2

THE EVOLUTION OF COOPERATION IN THE FINITELY REPEATED PRISONER'S DILEMMA

John H. Nachbar

September 1989

P-7591

The RAND Corporation, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

THE EVOLUTION OF COOPERATION IN
THE FINITELY REPEATED PRISONER'S DILEMMA

John H. Nachbar
RAND Corporation
Economics and Statistics
1700 Main Street
Santa Monica, California 90406

September 18, 1989

This paper is a revised version of Chapter 2 of my PhD thesis. It has greatly benefitted from the help of Dilip Abreu, Vijay Krishna, Eric Maskin, Andreu Mas–Colell, William Nachbar and several anonymous referees. Robert Axelrod was most gracious in answering questions concerning his article and book. Of course, all errors are the author's.

Abstract

This paper examines "evolutionary" dynamic behavior in the finitely repeated prisoner's dilemma. It is first noted that the "fitness" of cooperation found in the best known simulation of this type, that by Robert Axelrod, stems from strategy set restrictions that altered Nash equilibrium behavior: Axelrod's restricted game has a continuum of pure cooperation equilibria and no pure defection equilibrium. New simulations, maintaining the finite game's equilibrium structure, are presented here. It is found that although cooperation is ultimately exploited and extinguished, dynamic paths can "pseudo converge" in ways that allow partial cooperation to flourish for extended periods of time.

1. Introduction

This paper examines "evolutionary" dynamic behavior in the finitely repeated prisoner's dilemma. Under the dynamics considered, if there is convergence to a limit (in general there need not be), then that limit must correspond to a Nash equilibrium of the game formed from the strategies in the initial distribution. This will be verified in the text. For the finitely repeated prisoner's dilemma, the implication is that if all strategies are present in the initial distribution then ultimately only defection can "evolve".

This observation appears to be contradicted by a well–known dynamic simulation of Robert Axelrod in which the dynamic path seems headed for a purely cooperative limit. As will be discussed, the discrepancy is resolved by noting that the strategies employed in that simulation yield a restricted game with, on the one hand, no purely defecting equilibria at all, and on the other, a continuum of cooperative equilibria. The fundamentally cooperative nature of Axelrod's game appears in turn to stem from his having used the strategies submitted for the second of his computer tournaments, which was constructed to be viewed by the participants as an <u>indefinitely</u> repeated game.

This paper takes up the question of what happens in the simulation of a "true" finitely repeated prisoner's dilemma with known endopoint. Of course, ultimately we expect cooperation to be extinguished, but this leaves unanwered how the dynamics behave short of the limit. For practicality, this paper, like Axelrod's, must typically consider restricted strategy sets. However, care is taken here that the corresponding restricted games share the finitely repeated prisoner's dilemma's main qualitative features. In particular, in the games considered here, mutual defection is the only Nash equilibrium play and "defect always" is the unique dominance solution. It will then be demonstrated by example that dynamic paths can "pseudo converge" to distributions in which partial cooperation can prosper for extended periods of time. It will be noted also that this same behavior can create computation problems making it imprudent in general to try to infer limiting behavior from a necessarily finite run. Finally, it will be pointed out that, in the 2–stage game at least, path limits all correspond to imperfect equilibria: the weakly dominanted strategy "Defect Tit–for–Tat" (defect in the first stage; in stage 2 play whatever one's opponent played in stage 1) is never driven to extinction.

The organization of the remainder of this paper is as follows. The next section reviews briefly the prisoner's dilemma. Section 3 introduces, and to a degree motivates, the dynamics. Section 4 discusses some fundamental properties. Section 5 reexamines the simulation run by Axelrod. Sections 6 and 7 discuss the new simulations run for this paper.

2. The Prisoner's Dilemma

The reader will recall that the payoff matrix for the 1–stage prisoners' dilemma is given by:

$$\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{\underline{\text{Player 2}}} \\
 & & C & D \\
\cline{3-4}
\underline{\text{Player 1}} & C & \begin{smallmatrix} & a \\ a & \end{smallmatrix} & \begin{smallmatrix} & c \\ b & \end{smallmatrix} \\
\cline{3-4}
 & D & \begin{smallmatrix} & b \\ c & \end{smallmatrix} & \begin{smallmatrix} & d \\ d & \end{smallmatrix} \\
\cline{3-4}
\end{array}$$

where $c > a > d > b$ and $2a > b+c$. C is to be read "cooperate" and D "defect". The 1–stage game has a unique Nash equilibrium in dominant strategies, namely (D,D), an outcome which is Pareto inferior.

As the stage game is repeated, the number of possible strategies grows explosively. For s stages there are

$$2^{2^{s}-1}$$

distinct strategies, which exceeds 10^{100} for $s \geq 9$.[1] The diversity of possible strategies suggests that it might be possible to enforce mutual cooperation in equilibrium by a threat to punish (by, say, a permanent reversion to "defect always", written DD) upon seeing a play of D. This approach can indeed elicit mutual cooperation in the infinitely or indefinitely repeated prisoner's dilemma (an instance of the "Folk Theorem") and also yields a Folk Theorem–like proliferation of (perfect) equilibria in many finitely repeated games (see Krishna 1988 for a survey of the Folk Theorems). However, as is well known (e.g. Luce and Raiffa 1957) the approach fails utterly with the finitely repeated prisoner's dilemma, at least when the terminal stage is common knowledge. While the diversity of strategies does generate a continuum of Nash equilibria (provided $s \geq 2$), qualitatively they are all alike: <u>within equilibrium</u> the only play is mutual defection at all stages. Moreover, although DD is no longer a dominant strategy for $s \geq 2$, iterated deletion of (weakly) dominated strategies yields DD as the unique dominance solution.[2] The absence of cooperation in equilibria of the finitely repeated prisoner's dilemma is a consequence of the fact that the punishment strategy and the dominant 1–stage strategy are one and the same, causing cooperation to "unravel" from the back: in equilibrium, players will always mutually defect in the last period; thus, since either will be "punished" in equilibrium in the last period, neither has reason not to defect in the second to last, and so on.

