Modelling How People Learn in Games

Ed Hopkins
Economics
University of Edinburgh
E.Hopkins@ed.ac.uk, http://homepages.ed.ac.uk/hopkinse/

Computational Thinking Seminar
6th Aug 2008

Game Theory and Nash Equilibrium

- Game theory is used in economics and other disciplines to explain and predict behaviour in situations where agents interact.
- Examples include
  - Pricing decisions by competing firms.
  - Cooperation in social situations (prisoner's dilemma, ultimatum and trust games).
  - Animal behaviour in zoology.
  - Choice of route in systems where congestion is a factor (roads, internet)

Nash Equilibrium and its Problems

- The main tool of game theory is Nash equilibrium (NE), first proposed by John Nash (1951).
- The standard approach is to calculate the NE and use that as a prediction for behaviour.
- Well-known major problems with NE:
  - Difficult to compute for professionals - what hope for real world agents?
  - Involves a great deal of coordination
  - Multiple answers: often many equilibria.

Learning in Games

- One possible answer is to assume that players learn using simple adjustment rules.
- These rules assume little or no knowledge of the structure of the game that is being played.
- In effect, the problem of calculating equilibrium is distributed amongst the different players.
- Rules/algorithms chosen on the basis of simplicity and realism not optimality.
- Nonetheless, theory shows that adaptive learning can often lead players to NE.
- Further, these learning processes reject some NE so reduces the effective number of equilibria to consider.

Today's Talk

- Outline shortcomings of Nash equilibrium.
- Show how learning theory potentially offers solutions to these problems in a reasonably realistic context.
- I offer two examples that involve both theory and laboratory experiments
  - In the first, learning supports Nash equilibrium.
  - In the second, learning generates behaviour that is entirely distinct from Nash.
- Highlight an important problem: How closely do existing models of learning really fit actual human behaviour? Is it close enough?

First Example: Congestion Problems

- These problems are well known in many disciplines.
- In economics, road pricing. Addressed in terms of learning dynamics by Bill Sandholm (2002, 2007).
- Investigated in many experiments (with human subjects) under the name of the "market entry game".
- Brian Arthur's "Santa Fe/El Farol Bar problem".
- In computer science, routing problems, for example, Roughgarden and Tardos (2003).

The Simplest Congestion Problem

- N players must make a choice between two routes (or resources or locations or markets)
- The payoff to all players to choosing the second route is constant

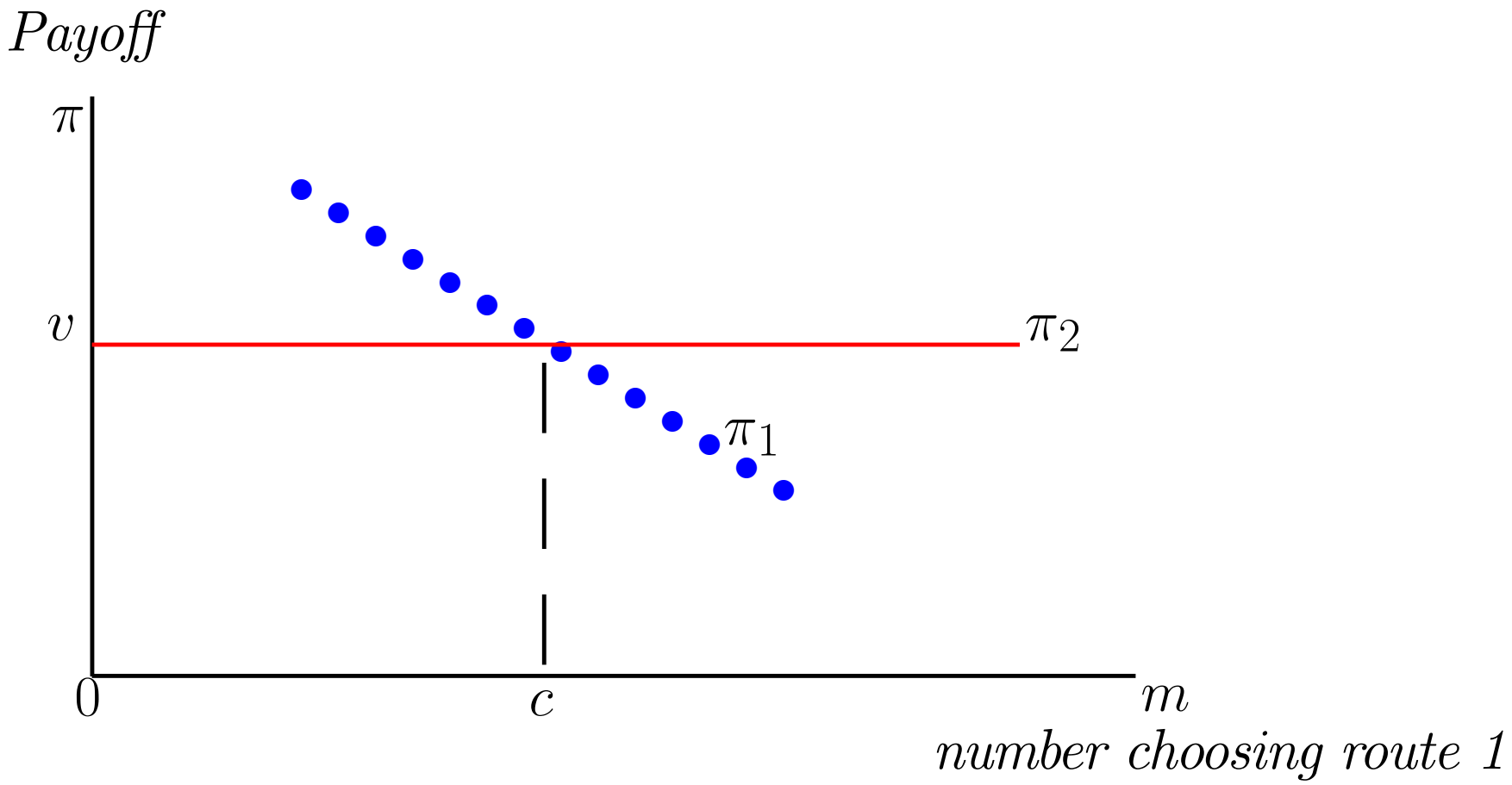
$$\pi_2 = v > 0$$

- The payoff to the first route decreases with the number of players choosing it, in the simplest case

$$\pi_1 = v + c - m$$

  where m is the number of players choosing the second route
- That is, c is the "capacity" of the first route: if more than c players use it, the payoff is worse than to choosing the 2nd route.

