




Adaptive Learning in Systems of Interacting Agents

Peyton Young

WINE

Rome, 2009




Some games have thousands or millions of interacting agents

- Commuters driving to work
- Traders bidding on exchanges
- Information processors routing packets through a network
- Dispersed sensors communicating information about potential targets
- Bees foraging for nectar from flowers


The usual premises in game theory

- The joint strategy space is common knowledge
- The payoffs, or at least the distribution of payoffs, is common knowledge
- People *predict* the behavior of others given their knowledge
- They *optimize* given their predictions



In large, complex systems these assumptions are implausible

- Agents may have no knowledge of the overall structure of the game
- They may be unable to observe the actions or payoffs of most of the other players
- Hence their attempts to predict are essentially useless



People can “learn” without predicting

- They *adapt* their behavior in response to their payoffs and to local conditions
- These adaptive behaviors can lead to system-wide equilibrium
- For some types of games convergence to equilibrium is quite rapid

Notation

Players $i = 1, 2, \dots, n$

Action spaces A_i

Joint action space $A = \prod_i A_i$

Utility functions $u_i : A \rightarrow R$

Learning rule #1 : Simple Experimentation

Time is discrete : $t = 1, 2, 3, \dots$

At time t the state of player i is a pair

$$z_i(t) = (\bar{a}_i(t), \bar{u}_i(t))$$

$\bar{a}_i(t)$ is a benchmark action

$\bar{u}_i(t)$ is a benchmark payoff level (aspiration level)

At time $t+1$

Agent i experiments with probability $0 < \varepsilon < 1$

Not experiment \Rightarrow play current benchmark action
benchmark utility = realized payoff

Experiment \Rightarrow play $a_i \in A_i$ drawn uniformly at random

realized payoff $> \bar{u}_i(t) \Rightarrow$ adopt experimental action
and payoff as new benchmarks

realized payoff $\leq \bar{u}_i(t) \Rightarrow$ keep old benchmarks

A game G on A is *weakly acyclic* if from every n -tuple of actions $\vec{a} \in A$ there exists a better reply path -- one player moving at a time -- to a pure Nash equilibrium of G .

G is a *potential game* if there exists a function $\phi : A \rightarrow R$ such that for all i, a_i, a_i', a_{-i}

$$\phi(a_i', a_{-i}) - \phi(a_i, a_{-i}) = u_i(a_i', a_{-i}) - u_i(a_i, a_{-i})$$

Every potential game is weakly acyclic

Theorem 1. *Let G be a finite n -person weakly acyclic game. Given any small $\varepsilon > 0$, if all players learn by simple experimentation with sufficiently small rate of experimentation $\varepsilon^* < \varepsilon$, then a Nash equilibrium is played at least $1 - \varepsilon$ of the time.*

Marden, Young, Arslan, and Shamma, "Payoff-based dynamics for multi-player weakly acyclic games," SIAM Journal on Control and Optimization, 48 (2009) 373-396


Learning rule #2: interactive trial and error

- Agents change their search behavior according to their *mood*
- When *content* they experiment occasionally
- When *discontent* they experiment frequently
- There are *transitional moods* between the content and discontent states

Example: foraging behavior by bees

- Sample a new patch of flowers with small probability when nectar is abundant
- Move far away from the current patch with high probability if no nectar was found in recent trials

*Thuijsman, Peleg, Amitai, and Shmida,
J.Theoretical Biology, 175 (1995), 305-316.*



Similar in concept to WoLF
(Win or Lose Fast)

Bowling and Veloso, Artificial Intelligence 136 (2002)

Player i 's *state* at a given time t has three parts:

(mood, benchmark action, benchmark payoff)

$$z_i = (m_i, \bar{a}_i, \bar{u}_i)$$

Four possible moods:

content, discontent, hopeful, watchful

Content 😊

Experiment with small probability $\varepsilon > 0$

If the experiment results in a higher payoff, adopt the new action and payoff as benchmarks; otherwise stay with the previous benchmarks.

