Center for Research on Economic and Social Theory
CREST Working Paper

Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake

Arthur J. Robson

1989
Number 89–22

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109

Efficiency in Evolutionary Games:
Darwin, Nash and the Secret Handshake

ARTHUR J. ROBSON

Department of Economics, University of Western Ontario, London, Ontario, Canada N6A 5C2

This paper considers any evolutionary game possessing a number of evolutionarily stable strategies, or ESSs, with differing payoffs. A mutant is introduced which will "destroy" any ESS which yields a lower payoff than another. This mutant possesses a costless signal and also conditions on the presence of this signal in each opponent. The mutant then can protect itself against a population playing an inefficient ESS by matching this against these nonsignallers. At the same time, the mutants can achieve the more efficient ESS against the signalling mutant population itself. This construction is illustrated by means of the simplest possible example, a coordination game. The one–shot prisoner's dilemma is used to illustrate how a superior outcome which is not induced by an ESS may be temporarily but not permanently attained. In the case of the repeated prisoner's dilemma, the present argument seems to render the "evolution of cooperation" ultimately inevitable.

RUNNING HEAD: EFFICIENCY IN EVOLUTIONARY GAMES

1. Introduction and Overview

The title "The Origin of Species" seems to suggest, at least to the unwary, that Darwin intended a group to be the unit of selection. The subtitle of Darwin's treatise reinforces this casual impression— "... or the Preservation of Favoured Races in the Struggle for Life". However, Darwin explicitly recognized that "the struggle for life (is) most severe between individuals and varieties of the same species". (Darwin, 1958, p. 84. Mayr, 1982, pp. 484–485, attributes Darwin's recognition of the importance of intra–species competition to his reading of Malthus, 1798. But see also pp. 491–493.) It is still not uncommon to hear popular explanations of animal behavior which rely upon an appeal to the interest of a species or, occasionally, even to the collective interest of several. A few modern biologists believe that there exist phenomena which cannot be explained without invoking a group selection mechanism. (See, for example, Wynne–Edwards, 1962.) Most other modern biologists, on the other hand, seem to find such explanations to be unattractive in the light of the logic of natural selection. Dawkins, (1976, pp. 8–9), for example, argues that natural selection operates most forcefully below the level of the group and, far from promoting the interest of the group, might entail its extinction. Indeed, Dawkins argues that the unit of selection is, in a certain sense, below even the level of the individual, is the gene. (See, in particular, 1982a, pp. 45–64).

Views of evolutionary biology which emphasize the role of the individual are, of course, highly congenial to economists. Further strengthening both the formal and substantive links between the two disciplines is that game theory has now also been applied to evolutionary biology. (See Maynard Smith and Price, 1973, and Maynard Smith, 1982.) Consider the interactions of the individuals of one particular species with one another. Suppose that these interactions involve two identical individuals at one time contesting some scarce resource, such as food. With a large population, the equilibrium outcome introduced by Maynard Smith and Price is a special kind of symmetric Nash equilibrium, involving what is designated an "evolutionarily stable strategy", or ESS. A population playing such a strategy is designed to be immune to invasion by a mutant population playing any alternative strategy. It need not be the case, of course, that such an ESS is unique, and indeed ESSs may have different payoffs. That is, an ESS can exist which yields an outcome which is inferior to that obtained under some other ESS. Such a situation has a well–known analogue within pure game theory. That is, it is often possible to Pareto rank some of the Nash equilibria of a game. A long–standing problem is how to construct a theory which predicts Pareto–efficiency within the set of Nash equilibria.

The intention of the present work is to suggest that Mother Nature might be less baffled by this problem than are game theorists. For there is a simple mechanism, based on the *individual* as the unit of selection, whereby such inefficient ESSs can be destroyed. That is, a mutation can be defined which will successfully invade a population which is playing any ESS which is payoff–dominated by another. This mutation must, of necessity, involve more than simply a different choice from the original set of strategies, for it is just these latter mutants which are considered in the definition of an ESS. Indeed, the mutation here entails the possession of a signal, that is, an observable characteristic which can be taken to have zero inherent cost. Mutants recognize the presence or absence of this signal in the other individual and condition their choice of strategy on this.

It should be emphasized that there is little doubt that animals actually use signals, for certain purposes at least. Thus, for example, Maynard Smith (1982, pp. 82–86) discusses the Harris sparrow, individuals of which vary in the color of their plumage and also in aggression towards other birds, with the darker birds being more aggressive. Dark birds painted pale continued to behave aggressively but were involved in a larger number of fights than normal dark birds. Pale birds painted dark tended to be persecuted by normal dark birds and were sometimes forced to feed away from the flock. It is not asserted that this will necessarily precisely fit the model here. What it does demonstrate is the reality of signalling as a natural phenomenon.

Consider then a population for some evolutionary game which is in a low–level trap, that is an ESS which is inferior to another ESS. The appropriate mutant uses the kind of signal discussed above as follows. Against the old non–signalling population the mutant plays the old inefficient EES, thus protecting itself from the consequences that would otherwise generally occur. On the other hand, mutants recognize the signal in other mutants and then can attain the more efficient ESS. It is assumed that the old non–signalling population remains blind to the signal. (This motivates the phrase the "secret handshake" of the subtitle.)

The argument can better be understood with the aid of the simplest possible example, an evolutionary game with two pure strategies, each of which is an ESS. This is presented in Section 3.1. The effect of the mutation is to add a third pure strategy which plays the inferior strategy against either of the original two strategies and the superior strategy against itself. Thus the entries in the new 3x3 matrix are derived in a straightforward way from the those in the original 2x2 matrix. It is then readily shown that the only ESSs now are the old superior strategy and the new mutant strategy, which are equivalent in payoffs. Furthermore, it is shown that the population must converge to one of these equivalent ESSs, given a generic initial point, and initial points near the original inefficient ESS converge in particular to the new efficient mutant ESS.

