



STRATEGIC LEARNING

Recent Advances and Open Problems

Peyton Young










The Question

When can rational players learn to play Nash equilibrium starting from out-of-equilibrium conditions?



The “classical” case

- The number of players is small
- The rules are common knowledge
- The payoffs, or at least the distribution of possible payoffs, is common knowledge
- Everyone is rational




Even in this high rationality world, learning is difficult because it is *interactive*: each player's learning process complicates what has to be learned by everyone else

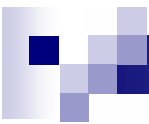
Kalai and Lehrer, Econ, 1993

Jordan, GEB, 1991, 1993

Nachbar, Econ, 1997, 2005




In fact, the learning behavior of a Bayesian rational agent can be so complex that it is essentially *unlearnable* by other rational players.



Theorem: *There are simple games of incomplete information (e.g., matching pennies with uncertain payoffs) such that, for any prior beliefs, Bayesian rational players will fail to come close to Nash equilibrium play with probability one.*

Furthermore at least one of the players will almost surely fail to learn how to predict the behavior of his opponent even approximately.


Foster and Young, Proceedings of the National Academy of Sciences, vol. 98, 2001.

- 
- Surely then it is *hopeless* to expect equilibrium to arise in situations where there are vast numbers of irrational agents who know little or nothing about the interaction structure or even who the other players are
 - But this is not the case
 - Play can equilibrate even though the agents are not rational, and do not learn about the system as a whole or even about each other



A “pinch” of rationality

Instead of taking a little bit away from high rationality, let's start with nothing and add a pinch



We'll demonstrate three types of adaptive rules that approximately equilibrate in the following situations:

- Potential games and weakly acyclic games
- Finite games with at least one pure Nash equil.
- Finite games with pure or mixed equilibria



Ground rules

- Agents adjust their behavior only in response to their own realized payoffs
- They have no knowledge of the overall structure of the game
- They cannot observe the actions or payoffs of most other players – perhaps of *any* other players
- They occasionally tremble and make mistakes



Applications

- Driving to work in a big city
- Trading commodities on the Chicago Board of Trade
- Routing packets between information processors in large networks

Notation

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

Action spaces A_i

Joint action space $A = \prod_i A_i$

Utility functions $u_i : A \rightarrow R$



Rule I : 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


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

Realized payoff $> \bar{u}_i(t) \Rightarrow i$ adopts experimental action and corresponding payoff as new benchmarks

Otherwise i keeps the previous 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}


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

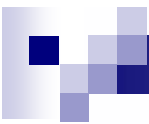
Every potential game is weakly acyclic



Examples of potential games

- Congestion games
- Cournot oligopoly games
- Schelling segregation models
- Decentralized control problems (mobile sensors, packet routing, etc.)

- 
- Better reply dynamic: one player moves at a time and chooses at random a better reply (if any) to the current strategies of the others
 - In weakly acyclic games, the better reply dynamic converges with probability one to a pure Nash equilibrium.
 - But this assumes that each player *knows* what actions the other players are taking.
 - Simple experimentation assumes only that players react to their *own* recent payoffs.



Theorem. *Given a finite n -person weakly acyclic game G , if all players use simple experimentation with experimentation rate $\varepsilon > 0$, then for all sufficiently small ε 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, forthcoming.



Rule II: Interactive Trial and Error Learning


Agents change behavior according to their “mood”

content

discontent

hopeful

watchful




Player i 's *state* has three aspects:
mood, benchmark action, benchmark payoff

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

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



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 with the current action and payoff as new benchmarks



