THE ECONOMIC JOURNAL

JANUARY 1996

The Economic Journal, **106** (*January*), 1–25. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA.

EQUILIBRIUM IN EVOLUTIONARY GAMES: SOME EXPERIMENTAL RESULTS*

Daniel Friedman

Evolutionary game theory informs the design and analysis of 26 experimental sessions using normal form games with 6–24 players. The state typically converges to the subset of Nash equilibria called evolutionary equilibria, especially under conditions of mean matching and history. Mixed strategy equilibria are explained better by 'purification' strategies than by homogenous independent individual randomisation. The risk dominance criterion fares poorly in some coordination game environments. With small player populations and large gains to cooperative behaviour, some players apparently attempt to influence other players' actions, contrary to a key theoretical assumption.

Strategic interaction over time can be modelled as an evolutionary game if the players do not systematically attempt to influence other players' future actions and if the distribution of players' actions changes gradually. Evolutionary games first appeared in the theoretical biology literature (Maynard Smith and Price, 1973; Maynard Smith, 1982) but in recent years several leading game theorists have used evolutionary games to address longstanding foundational issues of equilibrium selection and convergence (e.g. Binmore, 1987–8; Fudenberg and Kreps, 1988; Selten, 1989). Economists are beginning to notice that the evolutionary approach has unique implications for economic applications. For example, historical accidents may have permanent effects when there are multiple equilibria, and ordinary ('complete information') Nash equilibrium may describe economic outcomes even when decision makers know very little about others' payoffs or strategies (e.g. Crawford, 1991; Friedman and Fung, forthcoming). Empirical evidence clearly is required to assess the economic relevance of evolutionary game theory.

For more than 40 years, laboratory experiments have been the predominant empirical method for testing game theoretic propositions. The laboratory studies necessarily examine strategic interaction over time. In most contexts, the subjects seldom try to influence other subjects' future actions and the action distributions change gradually.[1] Evolutionary game theory therefore is the

* Support by the US National Science Foundation under grant SES-9023945 made this work possible. The final revision owes much to the hospitality of WiThI at the University of Bonn, and to the careful readings of two anonymous referees and the editor of this JOURNAL. I am grateful to Debbie Carson and Carl Plat for their patient research assistance, to Tim Kolar and Thanh Lim for programming assistance, and to seminar participants at Caltech and UCLA and at the University Pompeu Fabra, WEA and ESA sessions for useful comments and encouragement.

[1] Some studies investigate trigger strategies or other strategies designed to influence other players' future actions. The theory of repeated games is more appropriate than evolutionary game theory for such purposes.

natural theoretical framework. Nevertheless, laboratory studies up to the early 1990s relied mainly on orthodox static game theory to define the issues and to structure the design and data analysis.

The research reported here explores the ability of evolutionary game theory to explain laboratory behaviour. The central question is whether behaviour converges to Nash equilibrium for a variety of payoff matrices under various environmental conditions. The exploratory nature of the paper dictates a rather broad set of treatments and rather heavy reliance on descriptive (as opposed to inferential) statistics. The work includes some follow-up experiments and some inferential statistics, but the reader should expect mainly broad tentative findings rather than narrow definitive results.[2]

The most relevant previous research is reported in Van Huyck *et al.* (1992). The authors examine generalised two-person divide-the-dollar games, using laboratory procedures similar to the random-pairwise, No History, one- and two-population protocols defined below. They test the predictive power of evolutionary equilibrium (also defined below) as a Nash equilibrium selection criterion for their games, with generally positive results.

The current paper sketches the basic theory of evolutionary games in Section I, beginning with the simplest 1-dimensional case of single population, two-action linear games. Such games are classified into three types, each with a 2-dimensional analogue. The sketch then mentions theoretical and practical problems that may prevent convergence to either pure strategy (corner) equilibria or to mixed strategy (interior) equilibria. Readers familiar with evolutionary game theory may wish to skip this section.

Section II explains the basic laboratory procedures for the evolutionary game experiments, and introduces the main treatments: the payoff matrices for one and for two populations, the matching protocols (either random pairwise or mean matching), and the information regarding the distribution of other players' choices in previous periods (either provided to all players or to none). Section III presents the results, beginning with graphical summaries of two early sessions. A statistical summary of convergence behaviour in all 26 sessions then follows.

The results are largely consistent with theory and intuition about which treatments and payoff matrices best promote convergence. The limits of applicability for evolutionary game theory are probed, and apparent attempts to influence other players' future behaviour are documented for small populations (four or fewer players) and for extreme choices of payoff matrices. The results generally support the 'purification' view of Harsanyi (1973) and Fudenberg and Kreps (1993) that mixed strategy Nash equilibria are achieved mainly through heterogeneous individual behaviour rather than through homogeneous mixed behaviour. Perhaps the most surprising finding concerns the risk dominance criterion favoured in recent years by some theorists for equilibrium selection in coordination and other games (or economies) with

[2] A companion paper, Cheung and Friedman (1994), focuses on the dynamics of the adjustment process. It tests the explanatory power of a parametric learning model against several alternative adjustment models with generally positive results. It also surveys some of the recent empirical literature.

multiple equilibria. Risk dominance fares poorly in some of the environments tested here, apparently because players' deviations are not random trembles but rather may be deliberate attempts to persuade other players to seek Pareto superior outcomes.

The last section offers a summary and concluding remarks.

I. THEORETICAL BACKGROUND

This section provides a brief introduction to the theory of one and two dimensional linear evolutionary games. See Weibull (1995) and Friedman (1991, 1992) for more general introductions.[3] The essential elements are one or two populations of players, e.g. row players as buyers and column players as sellers; a payoff matrix (or bimatrix); and an adjustment dynamic that specifies how the state (i.e. the distribution of actions within each population) responds to current conditions. The theory identifies the (locally asymptotically) stable steady states, here called evolutionary equilibria (EE), and the basin of attraction of each EE, i.e. the set of states that converge to the EE.

I.A. *Linear One-dimensional Games*

Let $\mathbf{A} = ((a_{ij}))$ be a 2×2 matrix specifying the payoff to a player choosing action i ($= 1$ for top row and $= 2$ for bottom row) when matched with a second player choosing action j ($= 1$ for left column and $= 2$ for right column). Assume all players come from the same population and perceive the same strategic situation – they all think of themselves as choosing rows. Then the second player's payoff is a_{ji} and the bimatrix $(\mathbf{A}, \mathbf{A}')$ specifies the game. But it is redundant to write out the transpose $\mathbf{A}'$, so a single matrix $\mathbf{A}$ will specify a game when there is only one population.

The current state $\mathbf{s} = (p, 1-p)$ specifies the fraction p of players in the population currently choosing action 1 and the fraction $1-p$ choosing action 2. The state space is the one dimensional line segment (or simplex) $S = \{(p, 1-p) \in \mathbf{R}^2 : 0 \leqslant p \leqslant 1\}$. The expected payoff to a player choosing mixed strategy $\mathbf{r} = (x, 1-x)$ when matched with a random opponent given state $\mathbf{s}$ is $\mathbf{rAs}'$.

A central idea in evolutionary game theory is that higher payoff strategies become more prevalent over time.[4] That is, the direction of change in $\mathbf{s} = (p, 1-p)$ is governed by the payoff difference $d(\mathbf{s}) = (1,0)\mathbf{As}' - (0,1)\mathbf{As}'$ between the first action $\mathbf{r} = (1,0)$ and the second $\mathbf{r} = (0,1)$. If $d(\mathbf{s}) > 0$ then p increases and $\mathbf{s}$ moves towards the first pure strategy $(1,0)$, while if

[3] Some of the material in Subsections I.A and I.B below is adapted from the latter reference. For a good sample of recent theoretical research and some introductory material, see the special issues of *Games and Economic Behavior* on evolutionary games (3:1, 1991) and on adjustment dynamics (5:3–4, 1993). The reader should be warned that terminology is not yet standardised; for example, evolutionary equilibrium can mean different things to different authors.

[4] This 'survival of the fittest' principle is straightforward when there are only two alternative strategies, as in most of the games examined here. With three or more alternatives the principle can be interpreted in several different ways, e.g. that rates of change or growth rates of strategy prevalence have the same ordering as strategy payoffs, or perhaps only that they are positively correlated. See the general references for extended discussions.

$d(\mathbf{s}) > 0$ then p decreases and $\mathbf{s}$ moves towards the second pure strategy (0, 1), i.e. the dynamic is assumed to be sign preserving. Write the payoff differential as

$$D(p) = d[s(p)] = (1, -1)\,\mathbf{A}(p, 1-p)'$$
$$= (1-p)(a_{12}-a_{22})-p(a_{21}-a_{11}) = (1-p)\,a-pb,$$

where the reduced parameters are $a = a_{12}-a_{22}$ and $b = a_{21}-a_{11}$. Then the graph of $D(p)$ is a straight line with intercept a at $p = 0$ and value $-b$ at $p = 1$. The result is that (apart from the degenerate case $a = b = 0$ in which a player is always indifferent between her two actions) each payoff matrix falls into one of three qualitatively different types as shown in Fig. 1.

