system 5 for all θ and to system 6, all with identical initial conditions. Let a time $T < \infty$ and assume that $\theta \to 0$ and $\theta t \to T$. Then $P^\theta(t)$ converges in probability to $P(T)$.*

Proof. (Sketch.) The proof relies on an application of a standard theorem on stochastic processes (Norman 1977, th. 1.1.), and only requires differentiability, (Lipschitz) continuity and finiteness of the conditional expectation of $\left|P^\theta(t+1) - P^\theta(t)\right|^3 / \theta$. ■

Other learning models with low rationality *(4/7)*

Proposition 5 (Börgers and Sarin 1997) *The stochastic process $P^\theta(t)$ converges a.s. to a pure strategy state, and if the initial condition is completely mixed to all of them with positive probability.*

Proof. (Sketch.) An application of another theorem of Norman 1977 (2.3) which says that the stochastic process converges a.s. to one absorbing state. In this case, it is immediate that the only absorbing states are the pure strategy profiles. The second assertion can be shown from the methods in section 7.2 Bush and Mosteller 1955. ■

Other learning models with low rationality *(5/7)*

Yet another approach from Cabrales (2000).

Let $x^i_j(t)$ the proportion of i agents using j at t.

Time is discrete, periods of length τ.

A player i using j at t gets $u^i(s^i_j, x^{-i}(t))\tau + \epsilon^i(t)$.

$\epsilon^i(t)$ is i.i.d. uniform random shock with support $\left[-\frac{A}{2}, \frac{A}{2}\right]$.

Agents change strategies when total payoff is less than the acceptable level, normalized to 0. Assume:

$$\max_{i \in N, s \in S} u^i(s) \leq A, \min_{i \in N, s \in S} u^i(s) \geq -A$$

If the performance of a strategy is adequate, agents keep using it.

Other learning models with low rationality *(6/7)*

If it is not, they choose s_j^i next period with probability $x_j^i(t)$.

With these assumptions the probability of changing a strategy j by an agent i is

$$p_j^i(t) = \frac{A - u^i(s_j^i, x^{-i}(t))\tau}{A}$$

hen the dynamics are:

$$x_j^i(t+\tau) = x_j^i(t)(1 - p_j^i(t)) + \sum_{k=1}^{n_i} x_j^i(t)x_k^i(t)p_k^i(t)$$

This can be rewritten as

$$\begin{aligned} x_j^i(t+\tau) &= x_j^i(t)\frac{u^i(s_j^i, x^{-i}(t))\tau}{A} + \sum_{k=1}^{n_i} x_j^i(t)x_k^i(t)\frac{A - u^i(s_k^i, x^{-i}(t))\tau}{A} \\ &= x_j^i(t) + \frac{x_j^i(t)}{A}\left(u^i(s_j^i, x^{-i}(t))\tau - u^i(s_j^i, x^{-i}(t))\tau\right) \end{aligned}$$

By letting the period length τ go to zero by obtain again the replicator dynamics.

Other learning models with low rationality *(7/7)*

- We have drifted far from the Roth-Erev reinforcement learning.
- Bad because it is a simple model that captures well **observed** behavior.
- In fact, Roth-Erev claim that the long run is irrelevant (and their model probably not good there).
- Unfortunately for reinforcement learning they are not perfect either.

Fictitious play and best-response dynamics *(1/6)*

Brown (1951), Robinson (1951), Cournot (1971), Matsui (1991).

Each agent i forms beliefs about the probability that her opponent k will use strategy s_j^k, which we denote by $\widehat{s}_j^{ik}(t)$,

$$\widehat{s}_j^{ik}(t+1) = (1-\lambda(t))\widehat{s}_j^{ik}(t) + \lambda(t) I(s_j^k, s^k(t))$$

where $\lambda(t) = \phi + \rho/t$.