Classical game theory's sharp prediction of mutual defection in the finitely repeated prisoner's dilemma is at variance with the evidence: the "stylized fact" from experiments is that play is typically cooperative until towards the end. See Selten and Stoecker (1986) and also Roth (1988). The question raised by Axelrod's simulation and addressed here is: to what extent can the dynamical system used by Axelrod make non–Nash equilibrium behavior "plausible"?

3. The Dynamics

We imagine that there is an infinite population of players and that in each time period players are paired randomly. After pairs have played their repeated games, each player is informed about his payoff in his pairing, is given some information (possibly incomplete) about the payoffs of other players and using this information he revises his strategy choice.

It is conjectured that the strategy revisions aggregate into a function G on the space of strategy distributions, namely the simplex $\Delta^n = \{P \in \mathbb{R}_+^{n+1} \mid \Sigma P_i = 1\}$, where n+1 is the number of strategies.[3] P_i is simply the population fraction playing i, and if P^t is the round t distribution, $P^{t+1} = G(P^t)$ is the distribution at date t+1. Loosely, the hypothesis will be that the fraction of participants submitting strategy i will go up if i does well relative to the average, and down if it does poorly. To be more explicit, it is helpful to introduce some notation.

Let A be the matrix of payoffs to player 1. Thus, for the 1–stage game, A is given by

$$A = \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array}\,.$$

For an s–stage game A would, of course, be much larger. We will typically be considering a subset of the full strategy set, one with, say, only n+1 pure strategies. If the current strategy distribution is P, the expected payoff to an i player is $(AP)_i$, the i entry in AP. The average payoff in the population as a whole is $P \cdot AP$. If the round t distribution is P^t, the dynamics proposed is:

$$P_i^{t+1} = G_i(P_i^t) = P_i^t \frac{(AP^t)_i}{P^t \cdot AP^t}$$

which can be rewritten:

$$P_i^{t+1} - P_i^t = \frac{P_i^t}{P^t \cdot AP^t}\left[(AP^t)_i - P^t \cdot AP^t\right].$$

In words, under the dynamics the growth rate, $(P_i^{t+1} - P_i^t)/P_i^t$, is equal to the difference between the expected payoff to strategy i, given the current population, and the average expectation for the population as a whole (ignoring the renormalization term $1/P^t \cdot AP^t$). I will assume that A is normalized so that $A \geq 0$ with a strictly positive diagonal. This will assure, among other things, that the P_i remain non–negative.

The customary home of these dynamics is within mathematical biology, where they are known as the replicator dynamics and serve as a model of natural selection based on asexual reproductive success. Areas of application are varied, including biological evolutionary game theory, mathematical ecology (the replicator dynamics are formally equivalent to the Lotka–Volterra dynamics), and population genetics. Schuster and Sigmund (1983) provide an overview of the role played by replicator–type systems in mathematical biology. The book by Hofbauer and Sigmund (1988) provides a comprehensive technical survey.

The success of the replicator dynamics in biology does not argue its suitability for modeling learning/imitation environments. The position taken here is that the dynamics offer a relatively tractable "first approximation" to modeling those imitation environments in which the probability of a player switching to a new pure strategy, say i, depends not only on the performance of i (here measured by $(AP)_i - P \cdot AP$), but also on the probability of i's performance being observed (as measured by P_i). Nachbar (1989) provides somewhat more detail on this type of motivation. Stress should be placed on the reference to "first approximation". No claim is made in Nachbar, for example, that the replicator dynamics can be exactly motivated in this way. The reader should also refer to Crawford (1988), who argues that if individuals play mixed strategies and adjust their strategies according to the replicator equation, then the replicator dynamics cannot simultaneously serve as a description of dynamics in the population as a whole. There is an aggregation problem.

4. Fundamental Properties of G

Denote by suppP the set $\{i \mid P_i > 0\}$ (i.e., the set of strategies to which P gives positive weight).

PROPOSITION 1 From an initial point P^0, if $\lim P^t = P^*$ then (P^*,P^*) is a Nash equilibrium of the game with (pure) strategy set equal to $\mathrm{supp}P^0$.

The proof is supplied in the Appendix, where it is derived from four simple properties of G. The basic argument is that if i is an inferior strategy in the limit, meaning that for some j, $(AP^*)_j > (AP^*)_i$, then i must be inferior to j near the limit. It follows that P_i tends to zero, hence $\mathrm{supp}P^*$ contains no inferior strategies.

Two aspects of the proposition deserve special comment. First, the proposition does not say that (P^*,P^*) needs to be a Nash equilibrium for the game with all strategies present. This is a consequence of the fact that the replicator dynamics never introduces strategies not already in $\mathrm{supp}P^0$.[4] The second point is that we cannot conclude that P_i tends to zero if we merely have that i is inferior to j near the limit but $(AP^*)_i = (AP^*)_j$. This observation underlies the result on imperfect limits discussed in section 7.

Proposition 1 is a known "folk theorem" in evolutionary biology. It has also been noted independently by some economists (e.g. Samuelson 1988, and Hansen and Samuelson 1988). The result readily generalizes to <u>any</u> dynamical system that obeys the four properties used in the argument, whether the system is discrete or continuous, whether the game is 2–player symmetric (as here) or n–player asymmetric. A related but distinct result is worth mentioning but will not be proved here:

Proposition 2 If P^* is a stable fixed point then it is a Nash equilibrium.

Recall that P^* is said to be stable if any dynamical path that starts near P^* remains near P^* (the path need not actually converge to P^*). Again, this is a known result. A proof is provided in Nachbar (1989), among other places. Again, the result is very general.

The reader should note that the propositions provide no guarantee that in fact any dynamic path P^t converges. For arbitrary games, convergence need not occur. Nachbar

(1989) discusses conditions known to imply convergence. Briefly, while convergence can easily be established for strict dominance–solvable games, weak dominance–solvable games, including the finitely repeated prisoner's dilemma, pose subtleties. It can be proved that any path in the 2–stage prisoner's dilemma converges. The analogous result for the k–stage game remains an open question.