A Simple Congestion Problem



The Simplest Congestion Problem - Coordination

- Without a central planner, agents must decide independently which route to take.
- A classic example of strategic uncertainty: what is the best route depends on what others do. How do I predict behaviour of others, given they may be in turn trying to predict my behaviour?
- Possibility of failure of coordination, with too many or too few using route 1.
- But what will people actually do in such a situation?
- Does Nash equilibrium help us to predict?

The Simplest Congestion Problem - Nash Equilibrium

- Even this simple problem has very many Nash equilibria (NE).
- Assume c is not an integer (this makes it simpler!).
- Then there is a set of NE where exactly $\bar{c}$ (largest integer smaller than c) players choose 1, $N - \bar{c}$ choose 1.
- There is a NE where all players randomise with the same probability over choice of 1 and 2.
- There are NE where j players choose 1, k choose 2, and the remaining $N - j - k$ players randomise. The number j can be anywhere between 1 and $\bar{c}$.
- All NE involve a phenomenal amount of coordination.

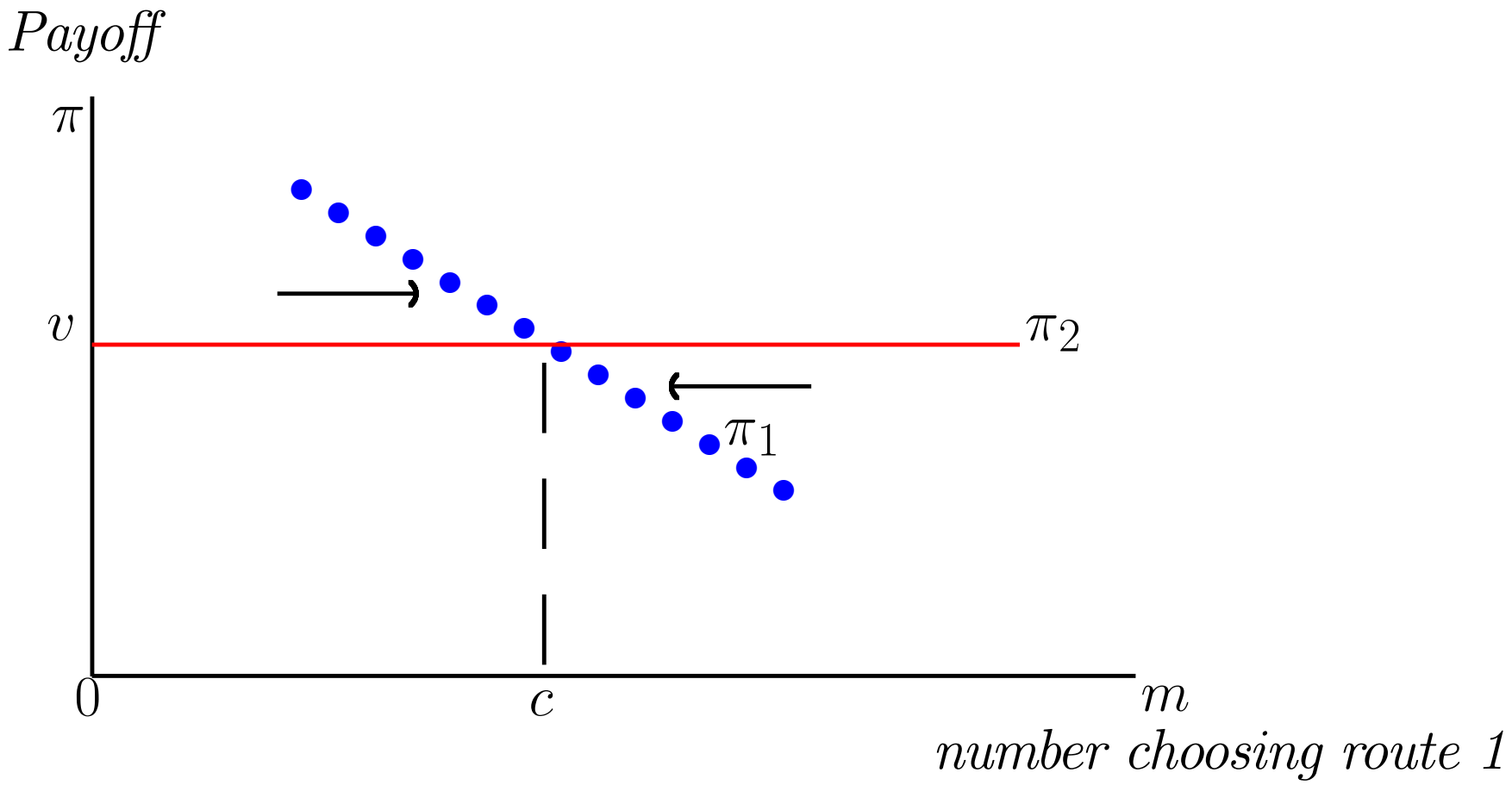
The Problem with Nash Equilibrium

- It is true that in all NE, expected number choosing 1 is between c and $c-1$, giving equalisation of returns to different routes.
- However, clearly different NE have very different variability, with NE where people randomise leading to the possibility of extreme outcomes.
- None of the NE are efficient (only $c/2$ should use route 1 to maximise total welfare).
- But to address this inefficiency (with e.g. congestion pricing), one first has to understand behaviour.
- Can people coordinate on a NE and, if so, which type?

A Simple Argument for Minimal Coordination using Adaptive Learning

- If players use any form of learning rule that tries different actions and adjusts frequencies in response to relative payoffs, this should lead to a minimal level of coordination in the simple congestion problem we consider.
- Simply, if the number choosing 1 is greater than c, its capacity, the return to switching to 2 is greater than staying with 1. If less than c choose 1, then there is an advantage to switching from 2.
- Simple adjustment should lead the number choosing 1 to approach c.

Adaptive Adjustment in a Congestion Problem



Can We Go Further Than This Simple Prediction?

- Even if the numbers choosing route 1 approach c, this does not imply that players are actually in Nash equilibrium.
- Can a more detailed learning model show convergence to Nash equilibrium?
- In fact, learning theory gives a surprisingly precise prediction about outcomes.

Summary of Duffy and Hopkins
Games and Economic Behavior, 2005

- I show that two types of adaptive learning (fictitious play, reinforcement learning) will converge to a pure Nash equilibrium where exactly $\bar{c}$ players choose route 1.
- That is, there is ***"sorting"***. Some players learn always to choose route 1, others always to use route 2.
- We ran experiments (with human subjects) and find that, if complete information is provided, indeed people do sort themselves between the two options.
- With lower levels of information, for example only one's own payoff is revealed, movement toward sorting can be seen in the data but is not complete by the end of the experiment.