Fictitious play and best-response dynamics *(2/6)*

$\widehat{s}^{-i}$ beliefs of player i about others, and $BR^i(\widehat{s}^{-i}(t))$ best responses of i to $\widehat{s}^{-i}(t)$.

$$P_j^i(t+1) = \begin{cases} 1 & \text{if } s_j^i = BR(\widehat{s}^{-i}(t)) \\ 0 & \text{if } s_j^i \notin BR(\widehat{s}^{-i}(t)) \\ \gamma_j^i & \text{if } s_j^i \in BR(\widehat{s}^{-i}(t)) \text{ and there is } k \neq j, s_k^i \in BR(\widehat{s}^{-i}(t)) \end{cases}$$

- Fictitious play strictly speaking when $\phi = 0$, $\rho = 1$.

- Best response dynamics when $\phi = 1$, $\rho = 0$.

Proposition 6 *A stationary state of $P_j^i(t+1)$ is a Nash equilibrium.*

Proof. Trivial, since in a steady state s^* we have for all $s_j^{*i} \in BR(s^{*-i})$. ■

Fictitious play and best-response dynamics *(3/6)*

Remark 7 *It is also trivial to show that iteratively strictly dominated strategies will not be played in the limit.*

The first round necessarily eliminates strictly dominated strategies (they are never a best response), then iterate.

Convergence is not guaranteed in general. For best response dynamics just think of matching pennies:

1/2	H	T
H	1,-1	-1,1
T	-1,1	1,-1

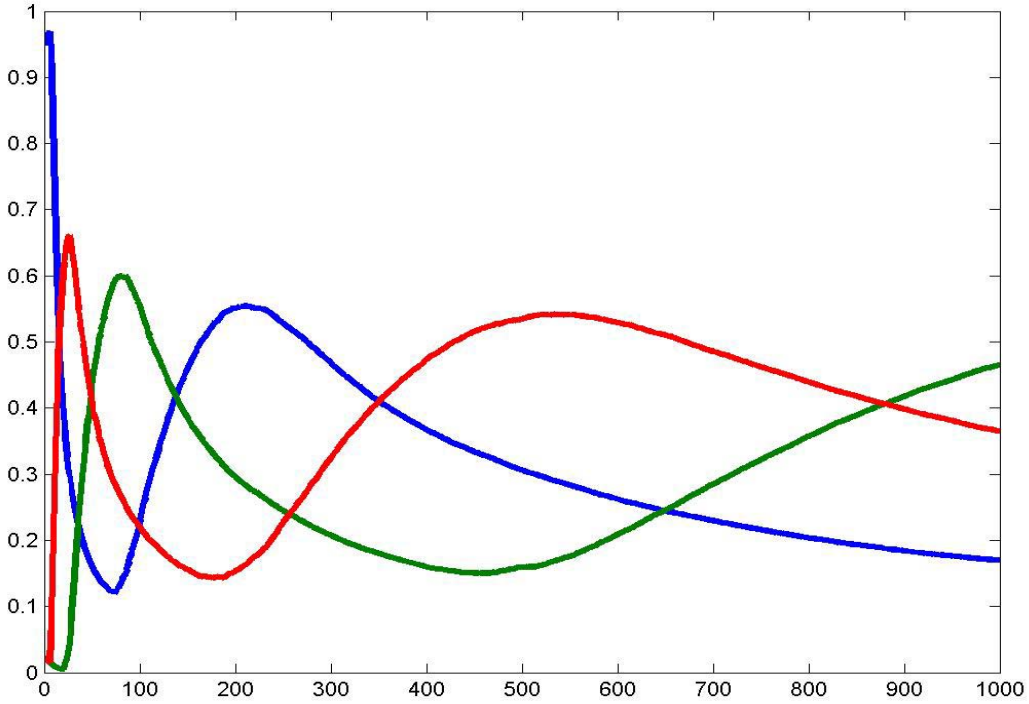
Fictitious play and best-response dynamics *(4/6)*

For fictitious play, the famous Shapley (1964) counterexample

1/2	L	m	R
T	1,0	0,0	0,1
M	0,1	1,0	0.0
B	0,0	0,1	1,0

Empirical frequencies cycle with ever increasing period.

Fictitious play and best-response dynamics *(5/6)*



Fictitious play and best-response dynamics *(6/6)*

In some games it does converge:

- Common interest games.
- Zero-sum games.
- 2X2 games.
- Like for other dynamics, strict equilibria are locally stable.