5. Axelrod's Simulation

Axelrod's simulation is presented in his 1984 book, "The Evolution of Cooperation".[5] The stated intention was to mimic his procedure of soliciting strategies, running a round–robin tournament, announcing the results, and then repeating the process. For his matrix **A**, Axelrod chose the payoff matrix generated in the second (and last) of the two actual tournaments. Despite this intimate link between simulation and tournament, the game of the simulation is fundamentally different from that of the second tournament. As this point has caused some confusion, I will be more explict.

From the perspective of tournament participants, the tournament game was indefinitely repeated: players knew that the actual length would be finite but they did not know the terminal stage.[6] Rather, the game length was drawn from a distribution characterized by a known probability ω that at any stage the game would continue for at least one more stage. As ω was set high ($\omega = .99654$), cooperation can be supported in Nash equilibrium in this game. This may in turn explain the highly cooperative nature of the submitted strategies.

In contrast, from the perspective of the dynamics the simulation was of a finite game with known endpoint, since the cells of **A** were computed for such a game.[7] If simulation of an infinitely or indefinitely repeated game had been intended, Axelrod would have needed either a different choice of **A** (with cells calculated as the ω–weighted sum of an infinite payoff stream) or a change in the dynamics to allow **A** to be stochastic.

Turning to the simulation, Axelrod took an initial distribution P^0 that gave equal weight to every strategy used and computed 1000 generations. The well-known finding was that the strategy "Cooperate Tit-for-Tat" (CTT: cooperate in stage 1 and thereafter play whatever one's opponent played in the previous round) was ranked strictly first from round 2 onwards, although only by a small margin (it is one of six strategies with shares between .1 and .15 at round 1000). The plot of frequency against generation demonstrates considerable monotonicity, with the ordinal position of the top 6 strategies not changing at all after the first few rounds.

According to Proposition 1, if the dynamic path converges to a limit and the round 1000 distribution is close to that limit, then the round 1000 distribution must be close to a Nash equilibrium. There is indeed a Nash equilibrium nearby. In fact, there is a continuum of them, supported by the 39 of the 63 submitted strategies that have the property that each is never first to defect (is "nice" in Axelrod's terminology).[8] The point in this equilibrium set that is closest to the round 1000 distribution lies .01 (in the Euclidean metric) away, in the set's strict interior.[9] Since the strategies are nice, play in every equilibrium from the set is mutual cooperation at all stages. The situation for defecting equilibria is diametric: Axelrod's game admits no purely defecting Nash equilibria whatsoever.[10]

It thus appears that the fitness of cooperation found in Axelrod's simulation derives predictably from the Nash equilibrium properties of his restricted game. These properties, in turn, are quite different from those that obtain in either the "complete" finitely repeated prisoner's dilemma (where defection is the only equilibrium outcome) or the "complete" infinitely or indefinitely repeated prisoner's dilemma (where pure cooperation is one of only many possible behaviors). However, there may nevertheless be substance in Axelrods' basic intuition that cooperation is dynamically fit, in some sense, even in hostile environments, an extreme being the "true" finitely repeated game with known endpoint. The remainder of this paper focuses on behavior short of the limit in the finite game. An alternative would be to investigate the replicator dynamics in infinitely repeated games with richer equilibrium possibilities. This approach is taken in Hirshleifer and Martinez Coll (1988) and in Blad (1986), both of which consider a variety of 3–strategy games related to the infinitely repeated game.[11]

6. Simulations of "Large" Finitely Repeated Prisoner's Dilemmas

The following two sections will provide evidence on dynamic behavior in prisoner's dilemmas with two or more stages (the reader can easily verify that behavior in the 1–stage game is trivial). The present section will focus on a 6–stage example which is a good qualitative representation of my experience with "large" prisoner's dilemmas. I will discuss at some length the degree to which the results found are robust. The next section will then turn to simulations of 2–stage games, which are of special interest because those games are small enough to run simulations with all strategies present.

For repeated games with even just a few stages, tractability requires restrictions on the number of strategies investigated. To retain the spirit of the finitely repeated prisoner's dilemma in the restricted game, the simulations I have conducted have included for any given strategy x, additional strategies of the form "play x until the last period, then defect for certain", "play x until the last two periods, then defect for certain", and so on. In such games, as in the full finitely repeated prisoner's dilemma, defection is the only play in Nash equilibrium and defect always (DD) is the unique dominance solution.

Consider then a 6–stage repeated prisoner's dilemma with 1–stage payoff matrix:

$$A = \begin{array}{|c|c|} \hline 4 & 0 \\ \hline 6 & 1 \\ \hline \end{array}$$

and strategies:

1) CTT (Cooperate Tit–for–Tat: C in stage 1; in stage t+1 play whatever one's opponent played in t)
2) CTT until stage 6 then D
3) CTT until stage 5 then D in the last 2 stages
4) CTT until stage 4 then D in the last 3 stages
5) CTT until stage 3 then D in the last 4 stages
6) C then D in the last 5 stages
7) DD

For an initial P^0 in the center of the simplex, the path diagram is given in figure 1. A partially cooperative strategy such as 4 is seen to prosper for a time, but eventually it is exploited and overtaken by the strategy that defects one period earlier. Ultimate convergence to $P_7 = 1$ can easily be shown based on the crude condition that if ever $P_7 > 15/16$ then from then on P_7 increases monotonically and all the other population fractions fall monotonically. It is evident that termination of the simulation at a round short of 700 could yield an erroneous prediction about the path limit even though it might appear to the researcher that the path had nearly converged. The path displays, in particular, "pseudo convergence" to $P_6 = 1$: there is a span of nearly 300 periods in which strategy 6 gets weight of .99 or higher.

