NON-EQUILIBRIUM ADJUSTMENT AND LEARNING IN GAMES*

LECTURE TO NORTHWESTERN UNIVERSITY GRADUATE STUDENTS

May 7 2013

*Informal talk without equations or references. See lectures 8 through 11 on http://isites.harvard.edu/icb/icb.do?keyword=k92794 for missing details and references and also see Fudenberg and Levine *Annual Reviews of Economics* [2009] .

Overview

When and why will observed play to approximate an equilibrium?
What sort of equilibrium?

Rationality, even common knowledge of rationality, is not sufficient for Nash equilibrium ("NE").

Common knowledge of rationality does imply NE in games where iterated strict dominance gives a unique solution.

But many games of interest have multiple NE, and all of these survive iterated strict dominance.

No reason for play to look like any of the equilibria without some explanation for why players all expect the <u>same</u> equilibrium.

Yet equilibrium seems a decent approximation of the outcomes of some (not all!) experiments, and has been useful in empirical analyses of field data. *(testing of the equil assumption in field data seems an open issue, equilibrium is typically built into the estimation procedure…)*

To understand equilibrium, look at long-run behavior of non-equilibrium dynamic processes.

"Learning in games" broadly defined: players play the game repeatedly, and adjust their play based on their experience. Then equilibria correspond to the long-run outcome (e.g. steady state, asymptotic limit, etc.) of the adjustment process.

This description applies to many sorts of adjustment processes, including biological evolution, and to myopic best responses to true state of the system- where nothing is being "learned."

Aside on evolutionary models:
Rationality not necessary for play to resemble equilibrium- a colony of bacteria can converge to NE.

In evolutionary biology models, the payoffs correspond to reproductive fitness, players are genetically programmed to play various actions, and the Nash equilibria are steady states of a model of how the share of the population evolves- the idea is that the fitter strategies have more offspring so that their share of the population increases.

The ***replicator dynamic*** is the standard model in evolutionary game theory; non-equilibrium dynamic process where the mass of agents using a given strategy grows at rate proportional to the strategy's current payoff.

Specialize to single population playing a symmetric two-player stage game.

Continuum of players, each of whom uses a pure strategy.

Standard motivation: individuals genetically programmed to play some pure strategy, and this programming is inherited. Round robin tournament or appeal to LOLN: total number of offspring of *s*-strategists is then number of agents using s times offspring per agent.

Replicator dynamics also describes the result of some types of imitation by economic agents and some of the properties of the replicator dynamic extend to various classes of more general processes that may correspond to other sorts of learning or emulation.

Under replicator dynamic, strategy's share is increasing whenever that strategy does better than the population average, even if the strategy is not the best response to the current state. But still steady states are closely linked to the Nash Equilibria**:**

- Every Nash equilibrium is a steady state: In the state corresponding to a Nash equilibrium, all strategies being played have the same average payoff, so the population shares are constant.

- Interior steady states, where all actions have positive probability, must be Nash equilibria, but steady states on the boundary need not be.

- Non-Nash steady states cannot be asymptotically stable: If the state is perturbed by introducing a small weight on an improving deviation, the share playing that deviation it will grow. So if play converges, it converges to a NE. (true for any continuous-time deterministic processes where the growth rates are a strictly increasing function of the payoff differences, and for many other models of learning in static games as well.)

- No guarantee the dynamics converges to a steady state- can cycle or be chaotic.

Replicator dynamic linked to notion of *evolutionarily stable strategy* (*ESS):*

Suppose population originally at some profile σ, and then a small share of "mutants" start playing some other strategy σ'.
ESS asks that the existing population gets a higher payoff against the resulting mixture than the mutants do.

Every strict Nash equilibrium is an ESS, and any strict Nash equilibrium is locally stable under the replicator dynamic. For generic static games, all pure-strategy equilibria are strict; for these games pure-strategy ESS same as pure-strategy NE.

ESS/NE useful in explaining a wide range of biological phenomena, including mutualisms and parasitess, animal behaviors such as territorialism, and genomic imprinting. This reinforces the point that rationality and equilibrium in games are distinct- and objections to rationality assumptions do not immediately translate into objections to game theory (/end aside.)

The models that seem most like (the narrow sense of) “learning”…

(a) Explicitly specify how individual agents use observations to change their behavior

(b) Self-interested play by the agent- not agents programmed to try to find equilibrium!

But some models are borderline- hard to say if “learning” or not- and not sure if having a precise definition is that useful.

Common themes in the learning-in-games literature:

1) Non-equilibrium adjustment.

We don’t want to explain equilibrium in a given game by assuming an equilibrium of some larger adjustment game.

If we don’t insist on “equilibrium in the adjustment game,” there can be players whose adjustment rule isn't a best response to the adjustment rules of the others and is thus “making a mistake.”

Such mistakes are a <u>necessary</u> component of any non-equilibrium explanation of equilibrium. So in evaluating these models it won’t be a relevant criticism to say that some player is playing suboptimally. Instead we'll ask that it not be obvious to the players that there is some simple alternative that would be better.

2) The learning-in-games literature has so far focused on LR play:

In short run initial beliefs/conditions matter, and needn't see equilibrium. But literature needs more results about speed of learning and learning in short or medium run.

3) Prefer conclusions that hold for a range of learning rules

Hard to pin down "the real rules" people use. *Though more info would be nice; discuss some attempts at this later.* To some extent rules that "point in the right direction" have similar long run properties; is this how people behave?

4) Sensible-seeming learning processes need not converge in some games. Conversely, in some sorts of games most sensible processes do converge.

The possibility of non-convergence isn't surprising, since lots of dynamic process don't converge to steady states. However we should expect learning to converge in games that are solvable by iterated strict dominance. And many processes converge on "potential games" even though these need not be dominance solvable and can have multiple equilibria.

