Learning and Equilibrium, II: Extensive-Form Games

Simons Institute Economics and Computation Boot Camp
UC Berkeley,August 2015
Drew Fudenberg

Extensive-Form Games

- Used to model games with sequential and/or multiple moves.
- Here strategies are "complete contingent plans" that specify the action to be taken in every situation (at every "*information set*") that could arise in the course of play.
- The *actions* that a player uses can depend on the actions of others, but we can think of players simultaneously choosing *strategies* before the game is played.
- Associate a unique strategic form with given extensive form.

- The <u>definition</u> of Nash equilibrium applies without change: A strategy profile such that no player can increase their payoff by changing their strategy, holding fixed the strategies of the other players.

- But Nash equilibrium is less satisfactory here:

  - NE doesn't incorporate predictions based on knowledge of opponents' payoff functions. This led to subgame-perfect equilibrium and other "equilibrium refinements." I will suppress this issue for most of this talk.

  - NE and its refinements describe situations where players know more than in guaranteed by learning. In some cases play can converge to a **self-confirming equilibrium (SCE)** that is not a Nash equilibrium.

- The effect of learning depends on what the players observe when the game is played.

- Assume they observe (at most) the terminal nodes that are reached in their own plays of the game.

- Don't observe how the opponents *would have played* at "off path" information sets- those that were not reached in that play of the game.

- Exploration/exploitation trade-off: people may choose to engage in "active learning" or "experimentation."

- W/o experimentation, incorrect beliefs about off-path play could persist.

Aside: other possible observation structures:

1) Players might not observe the realized terminal node, but instead a partition of them. For example, in a first-price sealed-bid auction players might observe the winning bid but not the losing ones. See Dekel, Fudenberg, Levine *GEB* [2004], Lehrer-Solan *JET* [2007], Esponda *AER* [2008], Fudenberg-Kamada *TE* [2015].

2) Agents might observe outcomes in other matches, or get signals about them (as in models of social learning).

Questions:

1) what sorts of outcomes can be steady states of learning processes? (various answers: SCE, rationalizable SCE, NE, and subgame-perfect equilibria.)

2) Some "experimentation" seems to be needed to rule out convergence to a non-Nash outcome**.** "How much" of this off-path play is needed for various equilibrium concepts? That is, how much information about off-path play is needed to imply that all steady states satisfy the equilibrium conditions, and how much experimentation with off-path actions does this require?

3) How much off-path play will occur under various models of learning?

Overview of the literature:

Fudenberg and Kreps [1988, *JET* 1995], Jehiel and Samet *JET* [2004], Laslier and Walliser *JET* [2004] look at "boundedly rational" learning: assumptions directly on the frequency of experimentation.

FK 88: "belief-based learning" in the spirit of fictitious play, with the extra assumption that agents experiment at rate $1/t$; i.e. there is a lower bound on the probability of each action, and the bound goes to 0 over time at rate $1/t$.

