Lecture 2: Learning and Equilibrium Extensive-Form Games

III. Nash Equilibrium in Extensive Form Games
IV. Self-Confirming Equilibrium and Passive Learning
V. Learning Off-path Play

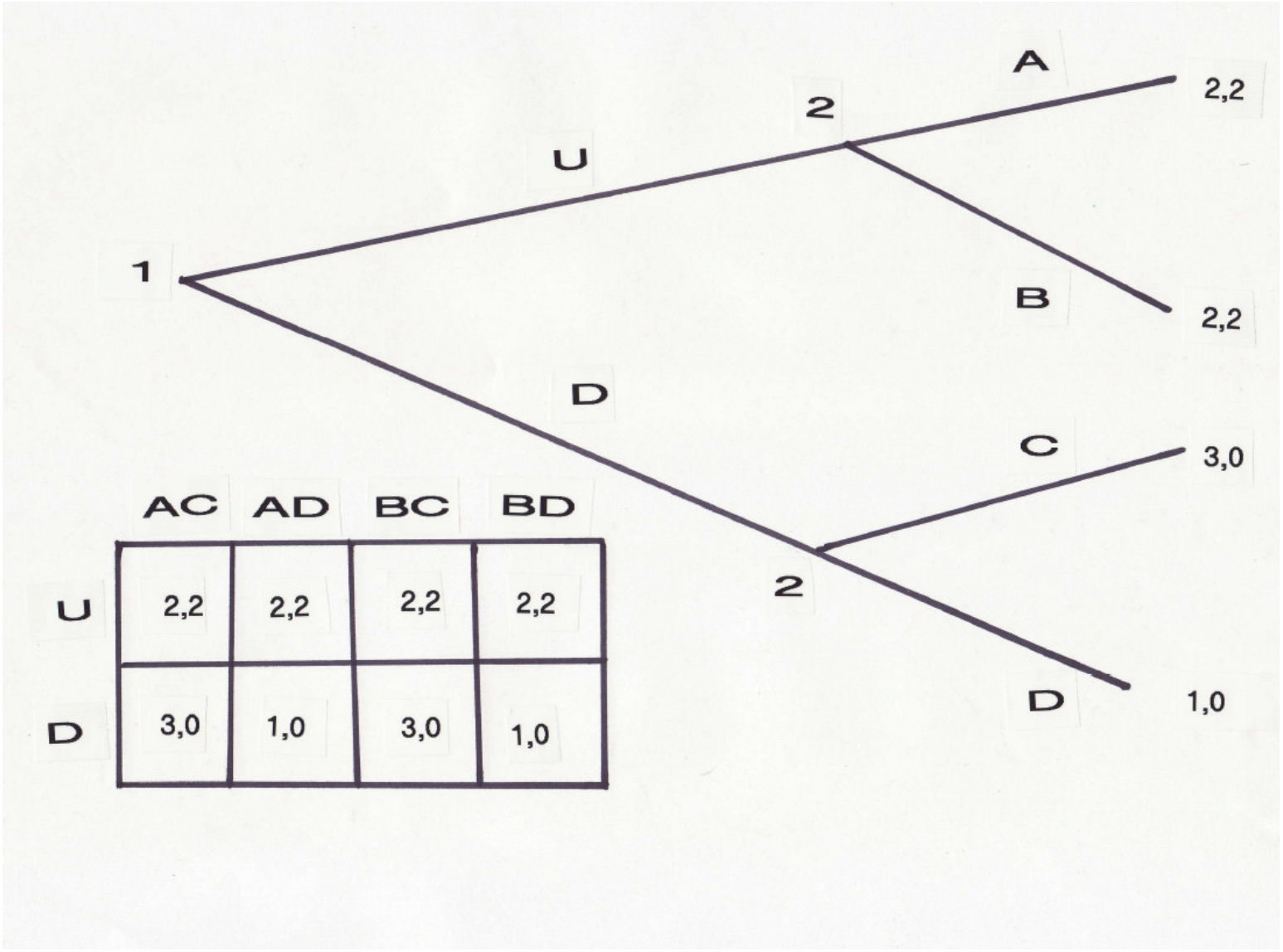
D. Fudenberg
Marshall Lectures 2009

III. Extensive-Form Games

- Strategies are "complete contingent plans" that specify the action to be taken in every situation (*i.e. at every "information set")* that could arise in the course of play.

- The *actions* that a player actually chooses may end up depending actions of the others, but we can think of players simultaneously choosing *strategies* before the game is played.

- Associate a unique strategic form with given extensive form.



	AC	AD	BC	BD
U	2,2	2,2	2,2	2,2
D	3,0	1,0	3,0	1,0

- The definition of Nash equilibrium applies without change:

It is a strategy profile such that no player can increase their payoff by changing their strategy, holding fixed the strategies (*not* the actions) of the other players.

- But the long-run outcome of learning depends on what the players observe when the game is played.

- In the strategic form player 1 observes player 2's strategy, and learning leads to Nash equilibrium as before.

- In the extensive form 1 sees only player 2's action. Here learning only leads to the larger set of **self-confirming equilibrium.** (Fudenberg and Levine [1993]).