5) **Adjustment processes may suggest equilibrium refinements- because they may tend to converge to some equilibria and not to others.**

6) P**layers play repeatedly without playing a "repeated game"**

That is, each time they play, they try to maximize the payoff in the "current game" and don't try to influence future play in other games.

This usually explained by an explicit or implicit appeal to a “**large population model"** with many "agents" in each "player role."

Leading cases:

6a) anonymous random matching: each period all agents are matched to play the game, and told only play in their own match. Agents are unlikely to play their current opponent again for a long time, even unlikely to play anyone who played anyone who played him.

So if the population size is large enough compared to the discount factor, it's not worth sacrificing current payoff to influence this opponent's future play.

This treatment is used in most experiments.

6b) A**ggregate statistic model:**

Each period all agents play. At the end of each period, subjects learn the aggregate distribution of play, and each player 1's payoff depends on the whole distribution of player-2 actions.

Each agent's action influences the aggregate, and the aggregate is seen by everyone, so in principle it could impact next period's play. But if there are 20 player 1's, each player has influence .05 on the aggregate; that may not have any effect on the play of the other half the class.

Appealing to a large population limits the domains where the theory applies. But by how much? …people may never play the <u>exact</u> same game very often, but they may also extrapolate between games.

We don't have good models or data on cross-game extrapolation.

Modeling non-equilibrium 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

What can agents observe ?

- Their own payoffs
- Realizations of moves by Nature
- Actions and/or payoffs of agents they interact with
- Actions and payoffs of other agents (i.e. market share, word-of-mouth, and other forms of “social learning.”)

Do players have any influence on what they observe?

- If not, "passive learning."
- If so, there may be an incentive to "experiment," i.e. to sacrifice short-run payoff to get information. (In computer science, this is called "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.
- The interaction of non-equilibrium "learning rules."

What determines players learning rules and behavior?

- Worst case/minmax considerations (popular in the statistics and CS literatures, but see also papers on “calibration,” “universal consistency,” and “checking rules”)
- Bayesian, Savage-rational maximization in a non-equilibrium setting
- Exogenously specified, “boundedly rational.”   (this may or may not differ from Bayesian… w.o. constraints on the prior and likelihhod function Bayesianism is extremely general…)

Passive Learning in Static Games

Fictitious Play

Introduced by Brown [1951] as a way to compute equilibrium.
(hence "fictitious" play) but can also be used as a stylized model of learning.

Easy to motivate and analyze, too simple to match experimental data..

Motivation: Suppose that an agent is going to repeatedly play a fixed strategic-form game. The agent knows the structure of the game (the strategy spaces and payoff functions) but not how the other side is going to play. (*can be adapted to less prior knowledge.)*

All that the agent observes is the outcome of play in their own matches.

The agent doesn't observe:

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

The agent believes she is playing against a randomly drawn opponent from a large population, so doesn't try to influence opponent's play.

Because what the agent observes is independent of own action, there is no incentive for "experimentation."

In FP agents act like Bayesian expected utility maximizers facing unknown distribution of opponents' strategies. (Corresponds to "Dirichlet prior," a conjugate prior for multinomial sampling.)

Key isn't functional form, but the implicit assumption that the player treats the environment as stationary.

Implies all observations equally informative, so beliefs converge to the empirical marginal distributions.

Plausible? Stationarity seems a reasonable first hypothesis in many situations, and players might stick with it when it is approximately right.

Perhaps they would reject stationarity given sufficient evidence to the contrary- as when there is a strong time trend or a cycle. And people sometimes give less weight to older data, as if facing a hidden Markov process- "weighted FP."

The Interpretation of Cycles in Belief-Based Learning

	A	B
A	0,0	1,1
B	1,1	0,0

Suppose this game is played according to FP, with only 1 agent per side and initial weights (1, sqrt (2)) for each player.
First period: both players think the other will play B, so both play A. The next period the weights are (2,sqrt (2)) and both play B; the outcome is the alternating sequence ((B.,B),(A,A),(B,B), etc.)
The empirical frequencies of each player's choices converge to (½, ½), which is the Nash equilibrium. So FP “works” for the purpose of computing equilibrium, but this isn’t a good model of learning: Realized play is always on the diagonal, and both players receive payoff 0 in every period even though each can guarantee a payoff of ½.

Reason: the empirical joint distribution on pairs of actions does not equal the product of the two marginal distributions, so that the empirical joint distribution corresponds to correlated as opposed to independent play.

We could? Should? expect players to notice the cycle and form more sophisticated beliefs.

So in general we won't want to identify a cycle with its average.

And we may want to worry about how sensitive the players are to correlations in the data.

Multi-Player Fictitious Play

Related modeling issue: a player's assessment (expected play of opponents this period) always correspond to a mixed strategy profile, or should correlated assessments be allowed?

Answer: under the interpretation of fictitious play as the result of Bayesian updating, a player's current assessment of his opponents' play can correspond to a correlated distribution even if the player is certain that his opponents in fact randomize independently.

If 2 and 3 randomize independently then

$$\gamma_t^1(s^2, s^3) = \int_{\Sigma^{-i}} \sigma^2(s^2)\sigma^3(s^3)\mu_t^1[d(\sigma^2, \sigma^3)];$$

independent mixing is reflected in the fact that the integrand uses the product of $\sigma^2(s^2)$ and $\sigma^3(s^3)$. This ntegral need not be a product measure, it typically won't be unless player 1's subjective uncertainty about his opponents is also uncorrelated, that is, unless μ_t^1 is a product measure.

To elaborate, in learning model, there is a difference between

(a) believing that your opponents are playing the correlated strategy
(1/2 (A,A), 1/2 (B,B))
and

(b) thinking they are either playing (A,A) or (B,B).

These two beliefs have the same (correlated), marginal over first period observations. But they get updated differently because they imply different correlations between period 1 and period 2. With beliefs (b), after one observation you know what they will be playing at all future dates.