Two Learning Rules

- The two most commonly considered forms of learning (in economics at least) have been **reinforcement learning** and **fictitious play**.
- They differ considerably in the level of sophistication assumed and the information that they use.
- Fictitious play (FP) assumes that players know they are playing a game, keep track of payoffs accruing to all strategies and optimise given this information.
- Reinforcement learning (RL) assumes that the probability a strategy is chosen is proportional to past payoffs from this strategy.
- NB "reinforcement learning" appears in many contexts and has many forms.

Modelling Learning Rules with Propensities

- It is possible, nonetheless, to model both using a similar mathematical framework.
- Assume each player has a "propensity" for each possible action, here route 1 or 2. Relative size of propensities determine the probability of taking each action.
- Under FP, in each period propensities for both routes are updated with the realised payoffs to each route whichever route was chosen. If route 2 was chosen, requires construction of hypothetical - what would I have got if I had chosen 1?
- Under RL, propensities only updated with payoff to action actually chosen. No hypothetical reasoning.

Updating Rules

Player i has a propensity in period n for route 1 q^i_{1n} and for 2 q^i_{2n}. $\delta^i_n = 1$ if player i chooses 1 in period n, zero otherwise.

Simple Reinforcement

$$q^i_{1n+1} = q^i_{1n} + \delta^i_n(v + c - m_n),\ q^i_{2n+1} = q^i_{2n} + (1 - \delta^i_n)v,$$

where m_n is the actual number of entrants in period n.

Hypothetical Reinforcement

$$q^i_{1n+1} = q^i_{1n} + v + c - m_n - (1 - \delta^i_n),\ q^i_{2n+1} = q^i_{2n} + v.$$

Choice Rules for FP and RL

y_n^i is a player's probability of choosing route 1 in period n.

- **The reinforcement rule:** randomise proportionally

$$y_n^i = \frac{q_{1n}^i}{q_{1n}^i + q_{2n}^i}$$

- **Traditional FP rule:** choose the best.
  If $q_{1n}^i > q_{2n}^i$, then $y_n^i = 1$,
  if $q_{1n}^i < q_{2n}^i$, then $y_n^i = 0$.

Sorting Results

- For both fictitious play and reinforcement learning, we have a sorting result.
- Under either process, eventually players will play a pure Nash equilibrium where exactly $\bar{c}$ choose route 1 and $N - \bar{c}$ choose route 2.
- Thus, in the long run, there can be exact coordination on a Nash equilibrium, even with minimal information or sophistication on the part of players.

Experimental Procedures

- $N = 6$. That is, groups of six subjects played the market entry game repeatedly for 100 rounds over a computer network.
- Inexperienced subjects.
- Actions labelled "Action X" and "Action Y".
- Capacity $c = 2.1$, payoffs were

No. choosing X	1	2	3	4	5	6
Payoff to X	10.20	8.20	6.20	4.20	2.20	0.20
Payoff to Y	8.00	8.00	8.00	8.00	8.00	8.00

- Every 25 rounds, one round drawn at random to count as the payoff round.

Experimental Treatments

1. Limited Information
2. Aggregate Information
3. Full Information

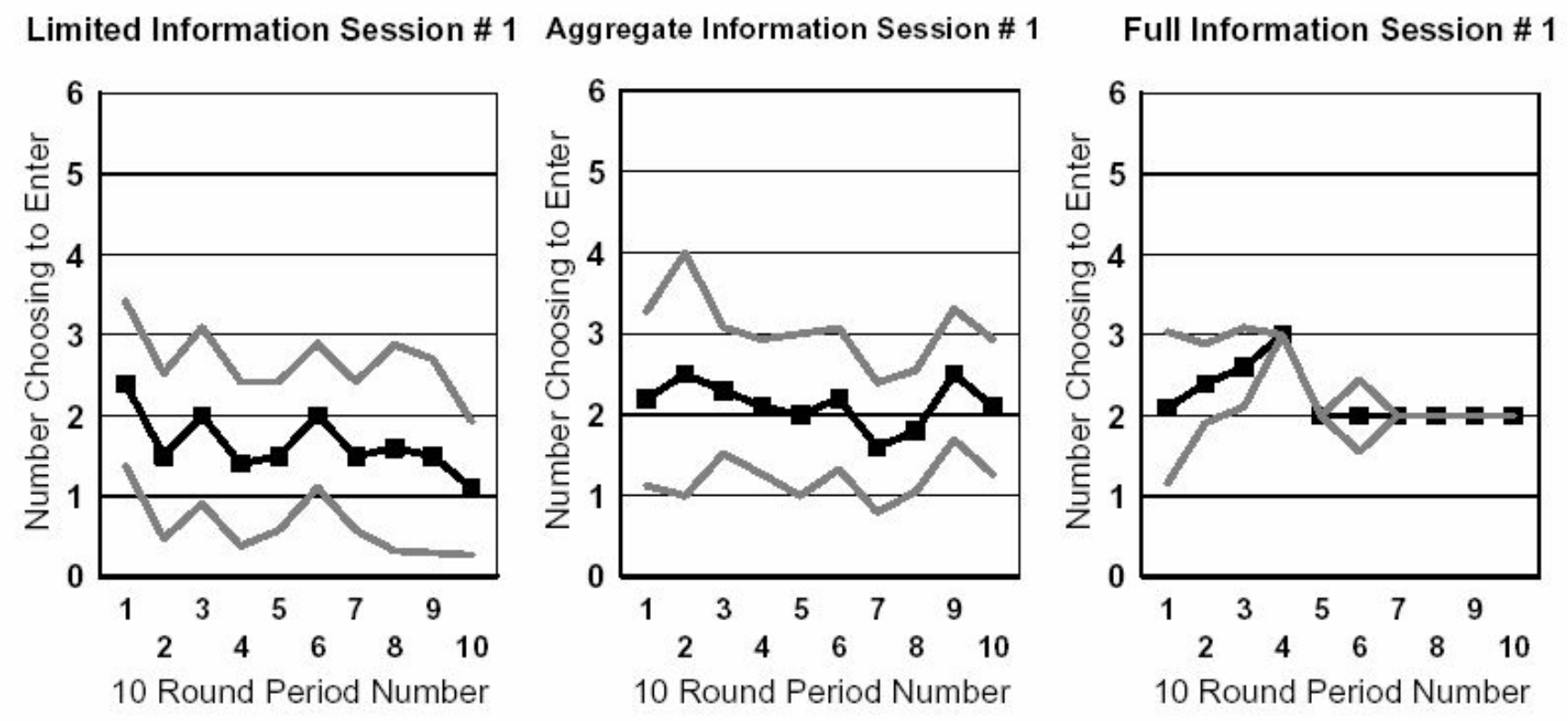
After making a choice, what information do subjects receive?