3, 2

2, 3

1, 0

2, 3

3, 2

1, 0

0, 1

0, 1

1, 1

3, 2


2, 3

1, 0

2, 3

3, 2

1, 0

0, 1


0, 1

1, 1

3, 2

2, 3

1, 0

2, 3
☹️ ☹️

3, 2

1, 0

↑
 $O(\epsilon)$

0, 1
☹️ ☹️

0, 1

1, 1

3, 2

2, 3

1, 0

2, 3
☹️ ☹️

3, 2

1, 0

↑
 $O(\epsilon)$

0, 1
☹️ ☹️

0, 1

1, 1

3, 2
☹️ ☹️

↑
 $O(\epsilon)$

2, 3
☹️ ☹️

↑
 $O(\epsilon)$

0, 1
☹️ ☹️

2, 3

3, 2

0, 1

1, 0

1, 0

1, 1

3, 2
☺ ☹

↑
 $O(\epsilon)$

2, 3
☺ ☺

↑
 $O(\epsilon)$

0, 1
☺ ☺

2, 3

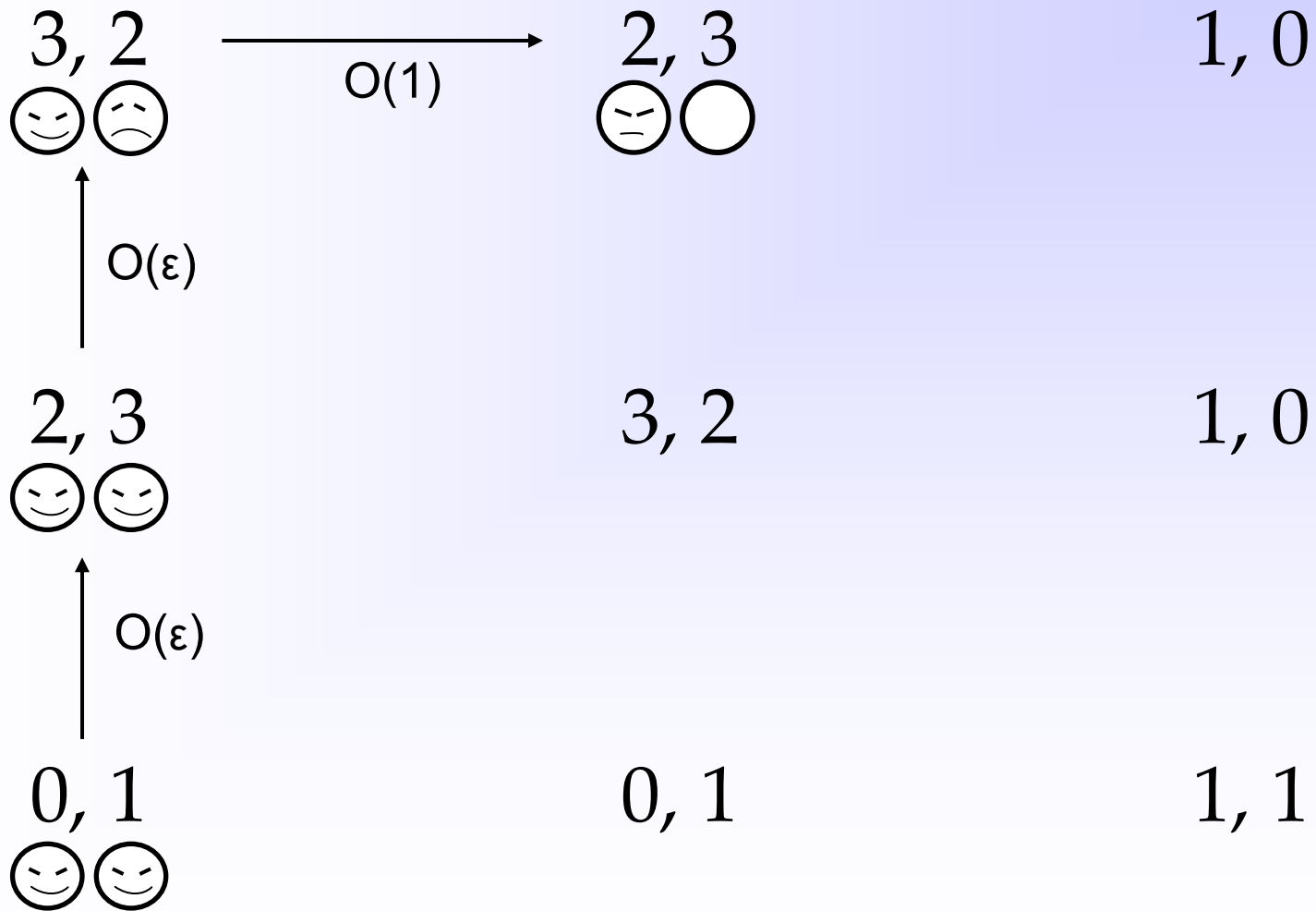
3, 2

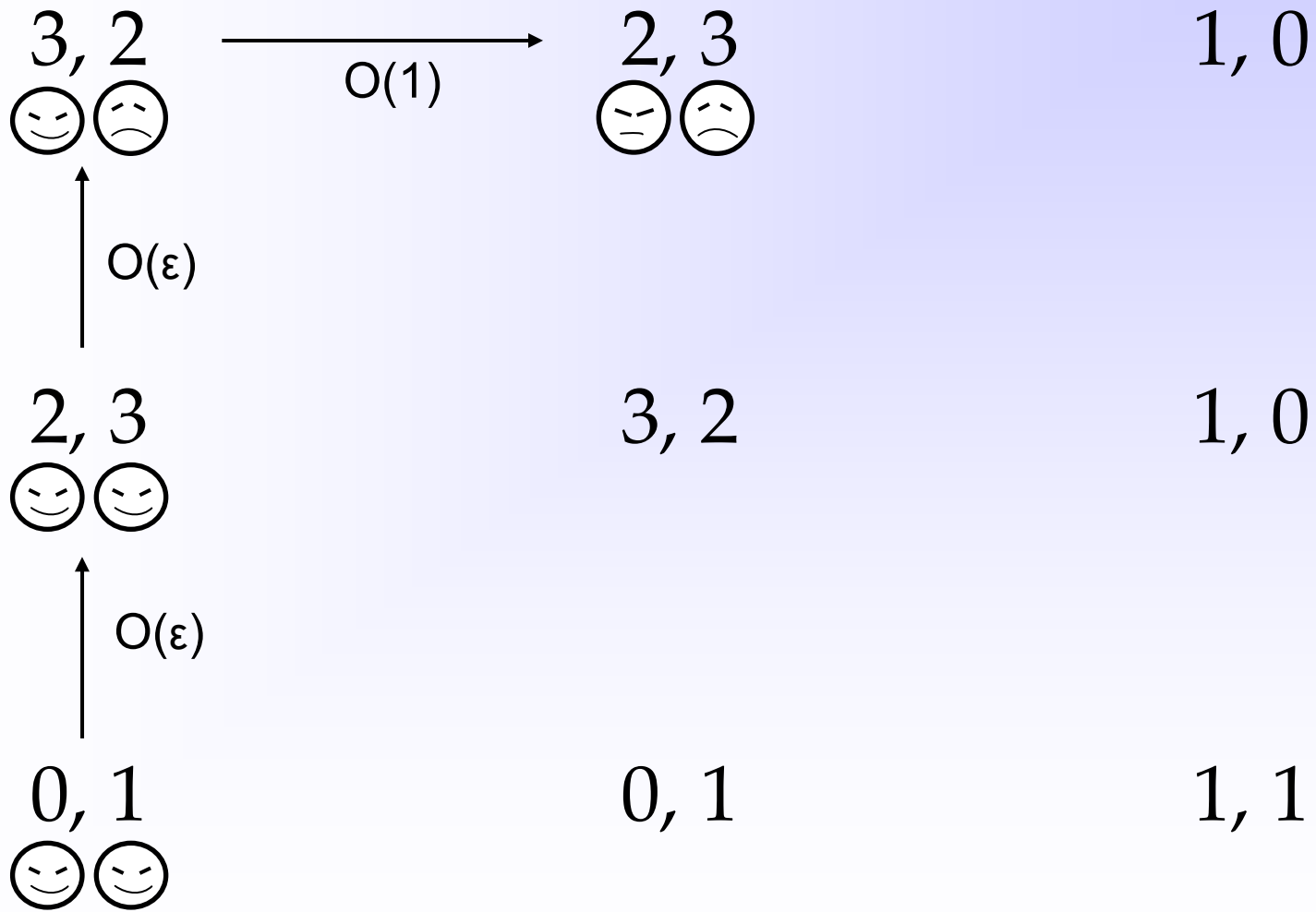