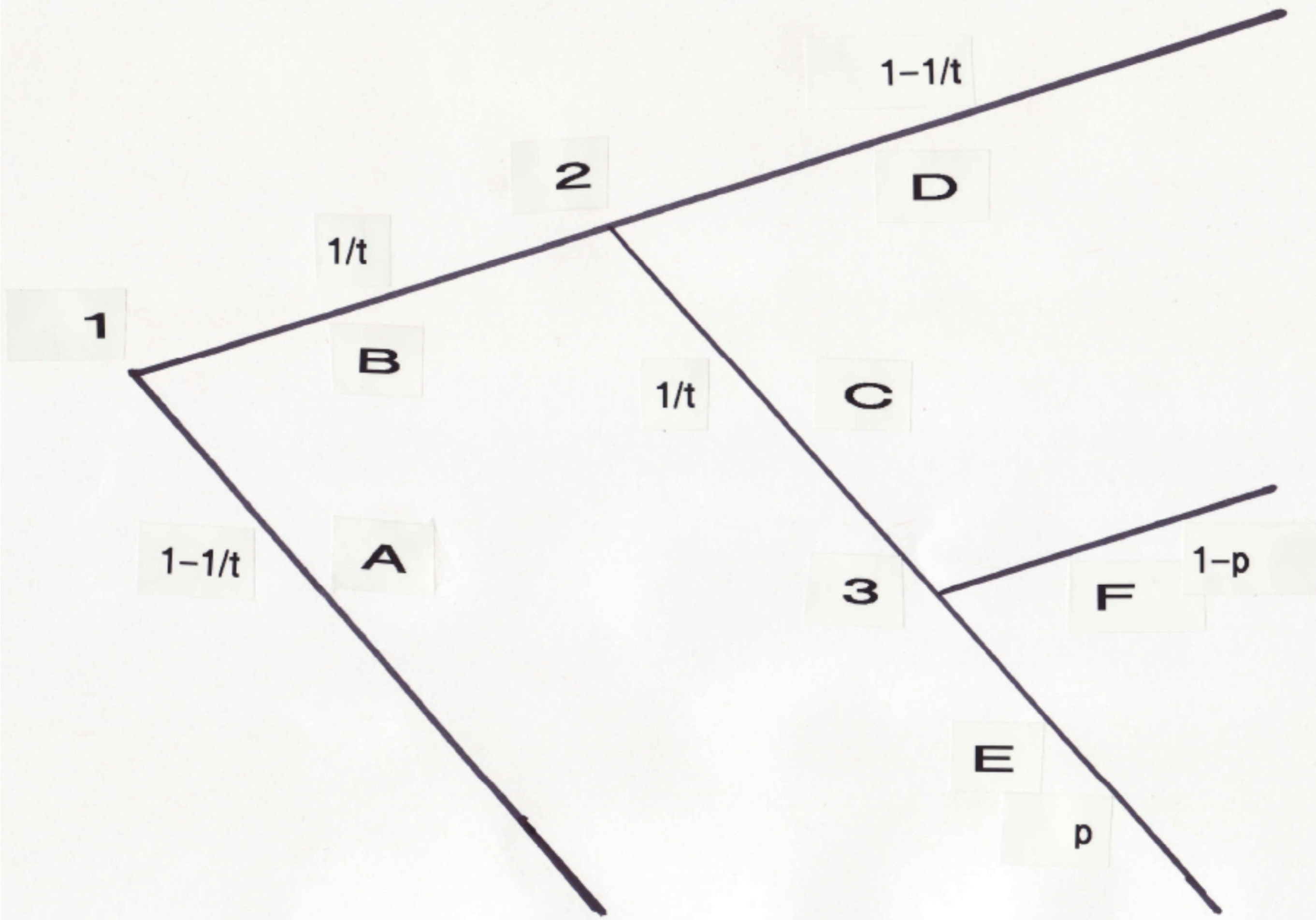
This experimentation condition rules out convergence to non-Nash outcomes when beliefs are asymptotically empirical.

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

- With "$1/t$ experimentation" these relevant information sets are reached infinitely often, because $\sum_{t=1}^{\infty} 1/t$. *(0-1 laws)*

- So from LOLN and asymptotic empiricism, beliefs at relevant information sets become correct.

- But $1/t$ experimentation needn't lead to correct beliefs at nodes that take 2 or more deviations to reach, because $\sum_{t=1}^{\infty} 1/t^2 \neq \infty$.


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

- Raises, but doesn’t answer, the question of how much experimentation players will actually do.

- Remember that rational decision makers typically won’t randomize.

- If players don’t experiment at all , can converge to SCE that are not Nash. (Fudenberg Levine Ema [1993], Fudenberg Kreps GEB [1995]).

Now define and analyze SCE..

Self-Confirming Equilibrium

- $I+1$ players in the game, player $i = I+1$ is nature.
- Finite game tree with nodes $x \in X$, information sets *h.*
- Terminal nodes $z \in Z$; player $i's$ payoff function u_i is a function of *z.*
- S_i is the set of pure strategies for player *i,* $s \in S$ denotes a strategy profile for all players including nature.
- Each strategy profile *s* determines a probability distribution $p(\cdot| s)$ over terminal nodes.

- Players know the extensive form of the game, except that they may not know the distribution of Nature's move. (If Nature's move is unknown, players' beliefs about it are treated in the same way as their beliefs about the strategies of other players- need to be learned from observations.)

- A probability measure μ_i over Π_{-i}, the set of other players' behavior strategies, describes player i's beliefs about his opponents' play.

- For a fixed mixed strategy profile σ, let $\underline{\pi}$ be the unique equivalent behavior strategy. Player i's beliefs are correct at information set h if they assign probability 1 to strategies that match the objective distribution at *h.*

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

(a) s_i is a best response to $\mu_i(s_i)$, and

(b) $\mu_i(s_i)$ is correct at every h that has positive probability under (s_i, σ_{-i}).

- Reduces to Nash in one-shot simultaneous-move games.
- Each s_i in the support of σ_i may be justified by a different belief.