	Own payoff	Aggregate info	Individual info
Limited	✓	×	×
Aggregate	✓	✓	×
Full	✓	✓	✓

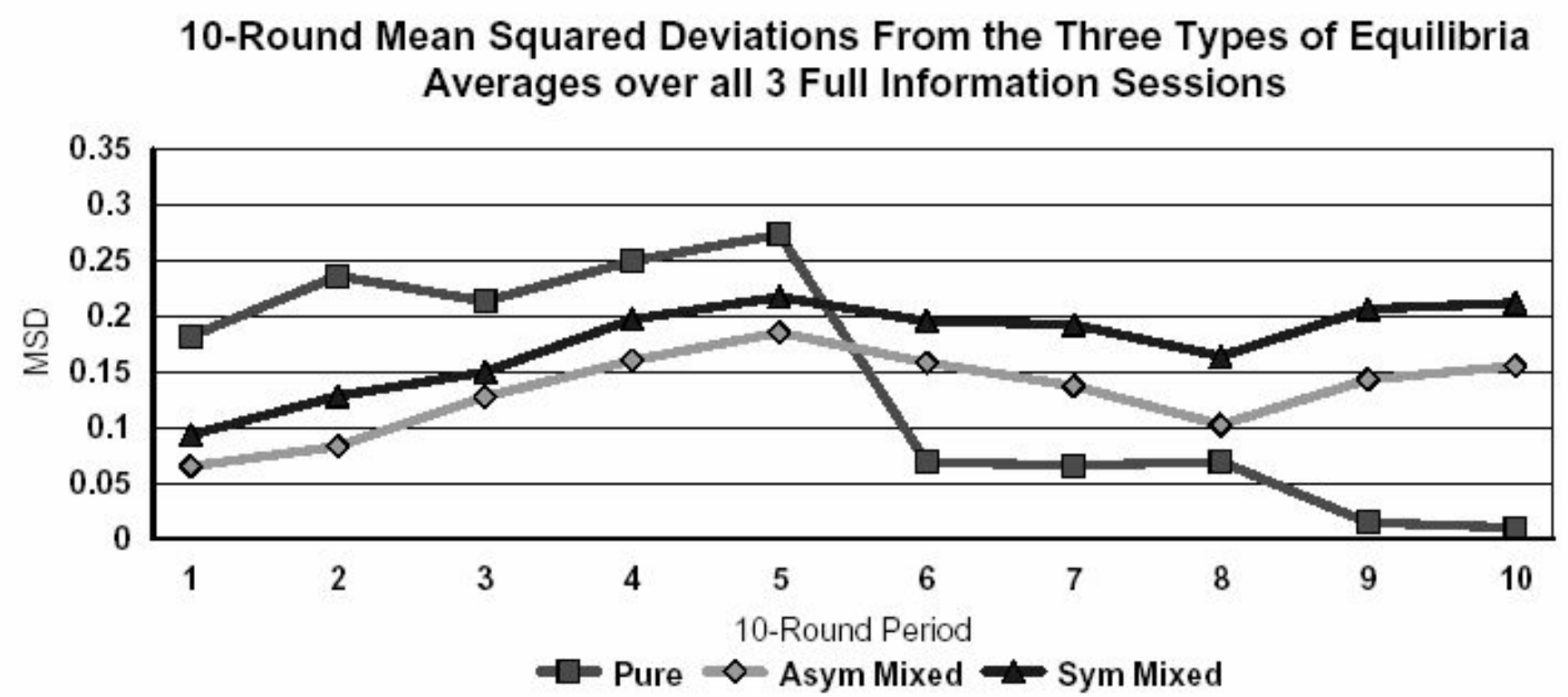
Aggregate: e.g. 2 players chose X, 4 chose Y

Individual: player 1 chose X, player 2 chose Y etc.

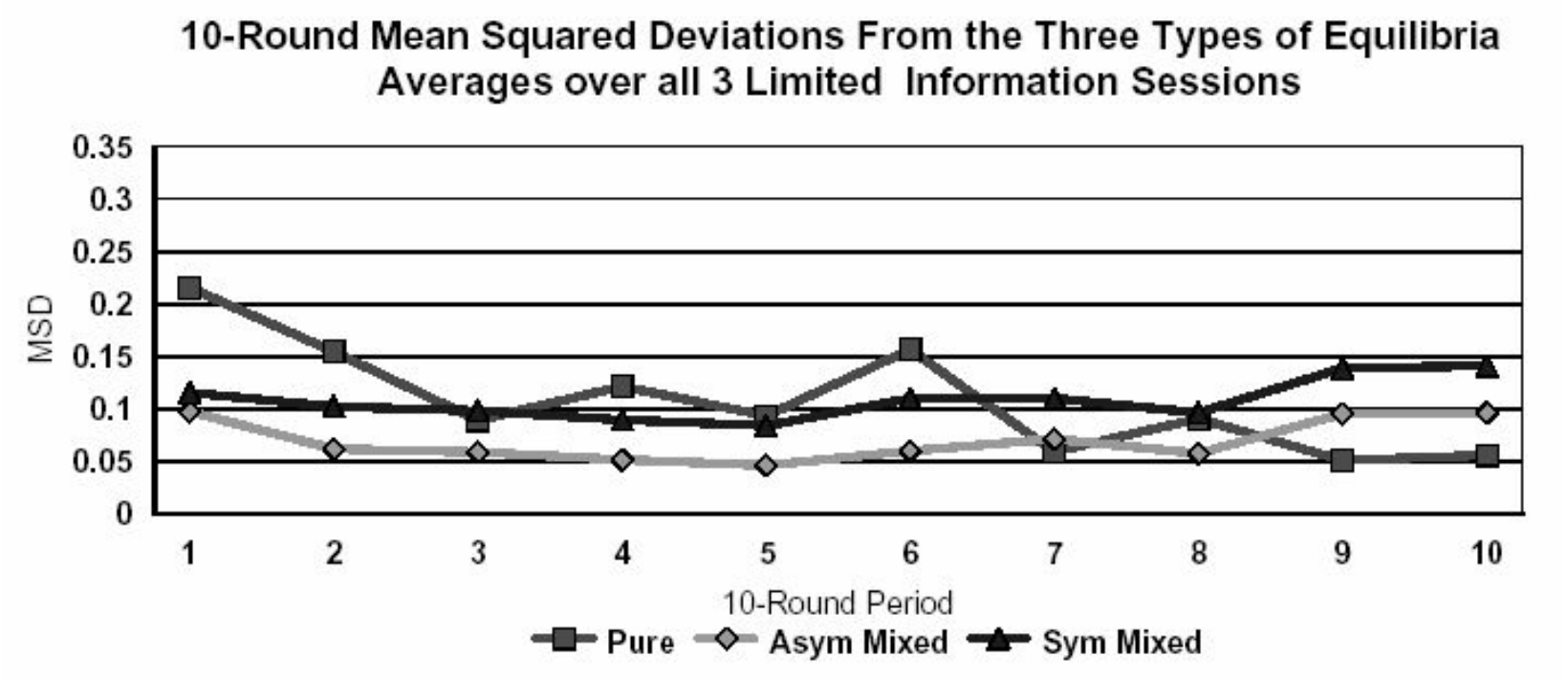
Sorting Depends on Information



Sorting under Complete Information



Sorting under Limited Information



Evaluation of Experiments

- Sorting happens only when players can see the pattern of play of others.
- This particular use of information is not included in current models of learning.
- Without this information, there is movement towards, but not complete sorting.
- In the short run, there is only the broad outline of Nash equilibrium: rough equalisation of return to the two options.
- Both NE and learning approximate human behaviour, but do not capture its finer points.