The mathematical intuition for why this behavior arises is straightforward. The weight on strategy 6 persists at high levels because the weight on strategy 7 (DD), the strategy that exploits 6, falls to roughly 10^{-30} by round 288. Since strategy 7 does only slightly better in a population of 6s than does 6 itself, it takes roughly 300 rounds for the population weight on 7 to recover. In turn, the magnitude of the fall in P_7 stems from the fact that DD is a poor response, indeed the worst (available) response, to any population distribution where mass is concentrated on strategies other than 6 or 7. Consequently, if P^t is a distribution where

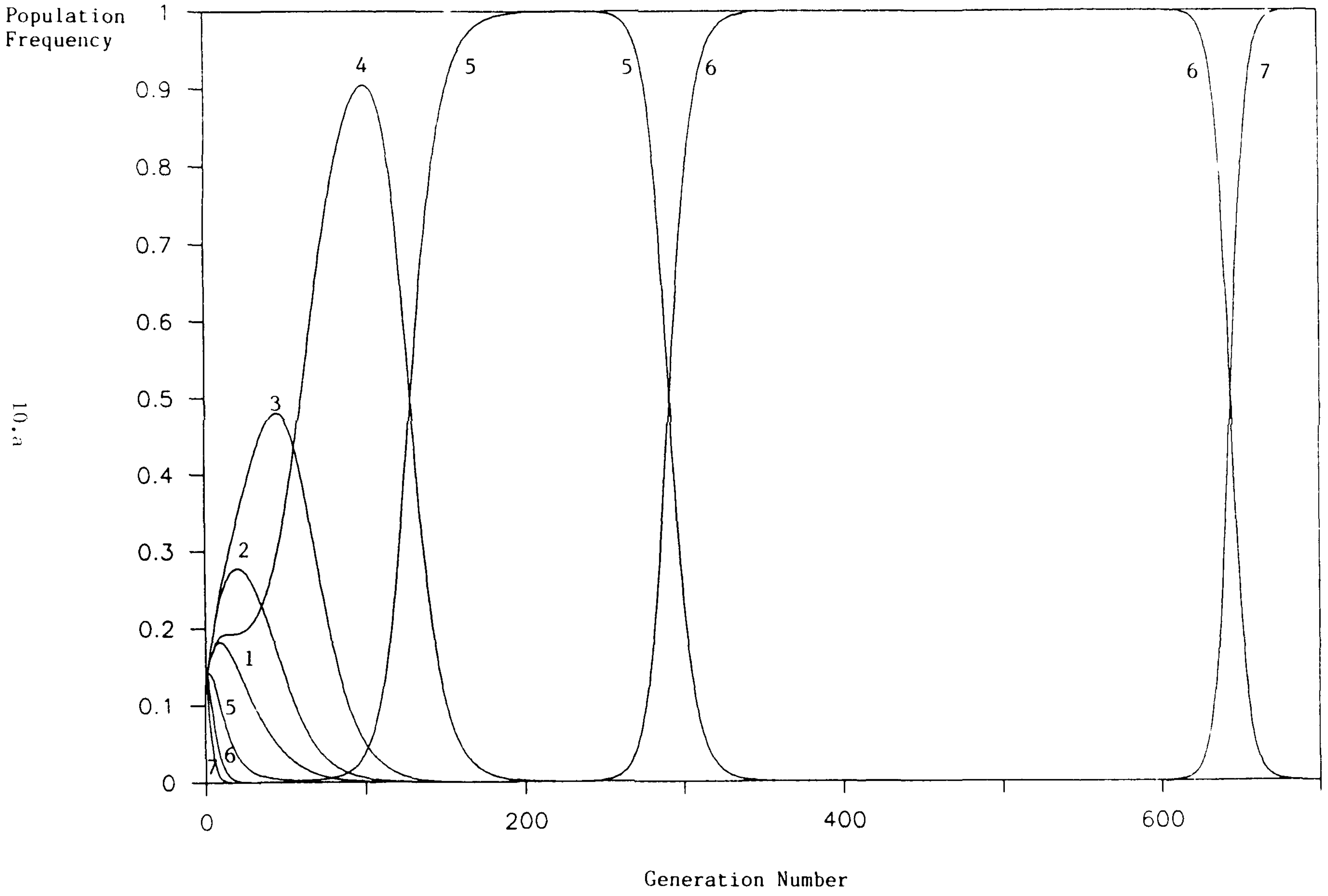

Figure 1 - 6-Stage Path Diagram

strategies 1 through 5 predominate, then the ratio:

$$r_7(P^t) \equiv \frac{(AP^t)_7}{P^t \cdot AP^t}$$

is less than one. But $P_7^{t+1} = P_7^t r_7(P^t)$, so if for all v in some time interval $\{0,...,V\}$ we have $r_7(P^{t+v}) < K < 1$, for some positive K, then P_7^{t+v} is less than $K^v P_7^t$, which can rapidly become very small. As it happens, $r_7(P^t)$ falls to a low of roughly .59 at round 24, climbs only to .73 by round 100, and was still only .92 at round 200.

Is figure 1 robust? This question will occupy the remainder of the present section. The short answer is "definitely yes" with respect to changes in basic data such as the initial distribution, "more or less yes" with respect to certain changes in the dynamics and the associated learning story.

Taking the latter point first, the current story states that during the 300 rounds in which strategy 6 predominates, the superior performance of strategy DD remains unknown to the vast majority of players for the simple reason that DD players are rare and hence infrequently observed. One might conjecture, as an alternative, that at least some players, having met strategy 6 players exclusively for tens if not hundreds of consecutive rounds, would analyze the normal form directly and switch to DD without benefit of example.

Suppose we change the dynamics by assuming that, while the behavior of most of the population remains described by G, a fraction ε of the population is composed of well–informed and computationally proficient "analyzers" who play in period t+1 the best pure strategy response to P^t (i.e. these players act according to the Cournot tâtonnement, on which see Moulin (1986)). The effect of the analyzers is to ensure that the population weight on the best current strategy is never less than ε. The consequence of this change is what one expects: for $\varepsilon > 10^{-30}$, intervals in which strategies 5 or 6 are close to 1 are shorter (if they exist at all) and approximate convergence to $P_7 = 1$ occurs earlier. For example, setting ε at 10^{-6} reduces the interval during which $P_6 > .99$ from nearly 300 rounds to only 4; consequently, the first round in which $P_7 > .99$ drops by more than half (310 versus 667).