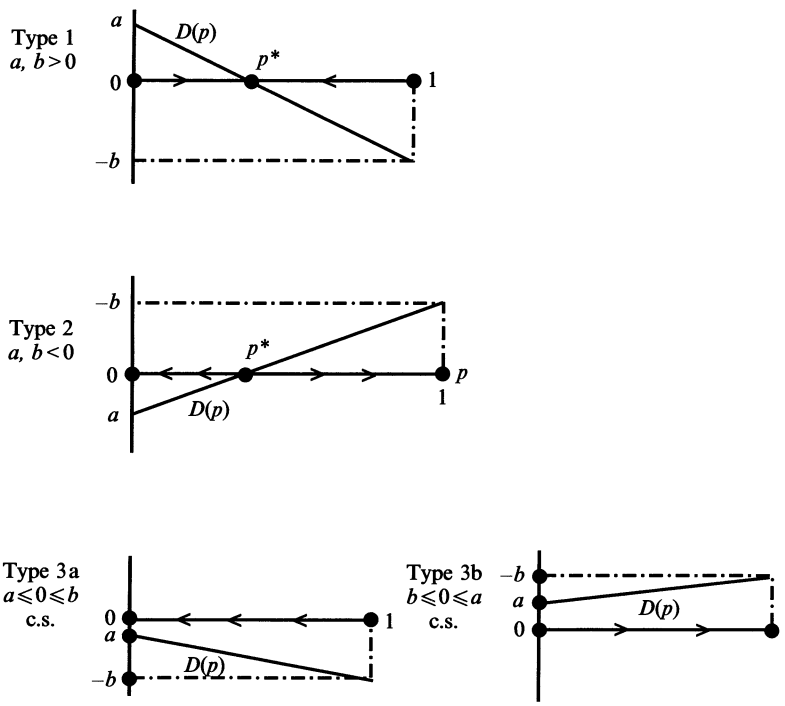

Fig. 1. Linear One-dimensional Evolutionary Games. *Notes:* For payoff matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

define $a = a_{12}-a_{22}$ and $b = a_{21}-a_{11}$. The point $s(p) = (p, 1-p)$ represents the current evolutionary state for $0 \leqslant p \leqslant 1$. The current payoff difference between the two pure strategies is $D(p) = (1-p)\,a-pb$.

Type 1: If $a, b > 0$ then the unique root of $D(p) = 0$ is $p^* = a/(a+b)$. It is immediate from the definitions that p^* is a symmetric mixed strategy Nash equilibrium (NE) of the 2-player bimatrix game $(\mathbf{A}; \mathbf{A}')$, and it is the only NE. More importantly for present purposes, D is downward sloping so p increases (decreases) whenever it is below (above) p^*. Hence p^* is the unique

evolutionary equilibrium. That is, for any sign-preserving continuous-time dynamic $\dot{\mathbf{s}} = F(\mathbf{s})$ we have $\mathbf{s}_t \to \mathbf{s}^* = (p^*, 1-p^*)$ as $t \to \infty$ from any initial state $\mathbf{s}_0$. The same conclusion also holds for discrete time ($t = 0, 1, 2, \ldots$) dynamics $\Delta\mathbf{s}_t = \alpha F(\mathbf{s}_t)$ on S if we add the proviso that the adjustment rate parameter $\alpha > 0$ is not too large. (We can get unstable oscillations if p_{t+1} jumps too far over p^*.) Familiar type 1 games include versions of Matching Pennies

(e.g., $\mathbf{A} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ so $a = b = 2$) and Hawk-Dove

(e.g. $\mathbf{A} = \begin{pmatrix} -1 & 2 \\ 0 & 1 \end{pmatrix}$ so $a = b = 1$).

Type 2: For $a, b < 0$, the root $p^* = a/(a+b)$ of $D(\mathbf{p}) = 0$ is still a NE of the associated bimatrix game, but now both pure strategies $p = 0$ and $p = 1$ are also NE. As Fig. 1 makes clear, $D(p)$ slopes upward and is negative (positive) for $p < p^*$ ($p > p^*$), so p^* is an unstable 'source' separating the basins of attraction of the two evolutionary equilibria $p = 0$ and $p = 1$. An economic interpretation is that each pure strategy has increasing returns in type 2 games and decreasing returns in type 1 games. Type 2 games are often called symmetric coordination games.

Type 3: If $D(p)$ lies above (below) the p-axis for all $p \in (0, 1)$, then the second pure strategy $p = 1$ (the first pure strategy $p = 0$) is dominant. Of course, the dominant strategy is the unique NE of the bimatrix game and the unique evolutionary equilibrium for any sign-preserving dynamic F. This type of game is characterised by $ab \leqslant 0$ (and $|a| + |b| > 0$). The most interesting example is Prisoner's Dilemma, in which payoffs decrease as the dominant strategy becomes more prevalent, e.g.

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix} \text{ so } a = -b = -1.$$

I.B. *More Complex Evolutionary Games*

There are two different ways to get a two dimensional state space. If each player in a single population of strategically identical players has three alternative actions, then the payoff matrix $\mathbf{A}$ is 3×3 and the current state is a point in two dimensional simplex

$$S = \{(p, q, 1-p-q) \in \mathbf{R}^3 : p, q \geqslant 0, p+q \leqslant 1\}.$$

Fig. 2 illustrates a version of the 'Hawk–Dove–Bourgeois' game, which has a corner NE at $(p, q) = (0, 0)$ and an edge NE at $(\frac{2}{3}, \frac{1}{3})$. Under standard dynamics (e.g. replicator dynamics; see Weibull for a simple exposition), the corner NE is an evolutionary equilibrium and the edge NE is a saddle point.

Much of the work reported here uses a second way to get a two dimensional state space. Suppose there are two strategically distinct populations, each with two alternative actions. Then the state for each population can be represented by a point in the unit interval [0, 1] so the overall state for the two populations

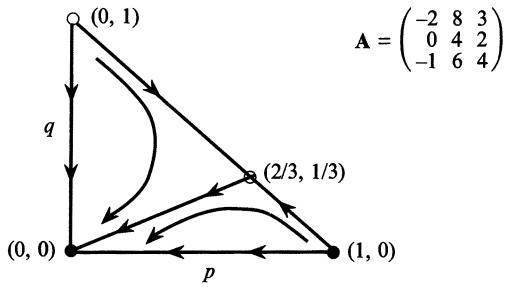

Fig. 2. The Hawk–Dove–Bourgeois game. *Notes:* Unstable (source), saddle and stable (evolutionary equilibria) are indicated respectively by open (○), crossed (⊗) and solid (●) dots. The equations $\dot{\mathbf{p}} = (1,0,0)\mathbf{As} - (\frac{1}{3},\frac{1}{3},\frac{1}{3})'\mathbf{As}$ and $\dot{\mathbf{q}} = (0,1,0)'\mathbf{As} - (\frac{1}{3},\frac{1}{3},\frac{1}{3})'\mathbf{As}$ characterise the dynamics in the interior of the simplex.

can be represented by a point (p, q) in the unit square $[0, 1] \times [0, 1]$. If the expected payoff to a player in the first (respectively second) population depends only on the distribution of actions $\mathbf{s}_2 = (q, 1-q)$ in the other population (respectively $\mathbf{s}_1 = (p, 1-p)$), then the payoffs have a bimatrix representation $\mathbf{rAs}_2'$ for player 1 (respectively $\mathbf{rBs}_1'$ for player 2), where $\mathbf{r}$ is the player's own mixed strategy and $\mathbf{A}$ and $\mathbf{B}$ are given 2×2 matrices.

For example, consider an asymmetric version of the Battle of the Sexes game with

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}.$$

The arrows in Fig. 3*a* are vectors $(D(q), D(p))$ for this payoff bimatrix. Battle of the Sexes is a two-dimensional analogue of a coordination (Type 2) game in that it has two pure-strategy NE (at $p = q = 0$ and at $p = q = 1$) both of which are also evolutionary equilibria for any sign-preserving dynamic, and one mixed NE whose stable saddle path separates the two basins of attraction.

Fig. 3*b* illustrates a Buyer-Seller game (Friedman, 1991, p. 641) with bimatrix

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 3 & -1 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix}.$$

This two dimensional game is analogous to Hawk–Dove (Type 1) because it has a unique, interior NE. It can be shown that this NE is an EE under some reasonable dynamics (e.g. fictitious play) but not under others (e.g. Cournot).

Fig. 3*c* illustrates the bimatrix game

$$\mathbf{A} = \begin{pmatrix} 1 & 4 \\ 2 & -1 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 3 & 1 \\ 2 & 0 \end{pmatrix}.$$

The first action (top row) is dominant for population 2 players, and the second action (bottom) is the best reply by population 1 players. Hence the corner $(p, q) = (0, 1)$ is the unique NE, analogous to type 3 games. This NE is automatically an EE because it is a solution by iterated elimination of dominated strategies (e.g. Weibull 1995).

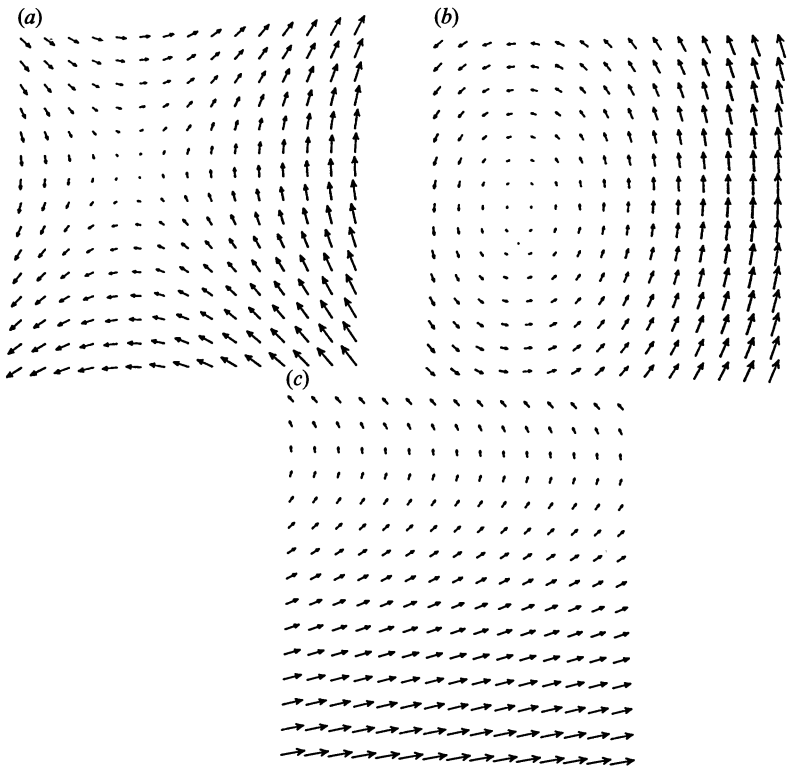

Fig. 3. Some two dimensional evolutionary games. (*a*) Battle of the sexes. (*b*) Buyer–seller. (*c*) Iterated dominated strategies. *Note:* The horizontal axis is p, the fraction of population 1 players choosing their first action, and the vertical axis is q, the fraction of population 2 players choosing their first action. The arrows are vectors with components $(D(q), D(p))$, where $D(q) = (1, -1)$ $\mathbf{A}(q, 1-q)'$ and $D(p) = (1, -1)\mathbf{B}(p, 1-p)'$ are the payoff differences for population 1 and population 2 respectively. The matrices **A** and **B** are listed in Table 1.