Second Example: Failure of Convergence to Mixed (Random) Nash Equilibrium

- A support for Nash equilibrium is that learning theory shows that simple adjustment rules can lead players to coordinate.
- However, there are negative as well as positive results: games in which the only Nash equilibrium is unstable under learning. Play should diverge not converge.
- So, in many games, some of significant economic importance, we seem to have no prediction about how people might play.
- Benaïm, Hofbauer and Hopkins (2006) fill in this gap with a precise prediction about what happens when there is divergence from equilibrium.

Experiments to Test the New Theory

(Joint Work with Tim Cason and Dan Friedman)

- We report experiments designed to test between Nash equilibria that are stable and unstable under learning.
- Drawing on recent theoretical results, we have a new, simpler way to test between stable and unstable play.
- We use two games each with a unique mixed Nash equilibrium, one stable and one is unstable.
- Subjects randomly matched in pairs to play one of the games.
- In our experiments there is a difference between the stable and unstable treatments which supports the new theory of non-equilibrium behaviour, but much remains unexplained.

Some Theory in a Random Matching Framework

- Single large population of players
- Time periods $n = 1, 2, 3, ...$
- In each period, players randomly matched into pairs to play a 2 player normal form game
- Approximates random matching protocol in experiments

RPS Games

Two examples of the generalized Rock-Paper-Scissors game

$A =$

	R	P	S
Rock	0, 0	-1, 2	3, -1
Paper	2, -1	0, 0	-1, 3
Scissors	-1, 3	3, -1	0, 0

$B =$

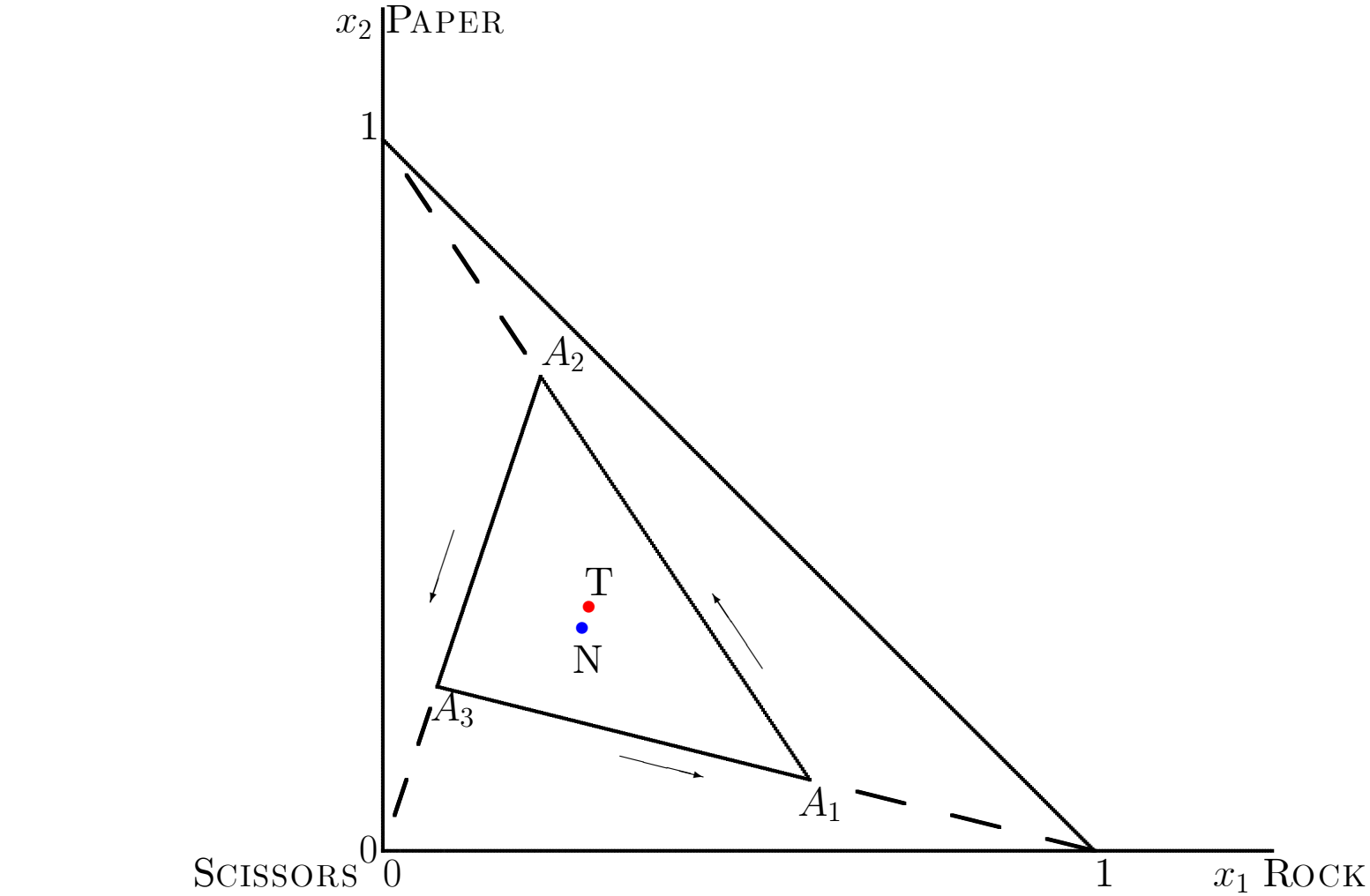
	R	P	S
Rock	0, 0	-3, 1	1, -3
Paper	1, -3	0, 0	-2, 1
Scissors	-3, 1	1, -2	0, 0

- Unique mixed Nash equilibrium (NE) for both games
- NE is stable under most forms of learning in game A
- NE is **unstable** in game B, under fictitious play, reinforcement learning, the replicator dynamics etc.