In both cases stay content

If payoff increases *without experimenting*,
become **hopeful** 😊 but don't change
benchmark action right away

If payoff stays up become **content** 😊 again
with new higher payoff as benchmark


If payoff *decreases below benchmark without experimenting*, become **watchful** 😞 but don't change benchmark action right away

If payoff stays below benchmark become **discontent** 😞

If payoff goes back above benchmark become **hopeful** 😊

Discontent ☹️

- Flail around: try a new action at random and with probability $0 < p < 1$ stay discontent
- With probability $1 - p$ spontaneously become content again with the current action and payoff as new benchmarks



A game G is *interdependent* if every proper subset of players S can influence the payoff of at least one player not in S by appropriate choice of their actions.

Note: any game with generic payoffs is interdependent

Theorem 2. *Let G be an n -person interdependent game on a finite joint action space A , such that G has at least one pure Nash equilibrium.*

Given any small $\varepsilon > 0$, if all players use interactive trial and error learning with sufficiently small experimentation rate $\varepsilon^ < \varepsilon$, then a Nash equilibrium is played at least $1 - \varepsilon$ of the time.*

Learning rule #3: logit response to actions of neighbors

Agents are located at the vertices
of an undirected graph

$E = \text{edges}, N = \text{vertices}$

$N_i = \{\text{vertices } j \text{ linked to } i \text{ by an edge}\}$

*Utility results from coordinating actions with neighbors
modified by idiosyncratic preference for each action*

$c(a, a) =$ payoff when you and your neighbor play a

$c(a, a') = 0$ if $a \neq a'$

$v_i(a) =$ i 's idiosyncratic payoff from playing a

$$u_i(\vec{a}) = \sum_{j \in N_i} c(a_i, a_j) + v_i(a_i)$$

Spatial potential game with potential function

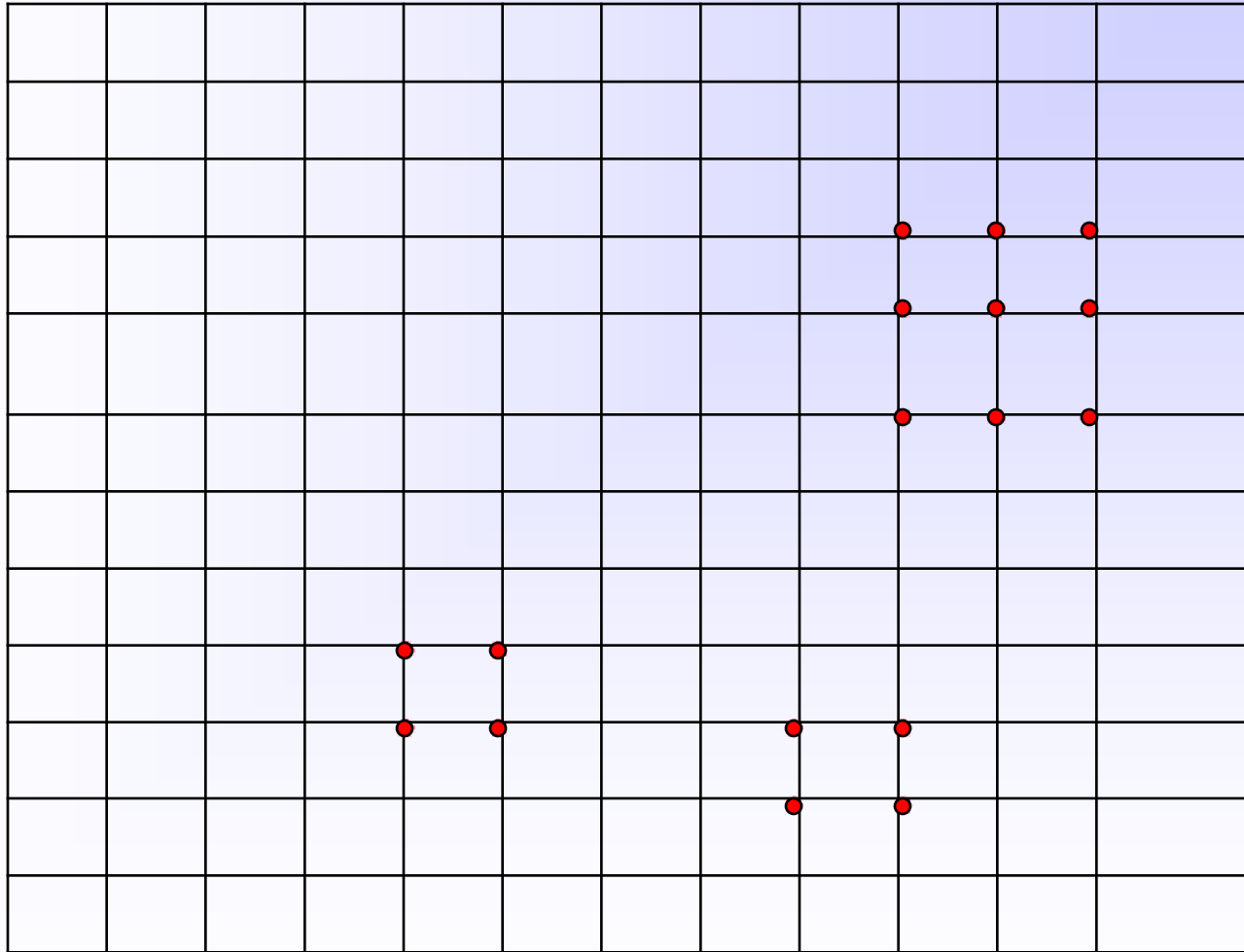
$$\phi(\vec{a}) = \sum_{\{i, j\} \in E} c(a_i, a_j) + \sum_{1 \leq i \leq n} v_i(a_i)$$