Table 1 summarises these linear 1 and 2 dimensional examples, which are the basis for the experiments reported below. Some confusion may be averted by noting that the two population game $(\mathbf{A}; \mathbf{A}')$ need not be the same type as the one dimensional game **A**. For example, the HD matrix in the first line of Table 1 defines a 1-dimensional game with a unique NE (and EE) at

$$p^* = a/(a+b) = [0-(-2)]/\{[0-(-2)]+(8-4)\} = 2/3.$$

Hence HD is of type 1. However, the 2-dimensional game HD2, defined in line 6 of the table, is not a type 1 analogue; it has an interior NE at $(p, q) = (2/3, 2/3)$ but also has two corner NE at $(p, q) = (1, 0)$ and at $(0, 1)$. Straightforward analysis discloses that for any sign-preserving dynamic on the square, the corner NE are both EE but the interior NE is not an EE. Hence HD2 turns out to be a Type 2 analogue. The intuition is that a stable mixture of hawks and doves will evolve in a single population, but with two interacting populations, one will become all hawks and the other all doves.

Many more general sorts of evolutionary games may be of interest in some applications, but will not be analysed here. For example, when there are own-

Table 1

Some Payoff Matrices

Name	Matrix	Type	NE	EE
1. Hawk–Dove (HD)	$\begin{matrix} -2 & 8 \\ 0 & 4 \end{matrix}$	1	$p = 2/3$	$p = 2/3$
2. Coordination (Co)	$\begin{matrix} 5 & -1 \\ 4 & 1 \end{matrix}$	2	$p = 2/3, 0, 1$	$p = 0, 1$
3. Weak Prisoner's Dilemma (WPD)	$\begin{matrix} 4 & 0 \\ 5 & 1 \end{matrix}$	3	$p = 0$	$p = 0$
4. Buyer–Seller (B–S)	**A** $\begin{matrix} 2 & 0 \\ 3 & -1 \end{matrix}$ **B** $\begin{matrix} 2 & 3 \\ -1 & 4 \end{matrix}$	1*a*	$(p, q) = 1/4, 1/2)$	$(p, q) = 1/4, 1/2)$?
5. Battle of the Sexes (BoS)	**A** $\begin{matrix} 1 & -1 \end{matrix}$ **B** $\begin{matrix} 3 & -1 \end{matrix}$	2*a*	$(p, q) = (1/3, 3/5)$, $(1, 0)$ $(0, 1)$	$(p, q) = (1, 0), (0, 1)$
6. Two-population HD (HD_2)	**A** $\begin{matrix} -2 & 8 \\ 0 & 4 \end{matrix}$ **B** $\begin{matrix} -2 & 8 \\ 0 & 4 \end{matrix}$	2*a*	$(p, q) = (2/3, 2/3)$, $(1, 0), (0, 1)$	$(p, q) = (1, 0), (0, 1)$
7. Iterated dominated strats (IDS)	**A** $\begin{matrix} 1 & 4 \\ 2 & -1 \end{matrix}$ **B** $\begin{matrix} 3 & 1 \\ 2 & 0 \end{matrix}$	3*a*	$(p, q) = (0, 1)$	$(p, q) = (0, 1)$
8. Hawk–Dove–Bourgeoise (HDB)	$\begin{matrix} -2 & 8 & 3 \\ 0 & 4 & 2 \\ -1 & 6 & 4 \end{matrix}$		$s = (2/3, 1/3, 0)$, $(0, 0, 1)$	$s = (0, 0, 1)$

Notes: The matrices appear here in the same format as on random pairwise (RP) screens: the player chooses the row and her opponent chooses the column. The usual convention in game theory literature shows the bimatrix as (**A**; **B**′). Matrix types are defined in Fig. 1 and in the text. Some variants of Co called Co1 and Co2 and variants of WPD called PD are also discussed in the text. Matrix 8 (HDB) has no one-dimensional analog. The NE and EE columns respectively list all Nash equilibria and evolutionary equilibria for the matrix. The '?' after the Matrix 4 indicates that this state is an EE for some but not all adjustment dynamics.

population effects (e.g. the payoff to players in population 1 depends on the distribution of actions in population 1 as well as in population 2) then the state space is still 2 dimensional but larger bimatrices are required to specify a linear game. Nonlinear games and higher dimensional games may also be of interest in applications; the interested reader should consult papers cited in Weibull (1995) and Friedman (1992).

I.C. *Convergence Issues*

The theory reviewed briefly in the previous two subsections uses a simple sign-preserving assumption to predict the asymptotic stability or instability of various NE. Several issues arise in applying the theory to laboratory experiments. First, only modest numbers of players and amounts of time are feasible in the laboratory, not the large populations and infinite time horizon typically assumed in the theory. Friedman (1992) argues that the main substantive reason for assuming a large population is to ensure that each player perceives that his current action has a negligible impact on other players' future actions, so in effect she is playing a 'game against Nature'. To have a 'large'

population in this sense does not necessarily require very many actual players. For example, 3 to 4 buyers and 3 to 4 sellers are large numbers in double auction market games (Smith, 1982), and 3 sellers appears to be a large number in the oligopolistic industries studied by Bresnahan and Reiss (1991).

How can we detect violations of the 'large numbers' assumption that players respond only to their own current payoff differential, e.g. to $D(p)$? An alternative assumption is that some players believe that if they try to increase the average payoff then others will follow suit. Specifically, given a 2×2 matrix $\mathbf{A}$ defending a 1 dimensional linear game, call a player Kantian (after Immanuel Kant's famous 'categorical imperative' to act as you wish everyone else to act) if she maximises the mean payoff $M(p) = (p, 1-p)\,\mathbf{A}(p, 1-p)'$, i.e. chooses the first (second) action when $M(p)$ is increasing (decreasing) at the current state $s(p)$. We can look for changes in the prevalence of Kantian play and other changes in behaviour as we vary the number of players in a population and as we vary the payoff matrix so as to alter the individual and group incentives, $D(p)$ and the slope of $M(p)$.

A second issue concerns convergence to pure strategy NE. Experimentalists going back at least to Siegel (1961) have noted that subjects resist excessive repetition. Recently McCabe, Michelitsch and Smith have begun to study performance on a task analogous to playing dominant strategies, and they find 5–13% deviant responses asymptotically, depending on the environment and rewards (Michelitsch, 1992). At mixed NE, deviations can have either sign and hence may largely cancel in the aggregate. At pure-strategy NE, by contrast, deviations are all of the same sign, so we might expect that here convergence will remain incomplete.

A final issue concerns convergence to mixed strategy NE. Since the 1950s experimentalists have reported difficulty in obtaining convergence to a mixed NE; see the Rappoport and Orwant (1962) survey, for example. More recently, J. Friedman and R. Rosenthal (1986) report behavioural steady states displaced from mixed strategy NE. Harsanyi (1973), Selten (1988) and Jordan (1991), among others, present theoretical arguments for the stability of mixed NE. Crawford (1985) and Jordan (1993), among others, present arguments against stability that appear to apply at least to mixed NE for two population games like Buyer–Seller. Harsanyi's argument is called 'purification' and, roughly speaking, says that players each choose pure strategy best responses after privately observing independent tiny random perturbations of their payoffs; typically an outside observer of one-shot games cannot distinguish the result of the perturbed games from the mixed strategy NE of the original game. Fudenberg and Kreps (1993) extend this idea to prove that a learning process can converge to mixed NE. The other authors propose vary different approaches that I will not attempt to explain here.

To summarise, there is a body of theory which identifies the behaviour one should eventually observe in simple strategic interactions among players belonging to one or two populations. Assuming the dynamic adjustment process is sign-preserving and convergent, the eventual behaviour is characterised by the evolutionary equilibria (EE), a subset of the Nash equilibria

(NE) associated with the payoff functions. The same set of ideas leads to the classification of bimatrix games into a few basic types. The empirical relevance of this evolutionary theory can be questioned on several grounds: time and populations are finite and discrete, not infinite and continuous; corner EE (i.e. pure strategy NE) may not quite be reachable; and interior EE (i.e. mixed strategy NE) may be unstable, especially in two population games. Empirical work clearly is in order.

II. LABORATORY PROCEDURES

II.A. *Basic Procedures*

The experiments consist of 60–120 minute laboratory sessions using profit-motivated subjects. Payoffs are calibrated to produce average earnings of about US$10 per hour per subject. Realised earnings depend sensitively on chosen actions and typically vary from $8 to $32 per subject in a two hour session. All subjects receive written instructions (available from the author on request) and about $\frac{1}{2}$ hour training on the computer prior to participation in a session. Each session consists of 60–200 periods of strategic interaction among 6–24 undergraduate subjects, the players. In each period the players, seated at visually isolated terminals, review historical data and the payoff function, and choose an action from a menu of two or three possible actions. The choices of all players are sent to a central processor (a Sun workstation) that computes the outcomes and notifies all players. Then players receive updated histories in preparation for the next period. All these features are publicly announced.

Fig. 4 illustrates the players' screen displays under the alternative matching protocols, RP in Panel A and MM in panel B, explained below. In either case, the player enters and confirms her current action (a or b) at the keyboard. The action is displayed on the screen in the lower left box, and the possible outcomes are highlighted in the payoff box on the right. When all players have chosen and confirmed their actions, the realised payoff appears at the intersection of the highlights in the payoff box and then is displayed in the upper right box along with other historical data.