Fictitious Play in RPS Games

- RPS games have unique interior/mixed equilibrium
- Cycle of best responses that converges or diverges
- NE is stable in game A
- **Unstable** in B
- If NE unstable, fictitious play approaches a limit cycle, named a "Shapley Triangle" in honor of Shapley.
- Hence, in B: under fictitious play, no convergence in behavior or in time average or in beliefs

A Shapley Triangle



The Shapley triangle for game B with the TASP (T) and the NE (N)

New Theory of the Unstable Case

Benaïm, Hofbauer and Hopkins (2006)

Suppose players place weight of $\rho \in [0, 1)$ on last period, ρ^2 on previous..., in constructing propensities that are the basis for choices.

Then, in unstable games,

1. Cycle is close to Shapley triangle
2. But speed is constant $\Rightarrow$ time average converges
3. As $\rho \to 1$, i.e. as greater weight is placed on past experience, this average approaches time average of a complete circuit of the Shapley triangle
4. That is, the time average $\to$ the **TASP:** "time average of the Shapley Polygon" - a new concept

Implications of the TASP

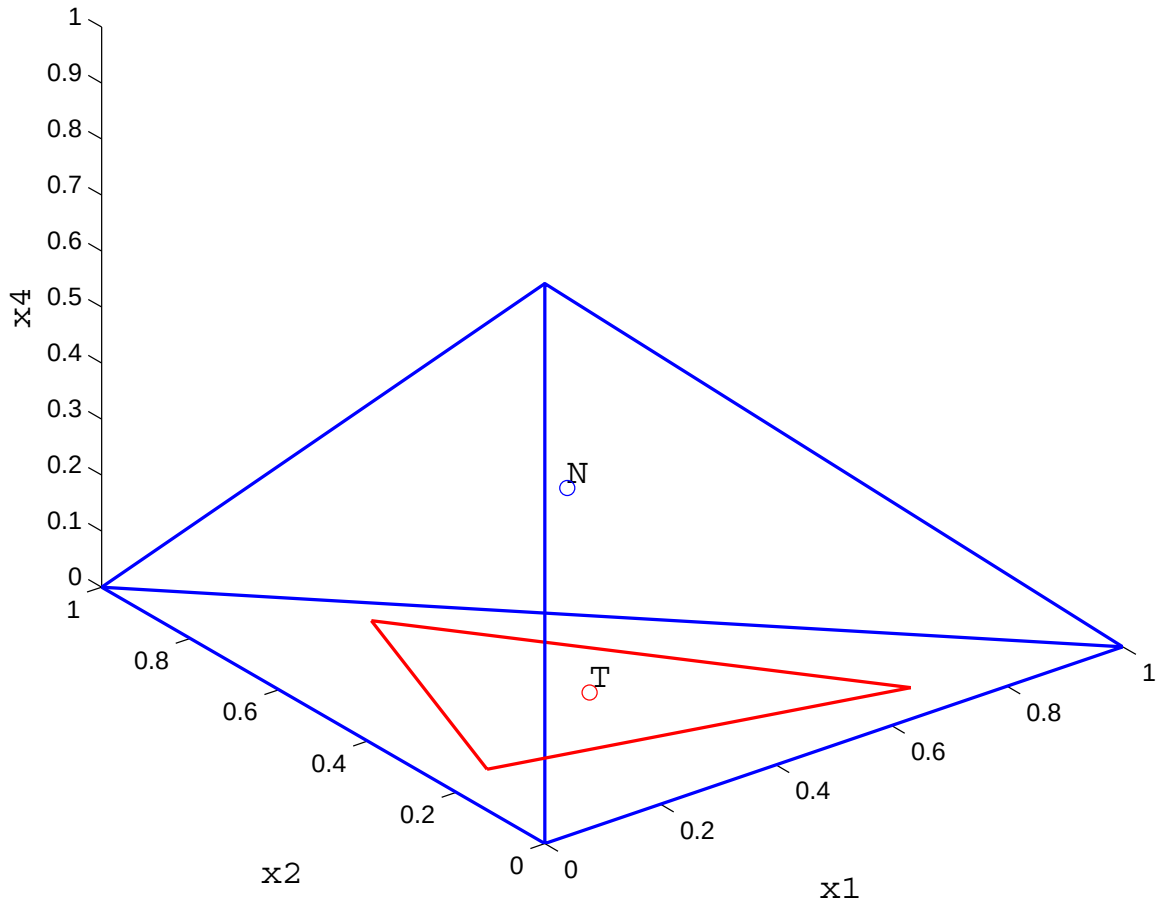
- It gives a point prediction for overall relative frequencies of different strategies even when there is no convergence to NE
  - This point can be close to NE as we have just seen
  - But in general not identical
  - Can be quite different (example to follow)
- It gives a prediction for the dynamics of play: they should follow a specific cycle.

A Game where Nash Equilibrium and TASP are distinct

$RPSD_U =$

	R	P	S	D
Rock	90, 90	0, 120	120, 0	20, 90
Paper	120, 0	90, 90	0, 120	20, 90
Scissors	0, 120	120, 0	90, 90	20, 90
Dumb	90, 20	90, 20	90, 20	0, 0

- RPS with the addition of a 4th strategy D: "Dumb"
- The unique Nash equilibrium is fully mixed and equal to $(1, 1, 1, 3)/6$.
- However "U" is for unstable, FP will approach a cycle which places no weight on D!
- The TASP is $(1, 1, 1, 0)/3$


x4
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
N
T
x2
0
0.2
0.4
0.6
0.8
1
x1
0
0.2
0.4
0.6
0.8
1

A Stable Version of RPSD

$RPSD_S =$

	R	P	S	D
Rock	60, 60	0, 150	150, 0	20, 90
Paper	150, 0	60, 60	0, 150	20, 90
Scissors	0, 150	150, 0	60, 60	20, 90
Dumb	90, 20	90, 20	90, 20	0, 0

