> The attainment of equilibrium requires a disequilibrium process.

Kenneth Arrow, 1987

## Objections To Equilibrium

- Presumption of Unreasonable Rationality
- Computationally Burdensome
- Multiplicity of Equilibria

## Could Boundedly Rational Players in Repeated Plays Learn (Stumble) Across Equilibria?

## The Bayesian Approach

- Players have priors over future outcomes.
- Players choose a best response to those priors.
- Things happen and priors are updated via bayes rule.

## Coordination Game

- 2 players, complete information, simultaneous move.
- If we both choose HEADS (or TAILS) we each win 1.
- If we choose different things, we each lose 1.

## A Difficulty

- I believe that you will choose H with prob. 1.
- My best response is to choose H.
- You believe that I will choose T with prob. 1.
- Your best response is to choose T.
- Outcome is (H, T). An event that our priors assign 0 probability to.
- How do we use Bayes rule to update on a 0 probability event?

## Three Fixes

1. Ensure that priors are 'sufficiently diffuse'.

   Nachbar ('97) shows under general conditions that this does not work.

2. Restrict priors.

   Kalai and Lehrer ('93) show that when priors are **absolutely continuous** wrt to each other get convergence to Nash equilibrium.

3. Forget bayes.

## Non-Bayesian Approaches

- Finite State Automata.

  Model players as bounded depth automata.

  (Gang of 4, Neymann, Rubinstein and Zemel)

- Evolutionary models.

  Fix a game $G$. Population of individuals each programmed to play a particular pure strategy of $G$. Players are randomly matched and the fraction playing a particular strategy waxes and wanes in proportion to accumulated payofs.

(Axelrod, Ellison, Epstein, Foster, Fudenberg, Harris, Kandori, Mailath, Maynard-Smith, Rob, Samuelson, Weibull, Young)

- Stimulus-Response Models

  Actions that have done well in the past are more likely to be selected in the future.

  (Borgers, Erev, Friedman, Roth, Rustichini, Sarin, Shenker)

- Forecasting/Adaptive procedures.

## The Set Up

1. A matrix game $G$.
2. 2 players.
3. Repeated interaction.
4. Each player forecasts what the other will do on the NEXT round ONLY.
5. Each player chooses a best response to that forecast.
6. ‘WE’ observe the proportion of times each pair of strategies is played (**the empirical frequency of play**).

## Where does the empirical frequency go?

# BEST RESPONSE

- On next round, rival will play what they did on the previous round.

- Due to Cournot.

- Converges to Nash Equilibrium for some games, eg. the classical quantity game.

- Cycles in some games, eg, the coordination game.

# FICTITIOUS PLAY

- Determine empirical frequency of strategies played by rival upto time $t$.
- Use observed frequency as forecast of distribution over rivals choices.
- Choose best response to that forecast.
- Proposed by George Brown ('51) as heuristic for solving zero-sum games.
- Julia Robinson ('51) shows that in zero-sum games FP converges to Nash Equilibrium.
- Miyasawa ('61), Monderer and Shapley ('96) show that in non-degenerate 2 by 2 games FP converges to Nash.

- Potential games (Monderer and Shapley ('96).
- Dominance Solvable (Milgrom and Roberts ('90).
- Strategic Complementarities (Krishna ('92)).

## Shapley Game ('64)

- Row and Col go to the same parties and pay attention to what other wears.
- Row is the follower of fashion and likes to wear what Col wears.
- Col is the fashion leader and likes to be *different* from Row.

| Row/Col | red | yellow | blue |
|---|---|---|---|
| red | 1, 0 | 0, 0 | 0, 1 |
| yellow | 0,1 | 1, 0 | 0, 0 |
| blue | 0, 0 | 0, 1 | 1, 0 |

1. Assume both wear RED in period 1.
2. Ties are broken as follows $Y > B > R$.

| t | Row | Col |
|---|---|---|
| 1 | R | R |
| 2 | R | B |
| 3 | B | B |
| 4 | B | B |
| 5 | B | Y |
| 6 | B | Y |
| 7 | B | Y |
| 8 | Y | Y |
| 9 | Y | Y |
| 10 | Y | Y |
| 11 | Y | Y |
| 12 | Y | Y |
| 13 | Y | R |
| 14 | Y | R |

Consider consecutive periods in which players do not switch.

(R, R) → (R, B) → (B, B) → (B, Y) → (Y, Y) → (Y, R)

Number of periods in each of these consecutive periods grows exponentially.