The reader should note that, strictly speaking, the path in figure 1 is nevertheless robust to the analyzer perturbation: although the path is sensitive to small ε, the impact of the perturbation all but disappears if ε falls below 10^{-30} (admittedly a very small number). Moreover, the degree of sensitivity to small ε should not be overstressed. With $\varepsilon = 10^{-6}$ as above, the path resembles a "compressed" version of that of figure 1; in particular, it remains true that there is an interval in which strategy 4 prospers, followed by a regime of 5, then one of 6, and finally convergence to 7. Thus something rather like figure 1 can still be obtained even with the addition of a "modest" number of analyzers.

As noted, with respect to changes in the basic data one can assert the robustness of figure 1 with considerably more vigor. In the case of the initial distribution, not only is the figure essentially unaltered by sufficiently small changes but we can even say that the behavior shown is, in a precise sense, probable. Specifically, if we take every starting point in the interior of the simplex as being equally likely then the probability is roughly 2/3 that 1) P_5 will exceed .99 at some point in the path and 2) following that, there will be a stretch of at least 200 rounds in which $P_6 > .99$. Conversely, the probability that P_7 will be less than, say, 10^{-6} after 300 rounds is roughly 75%, rising to 90% if we stop the simulation at round 100. (These probabilities are Monte Carlo estimates, for a sample size of 1000.)

The behavior expressed in figure 1 appears also to be robust to small changes in the payoff matrix, to the introduction of more complex strategy sets (although given the huge number of strategies this is a conjecture rather than an assertion), and to increases in the number of stages.[12] On the last, an increase in the number of stages yields a path that is analogous to figure 1 but more extreme. In particular, the fall in the population share playing DD can become so severe that problems arise with computer underflow, wherein one's software arbitrarily sets any number below some cutoff, roughly 10^{-300} in the case of the package I used, equal to zero. In effect, the software decides that any population that small becomes extinct, rendering further computation meaningless from the standpoint of a strict interpretation of the dynamical system.

Superficially, underflow may appear to be a non–problem since one might argue that the imposition of an "extinction threshold" is "realistic". However, to the degree that extinction makes a qualitative difference (i.e. substantially alters dynamic behavior), it works its effect by exploiting one of the replicator dynamics least appealing features, namely the "no creation" property that once a strategy i is extinguished it disappears forever. In the interpretation maintained here, "no creation" stems from an implicit assumption that no one in the population has the informational or computational resources to reinvent extinct strategies. As an example, suppose that in the game of this section we set an extinction threshold of between 10^{-10} and 10^{-30}. Then for an inital P^0 in the center of the simplex we get a limit of $P^*_6 = 1$, we get cooperation for one period. This is simply a consequence of the fact that strategy 7 dies off before it gets a chance to exploit 6. Not surprisingly, this kind of behavior does not persist under the introduction of an ε subpopulation of "analyzers" (naturally ε would have to be set above the threshold level to have an effect). An alternative would be to allow for the periodic introduction of small "mutant" populations. As should be intuitive, periodic mutution will restore dynamic behavior to something more or less like that displayed in figure 1, with eventual convergence to $P^*_7 = 1$.[13]

7. Simulations of the 2-Stage Prisoner's Dilemma

Figure 1's most striking feature, the "pseudo convergence" to a non–Nash equilibrium, can be obtained when there are only 2 stages. However, 2–stage examples seem to require starting points that are extreme (e.g., that place most weight on CTT). Rather than dwell on this point, it is more fruitful to observe that even with a 2–stage game it is easy to find circumstances in which dynamic paths remain far from their limits for extended periods of time. For a dramatic example, consider the game with stage payoffs:

$$A = \begin{array}{|c|c|}\hline 99 & 0 \\ \hline 100 & .5 \\ \hline \end{array}$$

These payoffs were chosen to keep the relative reward for defecting low, so that to a limited degree the 2–stage payoffs resemble those generated by a longer game. The two–stage game has 8 distinct pure strategies:

1) C always
2) CTT
3) C in the first stage then "reverse tit–for–tat": D if one's opponent played C and vice versa
4) C and then D
5) D and then C
6) D and then tit–for–tat (DTT)
7) D and then reverse tit–for–tat
8) DD.

The path for a P^0 that gives equal weight to each strategy is depicted in figure 2. Ultimately, the path converges (Proposition 2 in Nachbar 1989), and it is to the mixed strategy Nash equilibrium given, approximately, by $P^*_6 = .0029$ and $P^*_8 = .9971$. Nearby starting points give similar, though not identical, path patterns, as do nearby stage payoff matrices. In particular, the "braiding" seems to be robust. Perhaps of greater interest, the probability is high that