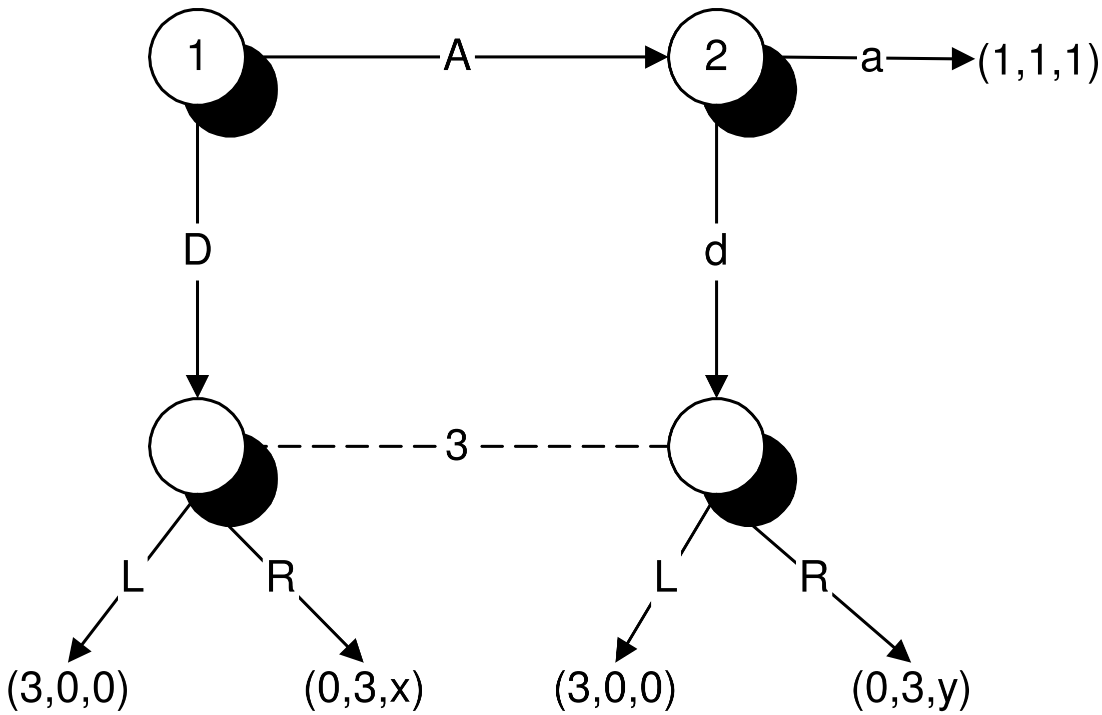
- <u>Interpretation</u>: multiple agents in the role of each player, and different agents in the role of player i may have observed play at different nodes. (*heterogeneity is a non-issue for NE- only one way for beliefs to be correct.)*

- *Unitary SCE*: one belief per player.

- Related concepts: subjective equilibrium (Kalai-Lehrer *Ema* [1993], conjectural equilibrium (Battigalli [1998], partially specified equilibrium (Lehrer-Solan *JET* [2007]).

- All of these are unitary.

- There can be unitary SCE that differ from Nash because 2 players disagree about the play of a 3rd (Fudenberg Kreps [1988]):