There seems no good reason (beyond convenience!) to suppose that the players' initial assessments correspond to uncorrelated randomizations.

Important facts about FP:

- If actions converge they converge to a pure strategy NE.
- If time averages of empirical marginals converge the joint distribution is a NE.
- The above holds in any belief-based learning model that is "asymptotically empirical" (beliefs converge to empirical frequencies) and "asymptotically myopic" (eventually players choose actions that are myopic best responses to their beliefs.)
- FP "behaves well" when there are infrequent changes of play: In this case it is "ε-consistent" (do as well as maximizing vs time average of opponents' play) and it also converges to the best response dynamic.

Fictious play: exact best respone to beliefs.
Discontinuous, can oscillate even as beliefs converge.
Causes technical problems, also leads to poor performance.
"Stochastic" or "smooth FP ("SFP") : smooth approximation to BR curves.

Better behaved, and better worst-case properties ("universally consistent").

Can be given a "Harsanyi-purification" foundation based on private payoff shocks.

Studied with s*tochastic approximation,* which relates limits of discrete-time stochastic system to invariant sets of the "mean field" which is a deterministic continuous-time flow: (slowing down +LOLN)

Here the mean fiekd is the continuous time best response dynamic.

Roughly speaking: SFP can only end up at steady states or cycles that are stable in this system.

Examples: In 2x2 coordination games pure equilibria are possible limits but not the mixed one; same true in potential games more generally. In matching pennies the mixed NE is stable. In some games cycles are stable: these cycles correspond to longer and longer cycles in the real discrete time game.

Extensions/Open Questions: Weighted SFP (less weight on old observations) ; SFP on a network (mostly play with friends/neighbors) SFP with 'trend-spotting" or cycle detection.

A little work on first two topics, see Benaim Hofbauer Hopkins *JET* [2009], Fudenberg-Takahashi *GEB* [2011]), even less known about the third.

Other Learning Rules in Static Games

REL

FP/SFP are models of “belief based learning" where agents know and use the structure of the game.

Contrast to reinforcement learning (REL) models.
Two steps to defing a REL process:
a) what is reinforced? Actions, strategies, rules?
b) how?

Stand case: agents update values to each action based only on received payoffs. Don’t think about payoff they would have received under different strategy.

Simple(st?) version is Cumulative Proportional Reinforcement (CPR)

Normalize so that all utilities are positive, and give initial weights to each action , update the score of the action that was played by its realized payoff, and do not update the scores of other actions.

Then set probability of action k = weight on k/sum of weights.

Here the response to “2 played R” depends on 1’s own action: if 1 played played U and 2 played R and this gave a good payoff then U is reinforced even if (D,R) would have been better.

Step size depends on payoff received but can still use stochastic approximation – use it to track empirical joint distribution, and let probability of actions at each time be given by the specified reinforcement rule- can characterize long run outcomes, somewhat similar results as SFP.

What better approximates actual behavior in static complete-info games?

If players are told structure of the game, and they are rational, they shouldn't condition on own action; in lab subjects *seem* to do so. OTOH REL agents don't respond at all to what they could have gotten by playing something else, and it seems that they do. And subjects play differently if told the play of others and not just own payoffs-which goes against standard REL models. Camer-Ho suggest the best fit is "in between".

Salmon [2001]: Little power in tests used to distinguish the learning models. Following CH assumed all agents use the same behavior rule.

Wilcox [2006] shows this leads to bias in favor of REL.

Intuition: With heterogeneous subjects, pooled estimation gives prediction errors that are correlated with past strategy choices, because those past choices carry idiosyncratic parameters. So players' own past choices have relevant into, and the past payoffs used in REL depend on past choices…

Imitation Processes: like REL don't require agents to know payoff matrix, and stochastic even w/o mutations.

Examples:

1) Imitation + aspiration model: realized payoff is the game payoff + noise term. Each period an agent is picked at random to reevaluate his choice; the agent sticks with his strategy if realized payoff exceeds an exogenous aspiration level and otherwise imitates at random.

2) Imitation model: When an agent reevaluates, he picks another randomly chose agent. If that agent's strategy is different, imitate with a probability that is increasing in the payoff difference.

3) "Imitate the most successful": Agent observes the payoff of others with noise, then picks the strategy with the highest observed payoff.

4) Best response to current state (*adjustment but not learning…)*

But

4') partial sampling of history: each period one set of agents chosen to play, they pick best responses to a sample of k of last m outcomes. (so no agent has any personal observations)

$k = m = 1$ a lot like BR dynamic, tracks it if no mutations and start with all agents doing the same thing

k, m large look like learning

Extensive form games and self-confirming equilibrium

Extensive form games: agents observe (at most) the terminal nodes that are reached in their own plays of the game.
(*"strategy method" results in a different game…*)

Agents don't observe how opponents would have played at information sets that were not reached in that play of the game, and they may not observe outcomes in other matches.

If agents only observe actual play in own matches, and an agent never plays a specific action, he'll never observe how his opponents react to it. So incorrect beliefs about off-path play could persist, and play might converge to a non-Nash outcome that is "self- confirming" unless agents "experiment."

Questions:

1) What sorts of outcomes can emerge from learning processes?

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 off-path play is needed to imply that all steady states satisfy the equilibrium conditions?

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

4) What if players observe don't observe exact terminal node? For example, in a first-price sealed-bid auction players might observe the winning bid but not the losing ones.

5) What if players have and use prior info on opponents' payoffs to restrict their beliefs?