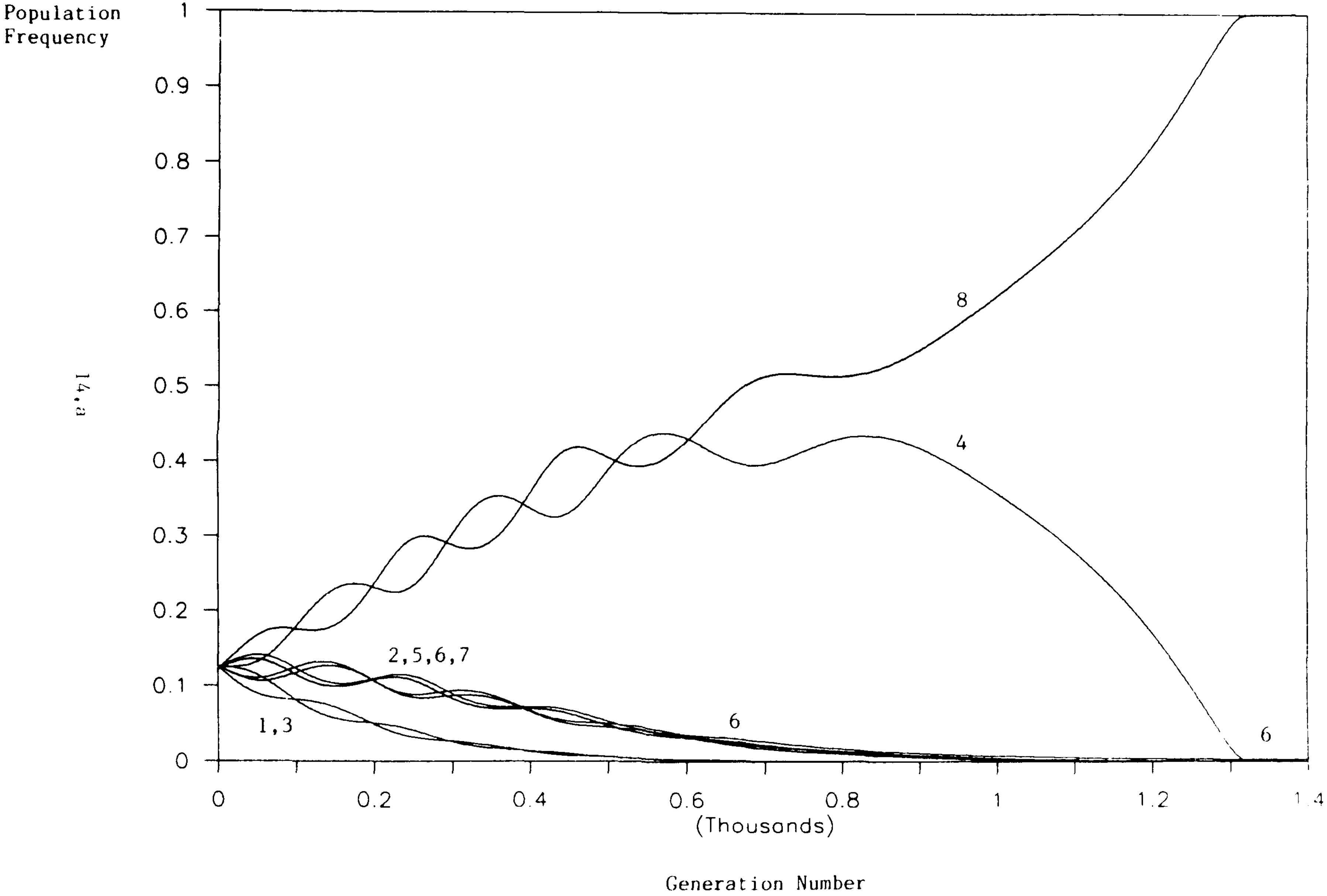

Figure 2 - 2-Stage Path Diagram

strategy 4 will maintain a large population share late into the simulation. For example, a Monte Carlo simulation gives a probability of roughly 80% that $P_4 > 1/3$ at round 800.

The limit of the path in this example is not subgame perfect: DTT allows the possibility of cooperation at stage 2 whereas the dominant strategy at that stage is to defect. One might conjecture, based on the 6–stage, 7 strategy example above, that as a general principle the evolutionary dynamic would drive weakly dominated strategies (such as DTT) to zero. However, this logic is invalid in the presence of payoff matrix ties. In matrices derived from extensive form games, of which the finitely repeated prisoner's dilemma is an example, such ties can be endemic. Here, they lead not only to imperfect limits but to the following:

FACT For any 1–stage payoff matrix which is admissible (i.e. is non–negative), $P_8 = 1$ is not the limit of any path from the interior of the simplex.

Because every dynamic path from the interior of the simplex converges (Nachbar 1989), the fact implies that for these particular dynamics DTT is "evolutionarily fit" in a strong sense: it is never driven (completely) to extinction from any starting point. A proof is provided in the Appendix. The basic reasoning behind the result is that although DD weakly dominates strategy DTT, it does so only if other strategies get positive weight, and under the dynamics used here those strategies are going to zero geometrically.

8. Conclusion

Ultimately, the merit of any "evolutionary" argument for games played by humans depends on whether the evolutionary story can be given a credible foundation. The present paper has largely finessed this issue, focusing instead on the consequences of such an approach. It has been shown that, in terms of limit properties, the replicator dynamics offer no relief if Nash equilibrium behavior is deemed undesirable. The dynamics offer more hope short of the limit: convergence to approximate Nash equilibrium behavior can be slow and there may be pseudo convergence to a distribution which is not a Nash equilibrium. A caution should be mentioned here. The system given by G is invertible,[14] hence every point in the interior of the strategy simplex is a possible population distribution at round, say, 1,000. Consequently, to be meaningful, statements about distribution after a finite number of rounds must be phrased in terms of the probability of hitting a particular region (unless, of course, one can be very confident about the choice of P^0).

Arguments based on evolutionary limits become attractive when the task is not to explain non–Nash equilibrium behavior but rather to discriminate among a profusion of Nash equilibria. One can argue for the "plausibility" of one equilibrium over another based on stability, size of basin of attraction (union of all paths converging to the given limit) and so on. This is the spirit of the analytical section of Axelrod's book, which deals exclusively with selecting among the plethora of equilibria available in the infinitely repeated prisoner's dilemma. This approach is taken also in Hirshleifer and Martinez Coll (1988), and in Blad (1986), already mentioned in the text. Finally, the reader is directed to the papers by Friedman (1988), by Samuelson (1988), and by the present author, which analyze broad classes of evolutionary dynamical systems, of which the replicator dynamics investigated here is but one representative.

APPENDIX

PROOF OF PROPOSITION The proof is easy after one has noted that, with the restrictions imposed on **A** (namely that it be non–negative with a strictly positive diagonal), we have the following properties.

1) G is continuous on Δ^n.

2) G is "responsive": $G(P) = P$ iff $(AP)_i = (AP)_j$ for all $i,j \in \text{supp}P$. Thus, the fixed points of G have the property that they are Nash equilibria with respect to themselves, but not necessarily with respect to the full strategy set. For example, any vertex of Δ^n is, trivially, a fixed point of the replicator dynamics, whether or not it corresponds to a Nash equilibrium.