IV. Self-Confirming Equilibrium

Preliminaries

- For a given strategy profile, an information set is "reached" if it has positive probability.
- It is "unreached" or "off-path" if it has probability zero.
- Let probability measure μ_i describe player i's beliefs about his opponents' play.
- Let $\sigma = (\sigma_1, \ldots, \sigma_I)$ denote a *mixed strategy profile*, where each σ_i corresponds to a probability distribution over the pure strategies of player *i*. *(either because player i is randomizing or because there are many agents in the role of player i and they don't all play the same way.)*

If players don't observe opponents' play at unreached information sets, they may not learn it, so their actions may not be optimal given the way that opponents respond to deviations.

Self-confirming equilibrium requires that each agent's beliefs are correct at the information sets that are reached given their play; but not necessarily correct; this is in the spirit of Hahn's [1977] "conjectural equilibrium."

More formally,

Definition: σ is a *self-confirming equilibrium* (SCE) if for each player i and each strategy s_i with $\sigma_i(s_i) > 0$ there are beliefs $\mu_i(s_i)$ such that

(a) s_i maximizes player i's payoff given his beliefs $\mu_i(s_i)$,

and

(b) $\mu_i(s_i)$ is consistent with what player i sees when he plays s_i. (*More formally,* $\mu_i(s_i)$ *is correct at every information set that is reached by* (s_i, σ_{-i}).)

Notes:

- Nash equilibrium : play maximizes payoff given opponents' strategies.

  This is equivalent to

  (a) (play maximizes payoffs given beliefs)

  and a more restrictive condition on beliefs

  (b') each player's beliefs are correct at every information set.

- SCE reduces to Nash equilibrium in static games, as there seeing an opponent's chosen action is the same as seeing the opponent's strategy.

- In some cases players might not even observe play at all of the information sets that are reached.

  For example, in a sealed-bid first-price auction players might observe the winning bid but neither the values of the other players nor the bids of the losers.

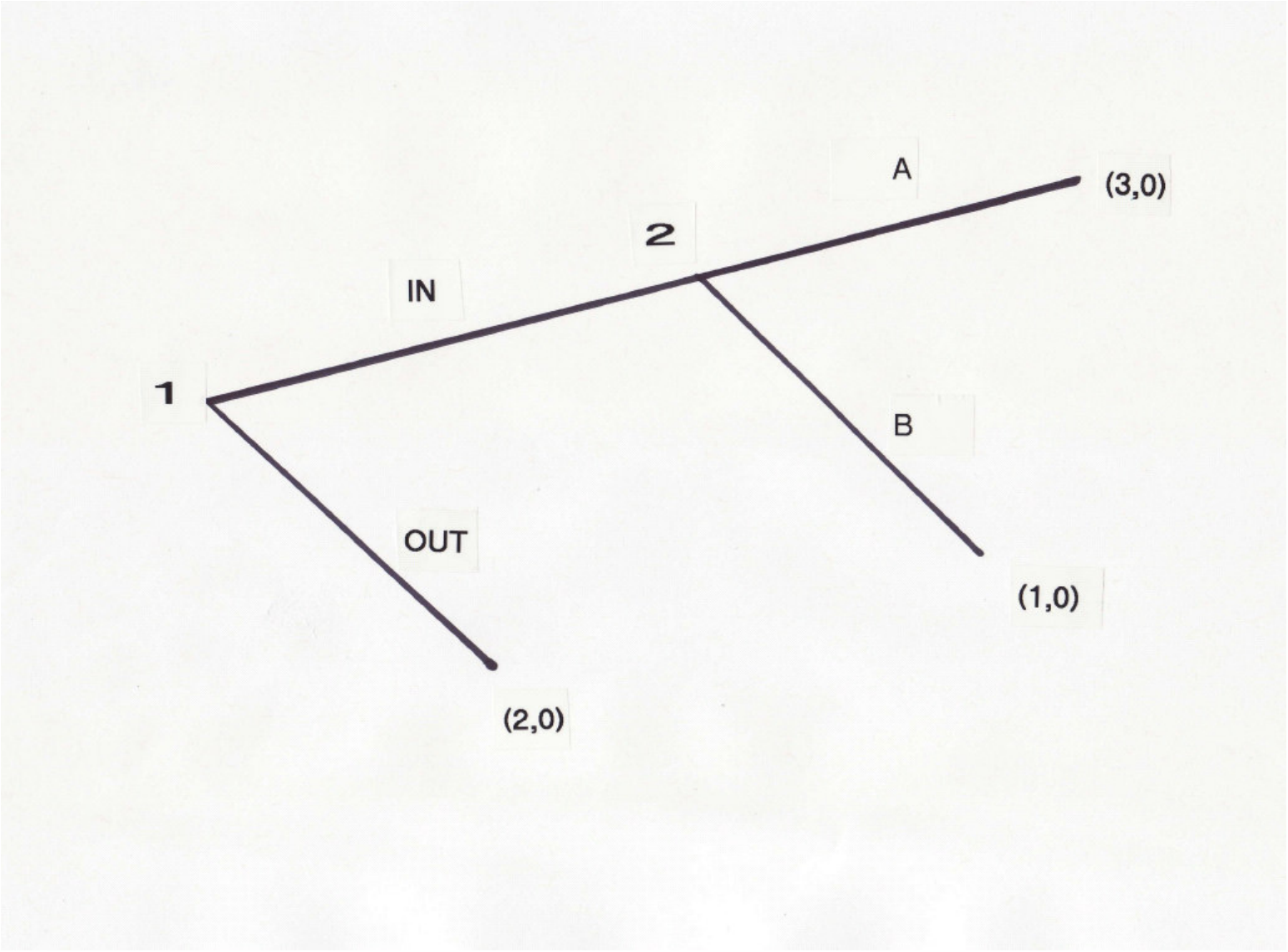
  This can be modeled with a generalization of SCE, that only requires beliefs to be consistent with a smaller set of observations; see Dekel, Fudenberg and Levine [2004].

