Evolution, Dynamics, and Rationality: The Limits of ESS Methodology

Simon M. Huttegger Kevin J. S. Zollman*

September 16, 2011

Abstract

In this paper we show that there are certain limits as to what applications of Maynard Smith's concept of evolutionarily stable strategy (ESS) can tell us about evolutionary processes. We shall argue that ESS is very similar in spirit to a particular branch of rational choice game theory, namely, the literature on refinements of Nash equilibrium. In the first place, ESS can also be viewed as a Nash equilibrium refinement. At a deeper level, ESS shares a common structure with other rational choice equilibrium refinements. An equilibrium is evaluated according to whether it persists under specific kinds of perturbations. In the case of ESS, these perturbations are mutations. However, from a dynamical point of view, focusing exclusively on perturbations of equilibria provides only a partial account of the system under consideration. We will show that this has important consequences when it comes to analyzing game-theoretic models of evolutionary processes. In particular, there are non-ESS states which are significant for evolutionary dynamics.

1 Introduction

Until quite recently, the study of games has developed almost independently in both economics and biology. In both fields, however, focus has been on developing *equilibrium concepts* – stable states that should be expected to emerge as a result of evolution by natural selection acting on populations of organisms (in biological game theory) or from the "rational" choices of individuals (in economics). While one could formalize both the dynamics of natural selection and notions of rationality, the history of game theory has relied more on intuitive characterizations of both rather than mathematically precise characterizations.

*This material is based upon work supported by the National Science Foundation under Grant No. EF 1038456. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

As an example, consider the canonical equilibrium concept, Nash equilibrium. In economics it is argued that if a player expects the other players to play a Nash equilibrium, then this player does best by playing her part of the Nash equilbrium. A Nash equilibrium thus represents a kind of stable state of the behavior and expectations of rational individuals. What comprises these beliefs and how they came about is often left to informal discussion. In biology, a population where all individuals play a Nash equilibrium is similarly stable, since each individual is doing the best she can given what the others are doing, it is assumed that natural selection will not imeadiately (without perturbation) take the state away from the Nash equilibrium. Again, the process by which natural selection operates, and the process by which a population comes to play a Nash equilibrium is left for informal discussion.

The absence of mathematically precise foundations was not due to ignorance or ambivalence but was instead an attempt at generality. It was believed that the intuitive notions on which the fields relied were so general as to be a feature of any formal characterization of either natural selection or rationality respectively. The result of this intuitive reflection was believed to be robust across many different potentially inconsistent precise specifications of the underlying notions. Because of this desire for robustness and also because of the difficulty of analyzing epistemic models, the study of game theory in economics proceeded by developing more and more sophisticated equilibrium notions creating a large scale project called the "equilibrium refinement" program.

More recently in economics, significant attention has been given to the foundations of equilibrium concepts used there. The notions of belief, that had previously been left informal, have been explicitly represented utilizing a variety of models. This "epistemic program" in economics focuses on formally explicating the notion of rationality and uncovers when rational individuals will indeed come to play equilibria (see, e.g. Brandenburger, 2007). In addition, many scholars have utilized models of bounded-rationality which attempt to determine the robustness of equilibrium notions to relaxing even the intuitive notion of rationality initially used to develop the equilibrium concepts.[1]

This foundational turn has yet to take hold in biology. In biology, the predominate equilibrium concept is the notion of an *evolutionarily stable strategy* (ESS) as introduced in Maynard Smith and Price (1973). The concept of an ESS is also a refinement of Nash equilibrium which better represents a state which cannot be eliminated by natural selection with the introduction of small perturbations. As with equilibrium refinements in economics, ESS provides a shortcut to analyzing the specific properties of the dynamics of evolution by natural selection; in fact, historically it precedes the introduction of dynamical systems to evolutionary game theory (Taylor and Jonker, 1978). ESS allows one to infer the stability of certain population states without

[1]It should be noted that John Nash justified his concept using this notion of learning among boundedly rational agents. However, this justification was largely lost until this methodology became popular much later.

recourse to some specific underlying dynamics. The claim is that the stability of ESSs will hold across a large range of evolutionary dynamics. Robustness comes for free.

ESS, therefore, is a powerful tool to analyze evolutionary phenomena. Though it may be tricky to find ESSs in particular cases, it is generally much easier than studying an underlying dynamical system. Furthermore, ESS goes beyond the claim of stability for a particular dynamical system. That is, ESS allows one to make a more general claim than mere dynamic stability. Thus, methods based on ESS appear to be more tractable and seem to deliver more general results than is possible by investigating stability in evolutionary dynamical systems. For these reasons, ESS is a much used concept in the biological literature on evolutionary games.