The analysis of the above example suggests that a similar mutant would be successful in invading a population playing some ESS if only there exists an *outcome* yielding a superior payoff. That is, the argument does not appear to use directly the assumption that the superior outcome be itself an ESS. In order to discuss this issue, the one–shot "prisoner's dilemma" is presented as a second example, in Section 3.2. It is noted firstly that, indeed, the addition of an appropriate mutant to the standard 2x2 prisoner's dilemma game results in a 3x3 game in which the group preferred "cooperative" outcome is generated by the unique ESS. Even if this is credible as a short–run outcome, there is an obvious objection to this in the long–run. That is, there is another mutant waiting impatiently in the wings, a mutant which avails itself of the signal and plays the old "fink" ESS against the old population, but also "finks" against the first mutant. This second mutant thus "finks" unconditionally. It is desirable, then, to consider the 4x4 game with both mutants. It is noted firstly that this 4x4 game does not, strictly speaking, possess any ESS. However, this is simply a technicality rather than a deep difficulty. It is shown, indeed, that the path of the population over time generally converges to some mixture between the old "fink" strategy and the mutant which signals but "finks" always anyway. The limiting payoff is uniquely determined as the usual payoff for the 2x2 game. In essence then, the conventional equilibrium prediction for the prisoner's dilemma is ultimately restored.

Section 2, following, defines the notion of an ESS and the dynamical system. Section 3, as noted above, discusses two key examples. Section 4 discusses related work and the implications of the present results. The Appendix shows how the analysis of the first example can be straightforwardly generalized whenever there are at least two ESSs with differing payoffs.

2. Definitions

Definition 1: Basic Evolutionary Game

There is a large (strictly infinite) population of individuals who interact two at a time. (Riley, 1979, discusses the complications needed in order to analyze small populations.) In each interaction, each individual can choose a pure strategy 1,...,n. The proportion of the population playing strategy i is denoted $x_i \geq 0$ where $\sum_{i=1}^{n} x_i = 1$ and so $x \in \Delta^{n-1}$, the (n–1) simplex in R^n. If a given individual chooses strategy i and his opponent chooses j, the payoff to the first individual is a_{ij} . The payoff to the second individual, given that the two individuals are identical, is therefore a_{ji} . (These payoffs measure "fitness" in the biological sense of determining the number of descendants, as is modelled explicitly in the dynamics below.) Given the symmetry between individuals, it is enough to specify $A = (a_{ij})$, a single nxn matrix.

The payoff to choice of strategy i against a population x is now given by

$$e_i^T Ax = (Ax)_i = \sum_{j=1}^{n} a_{ij}x_j$$

If a population y plays against a population x the payoff to the first population is then

$$y^T Ax$$

The notion of an "evolutionarily stable strategy", or ESS, then requires that the population x, say, be resistant to invasion by any population y. Suppose indeed that (1−ε) of the total population is x and ε is y, so that the total population is (1−ε)x + εy, where ε is small. The payoff to x is then

$$x^T A((1-\varepsilon)x + \varepsilon y) \;=\; (1-\varepsilon)x^T Ax + \varepsilon x^T Ay$$

and the payoff to y is

$$y^T A((1-\varepsilon)x + \varepsilon y) = (1-\varepsilon)y^T Ax + \varepsilon y^T Ay$$

This motivates the following:

Definition 2: Evolutionarily Stable Strategy, ESS

An ESS is $x \in \Delta^{n-1}$ such that, for all $y \in \Delta^{n-1}$, where $y \neq x$, *either*

(i) $y^T Ax < x^T Ax$

or (ii) $y^T Ax = x^T Ax$ *and* $y^T Ay < x^T Ay$.

The above conditions can be paraphrased in game theoretic terms as follows. It is required that the strategy x be a best–reply to itself and, further that any other best–reply to x, y, say, be a worse reply to y than is x. In particular, then, an ESS yields a symmetric Nash equilibrium. This can be described with just one strategy vector, x. The fractions x_i could be interpreted from this formal game theoretic point of view as probabilities of choosing i, so that x itself is then a mixed strategy. Not every symmetric mixed strategy Nash equilibrium induces an ESS, however, as taking A to be a matrix of zeroes shows. Indeed, this also shows that an ESS need not exist.

For some purposes it is useful to consider not only an appropriate static equilibrium concept, as above, but the evolution of the population over time. Thus:

Definition 3: Dynamical System

The fraction of the population playing a particular pure strategy is taken to evolve according to the difference between the "fitness" of this pure strategy and the average fitness of the population (Taylor and Jonker, 1978). Indeed, this yields a quite concrete interpretation of the payoff matrix, A. Thus,

$$\frac{\dot{x}_i}{x_i} = (Ax)_i - x^T Ax$$

or

$$\dot{x}_i = x_i[(Ax)_i - x^T Ax], \qquad i=1,\ldots,n$$

where the right–hand side is cubic in x. Clearly, if $x_i=0$, for any i, at t=0, then $x_i=0$ always. Furthermore, it is easily seen that every solution path remains on the unit simplex, given that it starts there. Hence the dimensionality of the system is n–1 rather than n.

A point $x \in \Delta^{n-1}$ is said to be a "point attractor" if it is the limit of the solution path of the above dynamical system for all initial values in some neighborhood of x. Zeeman (1979) shows that any ESS must be a point attractor, but that an attractor need not be an ESS. He recommends attractors as a more satisfactory equilibrium notion, when the dynamical system is given in the above pure strategy form. (Zeeman presents an example due to Hofbauer, Schuster, Sigmund and Wolff of a generic "Hopf bifurcation", but states that he does not know whether "strange attractors" and hence "chaos" can occur.) Hines (1987) indicates that being an ESS is both necessary and sufficient for being an attractor if each individual is permitted to use a mixed strategy and the appropriate corresponding dynamical system is considered.

3. Two Key Examples

3.1 A COORDINATION GAME.

This is, it seems, the simplest possible game yielding two ESSs which have different payoffs. It is given in Figure 1.

INSERT FIGURE 1 HERE.