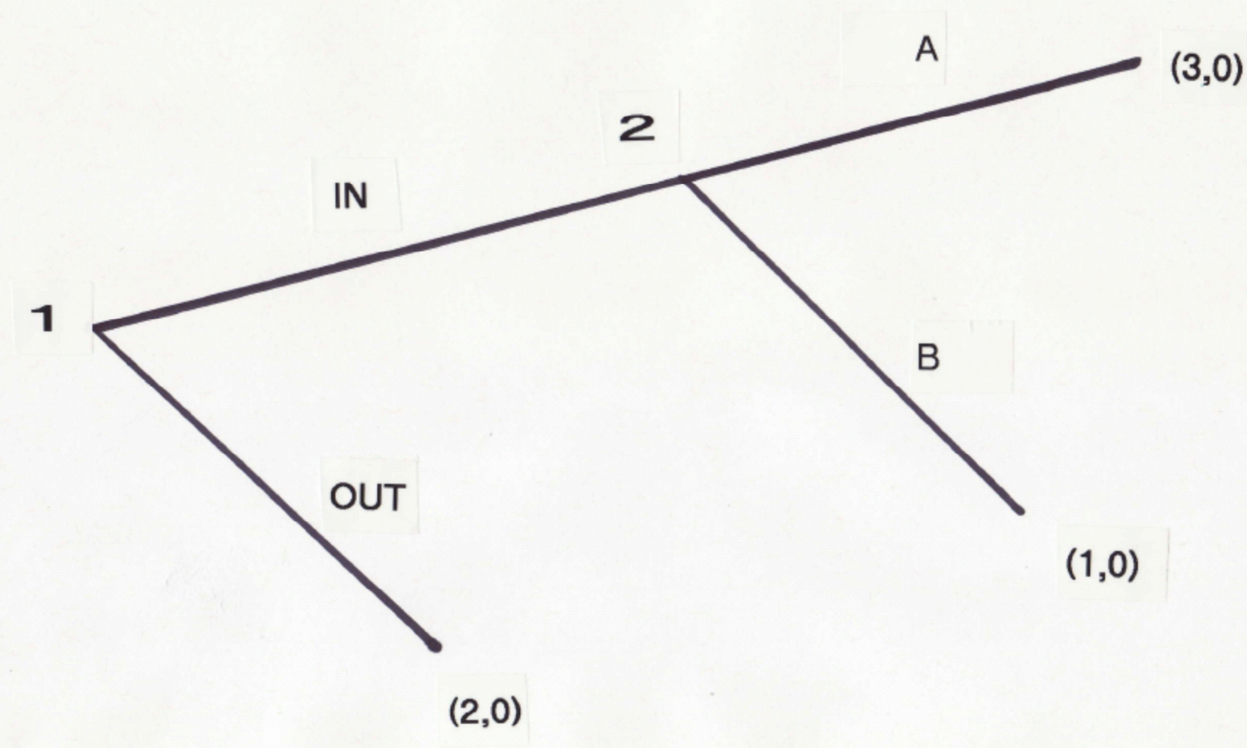


- (A,a,L) is a SCE with outcome (A,a).

- (A,a,L) can also arise from a Bayesian learning process: As long as both play A, they get no data on 3's play, and if beliefs are a product measure, they don't update- seeing player 2's action doesn't change 1's beliefs about 3's play.

- But (A,a) is not a Nash equilibrium outcome: Nash equilibrium requires players 1 and 2 to make the same (correct) forecast of player 3's play, and if both make the same forecast, at least one of the players must choose *D*.

Now allow heterogeneous beliefs.
In the following game, there is no unitary SCE (and hence no NE) with outcome distribution (½ *Out,* ½ (*In*, *A*)).


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

- Heterogeneous beliefs about off-path play are important in explaining data from game theory experiments.

- Set of Nash equilibrium outcomes is not convex, so observing a convex combination of NE typically rejects NE.

- In more complicated games, heterogeneous beliefs can generate outcomes outside co(NE).

- Heterogeneous beliefs especially important when subjects have doubts about each other's preferences.

- In lab settings where "social" or non-monetary preferences important, it is hard to see how the subjects could know the distribution of opponents' preferences, since even the experimenters don't.

What factors lead SCE to differ from Nash equilibrium?

- Heterogeneous Beliefs

- Correlated Beliefs: not a product measure. This can deter some actions that are best responses to any product measure. (*Note: in this model players believe opponents randomize independently- there are no correlating devices. The correlation here is* subjective)

- "Inconsistent" beliefs: 1 and 2 disagree about the play of 3 at some info set h. This only matters if both 1 and 2 can unilaterally cause *h* to be reached.

FL show that this list is exhaustive.
I'll state a related theorem simpler theorem instead.

Let $\bar{H}(s)$ be the information sets reached with positive probability when s is played.

Defn: A game has *observed deviators* if for all players *i*, all strategy profiles *s*, and all $\hat{s}^i \neq s^i$, $h \in \bar{H}(\hat{s}^i, s^{-i}) \setminus \bar{H}(s)$ implies that there is no $\hat{s}^{-i}$ with $h \in \bar{H}(s^i, \hat{s}^{-i})$.

- Implies that if a deviation by player *i* leads to an information set off the equilibrium path, there is no deviation by *i*'s opponents that leads to the same information set.

- Games of perfect information have observed deviators, as do all multistage games with observed actions.

- Includes all two-player games of perfect recall: With two players, both players must know who deviated.

- Holds in many economic examples but not the "horse" game.

- Defn: *Independent beliefs:* each player's beliefs about the others is a product measure. (even off of the path).

Theorem (FL *Ema* 93a, Kamada *Ema* [2009]**)** In games with observed deviators, the outcome of any independent unitary self-confirming equilibria is the outcome of a Nash equilibrium.

Ideas:

- Unitary beliefs: a single set of beliefs for each player *i.*

- Independent beliefs: these beliefs correspond to a mixed strategy profile. (*not true with correlation)*

- Observed deviators: only one player's beliefs about play at an off path information set are relevant- namely the beliefs of the player who could cause that information set to be reached.

- So construct profile where actual play at the info set corresponds to the beliefs of the relevant player.

Back to Learning Dynamics

Various learning models have SCE as long run outcome.

FK *GEB* [1995], [1996 unpublished] look at boundedly rational learning rules with exogenous experimentation rules (*at least at the time this was frowned on by economists…*)

FL *Ema* [1993] analyze rational learning and experimentation by Bayesians who maximize expected discounted utility.

FL 93

The Agent's Decision Problem

- Each "agent" in the role of player *i* expects to play T times
- Tries to maximize $$\frac{1-\delta}{1-\delta^T} E \sum_{t=1}^{T} \delta^{t-1} u_t \, .$$

- Each time the game is played, the agent observes only the terminal node.

- Agent believes that they face a fixed time- invariant probability distribution of opponents' strategies.

- Unsure what the true distribution is. Assume that prior beliefs are *non-doctrinaire*: *given by a continuous density function that is strictly positive at interior points.*
- Updates beliefs about strategies using Bayes rule.
- Non-doctrinaire prior implies non-doctrinaire posterior.
- Each agent faces a dynamic programming problem.
- Pick an optimal policy for each agent- a map from sequences of terminal nodes to strategies. W.l.og. we can take the policy to be deterministic.