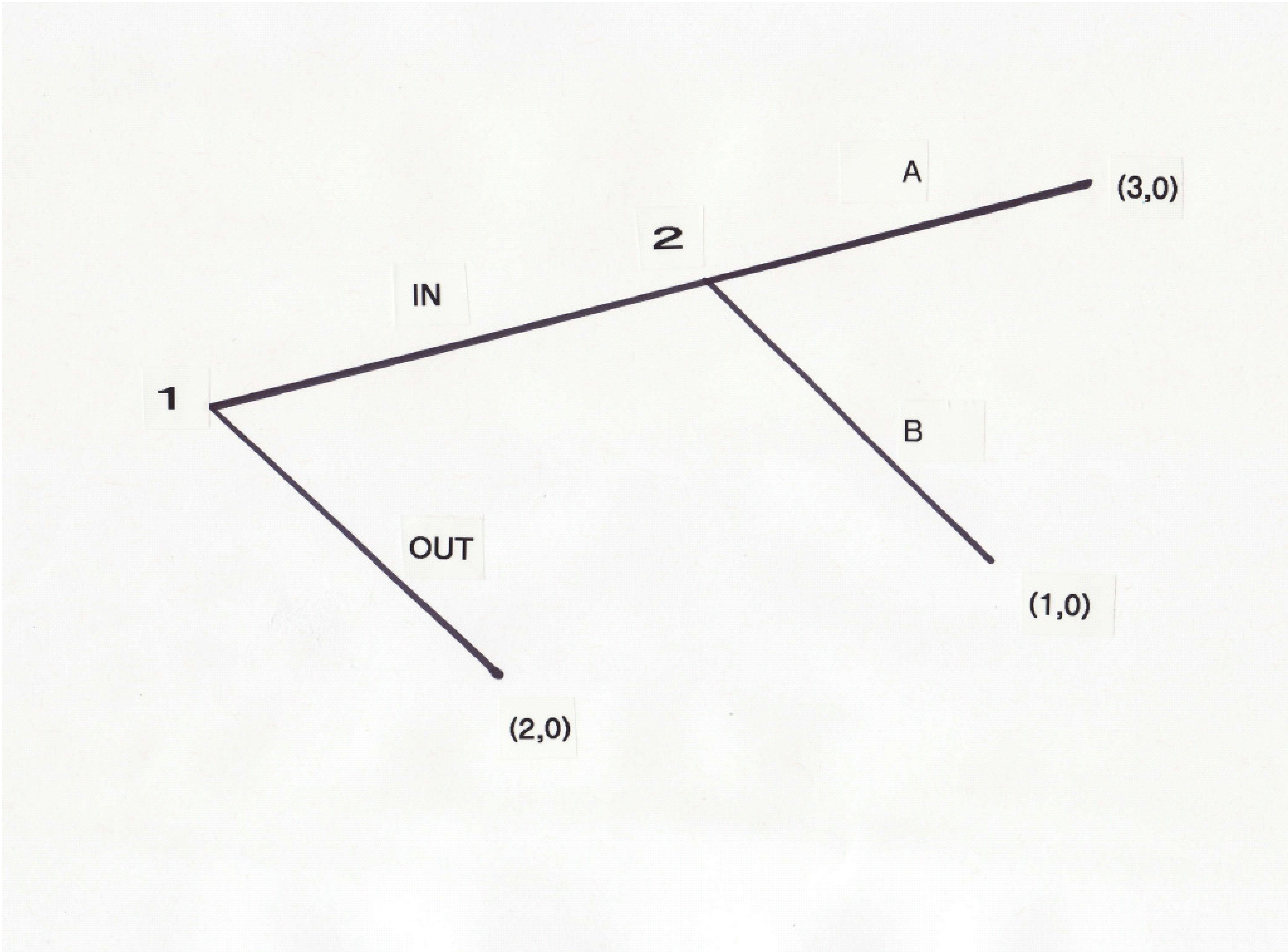
D<u>efinitions:</u> σ is a *self-confirming equilibrium* (SCE) if for each player i and each s_i with $\sigma_i(s_i) > 0$ there are beliefs $\mu_i(s_i)$ about other players' strategies such that

(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}).

This allows player *i* to rationalize each strategy in the support of σ_i with a different belief. <u>Reason</u>: there are many agents in the role of each player, and different agents in the role of player i may have observed play at different nodes.

"Unitary" SCE: one belief per player.



SC

SCE: beliefs are not restricted to reflect knowledge of opponents' payoffs.

"Rationalizable self-confirming equilibria" (RSCE): players' beliefs are consistent with what they observe, as in SCE, and players know the payoff functions of their opponents and expect them to play rationally-*provided* that the opponents haven't yet done anything "irrational."

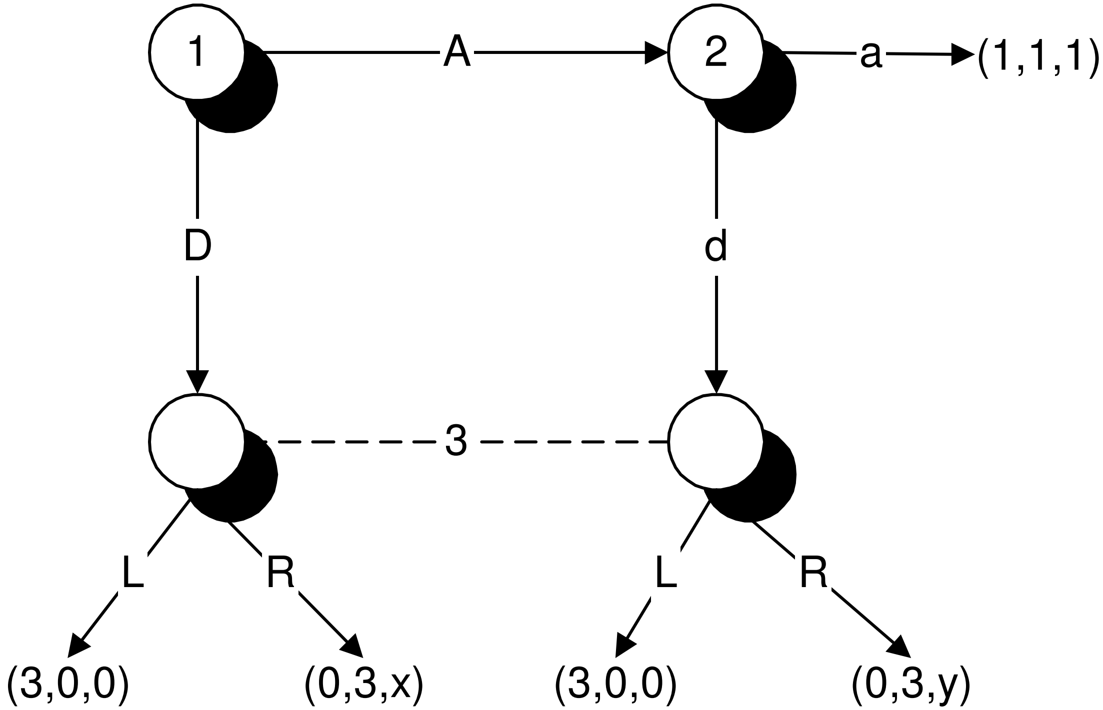
Unitary RSCE coincides with backwards induction in two-stage games of perfect information (1 can predict 2's play), but if a player can move twice on a path it is much weaker and more like SCE.