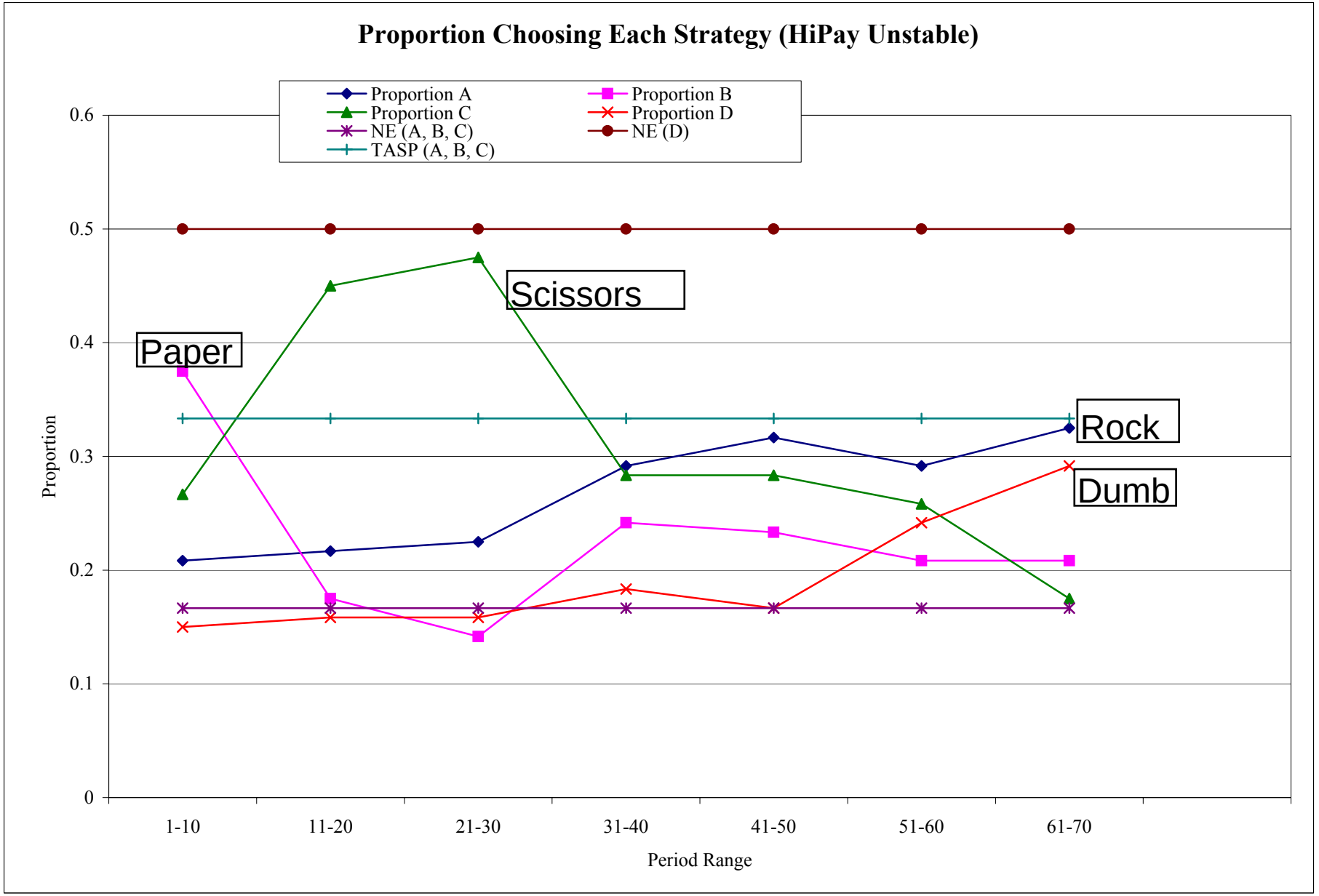
- $RPSD_U$ and $RPSD_S$ have the same fully mixed Nash equilibrium $(1, 1, 1, 3)/6$.
- However, in $RPSD_S$ the NE is an attractor for most learning dynamics including SFP