0, 1

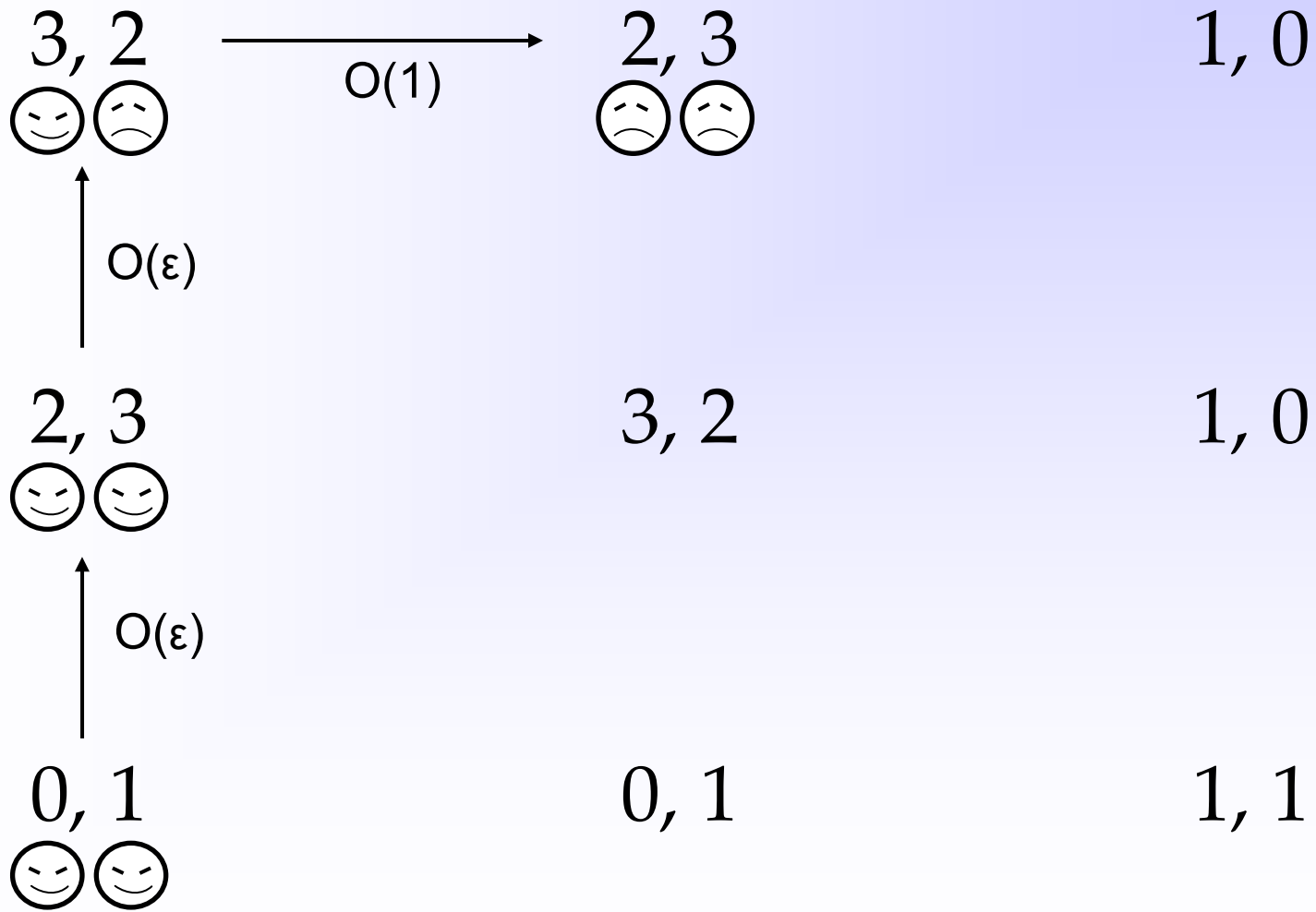
1, 0

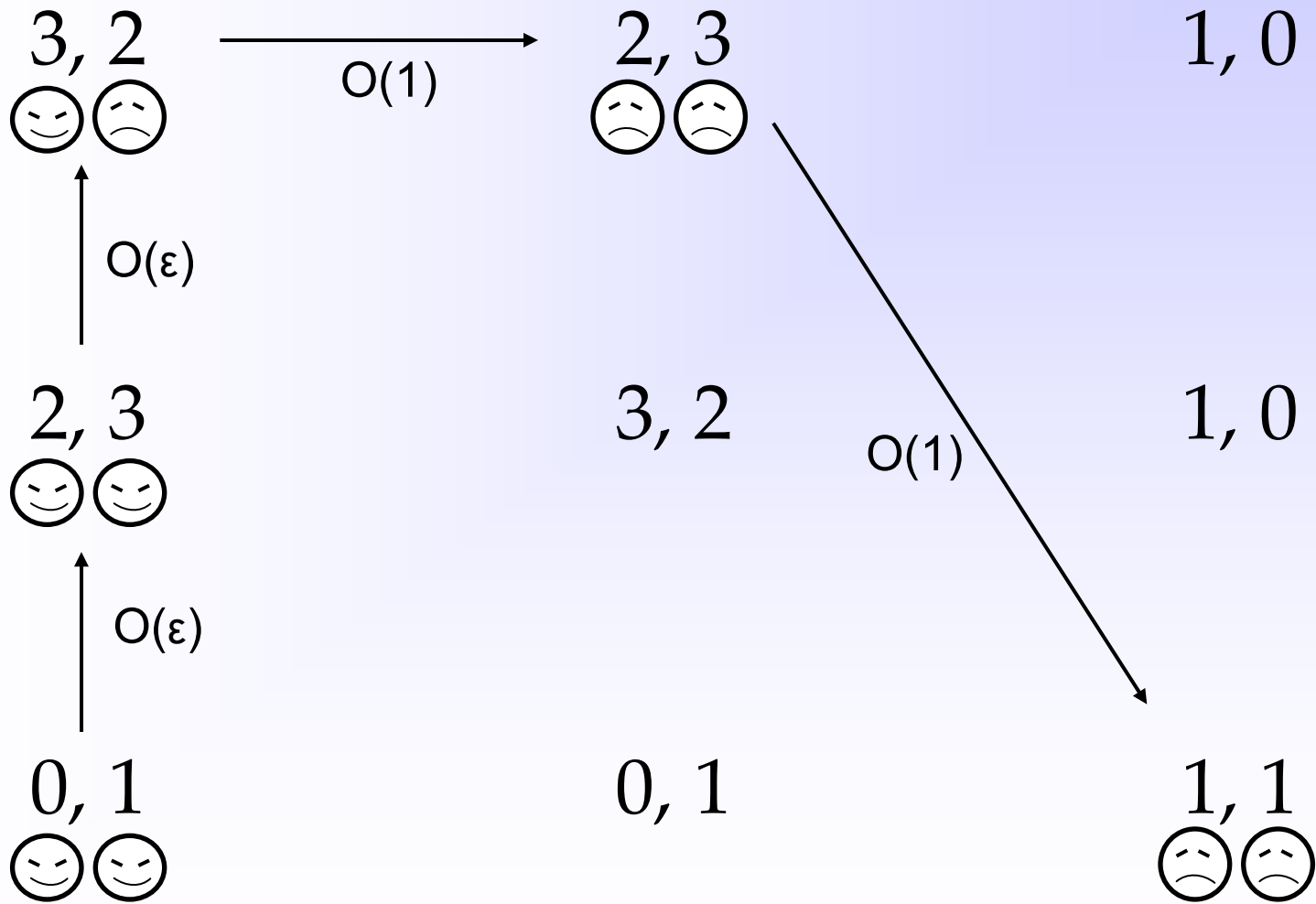
1, 0

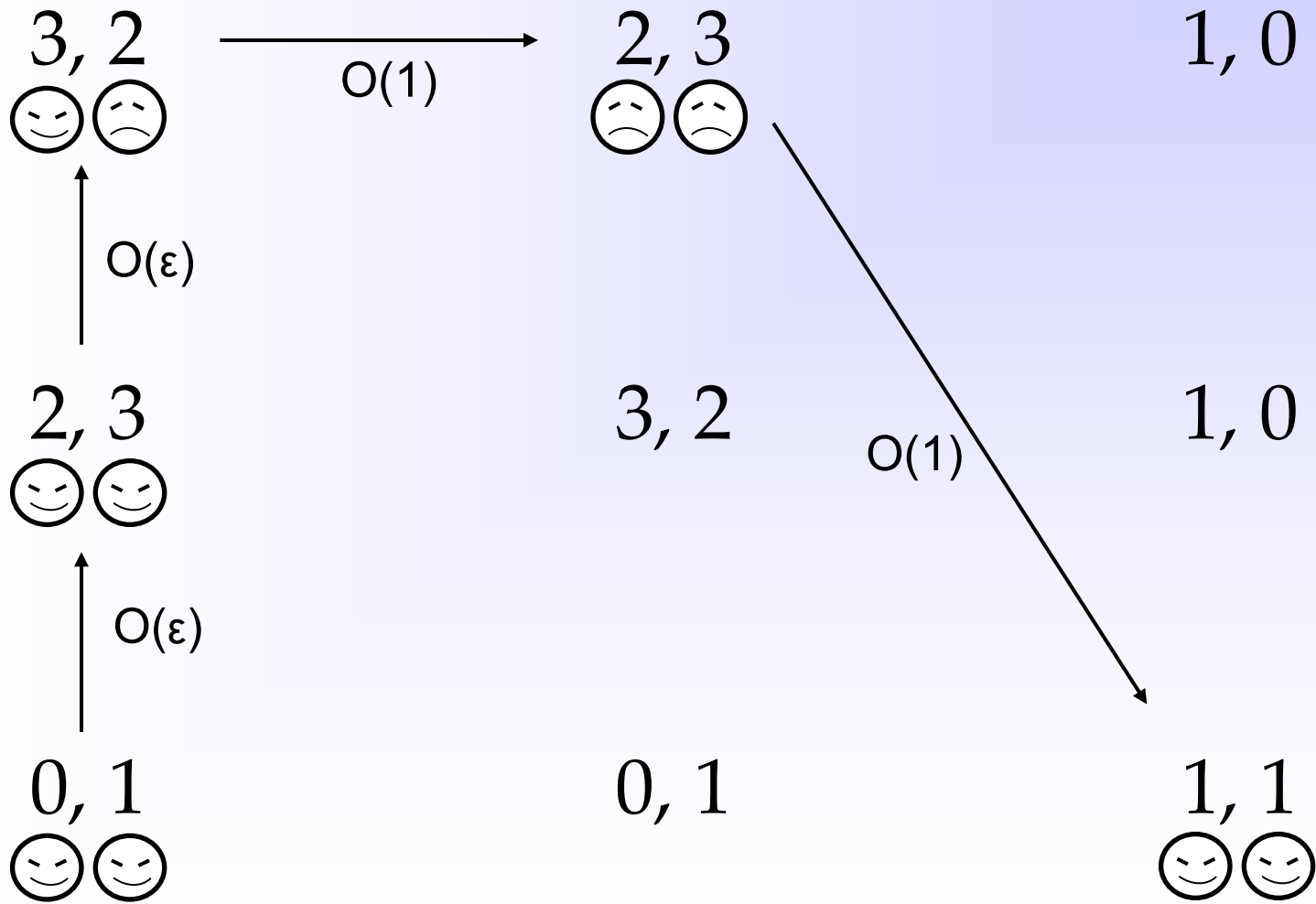
1, 1

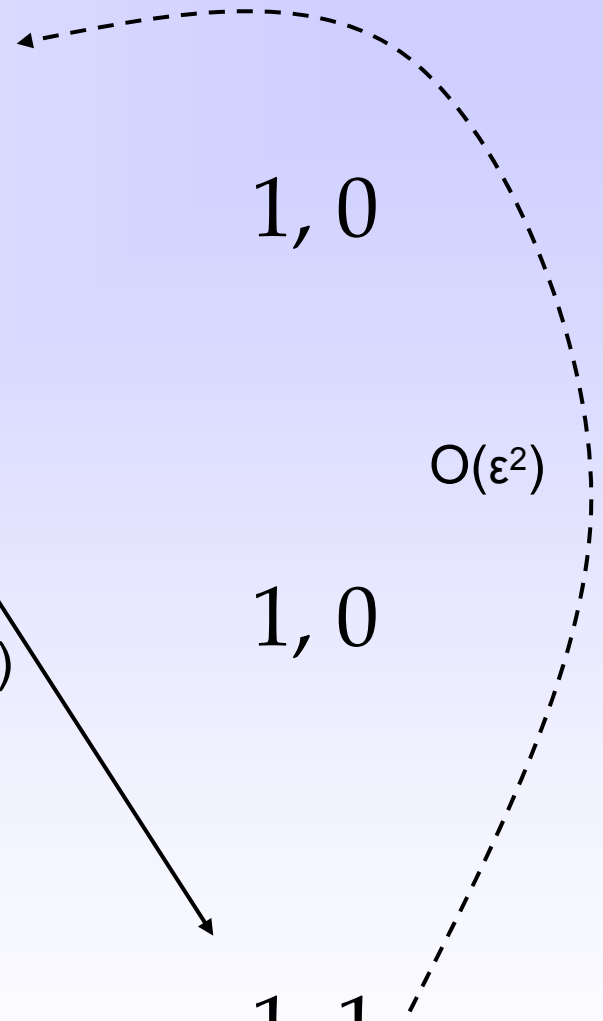
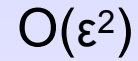
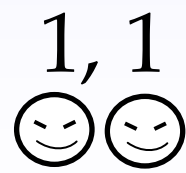
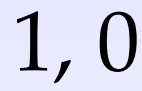
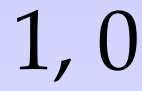
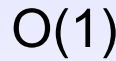
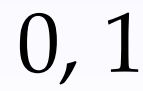
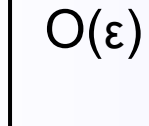
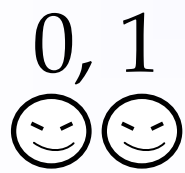
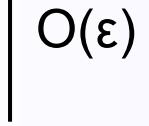