Shapley showed that empirical frequencies did not converge to Nash.

Not obviously a bad thing for FP. fashion is supposed to be in flux!

## Merry-Go Round Game (Foster and Young)

Players *want* to coordinate.

Miscoordination hurts them both *but* one is hurt more than the other.

FP cycles in this game.

## Variations of FP

- Exponential Smoothing/Weighted Average
- Average over last $K$ periods.
- Exponential FP

## Hannan Theorem's

- $a_{ij}$ payoff to ROW from playing $i$ when COL plays $j$.
- Suppose ROW knows in advance the proportion $y_j$ of times that COL will play strategy $j$.
- Best average payoff that ROW can achieve is

$$v(y) = \max_i \sum_j a_{ij} y_j.$$

- Can ROW achieve a time average payoff that is at least $v(y)$ *without* knowing $y$ ahead of time?
- Yes. Hannan ('59)
- Proved independently many times after using different approaches in Game Theory, Statistics, Computer Science,

Information Theory and Operations research.

- Player only needs to remember their strategy choices and the realized payoffs.
- One proof uses exponential FP.
  1. $A_i(t)$ = total payoff from using strategy $i$ in every period upto the $t^{th}$.
  2. Parameter $x > 1$.
  3. In period $t+1$ play strategy $r$ with probability
  $$\frac{x^{A_r(t)}}{\Sigma_i \, x^{A_i(t)}}.$$

If players use Hannan methods to play a 0-sum game, they converge to the Nash equilibrium.

Convergence faster than FP.

Some empirical support.

Convergence not guaranteed in non-0 sum case.

Unlike FP, security level is guaranteed.

## No-Regret (Foster and Vohra)

Look back at all times played strategy $i$.

What if, on *all* those occasions we had played strategy $j$ instead.

If we would be better off by doing this, then we feel regret.

Goal: play so as to *avoid* regret.

**Notation**

1. $L_i(t)$ = loss incurred at time $t$ from playing $i$.
2. $w_i(t)$ = probability of playing $i$ at time $t$.

3. Expected loss upto time $T$ is

$$\sum_{t=1}^{T} \sum_{k} w_k(t) L_k(t).$$

4. If we switch from $i$ to $j$ loss becomes

$$\sum_{i=1}^{T} \sum_{i} w_i(t) L_i(t) + \sum_{t=1}^{T} w_i(t)(L_j(t) - L_i(t)).$$

5. If

$$-\sum_{t=1}^{T} w_i(t) L_i(t) + \sum_{t=1}^{T} w_i(t) L_j(t) < 0$$

we feel regret.

6. The **pairwise regret** of switching from $i$ to $j$ is

$$R_T^{i \to j} = \sum_{t=1}^{T} w_i(t) L_i(t) - \sum_{t=1}^{T} w_i(t) L_j(t).$$

7. Regret is
$$R_T = \max_{i,j}(R_T^{i\to j})^+.$$

8. Goal is to choose $\{w(t)\}$'s using history of past plays so that regret is small, i.e.
$$R_T(S) = o(T).$$

## Properties of No-Regret

- No-regret procedures exist.
- Hanann theorem can be derived as special case of No-regret.
- If players use No-regret they converge to **Correlated Equilibrium**.

## Existence of No-Regret Procedures

1. Case of just two strategies, 0 and 1.
2. Define ancillary 2 by ? matrix game $M$.
3. Payoffs in $M$ are vectors.
4. We are row.
5. If at time $t$ we choose strategy '0' with probability $w_t$, payoff in $M$ is vector
$$v_t = (L_0(t) - L_1(t), 0).$$
6. If at time $t$ we choose strategy '1' with probability $1 - w_t$, payoff in $M$ is vector
$$v_t = (0, L_1(t) - L_0(t)).$$

7. Observe that

$$V_T = \frac{\Sigma_{t=1}^{T} v_t}{T} = (R_T^{0\to 1}/T, R_T^{1\to 0}/T).$$

8. If $V_T$ is in negative orthant then done.

9. If not, consider hyperplane that separates $V_T$ from negative orthant:

$$(\frac{R_T^{0\to 1}}{T})^+ x + (\frac{R_T^{1\to 0}}{T})^+ y = 0.$$

10. In period $T+1$ choose strategy '0' with probability $p$ so that (in expectation) $v_{T+1}$ will be on other side of separating hyperplane.