The experience-weighted attraction learning model *(1/5)*

Camerer and Ho (1999).

The probability that strategy $s_i^j \in S^i$ is chosen by agent i at time $t+1$ is given by

$$P_j^i(t+1) = \frac{e^{\lambda A_j^i(t)}}{\sum_{k=1}^{n_i} e^{\lambda A_j^i(t)}}$$

where $A_j^i(t)$ is the "attraction" of strategy j for agent i at time t, which is given by:

$$A_j^i(t) = \frac{\phi N(t-1) A_j^i(t-1) + [\delta + (1-\delta) I(s_j^i, s^i(t))] u^i(s_j^i, s^{-i}(t))}{N(t)}.$$


The experience-weighted attraction learning model *(2/5)*

$N(t)$ is a variable that is used to express the importance of past experience and is recursively defined by

$$N(t) = \rho N(t-1) + 1, \;\; t \geq 1.$$

the variables $N(t)$ and $A_j^i(t)$ are started with some initial values $N(0)$, and $A_j^i(0)$.

Let $F_j^i(s(t)) = [\delta + (1-\delta) I(s_i^j, s_i(t))] u^i(s_j^i, s^{-i}(t))$, then one can also write:

$$A_j^i(t) = \frac{\sum_{k=0}^{t-1} \phi^i F_j^i(s(t-k)) + \phi^t N(0) A_i^j(0)}{\rho^t N(0) + \sum_{k=0}^{t-1} \rho^k}.$$

The experience-weighted attraction learning model *(3/5)*

Summary:

- $\rho = \phi = 1$, $\delta = 1$, $\lambda = \infty$: fictitious play.
- $\rho = \phi = 0$, $\delta = 1$, $\lambda = \infty$: best reply.
- $\rho = \phi = 0$, $\delta = 0$, $\lambda = 1$: learning by reinforcement.
- In general $\rho = \phi$, and $\delta = 1$: geometric-weighted belief model.

The experience-weighted attraction learning model *(4/5)*

$\beta = (\rho, \phi, \delta, \lambda, N(0), A_i^j(0))$: vector of parameters of this model.

The vector β can be estimated by minimizing:

$$Q_n(s(t), A(t-1), \beta) = n^{-1} \sum_{t=1}^{n} q(s(t), A(t-1), \beta)$$

where the function

$$q(s(t), A(t-1), \beta) = \sum_{j=1}^{J} \sum_{i=1}^{m_i} [I(s_i^j, s_i(t)) - P_i^j(t)]^2$$

in the case of minimum quadratic deviations or

$$q(s(t), A(t-1), \beta) = - \sum_{j=1}^{J} \sum_{i=1}^{m_j} [I(s_i^j, s_i(t)) \log P_i^j(t)])]$$

in the case of maximum likelihood.

The experience-weighted attraction learning model *(5/5)*

- No theorems here. This model is vocationally descriptive (Roth-Erev philosophy).
- Cabrales and García-Fontes (2000):
  Consistent estimates and nice asymptotic distribution if $\phi < 1$.
- So strictly speaking fictitious play and reinforcement learning have too long memory.
- But, bad small sample properties (with up to 200 repetitions of the game).
- Solution: random parameters and use cross section.

Further themes *(1/2)*

1. **The futile search for convergence.**

   (a) Hart-Mas Colell (Econometrica 2000, JET 2001) show that regret matching leads to correlated equilibrium.

   (b) Hart-Mas Colell (American Economic Review 2003) show that deterministic "uncoupled" dynamics (like all the ones we have seen) cannot lead to Nash equilibrium in general.

   (c) Foster and Young (2003) Hart-Mas Colell (2004 WP), Germano-Lugosi (2005 WP) show that some "uncoupled" (and weird) stochastic dynamics guarantee convergence to Nash (basically through exhaustive search).

Further themes *(2/2)*

2. **The futile search for uniqueness.**

   (a) Stochastically stable sets - Kandori-Mailath-Young (Econometrica 1993), Young (Econometrica 1993).