- The definition of SCE allows player *i* to use a different belief to rationalize each strategy s_i that has positive probability under σ_i.

- These "heterogeneous beliefs" are important in many game theory experiments, possibly also important in the field.

- Following game allows a simple example of the impact of heterogeneous beliefs:

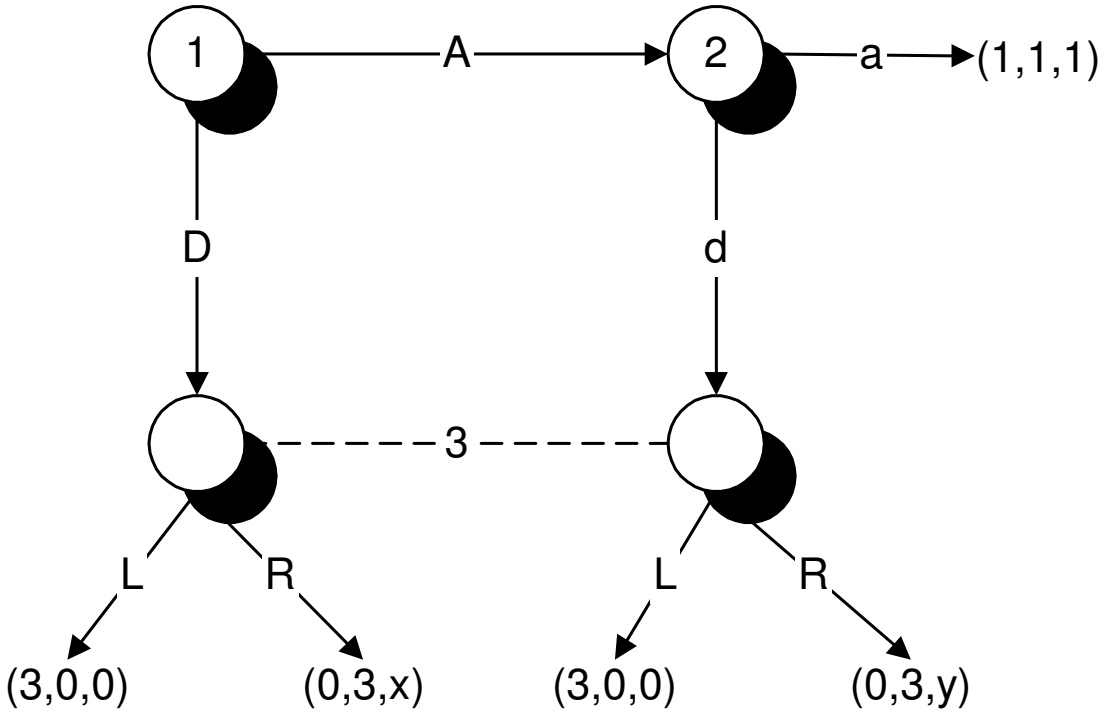
  No Nash equilibrium with outcome distribution
  (½ Out, ½ (In, A))


A
(3,0)
2
IN
1
B
OUT
(1,0)
(2,0)

- "Unitary" SCE requires a single belief μ_i for each player *i*.

(relevant with a single agent per role or if agents pool their observations)

- There can be unitary SCE that are not Nash equilibria because two players disagree about the play of a third…



- (A,a) is the outcome of a self-confirming equilibrium.
- It is not the outcome of a Nash equilibrium.
- Even if players 1 and 2 know x and y, they need observations to learn 3's play if x and y have opposite signs.

- SCE: the only constraint on player's predictions is their observations of play in the game.

- *Rationalizable SCE* (Dekel, Fudenberg, Levine [1997]) combines SCE with the idea that player know the payoff functions of their opponents and expect them to play rationally-*provided* that the opponents haven't yet done anything "irrational."

- Unitary RSCE coincides with backwards induction in two-stage games of perfect information, but in longer games it is much weaker and more like SCE.

Apply self-confirming equilibrium to the analysis of experiments:

Play can be depart from Nash equilibrium even if players are fully rational.