11. In expectation the $x$ component of $v_{T+1}$ is

$$(p[L_0(T+1) - L_1(T+1)]$$

and the $y$ compnent is

$$(1-p)[L_1(T+1) - L_0(T+1)]).$$

12. Choose $p$ so that $v_{T+1}$ lies on hyperplane:

$$(\frac{R_T^{0\to1}}{T})^+ p[L_0(T+1) - L_1(T+1)]$$

$$+(\frac{R_T^{1\to0}}{T})^+ [L_1(T+1) - L_0(T+1)] = 0.$$

13. Solve for $p$:

$$p = \frac{(R_T^{0\to 1})^+}{(R_T^{0\to 1})^+ + (R_T^{1\to 0})^+}.$$

14. Now apply Blackwell's approachability theorem.

15. In general case need to solve system of equations to find $p$.

16. Rate of convergence is $O(1/\sqrt{T})$.

## Hanan Theorem

1. $A_i(T) =$ accumulated loss from playing pure strategy $i = 0, 1$ in every period upto $T$.

2. $W^T$ = expected loss from playing strategy '0' in period $t \leq T$ with probability $w_t$.

3. Choose $w_t$'s so as to have no regret.

4. $W^T \leq \min\{A_0(T), A_1(T)\} + o(T)$.

To prove assume wlog that $A_0(T) \leq A_1(T)$.

If no regret:

$$(R_T^{1\to 0})^+ = o(T).$$

$$[\sum_{t=1}^{T}(1-w_t)(L_1(t)-L_0(t))]^+ = o(T).$$

$$\sum_{t=1}^{T}(1-w_t)(L_1(t)-L_0(t)) \leq o(T)$$

Add $A_0(T) = \sum_{t=1}^{T} w_t L_0(t)$ to both sides

$$\sum_{t=1}^{T}[w_t L_0(t)+(1-w_t)L_1(t)] \leq \sum_{t=1}^{T} L_0(t)+o(T).$$

# Regret Matching (Hart and Mas-Collel)

1. Fix parameter $x > 0$ sufficiently large to guarantee convergence.

2. Let $j$ be strategy chosen at time $t$.

3. In time $t+1$ play strategy $i \neq j$ with probability
$$\frac{(R^{j \to k})^{+}}{xT}.$$

4. In time $t + 1$ play strategy $j$ with probability

$$1 - \sum_{i \neq j} \frac{(R^{j \to k})^+}{xT}.$$

If *all* players use regret matching they converge to Correlated equilibrium.

Informational requirements for No-regret and regret matching are low.

Enough to know loss incurred on each round.

# 1 Notes

In the physical sciences every equilibrium is the natural 'rest' point of some dynamical system. Game Theory, in contrast, has equilibria, a multitude of them, but all divorced from any dynamical system. This absence of a natural dynamic associated with, say, Nash equilibria makes it impossible to draw conclusions about play that is not at a Nash equilibria. Are, for example, the players 'close' to a Nash equilibria? If we gave them enough time, would they settle upon a Nash equilibria?

The search for a plausible dynamic whose 'rest points' would be a Nash equilibria of the game has become, in recent years, a promising approach for justifying Nash or other kinds of equilibria. We can associate a dynamical system with a game in many ways, we describe only one.

Fix a two person game $G$ that will be played repeatedly over time. The players while rational, are only boundedly so.[1] They never look farther than the next round in choosing their strategies. On the eve of each round players make a forecast of what their opponent will play and choose a best response to that forecast as their play for that round. What makes the system dynamic is that the forecast depends on the sequence of outcomes observed by each player to date. What one keeps track of is the empirical (joint) distribution of plays and the question one asks is this: does the empirical distribution converge and if so, is it to something one would recognize as the Nash, say, equilibrium of the game?[2] Implicit in this question is the requirement that the forecasting scheme implemented by the players should be as simple as possible. Forecasting in this context is often called learning. Players form beliefs about the future on the basis of the past. Test them in actual play and subsequently revise those beliefs on the basis of the outcome. The obvious deficiency with this model of learning in games is the myopia of the players. They play $G$ infinitely many times and learn not the equilibria of the repeated game but the one shot game! What is more, the absence of discounting makes learning a very leisurely affair. Alternative models have been proposed to account for these drawbacks, but since this model is simple and easy to understand we stick with it for this exposition.[3]

[1] This is inspite of Shakespear's own views as expressed by Hamlet: 'What a piece of work is a man! how noble in reason! how infinite in faculties!'