It is clear that both "u" and "d" are ESSs and it is not difficult to see that there is no other ESS. (This follows also from Bishop and Cannings, 1978, p. 91.) Clearly the ESS "d", which yields a payoff of 2, is better for the population as a whole than is "u", which yields only a payoff of 1. Notice however that "u" is certainly immune to invasion by a small group of mutants playing "d". Indeed, suppose that the mutant "d" comprises a fraction ε of the total population with the remaining fraction $(1-\varepsilon)$ being still "u". In this case, each mutant obtains an average payoff of 2ε against the whole population, whereas the original "u" strategy yields $(1-\varepsilon)$. Thus the mutant will die out if $\varepsilon < 1/3$.

Now suppose that the mutation discussed in the introduction is introduced. This mutant carries a signal which is assumed to cost nothing to produce. Furthermore the mutant recognizes the presence of the signal in its opponent and conditions its choice of strategy on this. Suppose that the mutant here plays "u" against the non–signalling original population but plays "d" against other signalling mutants. The effect of this is to enlarge the original 2x2 game to the 3x3 game given in Figure 2.

INSERT FIGURE 2 HERE.

In Figure 2, the mutant signalling strategy is labelled "m". It should be emphasized that this enlarged 3×3 game still has just "u" and "d" as its underlying choices and that the payoff consequences of a given pair of these underlying choices also remain as in the 2×2 game.

It is easy to check that "u" is no longer an ESS, although (u,u) remains a Nash equilibrium. That is, "m" is also a best reply to "u" but "m" does better against itself than "u" does against "m". The only ESSs are now easily seen to be "d" and "m".

In order to more fully characterize the behavior of the population with these three strategies, the dynamic system derived from the above matrix, A, is treated. This is:

$$\frac{\dot{x}_1}{x_1} = x_1 + x_3 - W$$

$$\frac{\dot{x}_2}{x_2} = 2x_2 - W$$

$$\frac{\dot{x}_3}{x_3} = x_1 + 2x_3 - W$$

where, x_1, x_2, and x_3 are the components of the vector x, corresponding to "u", "d" and "m" respectively, and where,

$$W = x_1(x_1 + x_3) + 2x_2^2 + x_3(x_1 + 2x_3)$$

now denotes the average fitness of the entire population. Using the relation that $x_1 = (1 - x_2 - x_3)$ to eliminate x_1 and simplifying yields the following equations for x_2 and x_3:

$$\frac{\dot{x}_2}{x_2} = -1 + 4x_2 - 3x_2^2 - x_3^2$$

$$\frac{\dot{x}_3}{x_3} = x_2 + x_3 - 3x_2^2 - x_3^2$$

The phase diagram for these equations is readily derived and is sketched in Figure 3. (When there are two dimensions it is not possible for exotic behavior such as "chaos" to arise.)

INSERT FIGURE 3 HERE.

This diagram shows that generally initial points with strictly positive amounts of all three strategies have solution paths which tend to either "d" or "m", and there is a limiting payoff of 2 in all cases. All such initial points near "u" have solution paths which tend to "m" in particular. That is, the old inefficient ESS at "u" has, in this sense, been supplanted by a new efficient ESS at "m".

It should be noted that the above process is *not* reversible. That is, suppose instead the mutant plays the *efficient* ESS against non–signallers and plays the *inefficient* ESS against fellow signallers. It is easily shown that "u" an "d" remain the only ESSs of this augmented game. Hence "d", in particular, not destroyed as an ESS by the introduction of this mutant. Such a mutant indeed will die out.

3.2 THE ONE–SHOT PRISONER'S DILEMMA

This game is given as Figure 4.

INSERT FIGURE 4 HERE.

In this case it is easily seen that the unique ESS is "u", which yields a payoff of 2, despite the possibility of obtaining 3 by means of the entire population playing "d". The analysis of the previous example suggests that a mutant playing "u" against the old population and "d" against itself will be able to successfully invade a population playing "u". Indeed, consider Figure 5.

INSERT FIGURE 5 HERE.

The only ESS now is easily seen to be "m", which entails the payoff of 3. If this mutant and only this mutant were introduced, there is no reason to doubt the merit of this ESS. The difficulty is just that such an ESS is a "sitting duck" for the introduction of still another mutant, one which would prey on the first mutant. The second mutant should carry the signal, but play "u" against the first mutant as well as against the non–signalling population. (It would seem that this second mutant could evolve relatively easily from the first since all it involves is a switch in the underlying choice to be played against other signallers.) With the introduction of the second mutant in addition to the first the game is as in Figure 6.

INSERT FIGURE 6 HERE.

In Figure 6, the second mutant is labelled "f". It is not hard to see that there are now no ESSs, and hence the dynamical system associated with this matrix needs to be analyzed directly. This is a three–dimensional system on the tetrahedron, Δ^3. In general, three–dimensional systems can have markedly more complex behavior than that possible in two dimensions. However, the present example is rather simple. Notice, in fact, that the strategy "d" is unambiguously "dominated" not just by one of the other strategies but by all of them. (A strategy is said to be dominated by another strategy if it yields no more payoff for every choice of the other player and strictly less for some choice.) When strategy "d" is included in the dynamical system, it unambiguously decreases to zero along nontrivial solution paths. It can be shown that the limiting behavior of the full system is then determined by the limiting behavior of the two–dimensional system where "d" and its corresponding fraction "x_2", say, are simply omitted.

The following equations obtain in this case:

$$\frac{\dot{x}_1}{x_1} = 2x_1 + 2x_3 + 2x_4 - W$$

$$\frac{\dot{x}_3}{x_3} = 2x_1 + 3x_3 + x_4 - W$$

$$\frac{\dot{x}_4}{x_4} = 2x_1 + 4x_3 + 2x_4 - W$$

where x_1, x_3, and x_4, are the fractions of the population playing strategies "u", "m", and "f" respectively and where

$$W = x_1(2x_1 + 2x_3 + 2x_4) + x_3(2x_1 + 3x_3) + x_4(2x_1 + 4x_3 + 2x_4)$$

is the average fitness of the entire population. Using the fact that $x_1 = (1 - x_3 - x_4)$ to eliminate x_1 on the right–hand side, the system can be expressed as

$$\frac{\dot{x}_1}{x_1} = -(x_3 + x_4)x_3 \leq 0$$

$$\frac{\dot{x}_3}{x_3} = -x_4 + x_3(1 - x_3 - x_4)$$

$$\frac{\dot{x}_4}{x_4} = x_3 + x_3(1 - x_3 - x_4) \geq 0$$

The phase diagram for this essentially two–dimensional system in x_3 and x_4, say, is represented in Figure 7:

INSERT FIGURE 7 HERE

All solution paths of the dynamical system converge ultimately to a mixture of "u" and "f", and so have limiting payoff of 2 as in the usual equilibrium for the original 2x2 prisoner's dilemma game. However, the path takes a detour towards the vertex at which the first mutant "m" is the entire population and the average fitness is hence 3. Indeed, average fitness at first increases along a solution path but then decreases as the path heads back to a mixture of "u" and "f". Although the indeterminate nature of the mixture involved implies that there exists no ESS, that there exists indeed no point attractor, this is a technicality in that the payoff is determinate. Note that this example is clearly "non–generic" in that small *independent* perturbations of the payoffs in the 4x4 matrix here are likely to break the ties in the payoffs. However, these ties are produced endogenously by the signalling mutants and hence should be treated as ties.

4. Related Works, Implications

A large number of authors from several disciplines have considered how natural selection might be reconciled with a variety of alternative concepts of altruism. Trivers (1971) apparently coined the term "reciprocal altruism" to describe a process for attaining group efficient outcomes by means which essentially respect the individual as the unit of selection. Discussions of altruism, reciprocal or not, have become a central concern of sociobiology (See, for example, Wilson, 1975, p. 3).

Within biology, related kinds of signalling mutants to those here are discussed by Dawkins (1982b, Ch. 8). This comprises a theoretical discussion of how "outlaw genes" might promote their own survival at the expense of other closely related genes. Two of these hypothetical outlaw types discussed use signals to distinguish themselves from other genes. These two types are christened "armpits" and "green beards"! Within pure game theory, a related contribution is due to Matsui (1988). He presents an analysis of a two–player infinitely repeated game in which only Pareto–efficient outcomes can be equilibria. It is assumed that there is a small probability that one agent's entire "supergame" strategy is revealed to the other. The proof relies on a construction reminiscent of the signalling strategy here. For Matsui, of course, the two players are fully rational human beings. Finally, many of the ingredients used here can be found in Binmore (1988). He uses an evolutionary game argument to buttress the utilitarian outcome in a certain situation at the expense of the Nash bargaining one. His argument is reminiscent of that used here in Section 3.2 to analyze the one–shot prisoner's dilemma.

The game theoretic issue addressed here arises repeatedly in the previous literature. Dawkins (1976, pp. 197–202), for example, outlines a game–theoretic approach to mutual grooming. He finds two ESSs with differing payoffs, thus fitting the model here. Axelrod and Hamilton (1981) and Axelrod (1984) contain a detailed evolutionary game theoretic analysis of cooperation. (See also Maynard Smith, 1984, Ch. 13, for a summary of this.) Much of this work assumes that the two individuals play the prisoner's dilemma not once, but repeatedly, and there is some given probability of termination at each repetition. In this context, different pure strategies turn out to generate ties in payoffs in a manner which creates difficulties with the notion of an ESS as in Definition 2. Boyd (1989), however, shows that pure strategies can be reinstated as ESSs if individuals can make mistakes with certain probabilities. For such a "supergame", indeed, always finking remains an ESS in this extended sense. A modified version of the strategy "tit–for–tat", which cooperates initially but thereafter matches the opponents last move, can also be an ESS. If so it will yield a group preferred outcome. This model then essentially fits the framework developed here. In the original analysis the "evolution of cooperation" was hampered because "tit–for–tat" mutants were at a disadvantage playing against an initial population which always finked relative to this initial population itself. Thus Axelrod (1984, p. 175), for example, emphasized the need for geographical clustering of these mutants to provide a friendlier initial environment for themselves. The present analysis suggests how mutants might arise which are not at a disadvantage relative to the initial population, while still obtaining higher payoffs against one another. The "evolution of cooperation" would then be ultimately inevitable.

ACKNOWLEDGEMENT

I thank Tom Rutherford for computer programming assistance.

REFERENCES

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

Axelrod, R., & Hamilton, W.D. (1981). *Science* 211, 1390.

Binmore, K. (1988). University of Michigan, CREST Working Paper 89–05.

Bishop, D.T., and Cannings, C. (1978). *J. theor. Biol.* 70, 85.

Boyd, R. (1989), *J. theor. Biol.* 136, 47.

Darwin, C. (1958). *The Origin of Species.* New York: Mentor Books.

Dawkins, R. (1976). *The Selfish Gene.* Oxford: Oxford University Press.

Dawkins, R. (1982a). *Current Problems in Sociobiology.* Cambridge: Cambridge University Press.

Dawkins, R. (1982b). *The Extended Phenotype.* Oxford: Oxford University Press.

Hines, W.G.S. (1987). *Theor. pop. Biol.* 31, 195.

Malthus, T. R. (1798). *An Essay on Population.* London: J. Johnson.

Matsui, A. (1988). "Information Leakage Forces Cooperation", Northwestern University, MEDS: Preprint.

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

Maynard Smith, J. and Price, G.R. (1973). *Nature* 246, 15.

Mayr, E. (1982). *The Growth of Biological Thought.* Cambridge, Mass: Harvard University Press.

Riley, J.G. (1979). *J. theor. Biol.* 76, 109.

Taylor, P.D. and Jonker, L.B. (1978). *Math. Biosci.* 40, 409.

Trivers, R.L. (1971). *Quart. Rev. Biol.* 46, 35.

Van Damme, E. (1987), *Stability and Perfection of Nash Equilibria.* New York: Springer.

Wilson, E.O. (1975) *Sociobiology.* Cambridge, Mass: Harvard University Press.

Wynne–Edwards, V. C. (1962). *Animal Dispersion in Relation to Social Behaviour.* London: Oliver and Boyd.

Zeeman, E. C. (1979). *Global Theory of Dynamical Systems,* Lecture Notes in Math. **819.** New York: Springer.