SCE doesn't require beliefs over strategies are consistent with a common prior.

Aumann [1987]: common prior over strategies equivalent to correlated equilibrium; don't want to assume this here.

Instead ask when repeated observations will lead agents with different priors to a common posterior.

There can be unitary SCE that differ from Nash because 2 players disagree about the play of a 3rd:



(A,a,L) is a SCE.

It can also arise from a learning process where players 1 and 2 update beliefs about opponents' play as in fictitious play: As long as the prior makes them both play A, they get no data on 3's play, and hence don't update.

But (A,a) is not a NE 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*.

Let $\bar{H}(s)$ be the path of s.

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 $h \in \bar{H}(s^i, \hat{s}^{-i})$.

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

-Satisfied in games of perfect information, more generally multistage games with observed actions, and all two-player games of perfect recall: With two players, both players must know whether it was their deviation other opponents that led them to a particular information set.

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

In games with observed deviators, the outcome of any independent unitary self-confirming equilibria is the outcome of a Nash equilibrium.

Idea: specify that each player's off-path actions are exactly those that the player's opponents believe would be played. Unitary beliefs implies there is only a single set of beliefs associated with each player *i*, and the independence condition ensures that these beliefs correspond to a mixed strategy profile. Observed deviators implies that only one player's beliefs about play at an off path information set h_i are relevant- namely the beliefs of the player *j* who could deviate and cause that information set to be reached. So we can specify that the actual play at h_i corresponds to the beliefs of player *j*.

Various learning models have SCE as long run outcome.

But as in a one-armed bandit patient players may choose to experiment.

Classic One-armed bandit

Choose "Out" : get 1 with certainty.
"In" has fixed but unknown probability distribution, prob. p of payoff 2 and (1-p) of payoff 0. Objective is discounted sum of payoffs.

If expected value of p< ½, the risky arm never tried when $\delta = 0$.

But a patient player may experiment:

a) If prob .6 that p=0, "always out" has normalized payoff of 1; **"**try In once then switch to Out forever iff get 0" has higher payoff if $\delta > 1/3$.
b) If prior is that p is either .8 or .2, one observation not fully informative; use dynamic programming to decide whether to play In or Out.

Experimentation in discounted bandit problem ends in finite time with probability one: there is a T such that at all t>T the arm chosen is myopically optimal.

There can be positive probability that play “locks on” to the safe arm when it is suboptimal, but probability of this goes to 0 as $\delta \to 1$. (with time averaging can get first best payoff by experimenting infinitely often but a vanishing fraction of the time.)

Note that always a deterministic optimal solution to bandit problem.

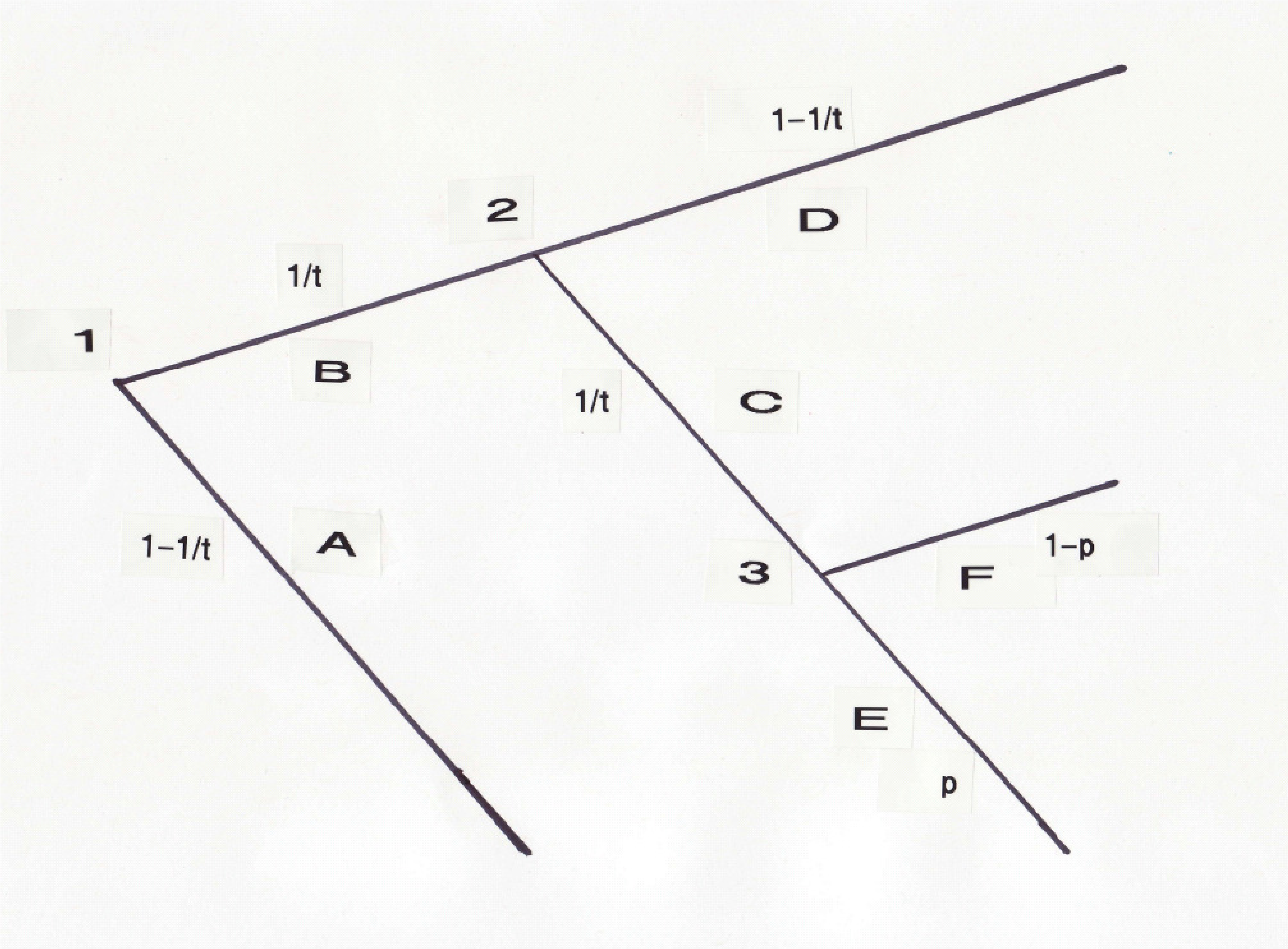
Now consider fictitious play in an extensive form game when there is a lower bound on the probability of each action that goes to 0 at rate $1/t$.