3) The system is forward invariant: $P_i^{t+1} = G_i(P^t) > 0$ iff $P^t_i > 0$. In words, strategies are neither created nor, except "in the limit", destroyed.

4) The system obeys a boundary rule: If $P^0_i > 0$, $P^0_j > 0$ and for some T, $(AP^t)_i > (AP^t)_j$ for $t > T$, then $\lim P^t_i = 0$ implies $\lim P^t_j = 0$. In words, if P^t_i gets driven to zero then so do all inferior strategies. The boundary property is proved below as a separate, and easy, Lemma.

These properties noted, suppose P* is not a Nash equilibrium with respect to $\text{supp}P^0$. Then there are pure strategies i,j such that $(AP^*)_i > (AP^*)_j$, with $i \in \text{supp}P^0$ and $j \in \text{supp}P^*$. Then by continuity, $(AP^t)_i > (AP^t)_j$ for $t > T$, some T sufficiently large (and by invariance $P^t_i > 0$ and $P^t_j > 0$). Hence by the boundary rule, $P^*_j > 0$ implies that $P^*_i > 0$. But then since G is responsive $G(P^*) \neq P^*$, hence P* is not a fixed point and so, since G is continuous, it is not a limit point. The proof follows by contraposition.▌

LEMMA The replicator dynamic obeys the boundary property.

Proof The replicator dynamic obeys:

$$\frac{P_i^{t+1} - P_i^t}{P^t_i} > \frac{P_j^{t+1} - P_j^t}{P^t_j} \quad \text{iff} \quad (AP^t)_i > (AP^t)_j$$

(as always, assuming $P^t_i > 0$, $P^t_j > 0$). Thus, the growth rate of strategy i exceeds that of j iff i has a higher payoff than j. Under the conditions of the boundary property, and thanks also to invariance, we then have that the ratio P^t_i/P^t_j is well–defined and strictly increasing for $t > T$. The result follows.█

PROOF OF FACT Let P* be given by $P_8 = 1$. It is easy to verify that the Nash equilibria of the game occupy a closed, nontrivial segment of the edge between 6 and 8, with P* occupying one endpoint. To show that P* cannot be an ecological solution, consider the ratio P^t_8/P^t_6. $P^t \longrightarrow P^*$ implies that this ratio tends to ∞. The proof will proceed by showing that if $P^t \longrightarrow P^*$ then this ratio remains bounded, a contradiction.

To show that P^t_8/P^t_6 is bounded, we majorize this sequence by another sequence that we know to be bounded. Note first that for every strategy i except 6 and 8, there is a number $\beta_i \in (0,1)$ such that $(AP^t)_i/(P \cdot AP^t)_i < \beta_i$, provided t exceeds some large number, say τ_i (this follows since by assumption $P^t \longrightarrow P^*$ and since $a_{8j} > a_{88}$ for all j other than 6 and 8; a_{ij} being the generic entry in A). We take a common $\beta < 1$ and a common threshold τ that serves for all such i. Next, denote by $\varepsilon^t > 0$ the total weight at t on all strategies other than 6 and 8. It is convenient to denote ε^τ simply by ε. Then $\varepsilon^{\tau+t+1} < \beta\varepsilon^{\tau+t} < \beta^{t+1}\varepsilon$. Finally, let B be the highest possible (expected) payoff to 8 conditional on 8 playing against one of the six strategies other than 6 or 8. Similarly, let C be the lowest possible (expected) payoff to 6 conditional on 6 playing one of the other six strategies. The precise values of B and C do not matter, only that they are independent of t and that $B > C > 0$. With this noted:

$$\frac{P_8^{\tau+t+1}}{P_6^{\tau+t+1}} < \frac{(F(1-\beta^s\varepsilon) + B\beta^s\varepsilon)}{(F(1-\beta^s\varepsilon) + C\beta^s\varepsilon)} \cdot \frac{P_8^{\tau+t}}{P_6^{\tau+t}}$$

where F (for "Fink"; > 0) is the payoff to 6 or 8 from mutual defection. Proceeding in this way we have:

$$\frac{P_8^{\tau+t+1}}{P_6^{\tau+t+1}} < \left\{ \prod_{s=0}^{t+1} \frac{(F(1-\beta^s\varepsilon) + B\beta^s\varepsilon)}{(F(1-\beta^s\varepsilon) + C\beta^s\varepsilon)} \right\} \cdot \frac{P_8^{\tau}}{P_6^{\tau}}$$

We need to show that the term in brackets is bounded. (This will show that in fact the term converges, since the bracketed term is non–decreasing, given that B > C.) Intuitively, it is clear that the term should converge, since as s goes to ∞ the ratio goes "geometrically" to F/F = 1. Rigorously, we show that the product is majorized by another product which converges. We need to establish some additional inequalities. Note first that we can assume without loss of generality that F = 1. Let then:

$$V(s) = \frac{1-\beta^s\varepsilon + B\beta^s\varepsilon}{1-\beta^s\varepsilon + C\beta^s\varepsilon}$$

Thus, in our new notation we are interested in lim ∏ V(s). Let $\delta = B - C > 0$. Then adding and subtracting $C\beta^s\varepsilon$ in the numerator we have:

$$= 1 + \frac{\delta\beta^s\varepsilon}{1+C\beta^s\varepsilon-\beta^s\varepsilon}.$$

Since $C > 0$ and since $\beta^s < \beta$, we have:

$$< 1 + \frac{\delta\beta^s\varepsilon}{1-\beta^s\varepsilon} \qquad < 1 + \frac{\delta\beta^s\varepsilon}{1-\beta\varepsilon}$$

$$= 1 + \delta^*\beta^s\varepsilon$$

where $\delta^* \equiv \delta/(1-\beta\varepsilon) > 0$, a constant.