Example

	<i>a</i>	<i>b</i>
<i>a</i>	3, 3	0, 0
<i>b</i>	0, 0	2, 2

Idiosyncratic payoffs = 0 \Rightarrow

$$\phi(\vec{a}) = 3(\#edges (a, a)) + 2(\#edges (b, b))$$



Logit learning

State at time t : $\vec{a}(t) = (a_1(t), \dots, a_n(t)) \in A^n$

At start of period $t + 1$ choose an agent i at random

Agent i selects action $a_i(t + 1)$ from A as follows

$$P(a_i(t + 1) = a) = \frac{e^{\beta u_i(a, a_{-i}(t))}}{\sum_{a' \in A} e^{\beta u_i(a', a_{-i}(t))}}$$

β large \Rightarrow best response with high probability

Theorem 3. *Logit learning on a spatial potential game is an ergodic process with unique stationary distribution*

$$\mu(\vec{a}) = \frac{e^{\beta\phi(\vec{a})}}{\sum_{\vec{b} \in A^n} e^{\beta\phi(\vec{b})}}$$

Hence the stochastically stable states (as $\beta \rightarrow \infty$) are the states that maximize the potential function ϕ


[*Blume, Games and Economic Behavior*, 1993;

Young, Individual Strategy and Social Structure, 1998]

Maximum potential state does not necessarily maximize total welfare

Potential $\phi(\vec{a}) = \sum_{\{i,j\} \in E} c(a_i, a_j) + \sum_{1 \leq i \leq n} v_i(a_i)$

Welfare $W(\vec{a}) = 2 \sum_{\{i,j\} \in E} c(a_i, a_j) + \sum_{1 \leq i \leq n} v_i(a_i)$



*Efficiency of the learning dynamics
depends on the topology of network*

Consider a coordination game with two actions $\{a, b\}$

Assume coordinating on a maximizes potential

Efficiency at level $\delta > 0$: the maximum expected waiting time T (over all initial states) such that at least $1 - \delta$ of the agents play a with probability at least $1 - \delta$ in all periods $t \geq T$

Given any two subsets of vertices $S' \subseteq S$

Let $d(S', S) = \# \text{ edges } \{i, j\} \text{ such that } i \in S' \text{ and } j \in S$