Aggregate Play

- Continuum population, unit mass of agents in each player role.
- Doubly infinite sequence of periods.
- Overlapping generations, with $1/T$ players in each generation.
- Every period, each agent is randomly and independently matched with one agent from each of the other populations. So probability of meeting an agent of a particular age equals $1/T$.
- Agents do not observe the ages or past experiences of their opponents
- Each population has a common prior and common optimal rule (this is just to lighten notation.)

- The state of the system: fractions of the population with each possible history.

- For a given distribution $\bar{\theta}$ on S, we can compute the fraction $\theta_i(y_i)$ of population *i* that would have each possible history y_i, and then compute the fractions playing different strategies

- Continuous map from the space of mixed strategy profiles to itself, so fixed point exists for *any* rule.

- These fixed points are the *steady states* of the system.

- FP is stochastic and non-stationary (updating slows down as players gain experience) so it only approaches a deterministic steady state in the limit.

- Existence of steady states here comes from modelling tricks: Individual agents have stochastic time varying beliefs, but the aggregate system is *deterministic* (due to continuum population) and *stationary* (due to finite lifetimes or "memory loss.")

- Steady state for T=1 is trivially determined by the priors.

- <u>Open problem</u>: characterize steady states for intermediate lifetimes in some interesting examples. Here long-lived, better informed players might be able to take advantage of the new entrants.

- Current results focus on limits as lifetimes long, so most players have lots of observations of play.

- Another open problem: stability of the steady states- dimensionality problem when lifetimes are long…

Theorem Fix a non- doctrinaire prior. Then any limit of steady states as $T \rightarrow \infty$ is a SCE.

Three step proof sketch:

- If a strategy has a positive share in the limit, then then for large lifetimes it is played by a positive fraction of the population a positive fraction of their life.

- Most agents who have played the strategy many times have approximately correct beliefs about what happens when they do. (From LOLN and Diaconis-Freedman *Annals Stat.* [1990]: posteriors converge to empirical distribution at a rate that depends only on sample size.)

- Agents eventually stop experimenting and play myopic BR to beliefs.

- *Agents also stop experimenting in the solution to discounted bandit problems. Additional complication here from the assumption that players know the extensive form- so they may know some samples are "unrepresentative." After such samples they may choose to continue to experiment, in contrast to the claim above, but these histories are rare.*

- This is the outline of how to show steady states are SCE.

Theorem: any iterated limit of steady states $\lim_{\delta \to 1} \lim_{T \to \infty}$ must be a NE.

- Easy to prove under "1/t experimentation."

- But for generic beliefs rational players don't randomize in decision problems. And it's not obvious whether players off the equilibrium path want to experiment at all. (in fact sometimes they won't, Fudenberg and Levine *AER* [2006].)

- Moreover, patient agents on the equilibrium path may never play some actions at all, because another action gives the same information more cheaply- correlated arms here as opposed to the independent arms of the Gittins problem.
  (*see chalkboard)*

- So instead of direct bounds on experimentation, proof uses an indirect approach: In a steady state, most players who use a strategy do so because it maximizes their current period's expected payoff: If they have played the strategy many times, they do not expect to learn much about its consequences, so its "option value" is low. *(This uses the order of limits…)*

- If the limit of steady states is not a Nash equilibrium, then along the sequence there is a strategy being played with non-negligible probability that is not optimal against the steady state. Show that this implies the strategy has a non-negligible option value, so players would continue to experiment, a contradiction.

Conclusion: patient players experiment enough to rule out non Nash states.

- This does not say that all NE are limits of steady states with patient players.
- FL *AER* [2006] examine this in games of perfect information with independent beliefs.
- Simplifies optimal experimentation for the same reason it simplifies inference: no reason to take action A to learn about payoff to action B.
- Show that for some non-doctrinaire priors there is no off-path experimentation.

- But off-path play isn't completely arbitrary: players one step off the path are reached infinitely often, and so play there looks like a SCE.

Defn: Node x is *one step off the path of* π if x is not reached under π and it is an immediate successor of a node that is reached with positive probability under π.

Defn: Profile π is a *subgame-confirmed Nash equilibrium* if it is a Nash equilibrium and if, in each subgame beginning one step off the path, the restriction of π to the subgame is self-confirming in that subgame.

Theorem In simple games with no own ties, any subgame-confirmed Nash equilibrium that is nearly pure is path-equivalent to a patiently stable state. (*may need to choose the priors carefully..)*

- No own ties: no player has a pair of actions that lead to the same payoffs for him. Implies unique BI solution, also implies that players will not randomize on the equilibrium path.