In this paper, we shall argue that in many relevant cases there are non ESS states that are nevertheless evolutionarily significant (Section 4). This implies that attempting to find ESSs will often not be sufficient to obtain an adequate understanding of an evolutionary process. In order to make this case, we shall distinguish between ESS as a concept and ESS as a methodology (Section 2). By the latter we roughly mean the maxim that finding the ESSs of the game allows one to derive all, or nearly all, relevant conclusions about the game in question.

The second topic of this paper is to relate the ESS concept to certain equilibrium concepts of economic game theory (Section 3). Like ESS, these equilibrium concepts are refinements of Nash equilibria. Our distinction between the concept itself and the methodology based on it (that its implications are the only relevant ones) carries over to the case of the other refinements. We conclude that, both in the case of ESS and in the case of other Nash equilibrium refinements, the methodology fails in important cases.

2 ESS and ESS methodology

We start by briefly reviewing the ESS concept (Maynard Smith and Price, 1973; Maynard Smith, 1982). Suppose a symmetric finite two-player game Γ is being played repeatedly in a panmictic population; i.e., the population is effectively infinite, interactions are happening at random (there is no population structure) and are taking place between two individuals whose roles in the game cannot be distinguished. Let $u(s, s')$ denote the payoff strategy s obtains when interacting with strategy s' (payoff may be understood as incremental fitness). If s^* is a strategy of Γ, then s^* is evolutionarily stable if for all strategies s of Γ other than s^* the following two conditions hold:

1. $u(s^*, s^*) \geq u(s, s^*)$

2. If $u(s^*, s^*) = u(s, s^*)$, then $u(s^*, s) > u(s, s)$

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

Figure 1: A Prisoner's Dilemma

The first of these conditions states that s^* has to be a symmetric Nash equilibrium (NE). The second condition requires s^* to obtain a higher payoff when interacting with s than s gets when interacting with itself whenever s is an alternative best response to s^*.

As an illustration consider the Prisoner's Dilemma, pictured in Figure 1. In this game, strategy D is an ESS, because it is a Nash equilibrium (satisfying the first condition) and because all other strategies do worse against D than D does against itself (rendering the second condition void). C is not an ESS because it is not a Nash equilibrium. There are many other applications of ESS found in the biological literature begining with the Hawk-Dove game (Maynard Smith and Price, 1973). Finding ESS in larger games can often be a non-trivial matter.

In an evolutionary setting ESS can be understood the following way. Suppose the incumbent strategy s^* is being played by the whole population. If a small number of mutants playing s arises and condition (1) does not obtain, then clearly the population state s^* cannot be stable for s will take over the population when it is given enough time. If, on the other hand, (1) obtains with a strict inequality sign, then s will not be able to displace the incumbent s^*. Condition (2) takes care of the third possibility, where $u(s^*, s^*) = u(s, s^*)$. Here, interactions with the incumbents have no effect on the population. Thus, for s^* to be stable in this case it needs to get a higher payoff against the mutant than the mutant gets when interacting with itself.

The ESS concept thus readily lends itself to an evolutionary interpretation. Formally, ESS is a refinement of NE. Condition (1) requires an ESS to be a NE, and condition (2) is not fulfilled by all NE. ESSs are attractive candidates for evolutionary outcomes because they are dynamically stable in several important evolutionary dynamics. For instance, in the replicator dynamics, a strategy i's per capita growth rate in a panmictic population is given by the difference of its fitness and the average fitness in the population:

$$\dot{x}_i = x_i(f_i - \bar{f}) \tag{1}$$

(f_i and $\bar{f}$ depend linearly on the payoffs of the game Γ; see Hofbauer and Sigmund (1998) for details.) This implies that strategies with above average fitness increase in frequency while strategies with below average fitness decrease in frequency. It is well known that an ESS is an asymptotically stable state of the replicator dynamics. This means that populations where a sufficiently high frequency of individuals already play the ESS will tend toward a state where all individuals will do so. It should be emphasized that this does not only hold for the replicator dynamics but for larger classes of dynamics that contain the replicator dynamics. Let us just mention monotone selection dynamics (Hofbauer and Sigmund, 1998). Monotone selection dynamics are given by a system of equations

$$\dot{x}_i = x_i g_i(x)$$

where x is the state of the population. For this system to be a monotone selection dynamics it is required that $g_i(x) > g_j(x)$ if, and only if, $f_i > f_j$; i.e., i's growth rate is higher than j's growth rate if, and only if, i's fitness is higher than j's fitness. It is well known that asymptotic stability of ESSs will also hold for monotone selection dynamics (Weibull, 1995).