[2] This is not the only question one could ask but is certainly an important one.

[3] The reader interested in a comprehensive treatment should see the forthcoming book by Drew Fudenberg and David Levine entitled 'Learning in Games'. Copies may be had

## 1.1 Best Response

The earliest model of learning of this type was proposed as far back as 1838 by Cournot.[4] Cournot's forecasting rule is very simple. Whatever your rival did on the last round is what they will do on the next round. Put differently, play a best response to your rival's actions on the last round. For some games this 'best response' dynamic, as it is called, converges to a Nash equilibrium of the game, but not in general. As an example, consider the matching pennies game. Suppose in the first round one player chose H and the other T. They would continue mismatching till the end of time.

## 1.2 Fictitious Play

The most famous of learning rules for games is called Fictitious Play (FP), first conceived in 1949 by George Brown. In a two person game it goes as follows: *Row computes the proportion of times up to the present that Column has played each of his/her strategies. Then, Row treats these proportions as the probabilities that Column will select from among his/her strategies. Row then selects the strategy that is his/her best response. Column does likewise.*

In 1951 Julia Robinson proved that FP converges to a NE in 2 person zero sum games.[5] After the Robinson paper, interest naturally turned to trying to generalize Robinson's theorem to non-zero sum games. In 1961, K. Miyasawa proved that FP converges to a NE in 2-person non-zero sum games where each player has at most two strategies. In 1964 Lloyd Shapley dashed hopes of a generalization by describing a non-zero sum game consisting of three strategies for each player in which FP did not converge to a NE. Here is Shapley's original example:

over the web from David Levine's home page at http://levine.sscnet.ucla.edu/

[4] Adherents of the gospel according to the Santa Fe Institute take note. There are very few new things under the sun and the search for emergent structure is not one of them.

[5] Zero sum games are ones in which what one player wins is exactly what the other loses.

Payoff Matrix for Shapley Game

| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 1\0 | 0\1 | 0\0 |
| 2 | 0\0 | 1\0 | 0\1 |
| 3 | 0\1 | 0\0 | 1\0 |

As observed by Shapley, FP in this game will oscillate between 6 states, (1,1) then (1,2), then $(2,2),(2,3),(3,3),(3,1)$, then repeat. FP stays longer and longer in each state, so the periods of oscillation get larger and larger.

Interestingly there is a variation of FP that does have robust convergence properties. It is called **exponential FP** and will be readily recognizable to those in machine learning as the exponential weighted majority algorithm.[6] One can show that in two person-zero sum games if both players use exponential FP, one gets convergence to the Nash equilibrium. In non-zero sum games this is no longer true. Nevertheless it is still possible to say something of interest.

Consider a two player game (not necessarily zero sum) which will be played repeatedly, where the payoff to the Row from playing strategy $i$ when Column plays her strategy $j$ is $a_{ij}$. Suppose that Row knew the proportion, $y_j$ of times that Column will play her strategy $j$. Knowing this, the best (average) payoff that Row can receive is

$$v(y) = \max_i \sum_j a_{ij} y_j.$$

Is there a forecasting/learning procedure that will allow Row to achieve a time average payoff that is at least $v(y)$ without knowing $y$ ahead of time? Yes, but it is exponential FP rather than FP.[7]

## 1.3 Bayesian Updating

A rather natural approach to the learning question, is to assume the players are Bayesians and that the learning rule is simple Bayesian updating. Players start with a prior distribution over what they think their rival will do and

[6] Variants of the majority algorithm have been developed independently in a variety of disciplines. A survey paper by ourselves entitled 'Regret and the On-Line Decision Problem' describes the history (as best we know it) as well as related results.

[7] In fact the existence of the desired learning scheme was first proved by James Hanan in 1957.

then choose a best response to this prior. After each round of play they revise their priors in accordance with the Reverend Bayes on the basis of what they observed. In the absence of any other conditions this is a dead end. To explain, consider matching pennies, yet again. Suppose I believe that with probability 1 you will play H. You on the other hand, believe that with probability 1 I will play T. Since I play a best response to my beliefs, I assign probability 1 to the event (H, H) and probability 0 to all other outcomes. Given your beliefs, you play T. So, the outcome is (H, T) something I assigned 0 probability to. Now, were I to religously follow the Reverend Bayes, I would find myself dividing by 0![8]

To avoid the difficulty above, one must ensure that no player is 'surprised' so to speak by any outcome. To do this requires that their priors must bear some relationship to each other. This raises the question of why two players would have beliefs that are related to each other. One such relationship is known as **absolute continuity** of the priors. Roughly speaking, absolute continuity of the players priors means that if one player thinks an outcome is possible (will occur with positive probability) then so does the other. Notice that absolute continuity does not require that the players agree on the actual probabilities only that they are zero (or non-zero).