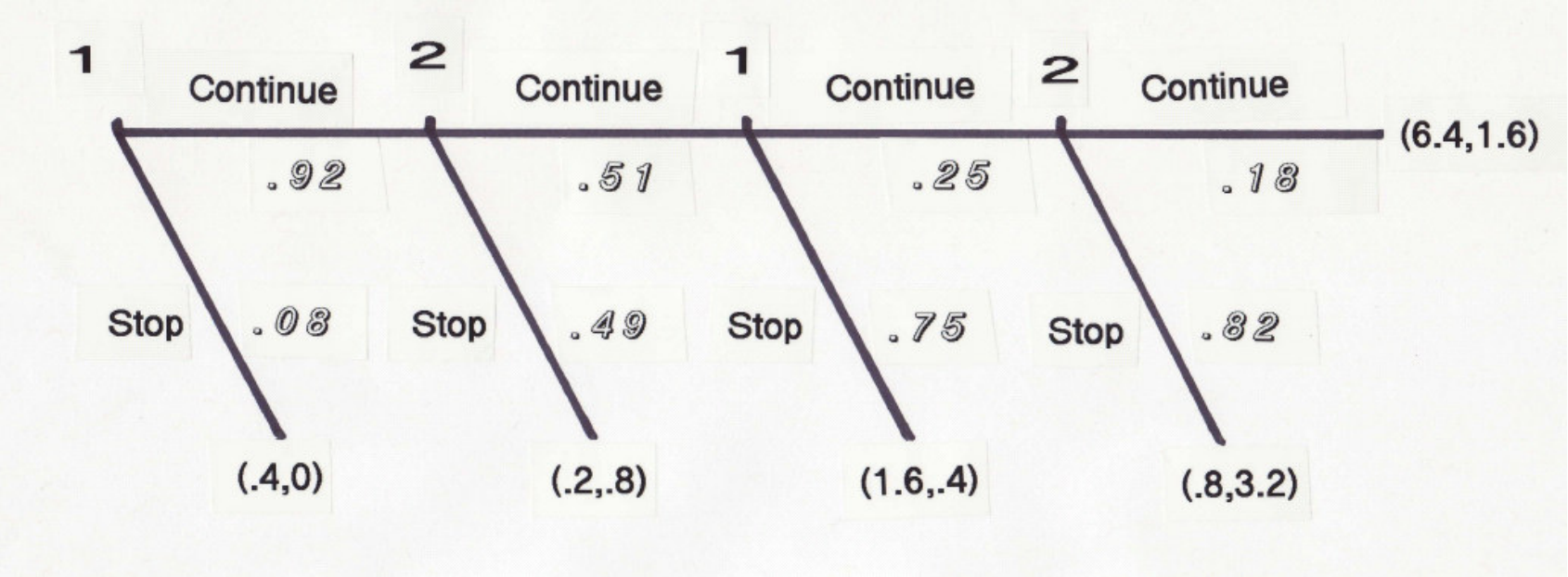
We should distinguish between play that is inconsistent with maximizing the agents' presumed utility functions and play that is optimal for self-confirming beliefs.

Ultimatum bargaining:

- In experiments, low offers are rejected: the rejecters are not maximizing money payoff in the experiment.
- But not all player 1's make the payoff-maximizing choice, perhaps because they don't know the exact probability that each offer is accepted. (which in turn suggests that here players do not start the experiment knowing the true distribution of preferences in the population.)
- Literature has focused on the losses of the player 2's.
- But the "self-confirmed" losses of the player 1's are on average 3-5 times greater than the losses of the second mover.
- These losses are not consistent with Nash or Bayesian equilibrium with correct beliefs about the distribution of player 2 preferences.

Centipede game: McKelvey-Palfrey [1992]

- Players take turns choosing "stop" or "continue."
- If they continue, the social value doubles.
- It is better to stop today if you expect the next player to stop.
- If the last player only cares about money payoffs, she will stop; backwards induction then implies that players should stop at every node.


1
Continue
2
Continue
1
Continue
2
Continue
(6.4,1.6)
.92
.51
.25
.18
Stop
.08
Stop
.49
Stop
.75
Stop
.82
(.4,0)
(.2,.8)
(1.6,.4)
(.8,3.2)

Given the play in the experiment, the best choice for player 1's is to continue at both information sets, and the player 2's should continue at their first information set.

Most player 1's continue at the first information set: the ones who stop (and so play according to backwards induction) are making a mistake given actual play.

The observed heterogeneous play in this game is not consistent with Nash equilibrium, nor with unitary SCE, but it is consistent with heterogeneous SCE. For example, a player 1 who always plays stop will not learn that continue is better.

General Point: Heterogeneous beliefs about off-path play are important in explaining data from game theory experiments, especially in cases where subjects have doubts about each other's preferences.

And in many lab settings it is hard to see how the subjects could know the distribution of opponents' preferences since even the experimenters don't.