APPENDIX: GENERAL CASE

As a minor matter of notation:

Definition 4: Support, Best-Replies

Consider an evolutionary game as in Definition 1. Suppose $x \in \Delta^{n-1}$, then define

$R(x) = \{i \in \{1,...,n\} \mid x_i > 0\}$ = support of x

and

$S(x) = \{i \in \{1,...,n\} \mid (Ax)_i = \max_j (Ax)_j\}$ = set of pure strategy best–replies to x.

Clearly, if x is an ESS as in Definition 2, $R(x) \subset S(x)$.

Suppose the following holds:

Assumption 1 Two EESs

Consider an evolutionary game as in Definition 1 with a nonzero number of ESSs as in Definition 2. (This number must be finite. See van Damme, 1987, p. 214). Suppose further that p is any ESS which yields the maximum over ESSs of average fitness, and that there is an ESS q with strictly lower average fitness:

$$q^T Aq < p^T Ap$$

It is desired, then, to introduce plausible mutants which destroy any such inefficient q as an ESS, or indeed, as an attractor of the pure–strategy dynamic. The obvious set of mutants to consider, perhaps, comprises each possible combination of a pure strategy to be played against non–signallers and a pure strategy to be played against fellow signallers. A difficulty with this approach is that it introduces a large number of ultimately irrelevant ties. Mutants which differ only in how they play the old non–signalling population will be indistinguishable when the old population dies out. Formally, then, it is not possible to obtain an ESS with the full set of these mutants. This would not seem to be an insuperable difficulty in that a slight generalization of the notion of an ESS is likely to fit the bill.

It is, however, more direct to simply hypothesize the following set of mutants.

Assumption 2 Form of Mutants

Suppose the evolutionary game given in Definition 1 is augmented by the introduction of n signalling mutants. Mutant i plays strategy i against fellow signallers, $i = 1,...,n$. Against non–signalling players, however, every mutant plays the inferior ESS q, as a *mixed strategy.*

Remarks

1. It is easy to see, given the proof below, that the number of mutants can be reduced to $|R(p)|$, that is, the number of elements in the support of p.
2. In the example of Section 3.1, both ESSs are in pure strategies, a special case in which $|R(p)| = |R(q)| = 1$.
3. The above form of mutation is particularly plausible in the case that the old ESS is played by a "monomorphic" population, each member of which then already uses q as a mixed strategy.

The effect of Assumption 2 is to augment the matrix A as follows

$$\underset{2n\times 2n}{\tilde{A}} = \begin{bmatrix} \underset{n\times n}{A} & \underset{n\times n}{B} \\ \underset{n\times n}{C} & \underset{n\times n}{D} \end{bmatrix}$$

where A is as in Definition 1. Now B = $(b_{i\ell})$ where

$$\begin{aligned} b_{i\ell} &= \text{payoff to old strategy } i \text{ against mutant } \ell \\ &= \sum_{j=1}^{n} a_{ij}q_j = (Aq)_i = b_i, \text{ say.} \end{aligned}$$

Furthermore, C = (c_{kj}) where

$$\begin{aligned} c_{kj} &= \text{payoff to mutant } k \text{ against old strategy } j \\ &= \sum_{i=1}^{n} a_{ij}q_i = (q^TA)_j = c_j, \text{ say.} \end{aligned}$$

Finally, D = $(d_{k\ell})$ where

$$\begin{aligned} d_{k\ell} &= \text{payoff to mutant } k \text{ against mutant } \ell \\ &= a_{k\ell}. \end{aligned}$$

Hence

$$\tilde{A} = \begin{bmatrix} A & b\ldots b \\ c^T & \\ \vdots & A \\ c^T & \end{bmatrix}, \text{ where } b = Aq, \text{ and } c^T = q^TA$$

As a notational convention, mutants will be numbered 1,...,n rather than n+1,,...,2n and 2n–vectors z, say, $z \in \Delta^{2n-1}$ will be written either as $[x^T,y^T]$ where x, $y \in R^{n+}$, or as $[\alpha r^T,(1-\alpha)s^T]$ where $r,s \in \Delta^{n-1}$, $\alpha \in [0,1]$.

The main result follows:

Theorem 1: Elimination of an Inferior ESS

Suppose an evolutionary game is described by $\tilde{A}$ as above, where Definition 1 applies to $\tilde{A}$. Now $[q^T, 0^T]$ is not an ESS of $\tilde{A}$. However, both $[p^T,0^T]$ and $[0^T,p^T]$ are ESSs of $\tilde{A}$. (ESSs are as in Definition 2.)

Proof

(a) $[q^T,0^T]$ is not an ESS of $\tilde{A}$.

Note that

$$[0^T,p^T]\tilde{A}\begin{bmatrix}q\\0\end{bmatrix} = [\,c^T,\,p^TA]\begin{bmatrix}q\\0\end{bmatrix} = c^Tq = q^TAq = [q^T,\,0]\tilde{A}\begin{bmatrix}q\\0\end{bmatrix}$$

That is, the mutant $[0^T,p^T]$ does exactly as well against $[q^T,0^T]$ as $[q^T,0^T]$ does. Further

$$[0^T,p^T]\tilde{A}\begin{bmatrix}0\\p\end{bmatrix} = [c^T,p^TA]\begin{bmatrix}0\\p\end{bmatrix} = p^TAp$$

whereas

$$[q^T,0]\tilde{A}\begin{bmatrix}0\\p\end{bmatrix} = [q^T,0]\begin{bmatrix}b\\Ap\end{bmatrix} = q^Tb = q^TAq < p^TAp$$

by Assumption 1. Thus the mutant $[0^T,p^T]$ does better against itself than $[q^T,0]$ does against this mutant. Hence $[q^T,0]$ is not an ESS of $\tilde{A}$.