II.B. *Treatments*

The experiments seek to identify conditions under which evolutionary game theory adequately characterises actual play in a diverse set of simple games. Diversity is achieved by varying the treatments, or environmental conditions, across and within experimental sessions. The main treatments are the payoff function, the matching rules, and the information conditions.

Table 1 lists the eight basic payoff functions used in the experiments. Note that all three types of linear 1-dimensional payoff functions are represented, as well as their 2-dimensional analogues. The Coordination matrix is a bit special in that the two pure strategy NE satisfy conflicting selection criteria (Harsanyi and Selten, 1989): $p = 1$ is payoff dominant (all players get 5 per period versus 1 per period in the other pure NE), while $p = 0$ is risk-dominant (an opponent's deviation actually increases a player's payoff by 3 versus a decrease by 6 in the first NE). The third matrix in the table, Weak Prisoner's dilemma (WPD), is

(a)

Average no. of (A, B, C) : (7·2, 4·8, 0)

Round	Total (A, B, C)	Your choice	Payoff
1	(9, 3, 0)	A	−2·0
2	(8, 4, 0)	A	−2·0
3	(8, 4, 0)	A	−2·0
4	(4, 8, 0)	B	4·0
5	(7, 5, 0)	B	0·0

Your total earnings: −2·00

Please choose *'a'* or *'b'*:

		Other player A	Other player B
You	A	−2·0	8·0
	B	0·0	4·0

Each period the other player you face is randomly chosen from the other 11 players in your group.

(b)

Average no. of (A, B, C) : (5·6, 6·4, 0)

Round	Total (A, B, C)	Your choice	Payoff
1	(5, 7, 0)	A	3·8
2	(4, 8, 0)	B	2·7
3	(8, 4, 0)	A	1·3
4	(6, 6, 0)	A	3·0
5	(5, 7, 0)	B	2·3

Your total earnings: 13·1

Please choose *'a'* or *'b'*:

No. of A's	A payoff	B payoff
0	8·0	4·0
1	7·2	3·7
2	6·3	3·3
3	5·5	3·0
4	4·7	2·7
5	3·8	2·3
6	3·0	2·0
7	2·2	1·7
8	1·3	1·3
9	0·5	1·0
10	−0·3	0·7
11	−1·2	0·3
12	−2·0	0·0

Fig. 4. Players' screens. (*a*) random pairwise matching. (*b*) Mean matching.

one of the more interesting Type 3 games in that the dominant strategy is just barely so (payoffs of 5 or 1 versus 4 or 0 for the alternative action) and the NE payoff of 1 is far less than the 'cooperative' payoff of 4. The two dimensional analogue IDS is also chosen to challenge the theory: the dominant strategy ($q = 1$) for the second population is just barely so, and $p = 0$ becomes a best response for the first population only when $q \geqslant 5/6$.

The experiments examine two alternative procedures for matching players. Under the random pairwise (RP) procedure, the computer randomly picks a matching scheme independently in each period, each admissible scheme being equally likely. For example, in a 1×12 Hawke–Dove game, the single population of players might be paired $\{0, 2\}$, $\{1, 9\}$, $\{3, 11\}$, $\{4, 5\}$, $\{6, 10\}$ and $\{7, 8\}$ in the third period. The payoff matrix appears on the right side of the player's screen in this treatment, as shown in Fig. 4*a*. The convention is that every player sees herself as choosing the row and her opponent as choosing the column.

Under random pairwise matching for payoff matrix **A**, a player's expected payoff is **rAs**′ if he chooses strategy **r** and the distribution of actions by potential opponents is **s**. However, his actual payoff depends on the action taken by his actual opponent, and so has some variance around its expectation. The variance is eliminated in the alternative matching procedure, called mean matching (MM). Here each player is matched once against each possible opponent in each round and receives the average (mean) payoff. Fig. 4*b* illustrates mean matching with 12 players and the Hawk–Dove payoff matrix

$$\mathbf{A} = \begin{pmatrix} -2 & 8 \\ 0 & 4 \end{pmatrix}.$$

For example, if 6 of 12 players choose the first action then the state is $(p, 1-p) = (0{\cdot}5, 0{\cdot}5)$ and the payoff is $(1, 0)\,\mathbf{A}(0{\cdot}5, 0{\cdot}5)' = 3{\cdot}0$ for the first action and is $(0, 1)\,\mathbf{A}(0{\cdot}5, 0{\cdot}5)' = 2{\cdot}0$ for the second action.

From the viewpoint of standard game theory, the matching protocols define quite different games. RP approximates a series of 2-player non-repeated games, while MM defines a single repeated n-player game, where n is the sum of the population sizes. From the viewpoint of evolutionary game theory, however, the protocols define games which are equivalent except that RP adds sampling error to the payoff function **rAs**′.

The third major treatment variable is the amount of historical information to appear in the upper left box on each player's screen. In the minimum level *No Hist* the player receives no historical information other than what she could tabulate herself: her own action and actual payoff in previous periods. In the usual level, *Hist*, the box contains a list of the actual state of the relevant population in previous periods. In Fig. 4*a* for example, the player can see that 9 of 12 players took action A in period 1, then 8 in period 2, and so forth, for an average of 7·2 in periods 1–5. The evolutionary game literature contains many papers that implicitly or explicitly assume information conditions corresponding to *Hist* and many others corresponding to *No Hist*, so both treatments are worth a look and the comparison is interesting. Of course, in every treatment the actions and identities of individual opponents remains confidential.

Two other treatment variables deserve brief mention. Population sizes are varied to test for the presence of small numbers effects such as Kantian behaviour in Prisoner's Dilemma or Coordination experiments. The number of players varies across sessions – e.g. perhaps 12 players in one session and 16 in another. Some sessions employ split groups in some periods – e.g. all 16 players belong to a single population in the first 40 periods, then are divided into two separate 8-player groups (no pairing or mean-matching across the two groups) for the next 80 periods, and reunited into a single group for the last 40 periods of a 160 period session.

The final treatment variable is *run length*. A run of several periods (all treatments held constant) is required to test for convergence for a given payoff matrix and player population. Runs are separated by obvious changes in the player population and/or the payoff matrix, the least significant being an

Table 2

Experimental Design Summary

Session	Payoff Matrix Name (type)	No. of populations × population size	Runs × run length	Other treatments
exp 1*	WPD	1 × 12	6 × 10	MM/RP
exp 2*	WPD	1 × 12	8 × 10	MM/RP
exp 3	HD, Co	1 × 12	12 × 10	MM/RP
exp 4	WPD	1 × 16/2	12 × 10	MM/RP
exp 5	B–S	2 × 6	12 × 10	MM/RP
exp 6	HD, Co	1 × 12	10 × 16	MM/RP
exp 7	BoS	2 × 5	12 × 10	MM/RP
exp 8*	Co, HD	1 × 12	12 × 14	MM/RP; Hist/No
exp 9	HD, Co	1 × 10	14 × 16	Hist/No
exp 10*	HD, $(HD)^2$	1 × 16/2, 2 × 8	12 × 16	Hist/No
exp 11	HDB	1 × 12/2	11 × 15	MM/RP, Hist/No
exp 12	WPD	1 × 12/2, 4, 6	24 × 10	Hist/No
exp 13	BoS, B–S	2 × 8	16 × 12	Hist/No,
exp 14	HD, $(HD)^2$	1 × 16/2, 2 × 8	6 × 16	MM/RP; Hist/No
exp 15	HD, Co	1 × 12	18 × 10	MM/RP
exp 16*	Co	1 × 12/2, 3, 4, 6	12 × 10	MM/RP
exp 17	HD, BoS	1 × 16/2, 2 × 8	15 × 10	none
exp 18	Co	1 × 16/2, 4, 8	16 × 10	MM/RP
exp 19	BoS, B–S, $(HD)^2$	2 × 6	18 × 10	MM/RP
exp 20	Co	1 × 12	12 × 10	MM/RP
exp 21	HD, Co	1 × 12/2	18 × 10	MM/RP; Hist/No
exp 22	HD, HD_2	1 × 12, 2 × 6	18 × 10	MM/RP, Hist/No
exp 23	HD, HD_2, B–S, BoS	1 × 12, 2 × 6	16 × 16	MM/RP; Hist/No
exp 24	HD, Co, PD	1 × 16/2	20 × 10	MM/RP; Hist/No
exp 25	HD, Co, PD	1 × 12	20 × 10	none
exp 26	HD, Co, PD, HD_2	1 × 12/2, 2 × 6	21 × 10	MM/RP; Hist/No
exp 27	B–S, BoS, IDS	2 × 8	19 × 10	MM/RP; Hist/No
exp 28	B–S, BoS, IDS	2 × 8	20 × 10	MM/RP; Hist/No
exp 29	B–S, BoS, IDS, HD_2	2 × 8	18 × 10	MM/RP; Hist/No
exp 30	B–S, BoS, IDS, HD_2	2 × 8	20 × 10	MM/RP; Hist/No

Notes: An asterisk after the session number indicates that some periods of the session had erroneous displays and therefore the data are excluded from the main analysis. The full matrix names, e.g. Weak Prisoner's Dilemma for WPD, are given in Table 1. The notation 1 × 16/2 in the third column means that the 16 players in a single population are split into two non-interacting 8-player groups in some runs. A run is a set of consecutive periods with no changes in payoff matrix, group composition or other treatments. For example, the entry 6 × 10 in column 4 means that the 60 periods in that session constituted 6 runs each 10 periods long. The default values of the other treatments are the mean matching protocol (MM) and the provision of historical information on population distributions (Hist); the alternative values of random pairwise matching (RP) and no historical information (No) are mentioned in the last column when applicable.

interchange of payoff matrix rows and columns. The history box also is erased at the beginning of a new run. If runs are too short then convergence will never be clear, but if runs are too long then players may respond to boredom (or possibly to repeated matching) rather than to current payoffs. Typical run lengths are 10 or 16 periods, and behaviour is also compared across half-runs of 5 or 8 periods.