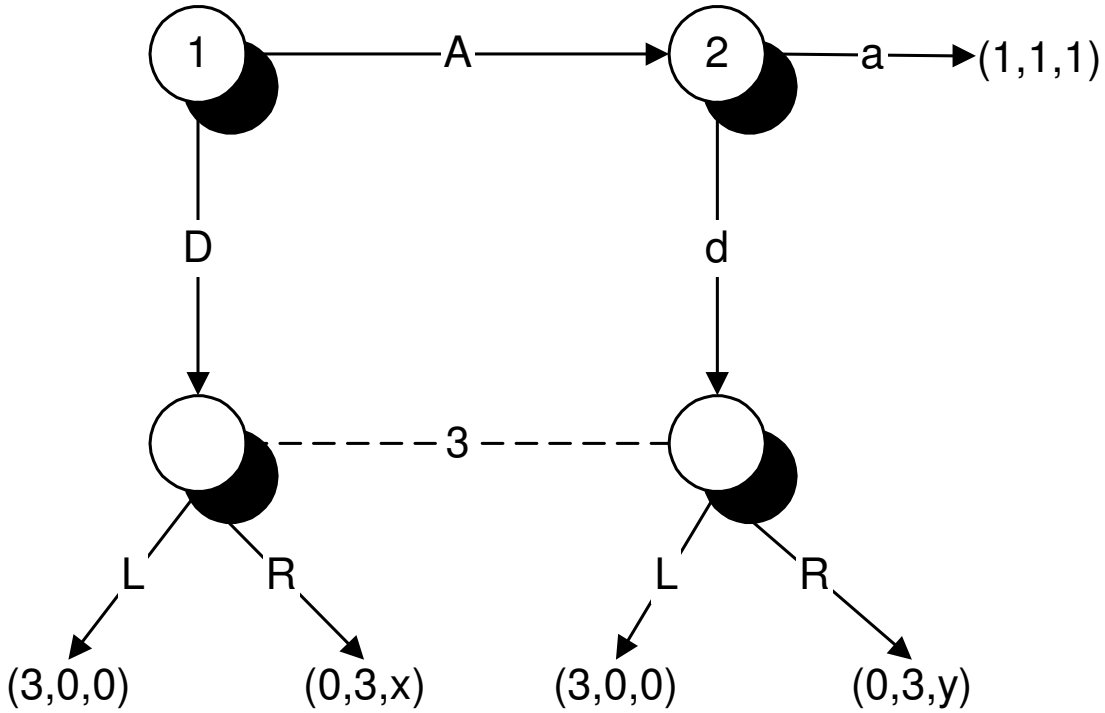
V. Learning Off-Path Play

- Self-confirming equilibrium is consistent with passive learning.
- Rational learning does lead to Nash equilibrium if subjects get "enough" information about off-path play, either exogenously or by experimenting" with off-path actions.
- One simple condition for learning to rule out non-Nash outcomes: suppose agents "experiment" at rate $1/t$ with actions that don't maximize short-run expected payoff given beliefs. (Fudenberg and Kreps [1988]).

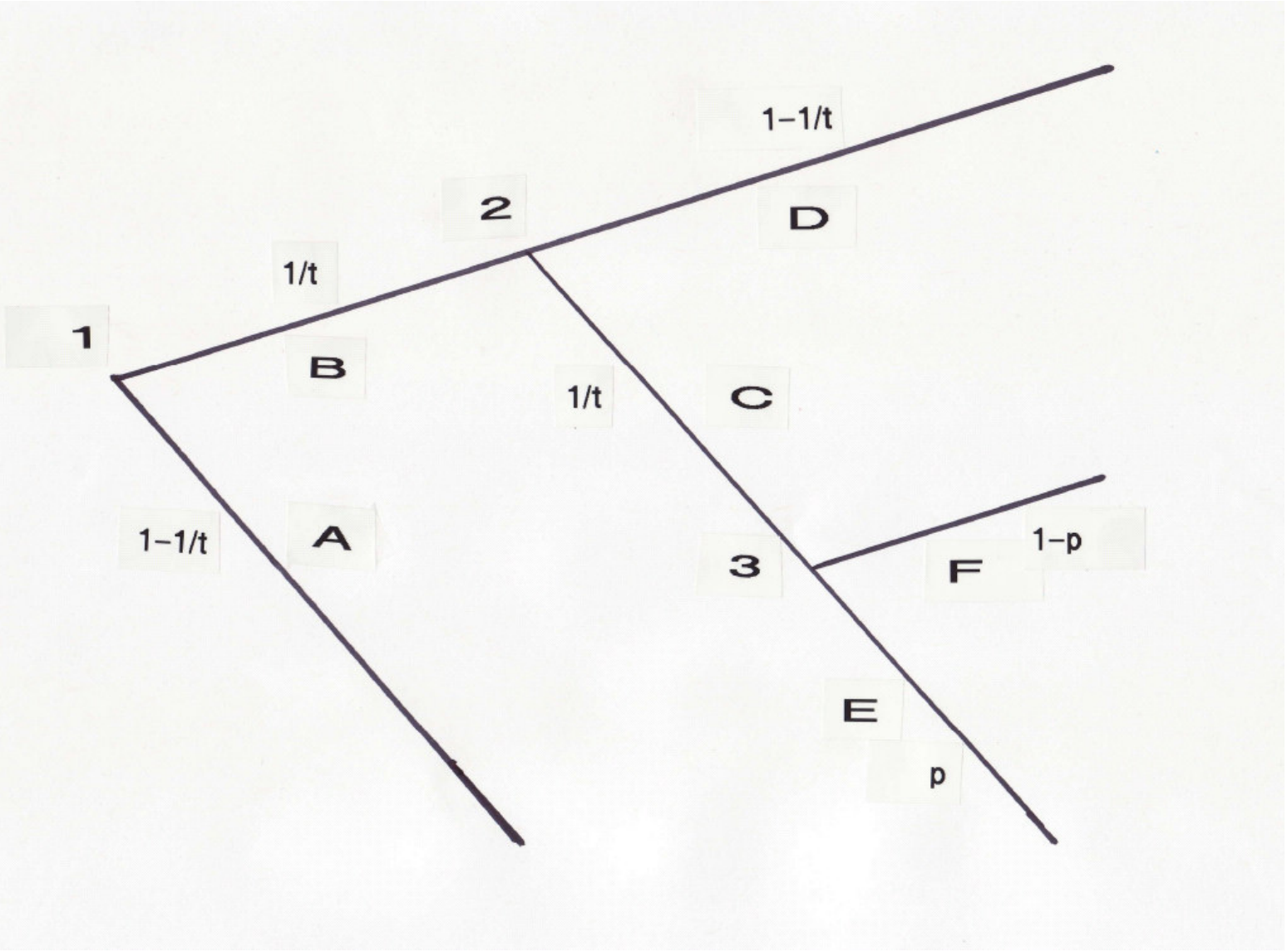
Reason: To rule out convergence to non-Nash outcomes, it is enough that players have correct beliefs about play at any "relevant" information set- these are the information sets that can be reached if any one player unilaterally deviates from the equilibrium path.