(b) $[p^T,0^T]$ is an ESS of $\tilde{A}$. Note that firstly since p and q are both ESSs of A

$$q^TAp < p^TAp$$

that is, q cannot be a best–reply to p. For suppose

$$q^TAp = p^TAp$$

Then since p is an ESS,

$$q^TAq < p^TAq$$

which contradicts q being an ESS. (This result is implied by Bishop and Cannings, 1978, p. 91.) Hence, $\forall s \in \Delta^{n-1}$

$$[0^T,s^T]\tilde{A}\begin{bmatrix}p\\0\end{bmatrix} = [c^T,s^TA]\begin{bmatrix}p\\0\end{bmatrix} = c^Tp = q^TAp < p^TAp.$$

Also, of course, $\forall r \in \Delta^{n-1}$

$$[r^T,0^T]\,\tilde{A}\,\begin{bmatrix}p\\0\end{bmatrix} = r^TAp \leq p^TAp$$

since p is an ESS of A. Altogether, then, $\forall \alpha \in [0,1]$

$$[\alpha r^T,(1-\alpha)s^T]\tilde{A}\begin{bmatrix}p\\0\end{bmatrix} = \alpha r^TAp + (1-\alpha)q^TAp \leq p^TAp = [p^T,0^T]\tilde{A}\begin{bmatrix}p\\0\end{bmatrix}$$

and equality is only possible if $\alpha=1$. In this case, then,

$$r^TAp = p^TAp$$

Since p is an ESS of A

$$r^T A r < p^T A r$$

so that

$$[r^T, 0^T]\tilde{A}\begin{bmatrix} r \\ 0 \end{bmatrix} < [p^T, 0^T]\tilde{A}\begin{bmatrix} r \\ 0 \end{bmatrix}$$

exactly as required in order that $[p^T, 0^T]$ be an ESS of $\tilde{A}$.

(c) $[0^T, p^T]$ is also an ESS of $\tilde{A}$.

Note that, of course,

$$[0^T, p^T]\tilde{A}\begin{bmatrix} 0 \\ p \end{bmatrix} = p^T A p$$

Now $\forall$ r, s $\in \Delta^{n-1}$ and $\alpha \in [0,1]$,

$$[\alpha r^T, (1-\alpha)s^T]\tilde{A}\begin{bmatrix} 0 \\ p \end{bmatrix} = \alpha r^T b + (1-\alpha)s^T A p = \alpha r^T A q + (1-\alpha)s^T A p$$

where

$$r^T A q \leq q^T A q < p^T A p$$

since q is an ESS of A and by Assumption 1. Further

$$s^T A p \leq p^T A p$$

since p is an ESS of A. Hence

$$\alpha r^T A q + (1-\alpha)s^T A p \leq p^T A p$$

with equality implying α=0. In this case, then,

$$s^T A p = p^T A p$$

and since p is an ESS of A

$$s^T A s < p^T A s$$

so that

$$[0^T, s^T]\tilde{A}\begin{bmatrix} 0 \\ s \end{bmatrix} < [0^T, p^T]\tilde{A}\begin{bmatrix} 0 \\ s \end{bmatrix}$$

exactly as required for $[0^T, p^T]$ to be an ESS of $\tilde{A}$.

Remark

1. It is not difficult to check that the above result is, in a sense, robust to the choice of mixed strategy played by mutants against non–signallers. That is, if mutants use $\tilde{q}$ for this, where $R(\tilde{q}) \subset S(q)$, as in Definition 4, and $\tilde{q}$ is sufficiently close to q, then the Theorem remains true.

The coordination game example of Section 3.1 has more structure than that indicated by Theorem 1, related to the introduction of a dynamical system. The following generalization applies:

Theorem 2: Dynamical System of Augmented Game

Suppose the evolutionary game is given by $\tilde{A}$ as above. Consider the pure–strategy dynamical system as in Definition 3. Now $[q^T,0^T]$ is not even an attractor. Indeed, there is a path from $[q^T,0^T]$ leading to $[0^T,p^T]$ in this dynamical system.

Proof

Suppose the initial point is given as

$[\alpha q^T, (1-\alpha)p^T]$, where $\alpha \in (0,1)$.

The payoff to the old strategy i is $(Aq)_i$ because *all* other strategies play q against the old population. The payoff to mutant j is

$$\alpha q^T Aq + (1-\alpha)(Ap)_j$$

because it obtains the ESS payoff for q a fraction α of the time and plays the mutant population p the remaining fraction of the time. Hence the average payoff overall is

$$\alpha(q^T Aq) + (1-\alpha)[\alpha q^T Aq + (1-\alpha)p^T Ap] = \alpha(2-\alpha)q^T Aq + (1-\alpha)^2 p^T Ap$$

It follows the pure strategy dynamic for $[x^T, y^T]$, as in Definition 3, is given by

$$\frac{\dot{x}_i}{x_i} = q^TAq-\alpha(2-\alpha)q^TAq-(1-\alpha)^2p^TAp = -(1-\alpha)^2[p^TAp - q^TAq] < 0$$

for all $i \in R(q)$. If $i \notin R(q)$ then, of course, $x_i \equiv 0$. Similarly,

$$\frac{\dot{y}_j}{y_j} = \alpha qTAq + (1-\alpha)\; p^TAp - \alpha(2-\alpha)q^TAq - (1-\alpha)^2p^TAp = \alpha(1-\alpha)[p^TAp - q^TAq] > 0$$

for all $j \in R(p)$. If $j \notin R(p)$ then $y_j \equiv 0$. Clearly then the solution is given by

$$x(t) = \alpha(t)q \qquad\qquad y(t) = (1-\alpha(t))p$$

where

$$\frac{\dot{\alpha}(t)}{\alpha(t)} = -(1-\alpha(t))^2[p^TAp - q^TAq] < 0$$

and $\alpha(0) = \alpha$. It follows that $\alpha(t) \to 0$ as $t \to \infty$, that is,

$$[x^T(t), y^T(t)] \to [0,p^T]$$

no matter how close $\alpha(0) = \alpha$ was initially to 1, as was to be shown.

Remark

1. The ESS of $\tilde{A}$, $[0^T,p^T]$, must be an attractor in this dynamical system. Its "basin of attraction" is then an *open* set which includes the above path. (See Zeeman, 1979). That is, ultimate convergence to $[0^T,p^T]$ is guaranteed if the initial point is sufficiently close to the above path.