Rules out convergence to non-Nash outcomes: sufficient 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 deviates from the equilibrium path, and wth "$1/t$ experimentation ," these relevant information sets are reached infinitely often (because $\sum_{t=1}^{\infty} 1/t = \infty$).

("zero-one laws")

So from the law of large numbers and asymptotic empiricism, if players get an infinite number of observations of play at an information set, their beliefs about play at that information set become correct.

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


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

This raises, but doesn't answer, the question of how much experimentation players will actually do.

And note that rational Bayesian decision makers typically won't randomize

Need a way to develeop "reasonable" conditions on experimentation.

FL: Continuum population of overlapping generations of Bayesians who maximize expected discounted utility.

<u>Conclusion:</u> patient players experiment enough to rule out non Nash states. But this doesn't say all NE are limits of steady states with patient players.

What about converse here? Can every NE be a steady state? Or will there be enough information about off-path play to eliminate some of them?

FL 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 1 to learn about payoff to action 2.

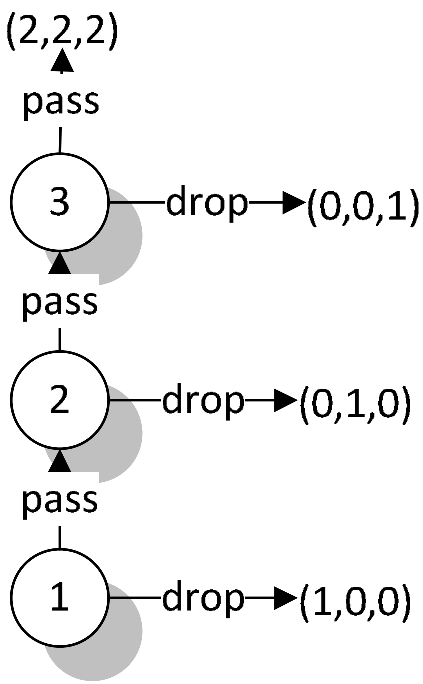
Show that for some non-doctrinaire priors there is no off-path experimentation.
But off-path play isn't completely arbitrary: players one step off the path are reached infinitely often, and so play there looks like a SCE.

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

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.

In simple games with generic payoffs, a subgame-confirmed Nash equilibrium that is "nearly pure" is path-equivalent to a patiently stable state.

Example (Three Player Centipede Game)



Unique subgame-perfect equilibrium: all players pass, but (drop, drop, pass) is subgame-confirmed. Since 2 drops, doesn't learn 3's play.

Open questions: how to extend analysis to more general extensive forms.

a) find right generalization of subgame-confirmed

b) hopefully provide learning-theory foundations

Learning perspective on refinements: instead of impose axioms on equilibrium concepts (such as invariance to allegedly inessential transforms) look to see what the learning model implies.
Out of equilibrium beliefs determined by the relative frequency of various experiments.

Don't get forward induction: best experiment to reach an info set may be strictly dominated…

Open question: implication of rational learning for refinements in specific classes of games such as signaling games: which are the more common e