With the "$1/t$ experimentation rule," these relevant information sets are reached infinitely often
($\sum_{t=1}^{\infty} 1/t = \infty$ and Borel-Cantelli).

Asymptotic empiricism and the law of large numbers:
If players get an infinite number of observations of play at an information set, their beliefs about play at that information set become correct, which rules out the non-Nash outcome here:.


1
A
2
a
(1,1,1)
D
d
3
L
R
L
R
(3,0,0)
(0,3,x)
(3,0,0)
(0,3,y)

Although $1/t$ experimentation implies that every "relevant" information set is hit infinitely often, it does not imply that imply *all* information sets are reached infinitely often, because $\sum_{t=1}^{\infty} 1/t^2 \neq \infty$.


1
2
3
1/t
1−1/t
1/t
1−1/t
1−p
p
A
B
C
D
E
F

Thus $1/t$ experimentation leads to correct beliefs about what happens following a single deviation from the equilibrium path, but needn't lead to correct beliefs at every information set.

Backwards induction/subgame perfection implicitly suppose that beliefs are correct at <u>every</u> information set.

In last example player 2 can learn player 3's play if 2 experiments "more" than $1/t$.

Note: Bayes-rational agents won't choose to experiment at random.

Question: How much experimentation will agents choose to do, and which information sets will they learn about?

Bayes-rational learning:

- Agents are utility maximizers.
- They know they will play the game many times, and try to maximize the expected discounted value of their payoffs.
- Know the extensive form of the game (except perhaps the distribution of Nature's moves) and are trying to learn the distribution of opponents' play

Strategic myopia: Agents believe the distribution of aggregate play is exogenous and they don't try to influence it.

They also believe that the distribution of play is in a steady state, and the prior over steady states is "non-doctrinaire." So beliefs are *asymptotically empirical.*

Each agent faces a sort of "multi-armed bandit."
(*with some additional structure from knowing the extensive form of the game. e.g. if 2 and 3 move simultaneously then 1 knows their play is uncorrelated.)*

In bandit problems a patient agent will do some "experimenting" with choices that don't maximize the current period's expected payoff.

Example of one-armed bandit:
"Out" give a sure payoff of 1;
"In" gives either always gives 2 or always gives 0.

If only play once, and the probability of "always 2" is .4, then choose "Out." $(.4 \times 2 = .8 < 1)$

Now suppose play repeatedly.
Only see the outcome if pull the arm.
So it might be worth playing "In" once to see.

If future payoffs are discounted by $\delta = .9$ per period then :

- **Always Out**" has payoff 1/.1=10;
- "**Always In**" has expected payoff .4x2/.1= 8,
- " **Try In once then switch to Out forever if get 0"** yields

$$.6 \times .9 \times 10 + .4 \times 20 = 13.4.$$

Here playing "In" is an experiment: it doesn't maximize the current period's expected payoff but it is worth doing for the information it provides.

Note that for given prior and almost all discount factors, the agent strictly prefers one of these choices- so will not randomize.

What if the true payoff distribution is stochastic?

Now a single observation doesn't reveal the true probability, and the solution is more complicated.

- Sufficiently patient agents will still do some "experimenting" with **In.**

- A sufficiently long string of bad outcomes will make agents so pessimistic they switch to Out- and once they do they receive no further information and so "lock on" to Out.

- When payoffs are stochastic agents can end up locking on to Out when In is optimal.

- The probability of this incorrect lock-on depends on the discount factor and the prior.

- When agent is more patient it experiments more, so there is lower probability of incorrect lock-on.

Now apply these ideas to learning in games:

We could expect that once agents have enough data the play a myopic best response to their beliefs (since they don't expect to learn much from future signals):
asymptotic myopia.

Could also hope that if agents are very patient, they will with high probability learn play at relevant information sets.

So conjecture that in general Bayesian learning leads to self-confirming equilibrium, and to Nash equilibrium if there is agents are patient.
To verify this, need to be more concrete about the model of the overall population of agents.

Steady State Learning (Fudenberg-Levine [1993b, 2006]

- A continuum population, with a unit mass of agents in each player role: aggregate behavior is deterministic.
- Agents live T periods, in overlapping generations: $1/T$ players in each generation.
- Every period, each agent is randomly matched with one agent from each of the other populations. (So the probability of meeting an agent of a particular age is equal to $1/T$.) Agents do not observe the ages of their opponents.)
- Each time they play the game, agents observe the terminal node- this is all the information they get.

- This system has steady states. (*learning and forgetting)*
- The steady state for $T = 1$ isn't very interesting.
- To analyze the effect of learning we study steady states when T is large.
- Here most agents will have played a long time and so have a lot of observations.

.

Results:

- A limit of steady states as lifetimes grow to infinity must be a SCE.

