

# Economists' Models of Learning

## What are agents trying to learn?

- Probability of exogenous events (“Nature’s moves”)
- Own preferences (can be a special case of the above)
- The distribution of opponent’s play (“learning in games”)

## What can agents observe ?

- Their own payoffs
- Realizations of moves by Nature
- Other players’ actions
- Other player’s payoffs

Do players have any influence on what they observe?

If not, “passive learning.”

If so, “active learning,” with a trade-off between “exploration and exploitation.”

What determines the environment players are learning about?

- Environment is directly specified as part of the model; it can be i.i.d. or given by a more complex stochastic process.
- Equilibrium play
- The interaction of non-equilibrium “learning rules.”

## What determines players learning rules and behavior?

- Equilibrium play
- Worst case/minmax considerations
- Bayesian, Savage-rational maximization in a non-equilibrium setting
- Exogenously specified “boundedly rational”

# Three Examples

## 1: Learning in Extensive Form Games

Starting point: Equilibria in games arise as the long-run outcome of a non-equilibrium dynamic process of learning, imitation, or evolution.

Suppose that agents are repeatedly matched to play the game against a sequence of anonymous opponents.

Agents don't know their opponents' strategies, they are learning about them. (The models mentioned below make varying assumptions about how this learning takes place.)

And suppose that all agents observe is the outcome of play in their own matches. They don't observe:

- what happens in other matches,
- opponents' past play,
- their opponents strategy.

If players are good at “passive learning,” and play converges to a steady state, we expect the players to learn the path of play.

But Nash equilibrium is “as if” players know the equilibrium path and the consequences of unilateral deviations from the equilibrium path.

In order to learn this, players must have “enough” observations of off-path play to learn the consequences of deviating.

Equilibrium refinements such as subgame-perfect equilibrium are “as if” players know play throughout the entire game tree. This requires “enough” observations of play at most information sets, not just those that can be reached by a single deviation.

How much off-path play is needed for various refinements, and how much off-path play should we expect to see?

Various answers in the literature:

Fudenberg and Kreps [88, 95, 98] make assumptions directly on the frequency of experimentation. The papers give various sufficient conditions on the “amount of experimentation” to get to Nash equilibrium.

Hart [2002], Jehiel-Samet [2004] analyze other sorts of exogenously-specified behavior rules, with enough exogenous experimentation to force the system to the subgame-perfect equilibrium in perfect information games.

Fudenberg and Levine [93, 2004]:  
Bayesian-rational learning by agents  
who solve the (fairly complicated) “multi-  
armed bandit problem.”

Findings: Impatient agents don't  
experiment enough to get to Nash  
equilibrium, patient ones get to Nash but  
not necessarily to subgame-perfect.

## Example 2: “Social Learning”

Players learn by talking to others.

When players talk, they don't communicate their entire sequence of observations, just their actions and their current or average payoff.

Usually studied in a “non-strategic” setting, where the payoff to an action (choice of pediatrician, movie, crop) is assumed to not depend on the number of other agents who choose the same action.

Also so far usually studied in models where each agent moves only once (so no experimentation!)

Questions: when will agents “herd” and all choose the same thing? When will the system asymptotically do as well as if players pooled all their information?

Banerjee [1992], Bikchandani et al [1992]: Inefficient herding can occur in Bayesian equilibrium when one agent moves at a time and each agent knows all past choices.

Ellison-Fudenberg [95,98] analyze simple “rules of thumb” in models where many agents move simultaneously.

## Examples of these rules:

- “contact N people at random and pick the action in the sample that had the highest average payoff;”
- “pick the action that had the highest average payoff at locations “nearby”,
- “pick the action that maximizes a weighted sum of average payoff and popularity.”

Long-run average payoff is higher with “popularity weighting,” because popularity acts as a form of “social memory:” If the action is really popular than it was probably good in the past.

Banerjee-Fudenberg [2004] look at a Bayesian equilibrium model of this “word of mouth,” and find that rational Bayesians also use a form of popularity weighting.

*Application to peer ratings?*

### **Example 3: Universal Consistency**

(an example of a non-Bayesian, worst-case rule, and also of the benefits of interdisciplinary conferences.)

Imagine that an agent has to take an action every period while he is learning about the environment. (The “on-line decision problem.”)

A learning rule is consistent if it does at least as well as the rule “play a best response to the long-run average of other players (opponents and/or Nature’s) play” *whenever* this play is drawn from an exchangeable distribution.

A Bayesian who believes the environment is exchangeable will use a consistent rule.

But in game dynamics, opponents' play is typically not exchangeable.

Say that a rule is "universally consistent" if the player does as well the rule "play a best response to the long-run average of other players (opponents and/or Nature's) play" regardless of whether opponents' play is exchangeable. (hence "universal.")

Universal consistency is a non-Bayesian, worst-case criterion.

No deterministic rule is universally consistent. (think of the game "matching pennies!")

But some randomized ones are.

This was first shown by Blackwell in the 1950's, then rediscovered by computer scientists (e.g. Nareeda and Turner) and by FL.

FL also showed that a very simple rule is  $\varepsilon$ -universally consistent, namely "exponential fictitious play." This rule says to compute the expected payoffs of each action versus the historical frequencies of opponent's play, and then play actions with probabilities equal to the logit transforms of these payoffs, so that the best action is played with probability near 1.

Note that a similar construction works when players only observe their own payoffs, and not the play of opponents; exponential fictitious play here yields a form of reinforcement learning.

