

An Evolutionarily Stable Strategy May Be Inaccessible

MARTIN NOWAK†

Institut für Theoretische Chemie der Universität Wien Währingerstr. 17, A-1090 Wien, Austria und Institut für Mathematik der Universität Wien, Austria

(Received on 18 July 1989, Accepted in revised form 10 October 1989)

The definition of an evolutionarily stable strategy (ESS) does not specify whether or not a population will evolve towards such a strategy.

If all members of a population use an evolutionarily stable strategy (Maynard Smith, 1972), then no mutant strategies can invade. However it is a common but misleading conclusion that evolution will tend towards such a strategy.

As Lessard (1988) and Uyenoyama & Bengtsson (1982) point out there is ambiguity in ESS approaches, because an ESS is often misunderstood as an ever invading strategy. This seems to reflect the double origin of this notion as Hamilton's (1967) "unbeatable" strategy and Maynard Smith's ESS concept. An "unbeatable" strategy has a selective advantage over any other strategy at all frequencies of the competing strategy. An ESS has a selective advantage only when the other strategy occurs in low frequency.

Several authors have investigated the question, whether genetic constraints will hinder a sexual reproducing population to evolve towards an ESS (Maynard Smith, 1981; Rocklin & Oster, 1976; Thomas, 1985*a,b*, see also Hines, 1987). In the various models dealing with the evolution of the sex ratio, it has been stressed that new mutations affecting the ratio will succeed only if they bring it closer to the evolutionarily stable ratio (Eshel & Feldman, 1982; Karlin & Lessard, 1984).

In this paper we emphasize that the same problems might appear even in the basic framework of evolutionary game theory: Will an asexual population evolve to an ESS in any case? In other words: Is an ESS equivalent to a strategy, towards which evolution will occur as a result of successive mutations?

In the context of evolutionary game theory the payoff for a strategy is related to the increase in fitness. Hence successful strategies will produce more offspring and thus spread in the population. Let us denote by $A(E, E')$ the payoff for strategy E in the contest with E' . In a homogeneous population adopting strategy E , each individual has—up to some constant—the "fitness" $A(E, E)$. A mutant strategy E' can invade this population by selection pressure, if $A(E', E) > A(E, E)$. Mutant strategies that fulfill the equation $A(E', E) = A(E, E)$ can invade by selection, if

† Current address: Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

$A(E', E') > A(E, E')$. If additionally this equality $A(E', E') = A(E, E')$ holds, they may spread in the population due to random drift. These considerations imply the following definition (Maynard Smith, 1972, 1982):

Definition: \hat{E} is said to be an evolutionarily stable strategy (ESS), if

$$A(E, \hat{E}) < A(\hat{E}, \hat{E})$$

or

$$A(E, \hat{E}) = A(\hat{E}, \hat{E}) \quad \text{and} \quad A(\hat{E}, E) > A(E, E)$$

for all $E \neq \hat{E}$.

Hence, if an infinite homogeneous population adopts strategy \hat{E} , then no single mutant strategy $E \neq \hat{E}$ can invade under the action of natural selection.

This explains how such a strategy is able to maintain itself, but not how it can get established. This can be quite a problem.

An illustrative example for an "inaccessible" ESS is found in the Prisoner's Dilemma game. In repeated encounters, two players are faced with the choice to co-operate or to defect (C or D). If both co-operate, their payoff R (reward) is higher than the payoff P (punishment) obtained if both defect. But if one player defects while the other co-operates, then the defector's payoff T (temptation) is higher than R , while the co-operator's payoff S (sucker) is smaller than P . Hence $T > R > P > S$. It is furthermore assumed that $R > \frac{1}{2}(S + T)$, so that joint co-operation is more profitable than alternating C and D . If the game consists of a single encounter, the best option is to defect, no matter what the other player does. But if the game is repeated an unknown number of times, the players may "learn" that it is in their interest to co-operate (Axelrod, 1984; Axelrod & Hamilton, 1981).

For the following arguments we consider the infinitely iterated Prisoner's Dilemma and restrict attention to "reactive strategies" given by the parameters p and q , which represent the conditional probabilities to co-operate, given that the adversary's last move was a C or a D respectively (Nowak & Sigmund, 1989a,b,c).

The payoff for strategy $E = (p, q)$ in the contest with $E' = (p', q')$ can be evaluated as;

$$A(E, E') = (R - S - T + P)ss' + (S - P)s + (T - P)s' + P$$

where,

$$s = \frac{q'(p - q) + q}{1 - (p - q)(p' - q')}, \quad s' = \frac{q(p' - q') + q'}{1 - (p - q)(p' - q')}$$

for $(p - q)(p' - q') < 1$.