Note: Different priors can lead to different steady states.

- As lifetimes grow and the discount factor tends to 1, any limit of steady states is a Nash equilibrium.

Again, different priors lead to different steady states.

Intuition/sketch for why steady states for long lifetimes are SCE:

a) If strategy *s* is played in the limit, it is played by a positive fraction of the population a positive fraction of their life.

b) Most agents who have played this strategy many times have correct beliefs about its consequences, because
   i) posteriors converge to the empirical distribution at uniform rate (Diaconis and Freedman [1990]) and
   ii) empirical distribution looks like theoretical one.

   c) "Lock on": Agents eventually stop experimenting and play myopic BR to beliefs. *Need to account for players knowledge of the extensive form…*

Intuition for why limits with patient players are Nash equilibria:

At steady state that isn't a Nash equilibrium, some agents who move on the path of play have a gain to deviating.

So observations are unlikely to make them believe the gain to deviating is very close to zero, yet patient players will keep on experimenting when they believe the probability of a profitable deviation is non-zero.

So when agents are patient, and have a gain to deviating, they are likely to keep on experimenting until they see that their beliefs are wrong.

The converse of this result is false: Not all of the Nash equilibria can be limits of steady states with patient learners.

Just which equilibria can arise depends on how much agents learn about play two or more nodes off of the path of play, and thus on how much experimenting the agents do at "off-path" information sets.

The answer is only known for *simple games*, which are games of perfect information (*every information set is a singleton)* where a no player moves more than once on any path.

Node x is *one step off the path of* strategy profile s if it is an immediate successor of a node that is reached with positive probability.

Profile s is a *subgame-confirmed equilibrium* if

1) It is a Nash equilibrium and

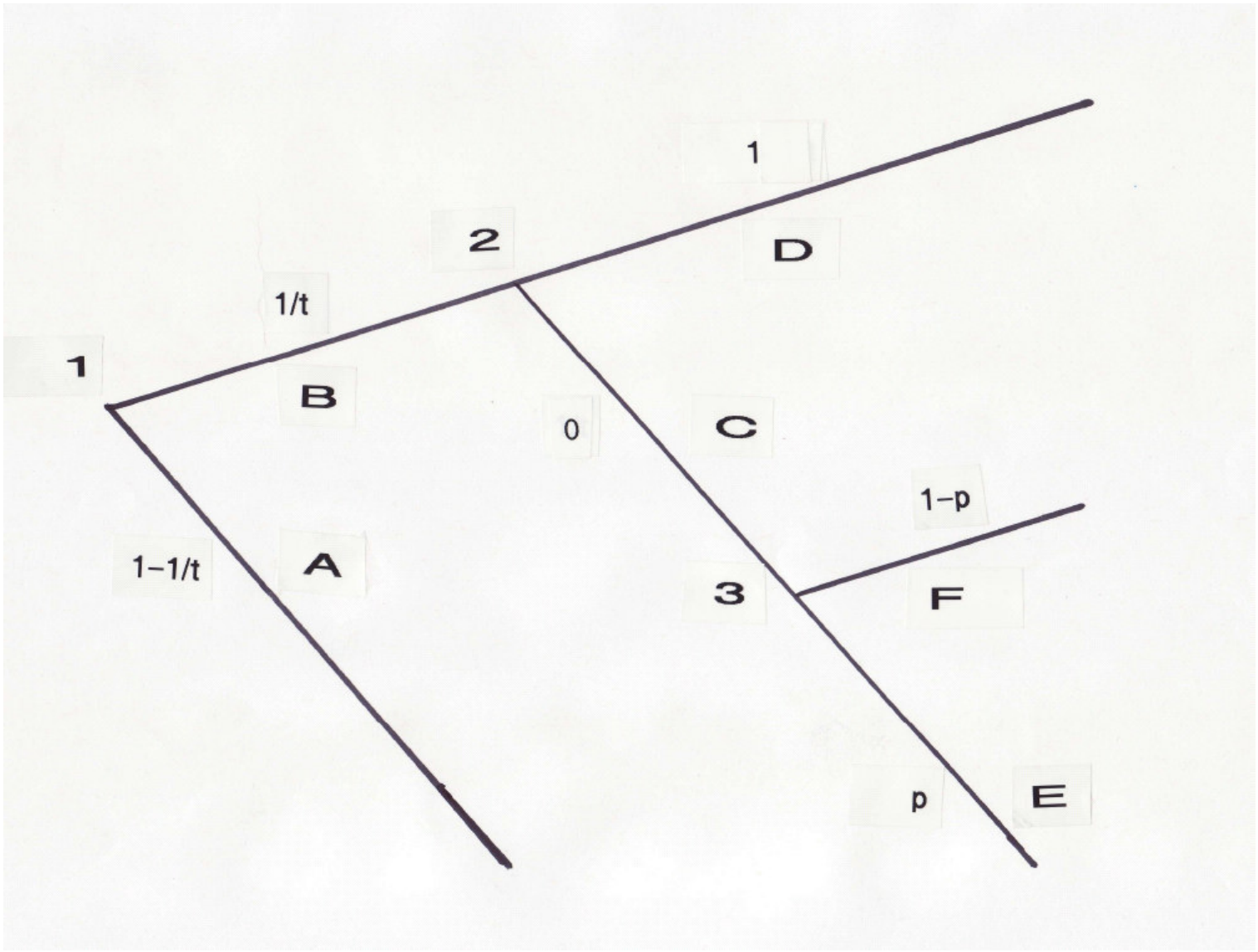
2) In each subgame beginning one step off the path, the restriction of s to the subgame is a self-confirming equilibrium in that subgame.