There is one important aspect of ESSs that we have not mentioned so far. The ESS concept does not just apply to the pure strategies of a game as presented above. It can be straightforwardly extended to mixed strategies. There is a certain ambiguity concerning mixed strategy ESSs that won't concern us here (the ambiguity has to do with whether one identifies a mixed strategy ESS as a population state or as a strategy being played by the population).

For our purposes it is more important that in many kinds of interactions ESSs can be identified with the strict Nash equilibria of a game. To be more specific, whenever we start with a two-player asymmetric game (i.e., a game where the two player roles are distinguishable) we may consider the *symmetrized* game based on it (Cressman, 2003). In the symmetrized game each player is assumed to be in one of the two player positions of the asymmetric game with equal probability. By taking the payoffs to be expected values, the resulting game is a symmetric game to which the ESS concept can be applied. It turns out that there is a simple relationship between the strict Nash equilibria of the asymmetric game and the ESSs of the corresponding symmetrized game: a strategy of the symmetrized game is an ESS if, and only if, the corresponding pair of strategies is a strict Nash equilibrium of the asymmetric game (Selten, 1980). Notice that this result is important for many biological examples. It is relevant whenever the situation at hand allows one to clearly distinguish between two player roles, such as male and female, owner and intruder, parent and offspring, and so on. We follow other authors and use ESS and strict Nash equilibrium interchangeably when talking about asymmetric games. (The two do have to be distinguished when talking about symmetric games. In what follows, no ambiguity should arise, however.)

The ESS concept is a purely formal object that can be applied whenever one has a finite symmetric two-player game. We think it is useful to distinguish what we call *ESS methodology* from the ESS concept proper. ESS methodology has to do with how to interpret the ESS concept rather than its mathematical definition. ESS methodology consists in (i) describing an evolutionary phenomenon in terms of a game, (ii) finding the ESS or the ESSs of the game and (iii) identifying the ESS or the ESSs of the game with the possible evolutionary outcomes of the situation in question.

While few have explicitly recommended the ESS methodology, it appears to be widespread and implicit in many applications of game theory in evolutionary biology. Maynard Smith cautions against the ESS methodology as a fully general methodology, but uses it himself in some particular cases that we believe are problematic (see Section 4). Although they never explicitly recommend the ESS methodology, Searcy and Nowicki (2005, Chapter 2) refer only to ESS results when surveying the literature on signaling between relatives. This is not to criticize Searcy and Nowicki, for there is very little else in the literature other than ESS results. Instead we wish to illustrate that whether or not it is ever explicitly defended, there is an implicit reliance in the biological literature on the ESS methodology.

Notice that some results that we have mentioned above lend some credibility to the ESS methodology. In particular, the stability of ESSs (or, for that matter, strict Nash equilibria in asymmetric games) for several standard evolutionary dynamics certainly seems to be a strong reason to trust step (iii). In this paper, we show that in certain situations this conclusion is not warranted. The line of argument we will use is to establish the evolutionary significance of non-ESS states (already for standard evolutionary dynamics); we also argue that these non-ESS states are not just odd cases but should be expected to be widespread. There are other limits of the ESS methodology that will be worked out in more detail elsewhere (Huttegger and Zollman, manuscript).

3 Equilibrium refinements

For what follows it will be important to point out a similarity between ESS and virtually all equilibrium refinements from economics. This similarity holds despite the fact that the motivation behind equilibrium refinements is quite different from the motivation behind ESS. As pointed out in the introduction, the general methodology is quite similar. One begins with an intuitive notion – of evolutionary stability in the case of biology and of rationality in the case of economics – and attempts to represent that intuitive notion formally. But the similarity is stronger than this. As we will see, both equilibrium refinements and ESS rely on some notion of perturbation from an equilibrium.