   (b) Add to (finite) dynamics like the ones we have seen some mutations. All populations states should be visited, but spend the longest time at equilibrium. But some equilibria require more mutations to get out than others.

   (c) Make mutation small. Then the (infinite) amount of time spent in each equilibrium depends on the likelihood of mutations that move you out. So at each equilibria you spend $1/\varepsilon^{mutations}$.

References

T. Börgers and R. Sarin (1997), "Learning through Reinforcement and Replicator Dynamics," *Journal of Economic Theory*, 77:1-14.

G.W. Brown (1951), "Iterative Solution of Games by Fictitious Play," in T.C. Koopmans, ed., *Activity Analysis of Production and Allocation*, New York: John Wiley and Sons.

R.R. Bush and F. Mosteller (1951), "A Mathematical Model for Simple Learning," *Psychological Review*, 58:313-323.

R.R. Bush and F. Mosteller (1955), *Stochastic Models for Learning* New York: John Wiley and Sons.

A. Cabrales (2000), "Stochastic Replicator Dynamics," *International Economic Review*, 41:451-481.

A. Cabrales and W. García-Fontes (2000), "Estimating Learning Models with Experimental Data," Working paper UPF 501.

C.F. Camerer, T.H. Ho (1999), "Experience-Weighted Attraction Learning in Games: A Unifying Approach," *Econometrica*, 67:827-874.

References *(2/4)*

A.A. Cournot (1971), *Researches into the Mathematical Principles of the Theory of Wealth*, New York: A.M. Kelley.

J.G. Cross (1973), "A Stochastic Learning Model of Economic Behavior," *Quarterly Journal of Economics* 87:239-266.

D.P. Foster and P.H. Young (2003), "Regret testing: A simple payoff-based procedure for learning Nash equilibrium," University of Pennsylvania and Johns Hopkins University WP.

D. Fudenberg and D.K. Levine (1998), *The Theory of Learning in Games*, Boston MA: MIT Press.

F. Germano and G. Lugosi (2005), "Global Nash convergence of Foster and Youngs regret testing," WP UPF.

S. Hart and A. Mas-Colell (2000), "A Simple Adaptive Procedure Leading to Correlated Equilibrium," *Econometrica* 68:1127-1150.

S. Hart and A. Mas-Colell (2001), "A General Class of Adaptive Strategies," *Journal of Economic Theory* 98:26-54.

References *(3/4)*

S. Hart and A. Mas-Colell (2004), "Stochastic Uncoupled Dynamics and Nash Equilibria," UPF WP.

M. Kandori, G.J. Mailath and R. Rob (1993), "Learning, Mutation and Long-Run Equi- librium in Games," *Econometrica* 61:29-56.

A. Ianni (2002), "Reinforcement Learning and the Power Law of Practice: Some Analytical Results," Southampton University WP0203.

A. Matsui (1991), "Best Response Dynamics and Socially Stable Strategies," *Journal of Economic Theory*, 57: 343-63.

M. F. Norman (1972), *Markov Processes and Learning Models,* Academic Press, New York, London.

M. Posch (1997), "Cycling in a Stochastic Learning Algorithm for Normal-Form Games", *Journal of Evolutionary Economics*, 7:193-207.

J. Robinson (1951), "An Iterative Method of Solving a Game," *Annals of Mathematics*, 54:296-301.

References *(4/4)*

A. Roth and I. Erev (1995), "Learning in Extensive Games: Experimental Data and Simple Dynamic Models in the Intermediate Term", *Games and Economic Behavior*, 8, 164-212.

L.S. Shapley, L. S. (1964), Some Topics in Two-Person Games in *Advances in Game Theory (Annals of Mathematics Studies, 52)* (M. Dresher, L. S. Shapley, and A. W. Tucker, eds.), Princeton: Princeton University Press, 128.

P.D. Taylor and L.B. Jonker (1978), "Evolutionarily Stable Strategies and Game Dynamics," *Mathematical Biosciences*, 40:145-156.

H.P. Young (1993), The Evolution of Conventions," *Econometrica* 61:57-84.

Evolutionary Games - Winter 2005

Chapter 2

From Psychology to Game Theory

Antonio Cabrales

February 2, 2005