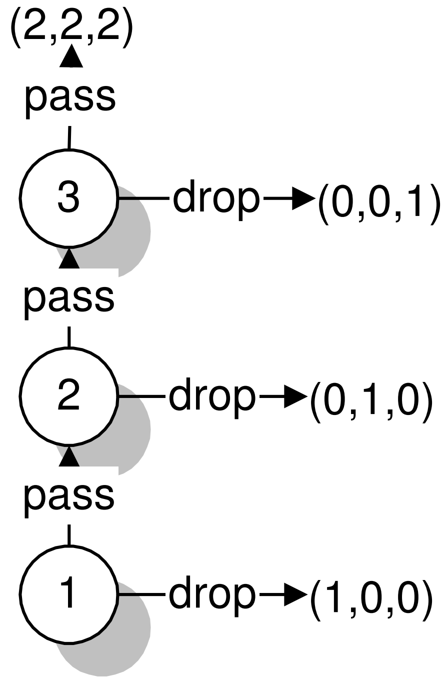
- Nearly pure: no randomization on path, only Nature randomizes off the path. Not necessary in games of length 3 or less, don't know if needed in general. (General result would need bounds on experimentation at off-path nodes when there is mixing on the equilibrium path)

Implications of subgame- confirmed equilibria:

- In a simple game with no more than two consecutive moves, self-confirming equilibrium for any player moving second implies optimal play by that player, so subgame-confirmed Nash equilibrium implies subgame perfection.

- This can fail when there are paths of length three:

Example (Three Player Centipede Game)



- Unique subgame-perfect equilibrium: all players pass.

- (drop, drop, pass) is subgame-confirmed: 1 is playing a best response to 2's action, and since 2 drops and doesn't experiment, doesn't learn 3's play.

How much patience and experimentation should we expect? (open)

- Discount factor/continuation probability in lab sessions bounded away from 1.
- In which field settings do very high discount factors seem plausible?
- To what extent can word of mouth, social learning, historical records serve to provide extra information about off path play?

Prior Information about Payoffs+ Rationality

- In SCE the only constraint on beliefs is what players observe about others' play- players aren't required to use information about opponents' payoff functions.
- May be a good approximation of some field situations and for experiments in which subjects are given no information about opponents' payoffs.
- In other cases, players do have some prior information about their opponents' payoffs.

- Experimental evidence that giving subjects information about other players' payoff functions makes a difference, and in some cases (e.g. Prasnikar-Roth *QJE* [1992]) this difference corresponds to the distinction between SCE and SCE+payoff information as modelled by "Rationalizable Self Confirming Equilibrium" or RSCE. (Dekel, Fudenberg, Levine *JET* 1999).

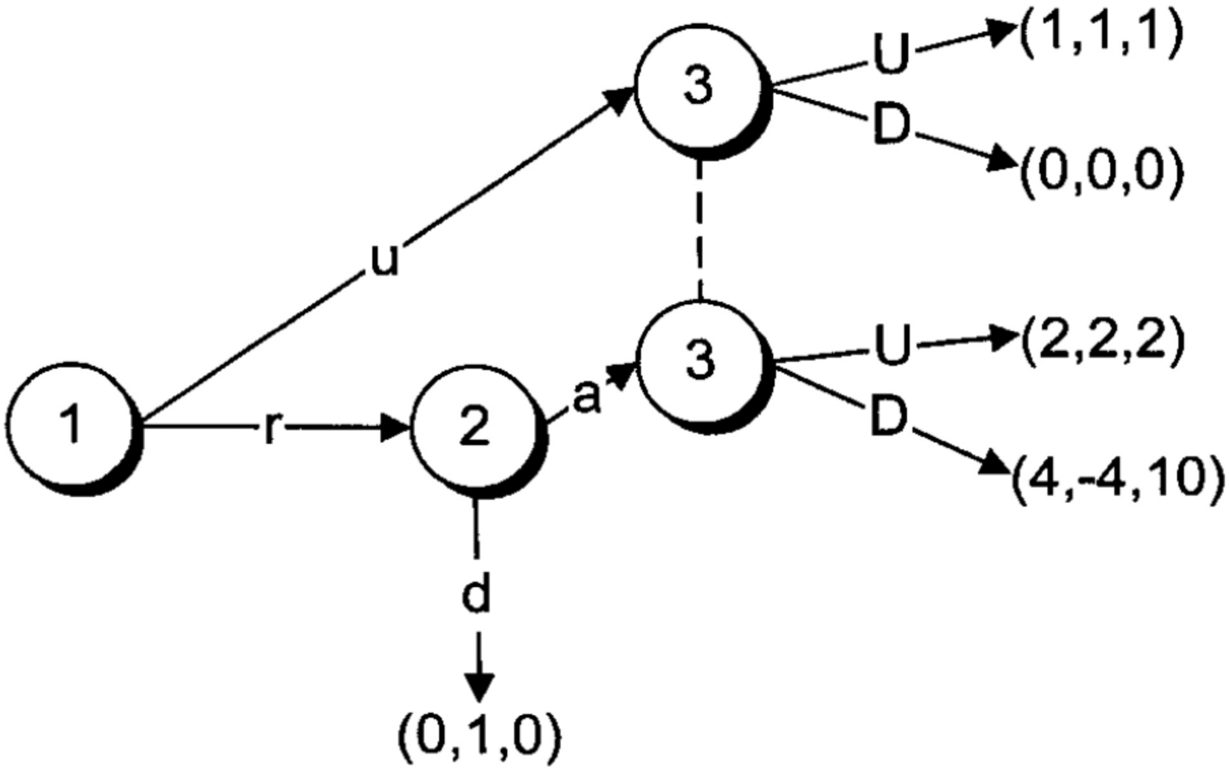
- RSCE is "unitary"- a single belief for all players, and all players see the same distribution on terminal nodes. Fudenberg-Kamada [in preparation] handles the heterogeneous case.