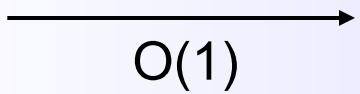














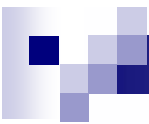


Let Γ_A be the set of all n -person games on finite action space A that possess at least one pure Nash equilibrium.



Theorem. *If all players use interactive trial and error learning and the experimentation rate $\varepsilon > 0$ is sufficiently small, then for almost all games in Γ_A a Nash equilibrium is played at least $1 - \varepsilon$ of the time.*


H.P. Young, Learning by Trial and Error, Games and Economic Behavior, 2008.



Can players learn mixed equilibria?

Rule III: Regret Testing

- An *experiment* consists of computing the average payoff over a large random sample of plays and comparing it with the average payoff from the current strategy
- If the sample average is better by some small amount $\tau > 0$ (the *tolerance level*), switch to the better strategy with high probability, and switch to an *arbitrary* new strategy with small probability



Theorem. *If experiments are rare, the tolerance τ is small, and the sample size is large, the players' behaviors constitute an ε -equilibrium of G at least $1 - \varepsilon$ of the time.*

Foster and Young, Theoretical Economics, 2006

Germano and Lugosi, GEB, 2007



Current research problems

- For what classes of games can one design algorithms that get close to equilibrium really quickly?
- What are the theoretical limits to the rate at which convergence to equilibrium can occur?
- Empirically what types of adaptive behavior do people use in large complex games?
- Is Nash equilibrium the right concept for predicting long-run behavior in such games?



For more on learning rules go to

- **Sergiu Hart's plenary talk Monday, 17:30**
- **Session 50: Monday, 8:30-10:50**
- **Session 199: Thursday 12:00-13:20**



Further reading

- Foster and Vohra, “Calibrated learning and correlated equilibrium,” *GEB*, 1997
- Hart and Mas-Colell, “Uncoupled dynamics do not lead to Nash equilibrium,” *AER*, 2003
- Foster and Young, “Learning, hypothesis testing and Nash equilibrium,” *GEB*, 2003
- H. P. Young, *Strategic Learning and Its Limits*, Oxford University Press, 2004