The goal of the equilibrium refinement research project is to rule out certain Nash equilibria as being unreasonable, in the sense that the corresponding Nash equilibrium profile would not be chosen by rational players. Consider *subgame perfect* Nash equilibria as a paradigmatic example of an equilibrium refinement. What is meant by a subgame perfect Nash equilibrium is best explained by example. In the Chain-Store

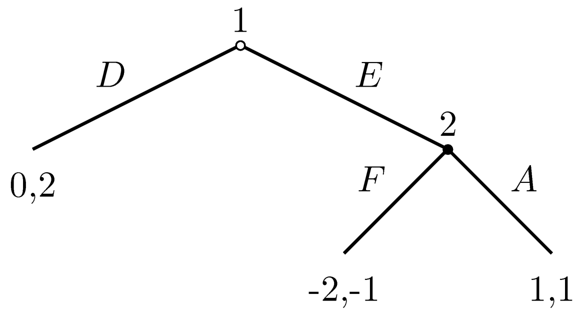

Figure 2: The Chain Store Game

game (Selten, 1978, see Figure 2), there are two players. The first player can enter a market that is dominated by a chain store (the second player). If player 1 enters the market, then player 2 has to decide between fighting player 1 or sharing the market. In this case, it is assumed that player 2 gets the higher payoff from not fighting (as does player 1). Player 1 can also choose not to enter the market in the first place. In this case, nothing changes. This is clearly the most preferred outcome from the point of view of player 2. Player 1 is assumed to have a higher payoff in this outcome than in the outcome where player 1 chooses to fight. But it would be best for player 1 if she enters the market and player 1 does not fight.

Why should player 1 decide not to enter the market? Well, player 2 might threaten player 1 that she will fight should player 1 enter the market. Is this threat credible? Selten (1965, 1975, 1978) thinks not. If player 2 had to decide between fighting and not fighting, she would choose not to fight since this outcome carries the higher payoff.

This kind of reasoning leads to the concept of a subgame perfect Nash equilibrium. There are two pure strategy Nash equilibria in the Chain Store game. The first one calls for player 1 not to enter the market and for player 2 to fight. In the second one player 1 enters the market and player 2 chooses not to fight. Only the latter Nash equilibrium is subgame perfect. It is subgame perfect because it calls for rational play in every subgame of the Chain store game. There are only two subgames, the whole game and the subgame where player 2 decides to fight or not to fight. Enter (player 1) and Not Fight (player 2) is rational in both subgames. The first Nash equilibrium is not rational in the second subgame, since it requires player 2 to fight if she had to decide.

Thus, subgame perfectness rules out the first Nash equilibrium as not being stable. The logic applied here requires us to imagine the two players are choosing according to some Nash equilibrium. This Nash equilibrium will be subgame perfect if all players act rationally when called upon to act even at unreached decision nodes (information sets). We may view this as a perturbation of the players' actions from equilibrium, similar to the population being slightly perturbed from a state in determining whether it is an ESS.

In fact, Selten has advanced a theory of ESSs in extensive form games (such as the Chain Store game) along these lines (Selten, 1983).

In the case of subgame perfect Nash equilibria, the perturbation from equilibrium is counterfactual. The players are to imagine what would happen if they were called upon to decide at a decision node that is not reached by equilibrium play. This can be turned into a real perturbation, leading to an equilibrium refinement that is closely related to subgame perfect equilibria. It is called a *trembling-hand* perfect equilibrium (Selten, 1975). Consider again the Chain Store game. Suppose player 1 decides not to enter the market, but that she cannot implement her strategy with absolute certainty. That is to say, she sometimes makes a mistake and chooses to enter the market (the arbitrary nature of the mistake being captured by the metaphor of a trembling hand). In this situation, it is not rational for player 2 to choose to fight with high probability. She should choose fighting with a sufficiently low probability. Given that player 2 is choosing to fight with very low probability, player 1 does better by entering the market. Thus, taking the probabilities of mistakes to zero, there is no sequence of perturbed Nash equilibria converging to the Nash equilibrium where player 1 does not enter and player 2 fights.

The logic of trembling hand perfect equilibria thus goes as follows. A player assumes that a certain Nash equilibrium is played. Instead of the original game, she imagines that a perturbed game is played where players can only choose mixed strategies. The original Nash equilibrium is trembling-hand perfect if there is a sequence of Nash equilibria of perturbed games converging to the original Nash equilibrium as the probabilities of mistakes go to zero. The subgame perfect equilibrium of the Chain Store game passes this test, while the other Nash equilibrium does not. For our purposes, it is important to notice that, again, the players are assumed to be at an equilibrium, and the stability of the equilibrium is determined by considering perturbations from equilibrium (in this case, players making mistakes).