Now suppose you and I are observing a sequence of numbers unfold. We each have a prior of what the full sequence will be and we update this prior in Bayesian fashion as each term of the sequence is revealed. If our priors are absolutely continuous with each other, then in the limit, the posteriors we have over the remainder of the sequence will approach each other; i.e., merge. This merging result was proved in 1962 by David Blackwell and Lester Dubins.[9] About 30 years later, Ehud Kalai and Ehud Lehrer applied the merging result to show that players in a repeated game who update their priors in Bayesian fashion will learn a Nash equilibrium provided their priors are absolutely continuous with respect to each other.[10]

[8] $Pr(A|B) = Pr(A \cap B)/Pr(B)$, but $Pr(B) = 0$.

[9] Ronald Miller and Chris W. Sanchirico give a beautifully intuitive proof in their paper entitled 'The Role of Absolute Continuity in Merging of Opinions and Rational Learning', Department of Economics Discussion Paper No. 9697-07, Columbia University. Suppose we observe a 0-1 sequence and there is no merging. Then there is some sequence of events I believe will happen with more than 50% probability and you think will happen with strictly less than 50% probability. It is precisely this difference of opinion that leads to a horse race. We are each willing to bet that we are right (on average) and the other wrong but this violates absolute continuity!

[10] Other forms of 'Bayesian' learning have been investigated, but the results are not as

## 1.4 Calibration

A popular objection to Nash equilibrium is that it is inconsistent with the Bayesian perspective. A Bayesian player starts with a prior over what their opponent will select and chooses a best response to that. To argue that Bayesians should play the NE of the game is to insist that they each choose a *particular* prior. Robert Aumann has gone further and argued that the solution concept consistent with the Bayesian perspective is not Nash equilibrium but Correlated Equilibrium (CE).[11] Given this, one might ask whether it is possible to learn a Correlated equilibrium. In this case there is a 'positive' answer and it depends on the notion of calibration.[12]

Calibration is one of a number of criteria used to evaluate the reliability of a probability forecast. It has been argued by A. P. Dawid[13] that calibration is an appealing minimal condition that any respectable probability forecast should satisfy. Dawid offers the following *intuitive* definition:

> Suppose that, in a long (conceptually infinite) sequence of weather forecasts, we look at all those days for which the forecast probability of precipitation was, say, close to some given value $p$ and (assuming these form an infinite sequence) determine the long run proportion $\rho$ of such days on which the forecast event (rain) in fact occurred. The plot of $\rho$ against $p$ is termed the forecaster's *empirical calibration curve.* If the curve is the diagonal $\rho = p$, the forecaster may be termed *well calibrated.*[14]

Roughly, calibration says that the empirical frequencies *conditioned on the assessments* converge to the assessments.[15]

The game theoretic importance of calibration follows from a theorem of Dawid's. Given the Bayesians prior look at the forecasts generated by

strong. Convergence to Nash equilibria is not guaranteed except in special cases.

[11] Aumann shows that if each player has a prior and that it is common knowledge that each player chooses the strategy that maximizes expected utility with respect to her prior than the strategies chosen constitute a correlated equilibrium.

[12] See our paper: 'Calibrated learning and Correlated Equilibrium', *Games and Economic Behavior*, forthcoming.

[13] 'The Well Calibrated Bayesian', *Journal of the American Statistical Association,* 77, #379, 605-613, 1982

[14] Dawid (1982) page 605. His notation has been changed to match ours.

[15] For a proof of the existence of forecasting schemes that are calibrated see our paper 'Asymptotic Calibration', *Biometrika*, forthcoming.

the posterior. The sequences of future events on which this forecast will not be calibrated, have measure zero. That is the Bayesian's prior assigns probability zero to such outcomes. Thus, assuming all players use a forecast that is calibrated, in repeated plays of the game, the limit points of the distribution of plays are correlated equilibria. The correlation device is the history of past plays.