Table 2 sketches the experimental design by session. For example, in the third session, exp 3, a single population of 12 individuals played 12 ten-period runs, some runs using Hawk–Dove payoffs and some using Coordination

payoffs. Some runs used the MM matching protocol and some used RP. Other sessions differed in various ways. Exp 4, for example, featured a single 16-player population in half the runs and two non-interacting 8-player populations ('split groups') in the other runs. In general, treatments were varied in each session in a balanced fashion to avoid confounding the variables.

II.C. *Testable Hypotheses*

The data analysis emphasises convergence behaviour. The state might converge to a population distribution $\hat{\mathbf{s}}$ – call this a behavioural equilibrium (BE) – that may or may not coincide with a theoretical equilibrium, Nash (NE) or evolutionary (EE). The formal data analysis begins by proposing empirical criteria for convergence and coincidence, and then uses the criteria to test the following hypotheses:

(1) Some BE is typically achieved by the second half of a (10–16 period) run.

(2) BE typically coincides with NE.

If a payoff matrix admits several NE, some of which are EE and some of which are not, then we have the more refined hypothesis:

(3) BE typically coincides with EE, especially those with larger basins of attraction and/or those whose basins of attraction contain the initial state of the run.

A second goal of the data analysis is to compare convergence behaviour across treatments. The relevant hypotheses include:

(4) No 'small-group' effects appear for population sizes six or larger.

(5) Convergence to BE is faster in the mean-matching (MM) than in the random-pairwise (RP) treatment.

(6) Convergence to BE is faster in the usual *Hist* treatment than in the alternative *No Hist* treatment.

In addition I test the convergence-related hypothesis:

(7) Individual behaviour at a mixed strategy BE is better explained by idiosyncratic 'purification' strategies than by identical mixed strategies.

III. RESULTS

III.A. *Overview*

Fig. 5 charts the time path of the state s_t in the first four runs of exp 3, a successful 12 player session consisting of 12 runs each 10 periods long.[5] The first four runs use Hawk–Dove payoffs, with a unique mixed strategy Nash equilibrium at $s^* = 2/3 = 8/12$. That is, in NE 8 of 12 players choose the first strategy (or 4 of 12 when the matrix rows and columns are interchanged as in runs 2 and 3). The graphs show a tolerance of 1 player in the band around NE, so 7, 8 or 9 players choosing the first strategy (or 3, 4 or 5 players when the matrix is interchanged) in any given period counts as a 'hit' for the Hawk–Dove NE. The time paths in the first four runs suggest that the NE

[5] Sessions exp1 and exp2 were invalidated by computer program glitches that scrambled the payoff matrices, so the data are omitted from subsequent analysis. The intended matrices were type 3 (Weak Prisoners' Dilemma). Most valid runs saw convergence to the NE.

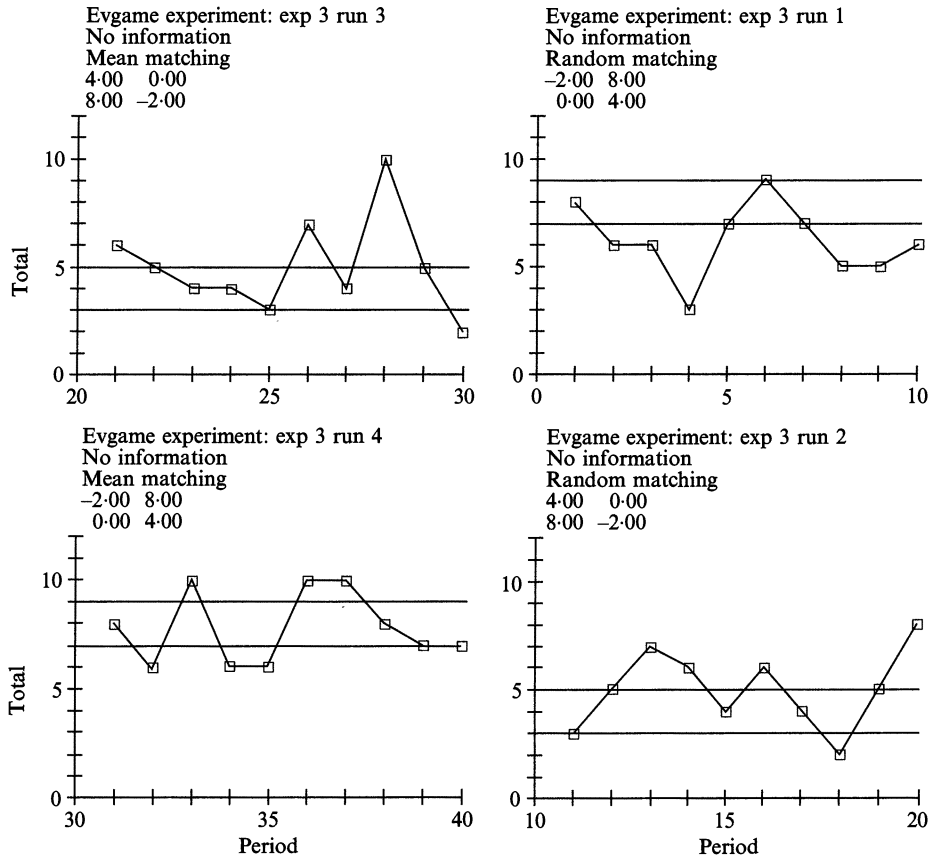

Fig. 5. Session exp 3: runs 1–4.

attracts states s_t outside the tolerance band $s^* \pm 1/12$, but there is considerable behavioural noise so hits occur in only about 50% of the periods.[6]

Fig. 6 graphs behaviour in the first four runs of exp 5, the first 2-population session. All periods use the Buyer–Seller matrix from Table 1 or its interchange, so the unique NE is at $(p, q) = (0{\cdot}25, 0{\cdot}50)$ or, for the interchanged version, at $(0{\cdot}75, 0{\cdot}50)$. The graphs show the time path of the state s_t in its space, the unit square. The time path is smoothed by a 2-period moving average, so the point graphed actually is $(s_{t-1} + s_t)/2$. The time path in first run looks like an unstable counterclockwise spiral diverging from the NE. The second run looks like a tidy counterclockwise double loop around the NE, neither converging nor diverging. The third run uses the RP matching protocol; at best there is a weak

[6] Graphs of the remaining runs, suppressed here to conserve space, can be summarised as follows. Run 5, the first of four Coordination runs, appears to represent slow, incomplete convergence from the theoretically unstable mixed strategy NE towards the risk-dominant (and Pareto inferior) NE $S^* = 0$. Run 6, quite surprisingly, appears to represent convergence to the theoretically unstable mixed strategy NE. Run 7 shows clear convergence to the risk-dominant NE $S^* = 12/12$ after the matrix interchange. Perhaps due to hysteresis, run 8 convergences quickly to $S^* = 12/12$, now the payoff-dominant NE because the matrix interchange was negated. The last 4 runs of the session again are Hawk–Dove. Now there seems to be less behavioural noise and most periods are hits. The session as a whole provides little evidence that the matching procedure (MM or RP) has any effect.

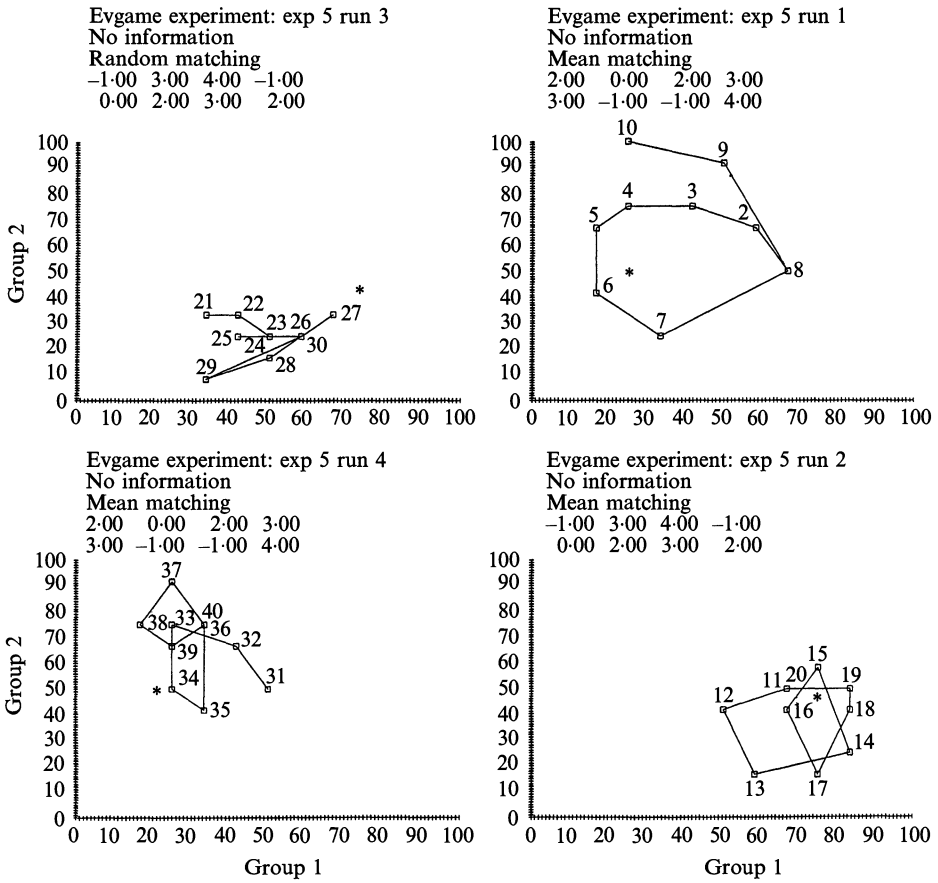

Fig. 6. Session exp 5: runs 1–4.

tendency to drift towards the NE. The fourth run reverts to MM and looks like a counterclockwise spiral converging to the NE.[7]

Looking at a large number of such graphs, one gets the general impression that behaviour tends to settle down to a BE under all treatments, more rapidly and closer to EE in some treatments than in others. None of the testable hypotheses seem grossly inconsistent with the graphs. The rest of this section examines the hypotheses more quantitatively.