The complete analysis of this game was carried out by Nowak (1990).

At first we assume that the parameter p is somehow evolutionarily fixed and that variation can only occur in q . If $R - T - S + P$ is negative, then there exists a value \hat{q} such that $A(\hat{q}, \hat{q}) > A(q, \hat{q})$ for all $q \neq \hat{q}$, which means that \hat{q} is a strict Nash equilibrium and in particular an ESS. Every population adopting strategy $q > \hat{q}$ can be invaded only by mutants $q' < q$, on the other hand strategies $q < \hat{q}$ can be invaded

only by mutants $q' > q$. There is always some tendency to approach the ESS \hat{q} . In this case \hat{q} might be called an "attracting" strategy, because only those mutants are successful that render the value q closer to \hat{q} . If $R - T - S + P$ is positive the tendency towards \hat{q} remains, although \hat{q} is no longer an ESS in this case. More precisely, \hat{q} can be invaded by any mutant strategy, but it remains attracting. We conclude that an attracting strategy need not be an ESS.

But if we consider the game with a constant probability q and variations only in the parameter p , the behaviour is opposite. If $R - T - S + P$ is negative and the fixed value q is sufficiently small but positive (precisely if $0 < q < 1 - (T - R)/(R - S)$), there exists a value \hat{p} between 0 and 1, which represents an ESS. If the whole population uses strategy \hat{p} , then no mutant strategy p can invade. However, there is no way for \hat{p} to become established in the first place. A homogeneous population using strategy p can be invaded only by mutants $p' < p$ if $p < \hat{p}$ and by mutants $p' > p$ if $p > \hat{p}$. Hence, only those mutants can invade that drive the value p away from the ESS \hat{p} . We conclude that an ESS need not be attracting. Note that no stable polymorphisms can exist and no homogeneous population can be invaded by the ESS \hat{p} . Therefore the ESS \hat{p} is inaccessible.

All these phenomena can be found in a more general framework, independent of the Prisoner's Dilemma. Let us consider the payoff function;

$$A(x, y) = ax^2 + bxy + cy^2 + dx + ey + f, \quad (1)$$

where a, b, c, d, e, f are some values given by the rules of the game. The strategies are described by the continuous parameters x and y , which may represent probabilities, hence $x, y \in [0, 1]$. If $a < 0$ then the strategy $\hat{x} = -d/(2a + b)$ is an ESS. It is attracting, if $2a + b < 0$, but inaccessible if $2a + b > 0$ (in this case the strategies $x = 0$ and $x = 1$ are ESSs, too). If $a > 0$ then \hat{x} can be invaded by any mutant strategy $x \neq \hat{x}$, but it remains attracting as long as $2a + b < 0$.

We are faced with two paradoxical examples: An ESS that is inaccessible and an attracting strategy that can be invaded by any mutant. Let us present a precise definition of "attracting":

Definition: A strategy x is said to be attracting if for all other strategies $y \neq x$ a value $\delta > 0$ exists such that for any strategy y' with $|y - y'| < \delta$, $A(y', y) > A(y, y)$ if and only if $|y' - x| < |y - x|$.

This definition is motivated by the following idea: An almost homogeneous population y is sprinkled by slightly different mutants y' that explore the strategy space. The population will evolve under selection in the direction which seems most promising. Thus the value y will tend towards the attracting strategy x , regardless of whether x is an ESS or not. An attracting strategy fulfills the m -stability condition of Taylor (1989).

An attracting ESS is a global version of Eshel's (1983) continuously stable strategy. However our examples show that ESSs need not be attracting, not even locally. Further examples can be found in Eshel (1983) and Grafen (1983). In addition an ESS might even be inaccessible if it cannot invade any homogeneous population or any stable polymorphism. Therefore the ESS condition is not sufficient for a strategy to provide a final outcome of an evolutionary game.

If an attracting strategy x can be invaded by a mutant y , then small deviations from y will shift the value back to x . However, there is a way to escape: If two mutants y_1 and y_2 arise, such that $y_1 < x < y_2$, a stable polymorphism between y_1 and y_2 might result. For the payoff function (1) this dimorphism can be invaded only by mutants y_3 , such that either $y_3 < y_1$ or $y_3 > y_2$. The final outcome is a polymorphism between $y_1 = 0$ and $y_2 = 1$, which is evolutionarily stable. But this represents a genetically polymorphic population that is completely different from a homogeneous population adopting an ESS.