2 1	u	d
u	1	0
d	0	2

Figure 1. A Coordination Game.

2 1	u	d	m
u	1	0	1
d	0	2	0
m	1	0	2

Figure 2. The Coordination Game with a Signalling Mutant

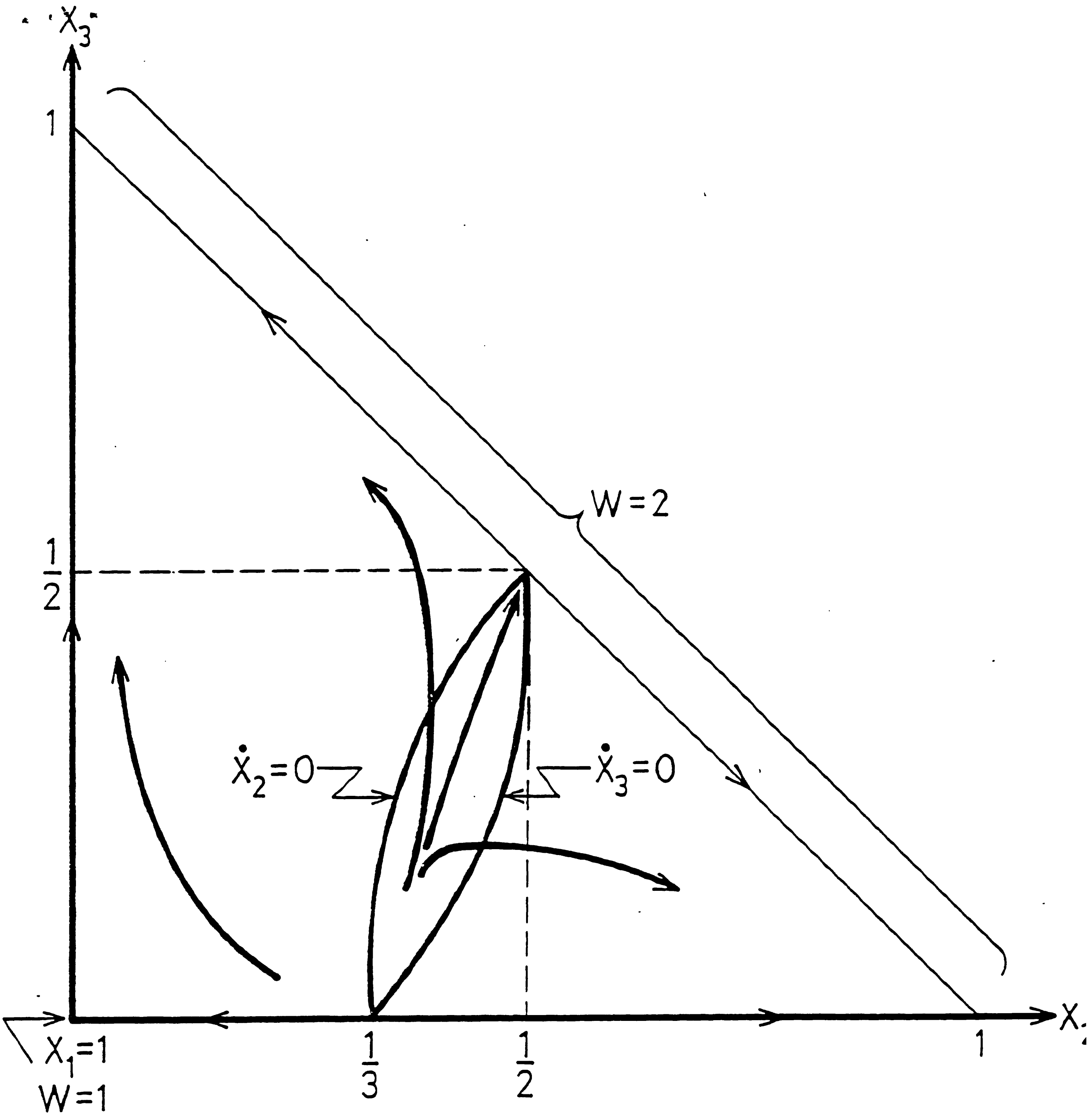

Figure 3. Phase Diagram for the Coordination Game with Mutant

2 1	u	d
u	2	4
d	1	3

Figure 4. One–Shot Prisoner's Dilemma.

2 1	u	d	m
u	2	4	2
d	1	3	1
m	2	4	3

Figure 5. One–Shot Prisoner's Dilemma with Mutant

2 1	u	d	m	f
u	2	4	2	2
d	1	3	1	1
m	2	4	3	1
f	2	4	4	2

Figure 6. One–Shot Prisoner's Dilemma with Two Mutants.