Adding Prior Information about Payoffs, Rationality, etc to SCE

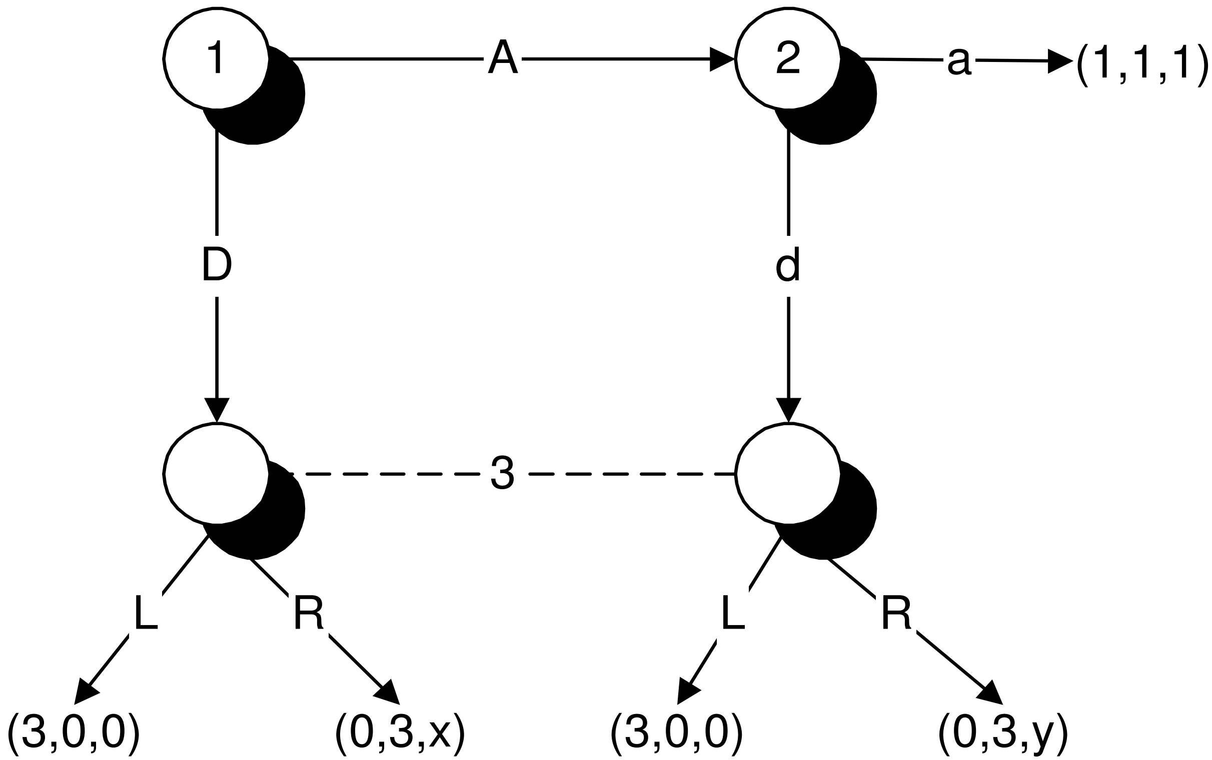
SCE allows beliefs about off-path play to be completely arbitrary- only constraint is what players observe about others' play.

May be a good approximation of some field situations and for experiments in which subjects are given no information about opponents' payoffs.

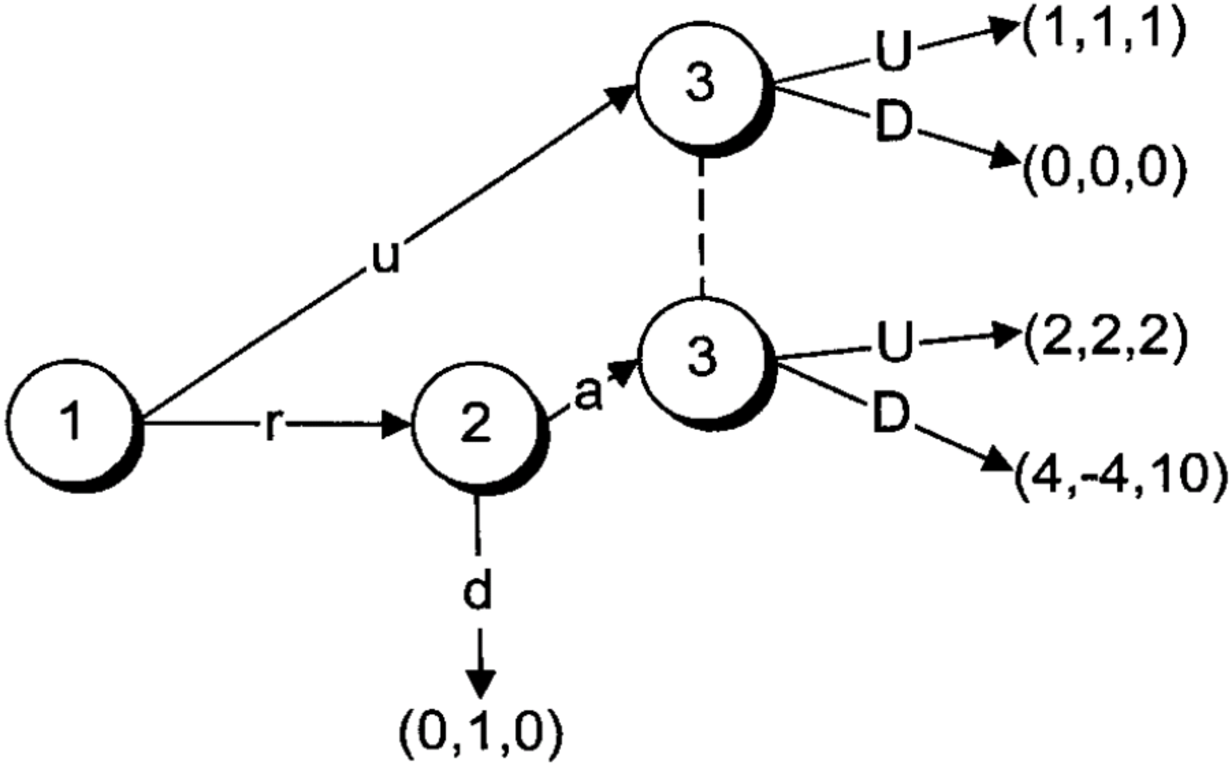
In other cases, players probably do have some prior information about their opponents' payoffs- this leads to rationalizable conjectural equilibrium" (RCE) and "rationalizable self-confirming equilibrium." (RSCE).