III.B. *Convergence Criteria*

The general criterion for convergence is that deviations from a given steady state s^* are small. Specifically, for a pre-selected tolerance bound $b > 0$, say that the state converged to s^* in run (or half-run) r with L periods if

$$\frac{1}{L}\sum_{t\in r} |s_t - s^*| \leqslant b, \tag{1}$$

[7] The remaining 8 runs in the session seem to tell much the same story: slow approximate convergence to the NE via counterclockwise spirals, with perhaps larger and more biased deviations in the RP runs than in the MM runs. Complete graphs of all sessions are available on request at reproduction cost.

i.e. if the mean absolute deviation from s^* does not exceed the tolerance bound in a given run or half-run. In 2-population games the norm $|\cdot|$ in inequality (1) is interpreted as the sup norm – i.e. the larger absolute deviation of the two populations. Behavioural equilibrium (BE) now can be defined operationally for any run or half-run as convergence to the median state, i.e. inequality (1) holds for s^* = the 50th percentile of states realised in the given run or half run. The operational definition for convergence to NE or EE is simply that inequality (1) holds with s^* equal to a given NE or EE state.

The summary data discussed below use two tolerance bounds. In the 'tight criterion' $b = 1/N$, and in the 'loose criterion' $b = 2/N$, where N is the number of players in each population. For example, in a two population game with 8 players in each population, the tolerance bounds are 0·125 (tight) and 0·25 (loose).

Table 3 reports instances of convergence by half-run. For example, in a 10 period run the two half-runs are periods 1–5 and periods 6–10. In the 355 first half-runs, BE is achieved over 92% of the time by the loose criterion and over 46% by the tight criterion. In the 353 second half-runs (computer crashes wiped out two second half-runs), the convergence percentage rises to over 98% (loose) and 70% (tight). The majority of BE coincide with NE; in second half-runs, for example, about 79% (= 77·5/98·3) of the loose BE are loose NE. Only a very few NE are not EE, but this empirical result is unsurprising because for many of the payoff matrices investigated here all NE automatically are EE.

The statistics in Table 3 are intended mainly to detect regularities for further investigation. A more detailed look at convergence for each type of game is now in order.

III.C. *Behaviour in Type 1 Games*

Recall that HD and other Type 1 matrices have a unique NE. It lies in the interior of the state space [0, 1], and is an EE. Evolutionary game theory predicts convergence to this NE since it is an EE, but some game theorists predict nonconvergence because the NE is in mixed strategies. Row 3 of Table 3 at first seems to give support to both views because it reports a loose convergence rate of over 87% but a tight convergence rate of less than 33%. The underlying data show that tight convergence is at least 50% more frequent in second half runs than in first half runs, and under MM than under RP. Convergence is also more frequent under Hist than under No Hist. My conclusion is that the state indeed converges to the unique NE but that convergence can be slow, especially under the RP matching protocol and the No Hist information treatment.

Does the state converge to the NE because individual players adopt the mixed NE strategy? In that case we would expect to observe loose but not tight convergence, as in fact we usually do.[8] A closer look at the raw data, however,

[8] To simplify the calculation a bit, suppose the symmetric equilibrium strategy is to play each pure strategy with probability 0·5 and that the median state is $s^* = 0{\cdot}5$, and suppose there are 12 players. We need the probabilities that the mean absolute deviation (MAD) of the states S_t from the median s^* is less than $1/12 = 0{\cdot}083$ and $2/12 = 0{\cdot}167$, when each state is the average of 12 independent Bernoulli trials with $p = 0{\cdot}5$. Back-of-envelope calculations give approximate mean and variance for

Table 3

Convergence Percentage by Half Run

Runs	Nobs	Tight criterion (%)			Loose criterion (%)		
		BE	NE	EE	BE	NE	EE
1. First-half	355	46·2	18·9	15·6	92·1	64·5	46·5
2. Second-half	353	70·4	39·4	36·7	98·3	77·5	65·0
3. HD	156	55·1	32·7	32·7	96·2	87·2	87·2
4. Coordinate	116	79·7	40·5	25·9	98·3	69·4	41·8
5. WPD	24	95·8	64·6	64·6	95·8	91·7	91·7
6. HDB	24	50·0	33·3	33·3	79·2	37·5	37·5
7. IDS	60	35·0	10·0	10·0	91·7	30·0	30·0
8. BoS	128	61·7	40·6	39·1	93·0	75·8	56·3
9. HD^2	92	52·2	16·3	14·1	98·9	81·5	26·1
10. B–S	84	27·4	8·3	8·3	88·1	56·0	56·0
11. MM	174	60·9	42·0	40·2	93·7	77·0	70·4
12. RP	206	71·8	30·1	23·3	98·5	77·7	57·0
13. Hist	180	66·1	33·1	31·4	98·3	72·2	61·9
14. No Hist	128	65·2	21·9	18·8	97·7	80·1	57·8

Note: The table is based on all data from sessions 3–30 of Table 2 with the exception of sessions 8 and 16 (scrambled matrices), the MM/split group runs of session 17 (data not recorded properly), and sessions 12 and 18 (small groups used). Session 11 involves a qualitatively different (3 action) environment and so is excluded from the first two lines of the table. The last two pairs of lines are based on (approximately) balanced subsamples drawn entirely from sessions in which both of the alternative treatments were used. Specifically, lines 11–12 (MM *vs.* RP) are based on sessions 3, 4, 5, 7, 14, 15, 19–24 and 26, and lines 13–14 (Hist *vs.* No Hist) are based on sessions 9, 10, 13, 14, 19–24 and 26. When players are split into two non-interacting groups, each group's outcome has weight 0·5 so the number of observations (Nobs) is 1·0 for each half-run.

casts doubt on that view; it appears that some players usually play 'Hawk', some usually play 'Dove' and others switch back and forth. This is exactly the pattern that Harsanyi's purification approach would suggest if players draw the main component of their idiosyncratic perturbations once per run. The idea is that players may have slight homegrown preferences for 'Hawk' or 'Dove' and that Harsanyi's argument shows how this can lead to the mixed NE in the population.[9]

A formal test of the purification approach employs the null hypothesis that all players independently choose 'Hawk' with NE probability $p = 2/3$ in each period, and examines the one-sided alternative hypothesis that players change their action less frequently across periods. Each 10 period × 12 player run, for example, gives 120 actions and $9 \times 12 = 108$ opportunities for a player to change her action. We use the standard (but not eponymous) runs test (Conover, 1980, p. 349) under the conservative convention that counts a

$|s_t - s^*|$ of 0·113 and 0·0115 in each period. Let $\Phi(x)$ be the unit normal cumulative distribution function evaluated at x. Then the probabilities of tight and loose convergence in a 5 period half run are approximately $1{\cdot}0 - \Phi[\sqrt{5}(0{\cdot}113 - 0{\cdot}083)/\sqrt{0{\cdot}0115}] = 1{\cdot}0 - \Phi(0{\cdot}626) = 1{\cdot}0 - 0{\cdot}734 = 0{\cdot}26$, and $1{\cdot}0 - \Phi[\sqrt{5}(0{\cdot}113 - 0{\cdot}167)/\sqrt{0{\cdot}0115}] = 1{\cdot}0 = \Phi(-1{\cdot}126) = 0{\cdot}87$. That is, if players independently randomise then we would expect to see tight convergence in only about one quarter of the half-runs, but would see loose convergence in about 7/8 of them.

[9] Fudenberg and Kreps (1993) assume that perturbations are independent each period. Clearly more persistent perturbations would work in a population (as opposed to individualistic) interpretation of their model.

change of action when the last period action of one player differs from the first period action of the next player. In each of the 99 Hawk–Dove runs we observe fewer changes of action than predicted under the null hypothesis; typically changes are about half as frequent as predicted. The null hypothesis can be rejected in favour of the purification alternative at the $p = 0{\cdot}01$ level in 75% of the runs and usually at the $p = 0{\cdot}0001$ level. Exceptions seem more frequent in non-convergent runs.

Could this result be due merely to player inertia rather than to small differences in players' tendencies to play Hawk or Dove? Cochrane's Q statistic (Conover p. 196) tests whether players' dichotomous actions are homogeneous random processes. In 10 of 13 sessions with Hawk–Dove runs the null hypothesis of player homogeneity is rejected at the 0·0001 level in favour of the purification alternative that some players are more likely than others to play Hawk. Even in the three exceptional sessions the evidence on balance favours the alternative hypothesis. I conclude that the Harsanyi purification approach explains the data much better than the classical mixed strategy approach.

The Buyer–Seller (B–S) game is a 2-dimensional analogue of Hawk–Dove. Line 10 of Table 3 shows a healthy 88% rate of loose convergence to BE of which about $\frac{2}{3}$ are loose NE. Tight convergence is much less frequent; BE is achieved in only about 27% of the half runs and of these less than $\frac{1}{3}$ are NE. To interpret these numbers, note first that the loose target has area $(2b)^2/b^2 = 4$ times the area of the tight target, and that the loose NE frequency (47 of 84) is more than four times the tight convergence frequency (7 of 84). Tight convergence is a bit more frequent in second half runs than in first half, and is quite rare under either RP or No Hist. Visual inspection of the graphs shows that typically the state spirals in counterclockwise towards the NE but there is little tendency to complete convergence once the state gets within loose tolerance. I conclude that the NE is behaviourally stable in the weak sense that the state typically converges to a $(2/N)$-neighbourhood of the NE. Perhaps decision costs (or lack of payoff dominance) preclude tighter convergence.

III.D. *Behaviour in Type 2 Games*

Recall that type 2 games (such as the Coordination game in Table 1) have three NE, two of which are endpoint EE and the third of which is interior, separating the basins of the EE. Line 4 of Table 3 indicates that in the experiments with such games behaviour usually settles down: BE convergence percentages are almost 80% with the tight and over 98% with the loose criterion. Convergence to (any) NE is surprisingly infrequent given that the three NE targets have total width $4b$ (width b for each EE and $2b$ for the interior NE), so the targets cover $4(2/12) = 2/3$ of the state space [0, 1] under the loose criterion and $4(1/12) = 1/3$ under the tight. Thus the actual NE convergence rates of 69% and 40% are close to what one would expect if the asymptotic state were uniformly distributed, and the EE convergence rates of about 42% and 26% are only slightly better. What is going on?