The payoff function (1) also includes the Hawk Dove game, if we choose $a = 0$, $b < 0$, $c = 0$, and $d > 0$. The parameter x denotes then the probability to play Hawk. We know that $\hat{x} = -d/b$ is ESS and attracting. As long as the payoff function is linear in x and y the results evaluated in the context with the game dynamical equation (Taylor & Jonker, 1979; Hofbauer & Sigmund, 1988) can be applied. In this dynamics every ESS must be asymptotically stable (Zeeman, 1979; Hofbauer *et al.*, 1979) and hence attracting in our sense.

Using the basic ideas of evolutionary game theory we have shown that evolutionarily stability does not imply that such a strategy will tend to evolve. A case by case analysis is needed to decide whether or not an ESS has some chance to become ever established in the population.

I would like to thank Karl Sigmund and John Maynard Smith for helpful discussions. Financial support from the Austrian Forschungsförderungsfond project P6866 is gratefully acknowledged.

REFERENCES

- AXELROD, R. (1984). *The Evolution of Co-operation*, New York: Basic Books.
- AXELROD, R. & HAMILTON, W. D. (1981). The evolution of co-operation. *Science* **211**, 1390–1396.
- ESHEL, I. & FELDMAN, M. (1982). On evolutionary genetic stability of the sex ratio. *Theor. pop. Biol.* **21**, 430–439.
- ESHEL, I. (1983). Evolutionary and continuous stability. *J. theor. Biol.* **103**, 99–111.
- GRAFEN, A. (1983). *The economics of evolutionary stability*. Doctoral thesis, University of Oxford.
- HAMILTON, W. D. (1967). Extraordinary sex ratios. *Science, Lond.* **165**, 477–488.
- HINES, W. G. S. (1987). Evolutionary stable strategies: A review of basic theory. *Theor. pop. Biol.* **31**, 195–272.
- HOFBAUER, J., SCHUSTER, P. & SIGMUND, K. (1979). A note on evolutionarily stable strategies and game dynamics. *J. theor. Biol.* **81**, 609–612.
- HOFBAUER, J. & SIGMUND, K. (1988). *Dynamical Systems and the Theory of Evolution*. Cambridge: University Press.
- KARLIN, S. & LESSARD, S. (1984). On the optimal sex-ratio: a stability analysis based on a characterization for one-locus multiallele viability models *J. math. Biol.* **20**, 15–38.
- LESSARD, S. (1988). Resource allocation in mendelian populations: further in ESS theory. In: *Mathematical Evolutionary Theory* (Feldman, M., ed.) Princeton: University Press.
- MAYNARD SMITH, J. (1972). *On Evolution*. Edinburgh: University Press.
- MAYNARD SMITH, J. (1981). Will a sexual population evolve to an ESS? *Am. Nat.* **117**, 1015–1018.
- MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games*. Cambridge: University Press.
- NOWAK, M. (1990). Stochastic strategies in the Prisoner's Dilemma. *Theor. pop. Biol.* **37**, in press.
- NOWAK, M. & SIGMUND, K. (1989a). Oscillations in the evolution of reciprocity. *J. theor. Biol.* **137**, 21–26.
- NOWAK, M. & SIGMUND, K. (1989b). Game dynamical aspects of the Prisoner's Dilemma *J. appl. Math. Comp.* **30**, 191–213.
- NOWAK, M. & SIGMUND, K. (1989c). The Evolution of Reactive Strategies in Iterated Games, preprint.
- ROCKLIN, S. & OSTER, G. (1976). Competition between phenotypes. *J. math. Biol.* **3**, 225–261.

- TAYLOR, P. (1989). Evolutionary stability in one parameter models under weak selection. *Theor. pop. Biol.* **36**, 125-193.
- TAYLOR, P. & JONKER, L. (1979). Evolutionarily stable strategies and game dynamics. *Math. Biosc.* **40**, 145-156.
- THOMAS, B. (1985*a*). Genetical ESS-models. I. Concepts and basic model. *Theor. pop. Biol.* **28**, 18-32.
- THOMAS, B. (1985*b*). Genetical ESS-models. II. Multi strategy models and multiple alleles. *Theor. pop. Biol.* **28**, 33-49.
- UYENOYAMA, M. K. & BENGTTSSON B. (1982). Towards a genetic theory for the evolution of the sex ratio. *Theor. pop. Biol.* **22**, 43-68.
- ZEEMAN, E. C. (1979). Population Dynamics from Game Theory. In: *Proceeding of Conference on Global Theory of Dynamical Systems*, Northwestern, 471-497.