Now $\prod V(s)$ tends to ∞ if and only if its log does. Thus we are interested in

$$\begin{aligned}\log \prod V(s) &= \sum \log(V(s)) \\ &= \sum \log(1+\delta^{*}\beta^{s}\varepsilon)\end{aligned}$$

But $\log(1+\delta^{*}\beta^{s}\varepsilon) < \delta^{*}\beta^{s}\varepsilon$. Hence:

$$\begin{aligned}\log \prod V(s) &< \sum \delta^{*}\beta^{s}\varepsilon \\ &= \delta^{*}\varepsilon \sum \beta^{s}\end{aligned}$$

The last is a geometric series, and thus converges.∎

Notes

1) This calculation counts as a single strategy any two strategies whose behaviors diverge only after a deviation, not by one's opponent but by oneself. Any two such strategies yield identical rows and columns in the normal form.

2) See Moulin (1986) for discussion of dominance solvability.

3) Implicit is a "Markov" property: only the current population distribution affects the choice of strategy for next period.

4) Precisely to emphasize this "no creation" property, Axelrod refers to the dynamics as "ecological" rather than "evolutionary" . However, usage here is consistent with that in (biological) evolutionary game theory.

5) The simulation I will be discussing appears on pp. 48–53. That material is in turn taken from Axelrod (1980), where the results of his second tournament were originally presented. The book also contains a "territorial" simulation which will not be examined here.

6) More accurately, Axelrod computed the payoff matrix A, used in both the tournament and the simulation, as the average of five finite games. As will be noted again in footnote 7, this complication is not of the essence.

7) The fact that A is actually an average of five games (see footnote 6) is not consequential from the point of view of equilibrium behavior. If all possible strategies are present then by the usual argument only defection obtains in Nash equilibrium, and defect always is the unique dominance solution.

8) Included in this set are all of the game's pure strategy Nash equilibria. CTT forms a symmetric pure strategy equilibrium with itself but so do 13 other (nice) strategies.

9) Extending Axelrod's simulation to 10,000 rounds brings the path to within 10^{-7} of a slightly different Nash equilibrium. This limit also has all 39 nice strategies in its support and continues to give weight of between one and two tenths to CTT and 5 other strategies. Axelrod's conjecture (p.55) that the fraction of the population playing CTT would converge to 1 is probably in error. Note, incidentally, that I have not claimed that the path in this simulation actually converges. While convergence seems highly likely, the size and complexity of the game makes analysis difficult. In the simulations of the next section, the games are simple enough that convergence can be formally demonstrated in every case.

10) In particular, DD was not in the strategy set. Conceivably, Axelrod's game may admit Nash equilibria in which defection is observed in some, but not all, stages (these would be mixed strategy equilibria; see footnote 8). However, I have yet to find any.

11) Blad considers the continuous version of the dynamics used here. Hirshleifer and Martinez Coll consider a number of dynamics, including a discrete system similar to the replicator dynamics but not quite identical to it.

12) When the number of stages is increased, the number of strategies must be increased as well to preserve dominance solvability to DD. See the discussion at the beginning of the section.

13) In the language of evolutionary game theory, $P_7 = 1$ is the unique evolutionarily stable strategy (ESS; Maynard Smith and Price 1973; see also Maynard Smith 1982 and Hofbauer and Sigmund 1988).

14) In fact, G is a diffeomorphism. See Losert and Akin (1983).

REFERENCES

Axelrod, R., "More Effective Choice in the Prisoner's Dilemma", *Journal of Conflict Resolution*, 1980, 24, 379–403.

Axelrod, R., *The Evolution of Cooperation*, New York: Basic Books, 1984.

Blad, M. C., "A Dynamic Analysis of the Repeated Prisoners' Dilemma Game", *International Journal of Game Theory*, 1986, 15, 83–99.

Crawford, V., "Learning and Mixed–Strategy Equilibria in Evolutionary Games", mimeo, 1988, University of California, San Diego.

Friedman, D., "On Evolutionary Games in Economics", mimeo, 1988, University of California, Santa Cruz.

Hansen, R. G., and Samuelson W. F., "Evolution in Economic Games", *Journal of Economic Behavior and Organization*, 1988, 10, 315–338.

Hirshleifer, J. and Martinez Coll, J. M., "What Strategies Can Support the Evolutionary Emergence of Cooperaton?", *Journal of Confict Resolution*, 1988, 32, 367–398.

Hofbauer, J., and Sigmund, K., *Dynamical Systems and the Theory of Evolution*, Cambridge: Cambridge University Press, 1988.

Krishna, V., "The Folk Theorems For Repeated Games", Harvard Business School Working Paper, 1988.

Losert, V., and Akin, E., "Dynamics of Games and Genes: Discrete versus Continuous Time", *Journal of Mathematical Biology*, 1983, 17, 241–251.

Luce, R., and Raiffa, H., *Games and Decisions*, New York: John Wiley and Sons, 1957.

Maynard Smith, J., *Evolution and the Theory of Games*, Cambridge: Cambridge University Press, 1982,

Maynard Smith, J., and Price, G., "The Logic of Animal Conflict", *Nature*, 1973, 246, 15–18.

Moulin, H., *Game Theory for the Social Sciences*, 2nd edition, New York: New York University Press, 1986.

Nachbar, J. H. "'Evolutionary' Selection Systems in Games: Convergence and Limit Properties", mimeo, 1989, the RAND corporation (forthcoming in the International Journal of Game Theory).

Roth, A., "Laboratory Experimentation in Economics: A Methodological Overview", *Economic Journal*, 1988, 98, 974–1031.

Samuelson, L., "Evolutionary Foundations of Solution Concepts for Finite Two–Player, Normal Form Games" in M. Y. Vardi ed., *Theoretical Aspects of Reasoning About Knowledge*, Los Altos: Morgan Kaufmann, 1988.

Schuster, P. and Sigmund, K., "Replicator Dynamics", *Journal of Theoretical Biology*, 1983, 100, 533–538.

Selten, R. and Stoecker, S., "End Behavior in Sequences of Finite Prisoners' Dilemma Supergames", *Journal of Economic Behavior and Organization*, 1986, 7, 47–70.