Virtually all equilibrium refinements assume that the players choose according to some particular Nash equilibrium and investigate whether the resulting strategy profile is feasible in terms of some particular perturbation (van Damme, 1991). Notice that this has a very similar structure to ESS. Firstly, ESS itself is a Nash equilibrium refinement in the sense that an ESS is also a Nash equilibrium, while not any Nash equilibrium is also an ESS. Moreover, the standard interpretation of ESS also assumes that players are at equilibrium, where now the players' choices are identified with the state of a population. A Nash equilibrium is ESS if it is stable under a particular perturbation, to wit, the introduction of a low share of a mutant strategy. Thus, ESS can be viewed as having the same structure as many other equilibrium refinements. The essential aspect appears to be a perturbation away from an equilibrium state.

4 Non-ESS states

As we have just mentioned, from the dynamic stability of ESSs one may be led to the conclusion that one should identify the set of significant evolutionary outcomes of a game with the game's ESSs. One way to undermine this position is by establishing that there are strategies that are not evolutionarily stable but are, nonetheless, stable for the replicator dynamics. Moreover, it is important that this does not only hold for exceptional games (i.e., games with very special payoff parameters), and that these games should be of biological importance. As a first step, we will present just such a game.

The example we have chosen is the well-known Sir Philip Sidney (SPS) game (Maynard Smith, 1991). This game was introduced by John Maynard Smith as a simple illustration of the handicap principle (Zahavi, 1975; Grafen, 1990). The SPS game is a signaling game that allows for conflicts of interest between the sender and the receiver. The sender can be in two states, needy or healthy. Given any state, she can send a message m or abstain from doing so. The receiver can transfer a resource to the sender or she can keep it to herself (after observing whether the sender has decided to transmit m). An extensive form representation of the Sir Philip Sidney game is pictured in Figure 3.

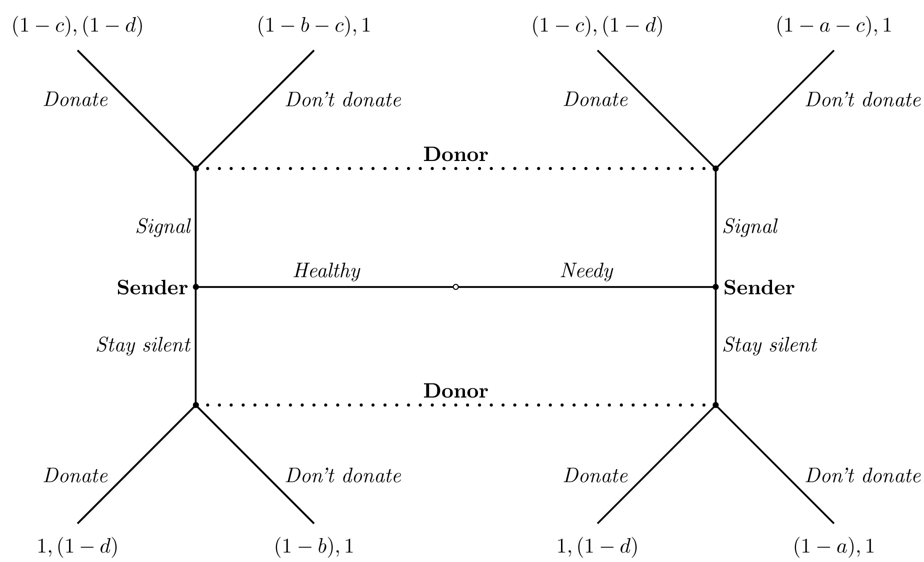

Figure 3: The Sir Philip Sidney Game. $1-a$ and $1-b$ represent the individual fitnesses of the needy and healthy types (respectively) if they do not receive the resource, c is the fitness cost to the sender for sending the signal. $1-d$ represents the individual fitness of the receiver if she transfers the resource to the sender.

In the version pictured here the receiver's dominant strategy is to keep the resource to herself, because she gains nothing from the transfer. To create the possibility for transfer we can assume that the sender and the receiver can be related to some degree $k \in [0,1]$. k determines the inclusive fitness of a player, i.e., a player's payoff in an interaction is the payoff from her own role plus k times the payoff from the role of the other player. If k is sufficiently close to 0, the receiver will never donate the resource to the sender since it's detrimental to her own payoff. However, if k is sufficiently large, then it becomes possible for the resource to be transferred in equilibrium.