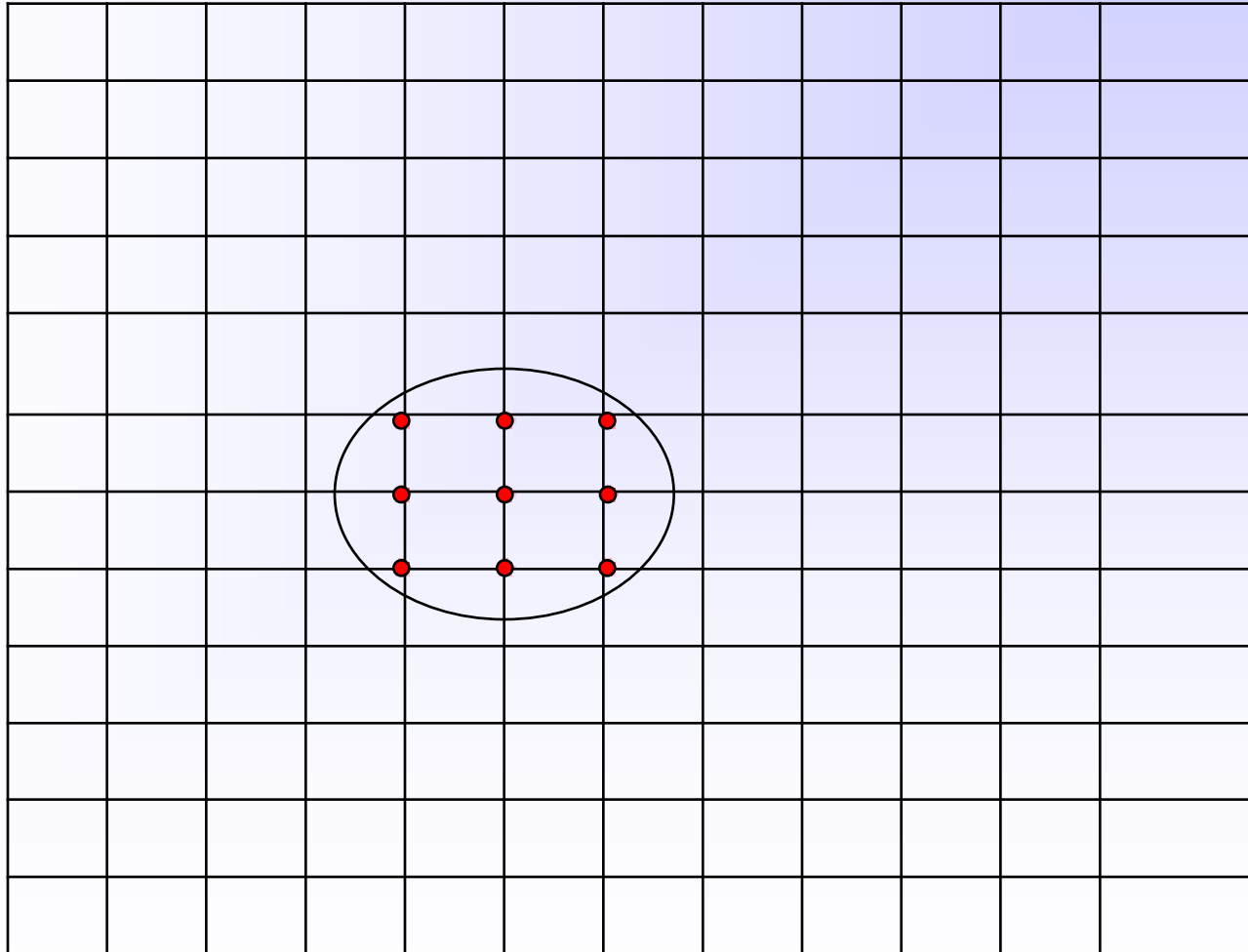
Let $d_i = \text{degree of vertex } i$. S is r -close-knit if