Paraphrase of Result: In simple games, a pure-strategy subgame-confirmed equilibrium has the same outcome as a limit of steady states for patient learners.

Intuition: Players who move on the equilibrium path experiment, so they learn play at nodes one step off the path.

Players one step off the path learn what happens from there on if no one experiments. (This is why the outcome is a SCE in the continuation game.)

But these off-path players needn't experiment, because they may get to play too rarely to make it worthwhile. So even with patient players, play does not need to be a Nash equilibrium one step off of the path.


1
2
1/t
1
B
A
1–1/t
0
C
D
1–p
3
F
p
E

This may help explain the apparent durability of the one of the laws of Hammurabi:

"*If any one bring an accusation against a man, and the accused go to the river and leap into the river, if he sink in the river his accuser shall take possession of his house. But if the river prove that the accused is not guilty, and he escape unhurt, then he who had brought the accusation shall be put to death, while he who leaped into the river shall take possession of the house that had belonged to his accuser.*"

- This code suggests there were incentive problems with false accusations.
- Seems to rely on the superstition that the guilty are more likely to drown than the innocent.

If there are no limits on superstitious, why didn't Hammurabi simply assert that those who are guilty will be struck dead by lightning, while the innocent will not be?

Our explanation: the wrong belief in the lightning-strike superstition is one step off of the path, and so is not robust to learning by patient players.

But the wrong belief in the trial-by-river superstition is two steps off of the path: Potential accusers have wrong beliefs about the river, yet they only get to act when a crime takes place. If the superstition is believed, only young experimenters commit crimes, so there are few crimes, and accusers only get to play infrequently, so there is little value to experimentation.

In practice, there may be other sources of experimentation in addition to rational learning.

For example, there might be an exogenous probability that crime does pay…

In the Hammurabi game, if the exogenous probability of crime is sufficiently low, then the probability of being called as a witness is also small, and for any fixed discount factor the incentive to experiment with false accusations is small.

In the U.S. today, the probability of being called as a witness at a trial is small; most people are not called even once in a lifetime. So witnesses don't have much incentive to experiment…

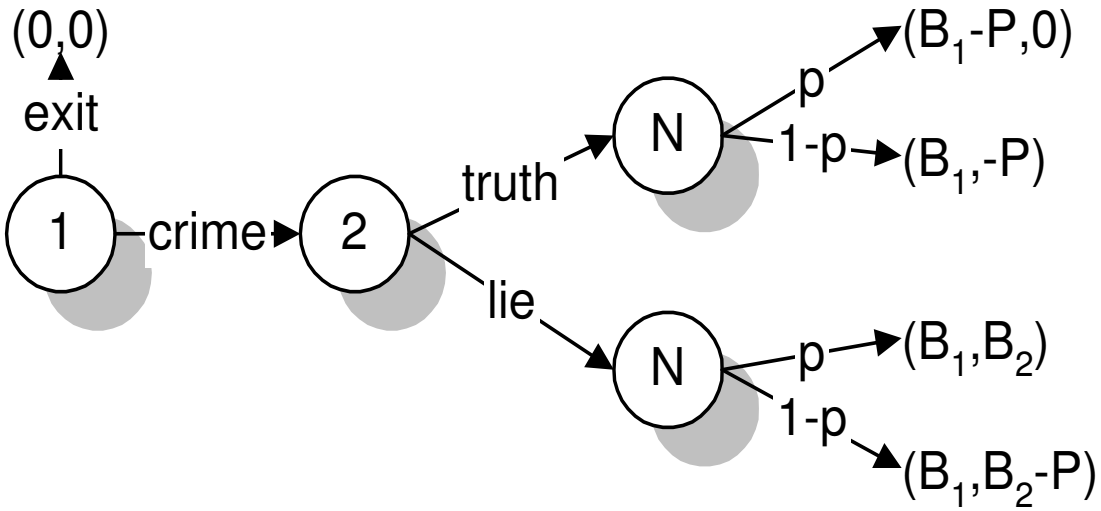
Summary: False beliefs two or more steps off of the equilibrium path are able to persist for much longer than false beliefs about play on or one step off of the path.

This may be linked to the durability of the "appeal to the river" superstition.

Even when long-run outcome involves enough experimenting to rule out non-Nash outcomes, this may take some time.

Conclusions:
If equilibrium is the result of learning, then

- Just what sort of equilibrium to expect depends on what players observe when the game is played.
- The nature of equilibrium should be a theorem- derived from assumptions about the learning process- and not an axiom.
- If players have little prior information about opponents' payoffs and do not experiment very much then learning theory points to self-confirming equilibria.
- Learning plus substantial experimentation with off-path play leads to Nash and then subgame-confirmed Nash equilibrium (*in simple games, but the analog for general games is open.)* But even in this case non-Nash but self-confirming play may persist for a while.


(0,0)
exit
1
crime
2
truth
lie
N
N
p
1-p
p
1-p
(B1-P,0)
(B1,-P)
(B1,B2)
(B1,B2-P)