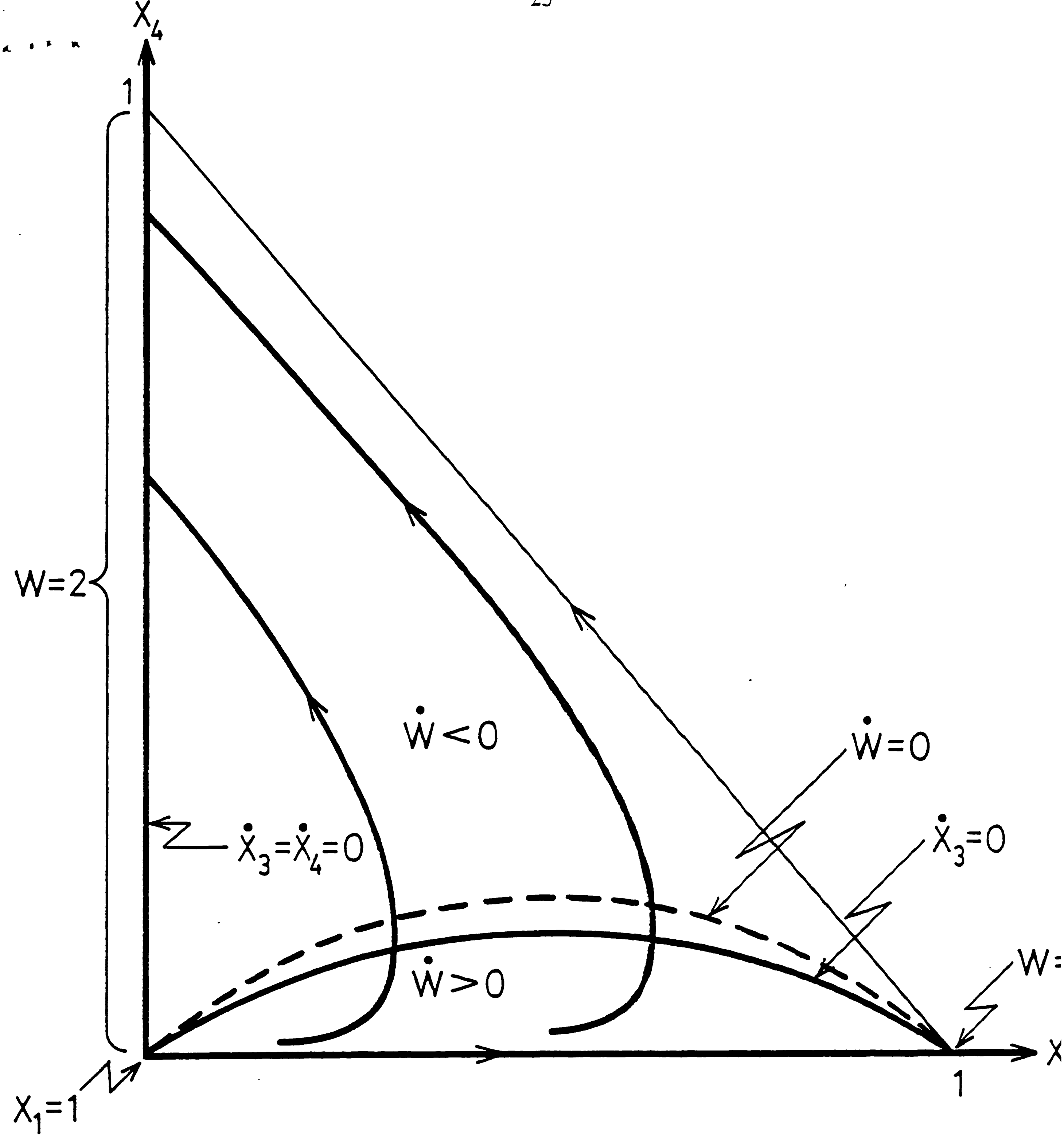

Figure 7. Phase Diagram for Reduced Form Prisoner's Dilemma with Two Mutants

Recent CREST Working Papers

89–01: Mark Bagnoli, Severin Borenstein, "Carrot and Yardstick Regulation: Enhancing Market Performance with Output Prizes," October, 1988.

89–02: Ted Bergstrom, Jeffrey K. MacKie–Mason, "Some Simple Analytics of Peak-Load Pricing," October, 1988.

89–03: Ken Binmore, "Social Contract I: Harsanyi and Rawls," June, 1988.

89–04: Ken Binmore, "Social Contract II: Gauthier and Nash," June, 1988.

89–05: Ken Binmore, "Social Contract III: Evolution and Utilitarianism," June, 1988.

89–06: Ken Binmore, Adam Brandenburger, "Common Knowledge and Game Theory," July, 1989.

89–07: Jeffrey A. Miron, "A Cross Country Comparison of Seasonal Cycles and Business Cycles," November, 1988.

89–08: Jeffrey A. Miron, "The Founding of the Fed and the Destabilization of the Post-1914 Economy," August, 1988.

89–09: Gérard Gaudet, Stephen W. Salant, "The Profitability of Exogenous Output Contractions: A Comparative-Static Analysis with Application to Strikes, Mergers and Export Subsidies," July, 1988.

89–10: Gérard Gaudet, Stephen W. Salant, "Uniqueness of Cournot Equilibrium: New Results from Old Methods," August, 1988.

89–11: Hal R. Varian, "Goodness-of-fit in Demand Analysis," September, 1988.

89–12: Michelle J. White, "Legal Complexity," October, 1988.

89–13: Michelle J. White, "An Empirical Test of the Efficiency of Liability Rules in Accident Law," November, 1988.

89–14: Carl P. Simon, "Some Fine-Tuning for Dominant Diagonal Matrices," July, 1988.

89–15: Ken Binmore, Peter Morgan, "Do People Exploit Their Bargaining Power? An Experimental Study," January, 1989.

89–16: James A. Levinsohn, Jeffrey K. MacKie–Mason, "A Simple, Consistent Estimator for Disturbance Components in Financial Models," April 25, 1989.

89–17: Hal R. Varian, "Sequential Provision of Public Goods," July, 1989.

89–18: Hal R. Varian, "Monitoring Agents with Other Agents," June, 1989.

89–19: Robert C. Feenstra, James A. Levinsohn, "Distance, Demand, and Oligopoly Pricing," July 17, 1989.

89–20: Mark Bagnoli, Shaul Ben-David, Michael McKee, "Voluntary Provision of Public Goods," August, 1989.

89–21: N. Gregory Mankiw, David Romer, Matthew D. Shapiro, "Stock Market Forecastability and Volatility: A Statistical Appraisal," August, 1989.

89–22: Arthur J. Robson, "Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake," 1989.

89–23: Mark Bagnoli, Ted Bergstrom, "Log-Concave Probability and Its Applications," September 7, 1989.

89–24: Gérard Gaudet, Stephen W. Salant, "Towards a Theory of Horizontal Mergers," July, 1989.

89–25 (evolved from 87-35): Stephen W. Salant, Eban Goodstein, "Predicting Committee Behavior in Majority-Rule Voting Experiments," July, 1989.

89–26: Ken Binmore; Martin J. Osborne, Ariel Rubinstein, "Noncooperative Models of Bargaining," 1989.

89–27: Avery Katz, "Your Terms or Mine? The Duty to Read the Fine Print in Contracts," February 19, 1989.