The data underlying Table 3 reveal that tight convergence to an EE is about twice as frequent in second half-runs as in first half-runs and under Hist than

under No Hist, and somewhat more frequent under MM than under RP. Still, tight convergence to EE remains rather infrequent under any of these conditions. A possible clue is that, despite a smaller basin of attraction, the payoff-dominant EE (the PDNE) accounts for 18 instances of tight convergence and the risk-dominant NE (the RDNE) accounts for only two instances. The RDNE does have almost as many instance of loose-but-not-tight convergence (5) as the PDNE (6), but the net result remains anomalous: convergence (especially tight convergence) is rare for the EE with the larger basin of attraction.

A second anomaly is that the state sometimes converges to the interior NE, an event not predicted by evolutionary (or traditional) theory.[10] Overall, 24 (resp. 18) of the 94 full group half-runs converge loosely (resp. tightly) to the mixed strategy equilibrium (MNE). A closer look at the graphs of half-runs deemed loosely but not tightly convergent to MNE suggests that many of these actually represent slow divergence from MNE. Likewise, many of the half-runs deemed BE but not NE seem to represent slow or incomplete convergence to an EE, usually the RDNE.

The first step in following up on these observations is to see if there is any empirical difference between convergence to the interior NE of a Type 1 game (theoretically stable) and of a Type 2 game (theoretically unstable). The regression $p_t - p_{t-1} = \beta(p^{MNE} - p_{t-1}) + \epsilon_t$ yields the parameter estimate $\hat{\beta} = 0{\cdot}65 \pm 0{\cdot}10$ and $R^2 = 0{\cdot}31$ for data from Coordination runs that converged tightly to MNE (D.F. = 96), but yields $\hat{\beta} = 0{\cdot}96 \pm 0{\cdot}06$ and $R^2 = 0{\cdot}49$ for data from convergent Hawk–Dove runs (D.F. = 273). Thus, even selecting the most favourable runs, we find that convergence to the interior NE in Coordination games is significantly slower and less reliable than convergence to the interior EE in Hawk–Dove games.

The fact that there is even occasional convergence to the interior NE in Coordination games suggests that there may be forces at work beyond the payoff differential recognised by evolutionary game theory. Do some players use forward-looking or altruistic strategies? Some responses to exit questionnaires in Exp 24, the last Co session, suggest that they might: '[I] chose the result that would be most beneficial to everyone...', and, from another player, 'I made choices that would raise the total score of the group.' Perhaps some players are Kantian and choose actions so as to increase the mean payoff $M(p) = (p, 1-p)\,\mathbf{A}(p, 1-p)'$. The standard Type 2 matrix from Table 1 is

$$\mathbf{Co} = \begin{pmatrix} 5 & -1 \\ 4 & 1 \end{pmatrix}$$

for which $M(p) = 3p^2 + p + 1$ is increasing in p. Kantian players therefore would avoid the low payoff RDNE action ($p = 0$) in this game. Risk-averse

[10] Hysteresis potentially provides an evolutionary game theoretical explanation, because sometimes the run immediately follows a Hawk–Dove run in which the interior NE is an EE. Of the 18 Coordination half-runs immediately following HD runs, only three were tightly convergent to the interior NE. This is about the same proportion as for all Coordination half-runs. I conclude that hysteresis plays at best a minor part in explaining the anomaly.

players, on the other hand, might stick with the RDNE action until well inside the PDNE basin, hence producing a BE near the MNE. Hence both behavioural anomalies potentially arise from some Kantian behaviour.

To test this qualification to evolutionary theory, I used the modified Coordination matrices

$$\mathbf{Co1} = \begin{pmatrix} 5 & -1 \\ 3 & 3 \end{pmatrix} \quad \text{and} \quad \mathbf{Co2} = \begin{pmatrix} 5 & -2 \\ 0 & 4 \end{pmatrix}$$

in some runs of sessions 20, 21, 24 and 26. The basins of attraction for the EE are the same in Co1 as in Co, but $M(p)$ is decreasing in the first half $[0, 1/3)$ of the RDNE basin of attraction $[0, 2/3)$. The RDNE basin $[0, 6/11)$ for Co2 is smaller than for Co, but $M(p)$ decreases in $[0, 5/11)$, most of the basin. Therefore the presence of Kantian players would make convergence to the RDNE more likely with Co1 and Co2 than with the standard coordination payoff matrix Co. The data strongly confirm the prediction: we have four half-runs loosely convergent to the RDNE and 19 to the PDNE with Co versus 25 and 12 with the modified matrices. The associated chi-squared statistic of 14·30 is significant at least at the 0·001 level.

Recall that BoS and HD2 both are two-dimensional analogues of Type 2 games. Each has three NE, two EE at diagonally opposite corners of the square and an interior NE at the saddle-point of the separatrix between the two EE basins of attraction. Line 8 of Table 3 suggests that evolutionary theory accounts well for the BoS data. Despite the small area b^2 of each corner EE relative to the target area $4b^2$ of the interior (non-EE) NE, 50 of the 52 half-runs that converged tightly to some NE actually converged tightly to an EE. Line 9 of the table suggests that HD2 runs had considerably more noise but roughly similar behaviour.

III.E. *Behaviour in Type 3 Games*

In WPD and other Type 3 games the players have a dominant strategy, so there is a unique NE (and EE) at one endpoint of the state space [0, 1]. Table 3 shows that in sessions where there are always 6 or more players in a group, the state virtually always converges tightly to a BE and loosely to the NE. Even the tight NE convergence frequency is an impressive 64%. The underlying data show that, unlike Type 1 and 2 games, the tight NE convergence frequency in WPD is lower under MM than under RP.

Group size appears to have a significant effect. Some WPD and PD sessions involve 12 players that always remain in the same group and some involve 16 players sometimes split into two player 8 groups. In these sessions the mean deviations from NE consistently were small, e.g. 0·12 in the 8 player split groups. Deviations were much more variable and usually much larger in sessions involving runs with smaller groups of 2, 4 and 6 players. Line 1 *a–c* of Table 4 summarise the results. The mean deviation from NE (i.e., the fraction of players choosing the dominated 'cooperative' action) rises to 0·28 with two 6 player groups, to 0·29 with three 4 player groups, and to 0·39 with six 2 player

Table 4
Significance Tests

Pools (X *vs.* Y) [Nobs] sessions	Means (s.d.)	Deviations from NE		Hit Frequency	
		Wilcoxon	Pooled t	χ^2	Pooled t
1. Group Size (WPD and PD) exp 4, 12, 24, 26					
(*a*) 6 *vs.* 8 player groups [182, 340]	0·28, 0·12 (0·23, 0·15)	8·84**	10·01**	11·44**	3·41**
(*b*) 4 *vs.* 8 player groups [120, 340]	0·29, 0·12 (0·29, 0·15)	5·78**	839**	0·04	0·19
(*c*) 2 *vs.* 8 player groups [546, 340]	0·39, 0·12 (0·42, 0·15)	6·52**	11·38**	4·79	−2·19
2. RP *vs.* mean matching exp 3, 4, 14, 15, 21, 22, 23, 24, 26					
(*a*) HD runs [420, 272]	0·14, 0·10 (0·12, 0·10)	3·77**	3·94**	14·40**	3·82**
(*b*) WPD and PD runs [260, 210]	0·14, 0·14 (0·16, 0·15)	−0·95	−0·53	3·82	−1·96
3. No history *vs.* history exp 9, 10, 14, 21, 22, 23, 24, 26					
(*a*) HD runs [426, 446]	0·16, 0·13 (0·12, 0·11)	3·08**	3·48**	3·64	1·91
(*b*) WPD and PD runs [130, 160]	0·16, 0·20 (0·16, 0·15)	−3·12**	−2·43	18·49**	−4·43**
4. No hist/RP *vs.* hist/MM exp 27–30 (B–S and IDS runs) [231, 222]	0·37, 0·27 (0.21, 0.23)	5·55**	4·62**	38·20**	6·44**

Notes: The statistics compare performance in Pool X to performance in Pool Y. For example in Line 1 *b*, Poql Y = runs with 16 subjects split into two non-interacting groups, each with 8 subjects and Pool X = runs with 12 subjects separated into three non-interacting groups with four subjects each. The number of periods (Nobs) appears in brackets. The second column records the mean absolute deviations from NE for Pools X and Y (and the standard deviation of the deviations). The next two columns report the statistics for the standard Wilcoxon and t-tests for the null hypothesis that both pools have the same distribution for NE deviations. The last two columns report statistics for the standard χ and t statistics for the null hypothesis that both pools have the same hit frequency, where a period is counted as a hit if the deviation from NE does not exceed 1/group size. Negative statistics indicate smaller deviations or higher hit frequency in pool X. Two asterisks (**) indicate that the null hypothesis is rejected at the p = 0·01 level.

(repeated matched pairs) groups.[11] These deviations are significantly larger than those for the 8 player groups according to standard t and Wilcoxon tests, as indicated in the middle columns of the table. The last two columns of the table compare the hit frequencies, i.e. the fractions of periods in which the deviation is within tolerance, using the conservative convention that the tolerance bound is b = 1/group size. This convention makes it easier for smaller groups to record a 'hit'. Even so, line 1*a* shows that the fraction of 6 player groups hitting NE is significantly lower than the fraction of 8 player groups. The hit rates differ insignificantly for 4 and 8 player groups, and are significantly higher for the two player groups. I conclude that small group

[11] Perhaps it should be mentioned that deviations remain large in these sessions even when the players are regrouped into a single 12 player group; the mean deviation in such runs is 0·24. This is a vivid illustration of the general tendency for behaviour to be influenced by all treatments employed in within-groups sessions.

effects, here in the form of playing the dominated 'cooperative' strategy, are definitely present in the 2, 4 and 6 player groups.