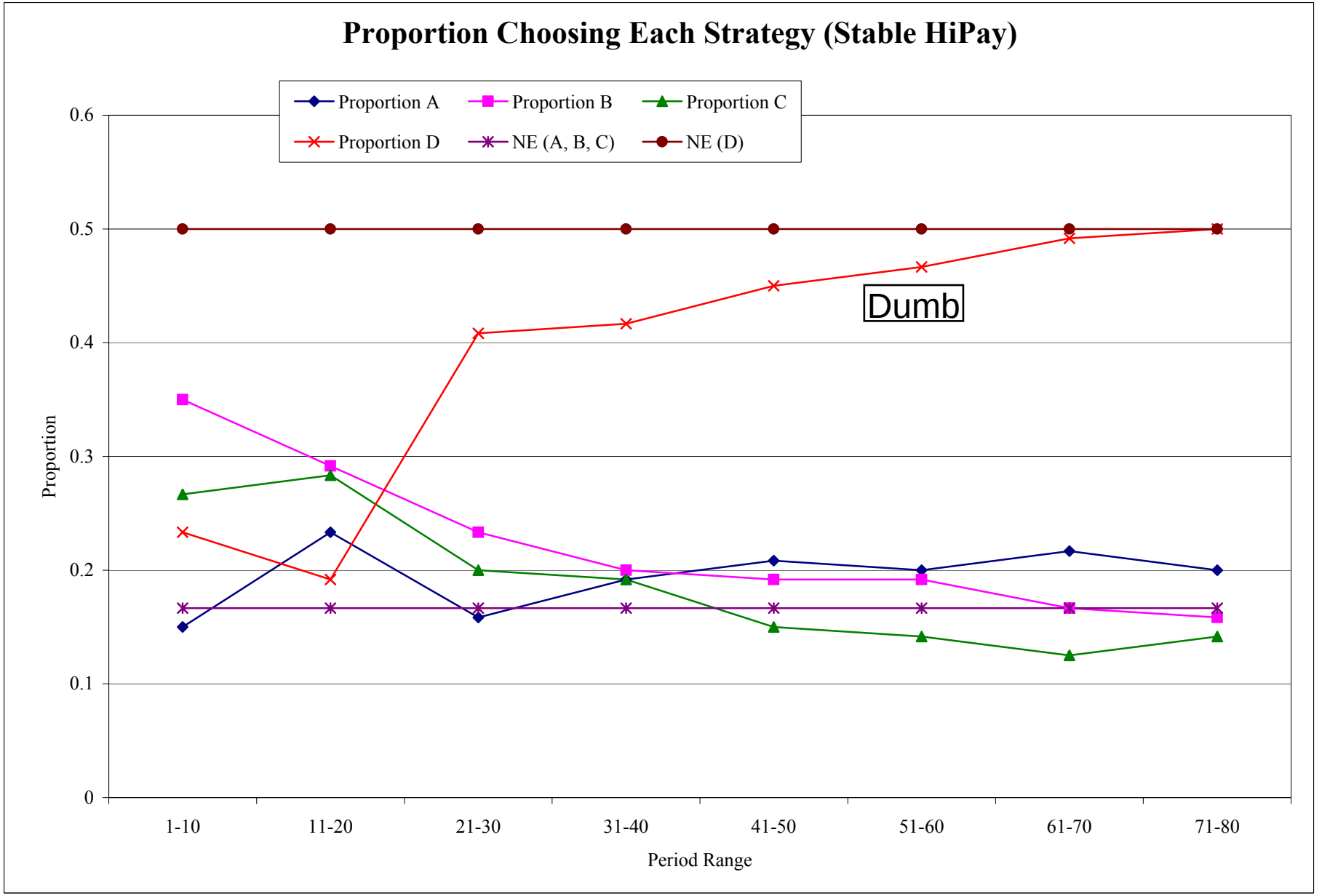
Theoretical Predictions

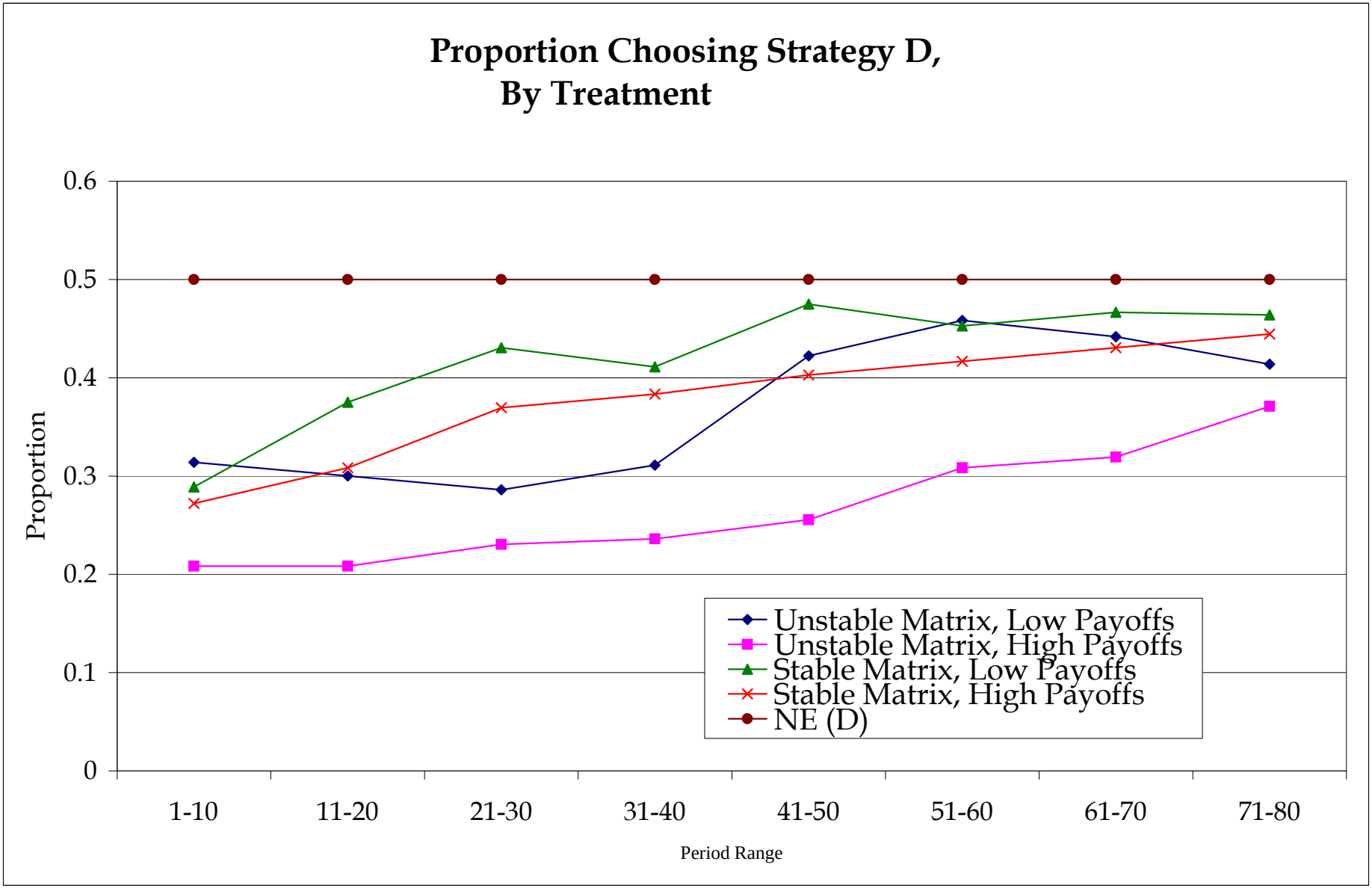
- So learning theory predicts 0 weight on strategy D in $RPSD_U$, and a weight of 0.5 on D in $RPSD_S$.
- Nash equilibrium predicts a weight of 0.5 on D in both games.
- So, the frequency of the strategy D is a ready reckoner for testing between stability and instability, and between learning theory and Nash equilibrium.

Experimental Procedures

- Experiments carried out at Purdue and at UCSC
- 2×2 design: Stable vs Unstable Game $\times$ Hi vs Low Payoffs
  - High payoffs: 100 experimental francs (EF) =\$5
  - Low payoffs: 100 EF =\$2, plus showup fee of \$10
- 3 sessions per treatment; in each, 12 subjects repeated randomly matched over computer network for 80 periods to play one game, matrix known to all subjects.
- Feedback: own action, action of opponent, payoff earned and actions of other subjects.


Proportion Choosing Each Strategy (HiPay Unstable)
Proportion A
Proportion B
Proportion C
Proportion D
NE (A, B, C)
NE (D)
TASP (A, B, C)
0.6
0.5
0.4
0.3
0.2
0.1
0
Proportion
Scissors
Paper
Rock
Dumb
1-10
11-20
21-30
31-40
41-50
51-60
61-70
Period Range


Proportion Choosing Each Strategy (Stable HiPay)
Proportion A
Proportion B
Proportion C
Proportion D
NE (A, B, C)
NE (D)
Dumb
0.6
0.5
0.4
0.3
0.2
0.1
0
Proportion
1-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
Period Range


Proportion Choosing Strategy D,
By Treatment
0.6
0.5
0.4
0.3
0.2
0.1
0
Proportion
1-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
Period Range
Unstable Matrix, Low Payoffs
Unstable Matrix, High Payoffs
Stable Matrix, Low Payoffs
Stable Matrix, High Payoffs
NE (D)

Experimental Results - Summary

- Basic comparative statics are supported by aggregate frequencies.
- Hi-Unstable treatment is further from Nash than Low-Unstable treatment
- All treatments show tendency to move toward Nash in the very long run.
- Movement towards Nash significantly slower in the Hi-Unstable treatment.

Evaluating these Experiments

- We test the new non-equilibrium concept, the TASP, that predicts play in games with unstable Nash equilibria.
- We look at a game for which the TASP and the Nash equilibrium are quite distinct.
- Overall frequencies show that there is a difference between stable and unstable treatments that cannot be explained by NE.
- However, NE does surprisingly well in the long run.
- Again, learning theory does not capture the full details of behaviour.

Conclusions

- Learning theory offers a somewhat more realistic approach to play in games than simply assuming people play Nash equilibrium.
- Nonetheless, it can offer support to Nash equilibrium, in that it allows for players to learn to play Nash even without strategic sophistication or much information.
- But in some cases, it predicts non-equilibrium behaviour quite distinct from traditional theory.
- Our current learning models have some empirical success but do not fully capture the variety and sophistication of actual human behaviour.