RCE is a strategic form concept and doesn't require optimization at off path info sets. RSCE imposes some off-path optimality restrictions. They both resemble rationalizability in having a belief-closed structure- so long as no deviations, each player thinks the other player thinks that …

Intersecting this with SCE gives some restrictions


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)

But RSCE has implications beyond the intersection of SCE+rationalizable:



(u,U) is Nash outcome (so self-confirming), and rationalizable at all nodes, but not RSCE: if 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.

Open: dynamic non equilibrium foundation for RSCE.

Related: SCE and RSCE with ambiguity-averse preferences; here too need learning foundations.

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

Some papers model the idea that players make systematic mistakes in inference, for example Jehiel's analogy-based expectations equilibrium or "ABEE:"

Players group the opponents' decision nodes into "analogy classes," and believe 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.

<u>Example</u> 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 receiving a zero payoff no matter what; suppose player 2 chooses A in state A and B 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 and in state B player 1 should play B1.

ABEE: suppose player 1 views all nodes of player 2 following a given move as belonging to an 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 A1 regardless of the state.

If player 1 observes and remembers the outcome of each game, then as he learns that player 2 plays A 2/3rds of the time, he will also get evidence that player 2's play is correlated with the state, so if he is Bayesian and assigns positive probability to player 2 observing the state, he should eventually learn that this is the case.

However, a Bayesian player 1 could maintain the belief that 2's play is independent of the state provided that he has a doctrinaire prior that assigns probability 1 to this independence.

One explanation: players are unable to remember all they have observed/

In the example this corresponds to player 1 only being 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 the player 1's end-of-stage observation is simply player 2's action, and includes neither Nature's move nor player 1's realized payoff. (*this implies imperfect recall)*

ABEE is closely related to Eyster & Rabin's notion of cursed equilibrium. This focuses specifically on Bayesian games, and assumes "analogy" classes of the form that "opponents' play is independent of their types. "

Assume beliefs convex combination of analogy-based expectations and truth.

Cursedness parameter equals zero is usual Bayesian equilibrium.
Cursedness =1 is a special kind of ABEE.

Open questions: What is ruled out by ABEE w/o constraints on the allowed analogy classes? What sorts of false analogies are relevant for which applications?

Esponda *AER* (2008): adverse selection in a "lemons problem".

Seller values object at s, buyer value is $v = s + x$
s uniform on [0,1], private info for seller, x is a known parameter.

Double auction: submit bids simultaneously, if seller's "ask" is <=buyer's price p then trade <u>at p</u>. (not split the difference)

Seller has weakly dominant strategy: ask=s.

NE: $p^* = \arg\max_p \Pr(s \le p)(E(v \mid s \le p) - p) = p(x - .5p)$,so $p^* = x$.
Fully cursed equilibrium: buyers maximize
$\Pr(s \le p)(Ev - p) = p(.5 + x - p)$,
so
$p^* = (.5 + x) / 2$: bid more than NE if $x < .5$, underbid if x >.5

Esponda: if buyer's beliefs wrong why equal unconditional average? Instead allow for errors but link them to observation structure.

Information structure: see ask each period, see v only when buy.

Esponda defines "naïve buyers" who act as if they

- know distribution of seller "asks" and
- view the quality distribution as independent of own bid.

Here equilibrium is fixed point where buyers optimize given belief that average value is the average they observe, ignoring the dependence of quality on p-
Thus buyers choose p to max $\Pr(s \le p)(\overline{v} - p) = p(\overline{v} - p)$ where

$\overline{v} = E(v \mid s \le p) = p/2 + x$ or $p = 2\overline{v} - 2x$

FOC for buyer is $p = \overline{v} / 2$, so fixed point is $\overline{v} = 4x/3, \quad p^* = \overline{v}/2 = 2x/3$

Underbid for any x.

So buyers pay –less- than in NE as opposed to more as in winner's curse.

In lab players are told typically told the distribution of values.

Esponda argues this is unrealistic as in most field settings players need to learn the distribution of values from their observations- excellent point!

What about SCE here?
Buyer observes own payoff so shouldn't end up worse off then not buying.
But SCE need not be Nash- as buyer can have any beliefs about how whether sellers accept/reject off-path prices.

Would like to apply RSCE, but if buyer only sees v when buys, don’t see full terminal node.

Esponda’s sophisticated buyers correspond to an extension of RSCE to Bayesian games. (Fudenberg-Kamada (2013a,b) do this more generally)

Sophisticated buyer: believes expected quality depends on price and moreover believes it is increasing in price (implied by RSCE and unknown distribution of s, not implied by SCE).

If equil price is p* the sophisticated buyer predicts that quality at a given out of equilibrium price will be given some function $\rho(p, p^*)$.

Buyer learns true expected value of goods sold at price p*, knows it will be no higher for lower prices, and no lower for higher prices.

So will bid at least as much as naïve buyer, maybe more.

Moreover- because buyer observes s each time she purchases- she learns the true distribution of s conditional on $s < p^*$ so she can compute the true expected value of goods at all prices $p < p^*$.

So sophisticated buyer will bid no more than in NE.

Sounds compelling (at least to me) but what really happens?

How descriptive are behavioral/cursed/self confirming in this game?

Hard to say in the field, doing experiments with Alex Peysakhovich..

Subjects get it wrong and overbid. But not an "equilibrium" phenomena- play of individuals doesn't converge. And can be more overbidding than even in cursed!

Explanation: overreaction to most recent history, "short memory."

If run blocks of 10 at a time and report average outcome per bid: subjects get it right.

Blocks of 10 and just show the raw data: little improvement.

So some combination of focusing attention and doing computation is the key.

Similar findings in other stochastic decision problems.

Possible take-away: rational learning models best describe settings where subjects have and use tools such as computers to record and analyze history, maybe less accurate for decisions made w.o aids.

Bigger picture take away: experiments as a useful adjunct to game theory.

Cost of entry is getting lower due to use of Mechanical Turk- lower payments, faster coding, less intrusive oversight.