Scrambled matrices provide unsought opportunities to investigate other Type 3 games. Experiment 8 uses a matrix that has a dominant strategy which also gives the highest mean payoff; players in this session chose the dominant/Kantian action a remarkable 99·6% of the time.

Recall that IDS is a 2-dimensional analogue of Type 3 games; it has a unique NE = EE at one corner of the square state space. The convergence rates reported in line 7 of Table 3, e.g. 30% loose and 10% tight EE convergence, at first might seem rather low. Recall, however, that the matrix entries for IDS were chosen to make convergence difficult and that the target area for a corner equilibrium is only b^2. Moreover, half the IDS runs used the RP/No Hist treatment for which convergence rates generally are low. Inspection of the time graphs under the more favourable treatment MM/Hist shows a consistent tendency for the state to converge towards the EE, interrupted by occasional loops back into the interior when a player in the second population chooses the dominated action. I conclude that the IDS data on closer examination are well explained by evolutionary game theory.

III.F. *Other Findings*

Only one session explored behaviour in HDB, a 1-population 3-action game with a triangular state space and with one corner NE (an EE with target area b^2) and one edge NE (not an EE but with target area $2b^2$). Row 6 of Table 3 indicates loose (tight) convergence to some BE in 19 (12) of 24 half-runs, tight convergence to the EE in 8 half-runs, and no loose or tight convergence to the edge NE despite its larger area. The data are sparse but consistent with evolutionary game theory.

The last four lines of Table 3 and the last three lines of Table 4 indicate the overall effects of the matching (MM or RP) and the feedback (Hist or No Hist) protocols. Table 3 indicates that tight NE and EE convergence is somewhat more frequent under MM and under Hist, and Table 4 confirms that hits are significantly more frequent and deviations from NE significantly smaller under MM. However, the size of the effects is not very impressive.

Do the evolutionary treatments MM and Hist together make much difference? Any tendency to speed convergence would be more noticeable in a two-population game, and should be confirmed in a balanced within-groups design. These considerations lead to the design of sessions 27–30. The last line of Table 4 shows that the IDS and B–S runs of these sessions had significantly smaller deviations from the unique NE and significantly higher hit rates under Hist/MM than under No Hist/RP. The BoS and HD2 runs are omitted from the table because their multiple NE make deviations more difficult to define clearly, but these data also appear to strongly support the same conclusion.

IV. DISCUSSION

Evolutionary game theory offers a simple classification of bimatrix games, suggests laboratory protocols, and suggests hypotheses regarding convergence

behaviour in bimatrix games. The evidence from a diverse set of laboratory games generally supports the seven hypotheses listed at the end of Section II. For all three types of one dimensional games and their two dimensional analogues, the states reliably achieve a loose behavioural equilibrium (BE) even within the first half-run of 5 periods. Most of the loose BE are also tight BE, the main exceptions occurring in two dimensional games with unique Nash equilibria (NE). Most BE coincide with NE, and most of the observed NE are indeed evolutionary equilibria (EE). In general,[12] the 'evolutionary' treatments of mean-matching (MM) and feedback (Hist) appear to improve convergence to EE. Thus the main tendencies of the convergence data are consistent with evolutionary game theory.

Two of the hypotheses deserve further discussion. The seventh hypothesis is concerned with the stability of mixed (or interior) NE. It states that such equilibria are achieved not by independent randomisations by each player, but rather by slight idiosyncratic preferences for pure strategies by individual players. The individual player data clearly favour this version of the 'purification' hypothesis. The group data also lend indirect support: as the hypothesis implies, we usually do see convergence to the interior NE = EE in one population games of Type 1, and less precise convergence in analogous two population games.

The fourth hypothesis is concerned with the range of applicability for evolutionary game theory. It states that players seldom will attempt to influence others' future behaviour ('small group effects') when there are at least 6 players in each group. The relevant data from Prisoner's dilemma experiments suggests that 6 is near the boundary. Cooperative ('Kantian') behaviour is considerably more prevalent in sessions which have runs splitting the players into groups of size 2 or 4, and it is especially prevalent in the runs with the smaller groups. Such behaviour is notably less frequent in sessions where the minimum group size remains above 6.

Perhaps the most surprising finding concerns another boundary for evolutionary game theory. Pilot experiments and other investigators had seemed to confirm the theoretical view that in simple coordination games with two pure strategy (corner) NE = EE and one interior NE, the 'risk-dominant' corner EE is most likely to be observed because it has the larger basin of attraction. (Indeed, Kandori *et al.* 1993, and Young, 1993 argue in influential theoretical papers that *only* the risk-dominant EE will be observed in the relevant limiting case.) My data strongly support the contrary theoretical view of Bergin and Lipman (1995) that one can bias convergence towards the other ('payoff-dominant') EE by increasing the potential gains to cooperation, even holding constant the basins of attraction for the two EE. The underlying behaviour can be regarded as Kantian. It remains to be seen whether other subject pools are as Kantian as mine, but it now appears that in some applications evolutionary game theory may have to be supplemented by a

[12] The main exception seems to be that when other conditions favour Kantian play (e.g. small numbers of WPD players), the MM treatment can further encourage this sort of deviation from NE.

theory of trembles (or 'mutations') that allows for forward-looking attempts to influence others' behaviour.

University of California, Santa Cruz

Date of receipt of final typescript: June 1995

References

Binmore, Kenneth (1987–8). 'Modelling rational players.' Parts I and II, *Economics and Philosophy* vol. 3 (2), 1987, pp. 179–214, and vol. 4 (1), 1988, pp. 9–55.

Bergin, J. and Lipman, B. L. (1995). 'Evolution with state-dependent mutations.' Draft manuscript. Queen's University Department of Economics.

Bresnahan, Timothy F. and Reiss, Peter C. (1991). 'Entry and competition in concentrated markets.' *Journal of Political Economy* vol. 99 (5), pp. 977–1009.

Cheung, Yin-Wong and Friedman, Daniel (1994). 'Learning in evolutionary games: some laboratory results.' UCSC Economics Department Working Paper no. 303.

Conover, W. J. (1980). *Practical Nonparametric Statistics* (Second Edition), NY: Wiley.

Crawford, Vincent P. (1985). 'Learning behaviour and mixed strategy Nash equilibria.' *Journal of Economic Behavior and Organisation* vol. 6, pp. 69–78.

Crawford, Vincent P. (1991). 'An "evolutionary" interpretation of Van Huyck, Battalio and Beil's experimental results on coordination.' *Games and Economic Behavior* vol. 3 (1), pp. 25–59.

Friedman, Daniel (1991). 'Evolutionary games in economics.' *Econometrica* vol. 59 (3), pp. 637–66.

Friedman, Daniel (1992). 'Economically applicable evolutionary games.' Center Discussion Paper no. 9226, Tilburg University, September.

Friedman, Daniel and Fung, K. C. (1996). 'International trade and the internal organization of firms: an evolutionary approach.' *Journal of International Economics* (forthcoming).

Friedman, James W. and Rosenthal, Robert W. (1986). 'A positive approach to non-cooperative games.' *Journal of Economic Behavior and Organization* vol. 7, pp. 235–51.

Fudenberg, Drew, and Kreps, David M. (1988). 'A theory of learning, experimentation and equilibrium in games.' Manuscript, July.

Fudenberg, D. and Kreps, D. (1993). 'Learning mixed equilibria.' *Games and Economic Behavior* vol. 5 (3), pp. 320–67, July.

Harsanyi, John C. (1973). 'Games with randomly disturbed payoffs: a new rationale for mixed strategy equilibrium points.' *International Journal of Game Theory* vol. 2, pp. 1–23.

Harsanyi, John C., and Selten, Reinhard (1988). *A General Theory of Equilibrium Selection in Games*, Cambridge, Massachusetts: MIT Press.

Kandori, Michihiro, Mailath, George and Rob, Rafael (1993). 'Learning, mutations, and long run equilibria in games.' *Econometrica* vol. 61 (1), pp. 29–56, January.

Jordan, James S. (1991). 'Bayesian learning in normal form games.' *Games and Economic Behavior* vol. 3, pp. 60–81.

Jordan, James S. (1993). 'Three problems in learning mixed-strategy Nash equilibria.' *Games and Economic Behavior* vol. 5, pp. 368–86.

Maynard Smith, John (1982). *Evolution and the Theory of Games*. NY: Cambridge University Press.

Maynard Smith, John and Price G. R. (1973). 'The logic of animal conflict.' *Nature*, vol. 246 (5427), pp. 15–8.

Michelitsch, Roland (1992). Economic Science Laboratory, University of Arizona, personal email communication, February.

Rapoport, A. and Orwant, C. (1962). 'Experimental games: a review.' *Behavioral Science* vol. 7, pp. 1–37.

Selten, Reinhard (1988 and 1991). 'Anticipatory learning in two-person games.' Discussion Paper B93, University of Bonn; a revised version appeared in (ed. R. Selten) *Game Equilibrium Models*, Vol 1, New York: Springer-Verlag.

Selten, Reinhard (1989). 'Evolution, learning and economic behavior: 1989 Nancy L. Schwartz Memorial Lecture.' University of Bonn Discussion Paper B-132. A revised version appeared in *Games and Economic Behavior* vol. 3 (1) (1991) pp. 3–24.

Siegel, Sidney. 'Decision making and learning under varying conditions of reinforcement.' *Annals New York Academy of Sciences* vol. 89 (5), pp. 766–83.

Smith, Vernon L. (1982). 'Microeconomic systems as experimental science.' *American Economic Review* vol. 72 (5), pp. 923–55, December.

Van Huyck, J., Battalio, R. Mathur, S. Ortmann, A. and Van Huyck, P. (1992). 'On the origin of convention: evidence from symmetric bargaining games.' Texas A & M working paper 92-05.

Weibull, Jorgen W. (1995). *Evolutionary Game Theory*. Cambridge, Massachusetts: MIT Press.

Young, H. Peyton (1993). 'The evolution of conventions.' *Econometrica* vol. 61 (1), pp. 57–84, January.