- RSCE imposes some off-path optimality restrictions. It coincides with backwards induction in two-stage games of perfect information, but in longer games it is much weaker and more like SCE.

- RSCE has implications beyond the intersection of SCE and rationalizability

- These come from the assumption that the outcome path is public information.

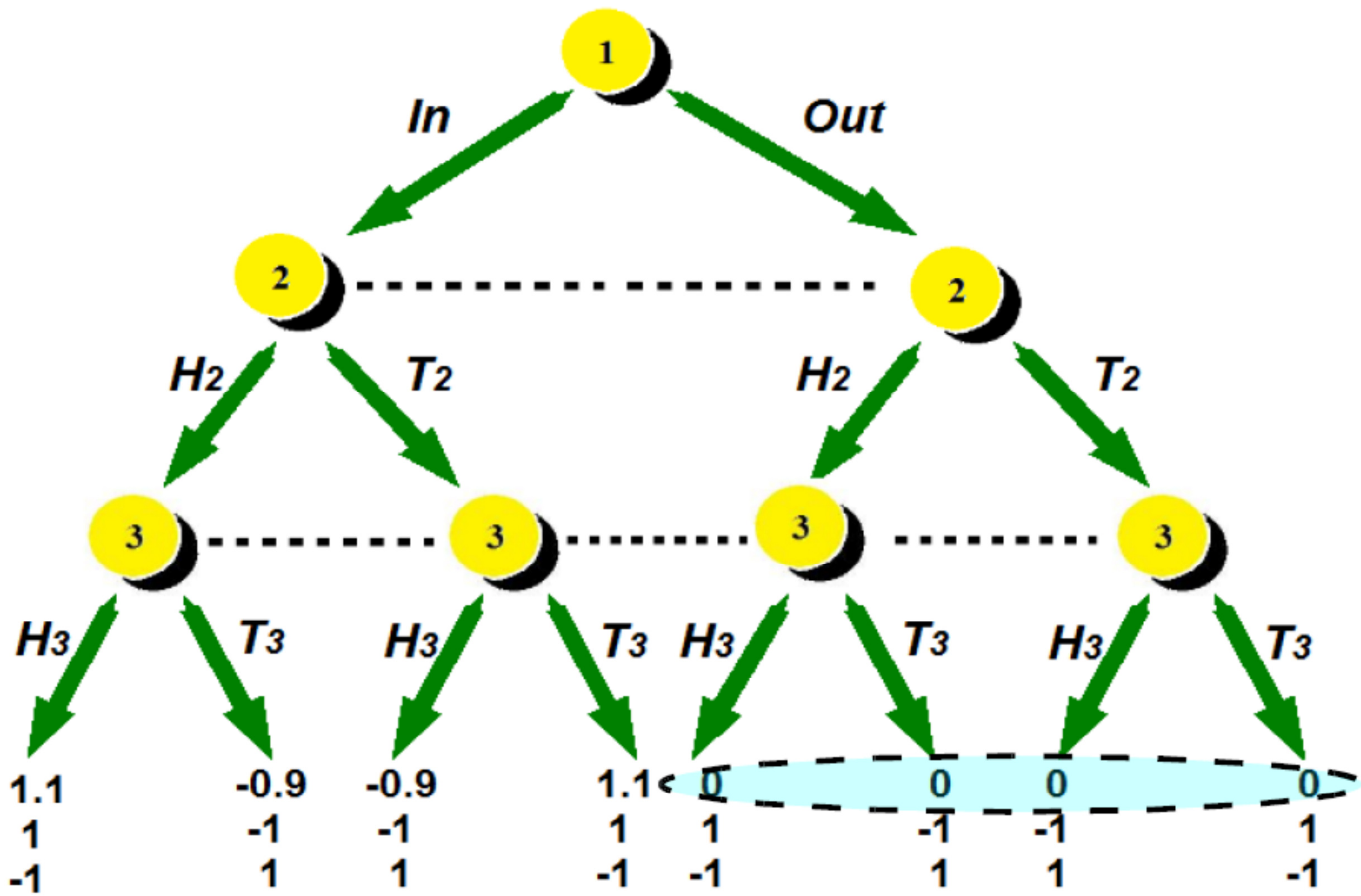
Try to explain this w/o the formal definition…



(u,U) is a Nash outcome (so self-confirming) not RSCE: Intuitively, if player 1 knows 2 knows 3 is playing up, can use this knowledge and his knowledge of player 2's payoffs to deduce that 2 will play a.