$$\min_{S' \subseteq S} \frac{d(S', S)}{\sum_{i \in S'} d_i} \geq r$$

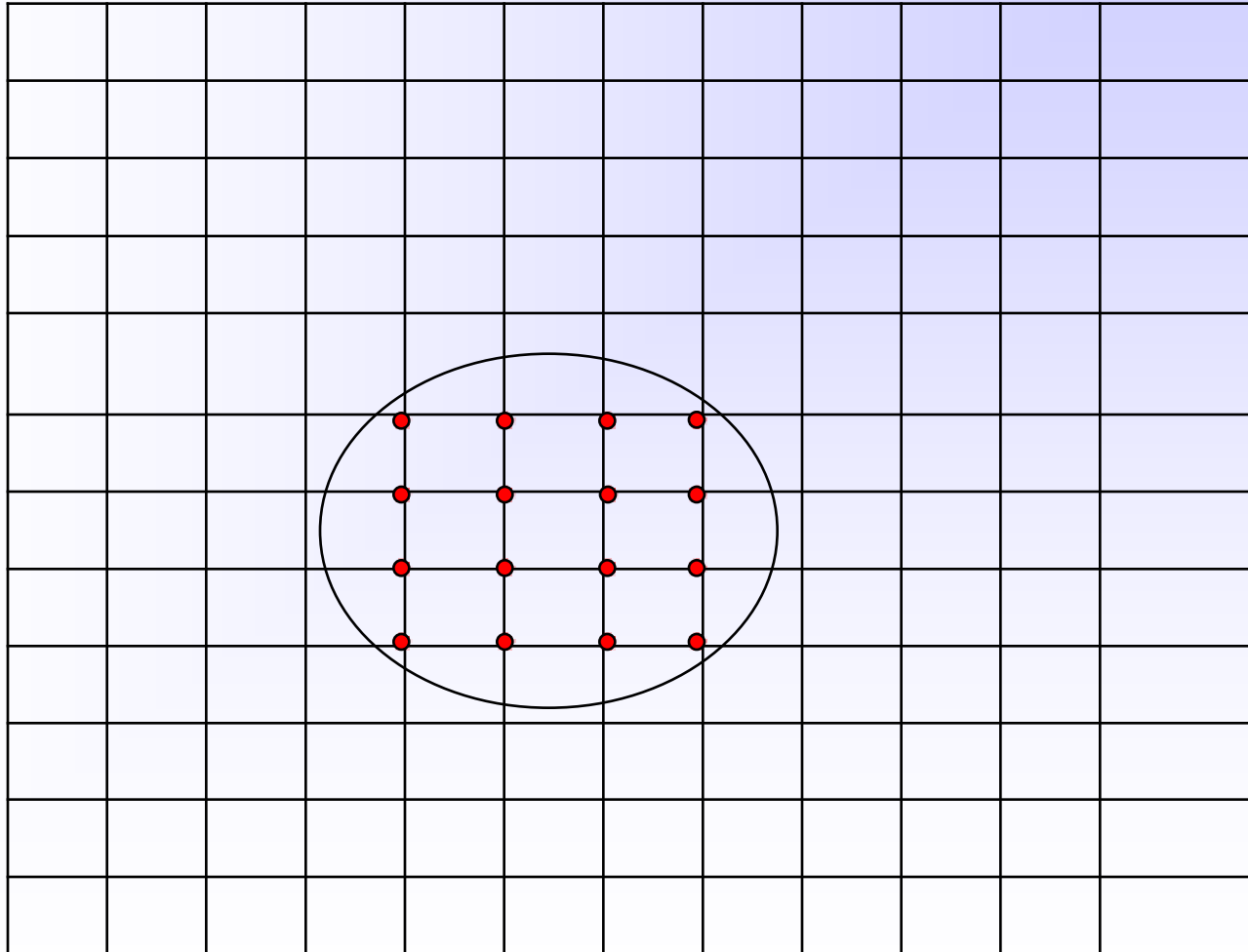
A family of networks \mathcal{F} is close-knit if for every $r < 1/2$ there exists $k(r)$ such that every network in \mathcal{F} , irrespective of size, is $(r, k(r))$ -close-knit

The family of two-dimensional lattices is close-knit : every square of side h is $(1/2 - 1/2h, h^2)$ -close-knit

$$\frac{d(S, S)}{\sum_{i \in S} d_i} = \frac{12}{36} = \frac{1}{3}$$



$$\frac{d(S, S)}{\sum_{i \in S} d_i} = \frac{24}{64} = \frac{3}{8}$$




Theorem 4. *Given any close-knit family of networks, and any 2 x 2 coordination game G , logit learning is scalable: for any small $\delta > 0$ the δ -efficiency is bounded independently of the number of agents n*

*"The Diffusion of Innovations in Social Networks,"
in Blume and Durlauf (eds.) The Economy as an Evolving
Complex System, vol III, Oxford University Press, 2006*

Open problems

- *Can one design decentralized learning algorithms that converge rapidly for important subclasses of games?*
- *Can one design algorithms that select the welfare-maximizing equilibrium?*
- *Is Nash equilibrium the right solution concept?*
- *Other forms of equilibria (coarse and 'coarse' correlated equilibria) are more easily achieved in decentralized learning environments.*



Further reading

Strategic Learning and Its Limits,
Oxford University Press, 2004

www.econ.jhu.edu/people/young/publications