Fudenberg-Kamada *TE* [2015]: players only see a partition of the terminal nodes, and different players have different partitions. .

Then there isn't a publicly observed outcome path, so the implications of common knowledge of the observation structure are less immediate and more complicated to formalize.


1
In
Out
2
2
H2
T2
H2
T2
3
3
3
3
H3
T3
H3
T3
H3
T3
H3
T3
1.1 1 -1
-0.9 -1 1
-0.9 -1 1
1.1 1 -1
0 1 -1
0 -1 1
0 -1 1
0 1 -1

- Can 1 play Out, fearing that In will give -.9?

- Out is only a best response if 1 thinks 3 will "beat" 2, that is if the outcome is likely to be (H2,T3) or (T2,H3).

- In RPCE, 1 correctly forecasts 2's observations.

    - 2 has a discrete terminal node partition, so 1 knows 2 sees actual play.

    - So 1 must expect that 2 is playing a best response to 3.

    - So 1 plays In.

- More generally, in a "player 1 participation game," 1's beliefs about play of other players must correspond to a Nash equilibrium.

Mistakes in Inference/Behavioral Game Theory

- SCE and RSCE assume that the players' inferences are consistent with their observations.

- Related literature assumes that players make systematic mistakes in inference, as in Jehiel's *JET* [2005] notion of analogy-based expectations equilibrium or "ABEE:"

  Players group the opponents' decision nodes into "analogy classes," with the player believing that play at each node in a given class is identical. Given this, the player's beliefs must then correspond to the actual average of play across the nodes in the analogy class.

Example: Perfect information.

Nature moves first, choosing state A with probability 2/3 or state B with probability 1/3.

Player 1 moves second, choosing either action A1 or action B1.

Player 2 moves last, choosing either action A2 or action B2.

Player 2 is a dummy who chooses A2 in state A and B2 in state B regardless of what player 1 did.

Player 1 gets 1if his action matches that of player 2 and zero if not.

Then in state A player 1 should optimally play A1 (forecasting 2 will play A2) and in state B player 1 should play B1.

ABEE: suppose player 1 views all nodes of player 2 as belonging to a single "analogy class."

Then he believes that player 2 will play A2 2/3rds the time, regardless of the state, and so player 1 will play always play A1.

Note: If player 1 observes and remembers each outcome, and if he is Bayesian and assigns positive probability to player 2 observing the state, he will eventually learn that this is the case.

Is the ABEE outcome here reasonable?

- Can be seen as approximation of cases where player 1 has a very strong prior conviction that 2's play is independent of the state, so that it will take a long time to learn that this is not true.

- Alternative explanation: players are unable to remember all that they have observed, perhaps because at an earlier stage they chose not to expend the resources required for a better memory. So in the example, maybe player 1 is only able to remember the fraction of time that 2 played A2, and not the correlation of this play with the state?

- Corresponds to SCE when player 1's end-of-stage observation is only player 2's action, and includes neither Nature's move nor player 1's own realized payoff.

- Relate ABEE to Eyster-Rabin's *Ema* [2005] "cursed equilibrium" of Bayesian games.

- "Fully cursed" players think opponents' play is independent of their types- thus in a lemons problem buyers don't realize that lower-value sellers are more likely to sell.

- Can represent cursed equilibria as ABEE (Miettinen *JET* (2007)). (note this doesn't mean that the cursed equilibrium makes sense, just that it can be interpreted as ABEE).

(Some of the) Open questions/problems:

- Characterize the implications of patient rational learning outside of simple games.

- Model what happens in mixed pools of experienced and novice players.

- Model how people extrapolate between "similar" games.

- Combine stochastic choice with learning in extensive-form games. If stochastic terms uniformly bounded away from 0, eventually players have many observations at every information set. This may take some time; what happens in the "intermediate run"?

- What are the implications of learning when agents use mis-specified models? (as *it seems sometimes they do)*

- In decision problems? In games?

- Older stat literature looks at mispecified learning when signals are exogenous; in a system of learners signals can depend on actions…

- See e.g Esponda-Pouzo [2015], Heidhues, Kozegyi, Strack [2015], Fudenberg, Romanyuk, Strack [2015].

- Strack is here at Berkeley…