One population state, the so-called *signaling ESS*, has attracted particular attention. Since the SPS game is an asymmetric game, the signaling ESS really is a strict Nash equilibrium. In the signaling ESS, the sender sends m if she is needy and does not send m otherwise. The receiver transfers the resource if she gets the signal and keeps it to herself otherwise. The signaling ESS is important because it illustrates the handicap principle, which asserts that signals need to be costly in order for signal reliability to be maintained when there is conflict of interest between a sender and a receiver. The conditions under which the signaling ESS exists reveal the handicap principle. If d is the amount a receiver can transfer, a and b are the gains from a transfer for a needy and healthy sender, respectively, and k is the degree of relatedness between the sender and the receiver, then the signaling ESS exists only if

$$a \geq c + kd \geq b.$$

The second inequality guarantees that a healthy sender will not send the signal in order to obtain the resource. The first inequality guarantees that the gain is sufficiently high for a needy signaler to pay the cost for sending the signal. (There is another condition for the signaling ESS that pertains to the receiver; see Bergstrom and Lachmann (1997) and Huttegger and Zollman (2010) for details.)

So far so good. But the SPS game has other Nash equilibria which do not correspond to ESSs, i.e., which are not strict Nash equilibria. Firstly, there are so-called *pooling equilibria*. These equilibria are characterized by there being no information transfer between the sender and the receiver. There are two kinds of pooling equilibria. In both kinds of pooling equilibria the sender never signals. In one kind of pooling equilibrium the receiver always transfers, while in the other kind she never transfers the resource. One kind of pooling equilibrium always exists (regardless of the parameter settings). But generically it is not the case that both kinds of pooling equilibria exist at the same time.

In Huttegger and Zollman (2010) we show that pooling equilibria cannot be strict because they are part of Nash equilibrium components. This is quite easy to see. In both kinds of pooling equilibria the sender never sends the signal regardless of whether she is needy or not. Therefore, the receiver's strategies of never transferring the resource or only transferring the resource if the signal is sent are behaviorally equivalent. Similarly, the receiver's strategies of always transferring the resource and only transferring the resource if she does not receive the signal are behaviorally equivalent in this case. This implies that each kind of pooling equilibrium is actually part of a convex set of Nash equilibria and can, therefore, not be strict.

Pooling equilibria provide a first case of non-ESS states in the SPS game. A second example is given by a so-called *hybrid* or *polymorphic equilibrium*. In a polymorphic equilibrium the sender mixes between two states, 'always signal' and 'signal only when needy'. This means that the sender sometimes reliably indicates her true state and sometimes she does not. The receiver also mixes between two strategies, 'never transfer' and 'transfer only upon receiving the signal'. As a reaction to the sender's dishonesty, the receiver must place some positive weight on not to transfer the resource under any circumstances. But, since the sender sometimes signals truthfully, she also has to place some positive probability weight on transferring the resource as a reaction to receiving the signal.

Using the notation introduced above, a polymorphic equilibrium exists only if

$$b - kd > c$$

(Huttegger and Zollman, 2010). This shows that the signaling ESS and the polymorphic equilibrium cannot exist together. In fact, the polymorphic equilibrium appears once the signaling ESS ceases to exist because the signal cost c gets too small. More specifically, it can be shown that $c^* = b - kd$ is the minimum signal cost that guarantees the reliability of the signal (Bergstrom and Lachmann, 1997). It follows that in the polymorphic equilibrium there is some information transfer with possibly low costs despite the fact that there is conflict of interest between the sender and the receiver.

It is clear that the polymorphic equilibrium cannot be an ESS. Since it is not a pure strategy equilibrium (both the sender and the receiver place positive probability weights on two strategies) it cannot be a strict equilibrium. So among the three kinds of equilibria that have been found in the SPS game, only one is an ESS. At this point the question arises whether following the ESS methodology is good advice, i.e., whether we should discard pooling equilibria and the hybrid equilibrium as evolutionarily insignificant.

In Huttegger and Zollman (2010) we show, firstly, that both pooling equilibria and hybrid equilibria are Liapunov stable (when they exist). Liapunov stability is a weaker stability condition than asymptotic stability. It requires solution trajectories that start nearby the state to stay close for all future times. Solution trajectories need not converge (as they do in the case of an asymptotically stable state). Pooling equilibria are not asymptotically stable because they are elements of a set of equilibria. It can be shown analytically that almost all nearby population states converge to some pooling equilibrium. The hybrid equilibrium is Liapunov stable in a quite different sense. One can show that it is a spiraling center. This means that almost all nearby solution trajectories approach the face on which the hybrid equilibrium can be found while also cycling toward it.

We also note that the signaling ESS does not have a large basin of attraction whenever there is significant conflict of interest between sender and receiver. Most initial populations converge to a state that corresponds to a partial pooling equilibrium. There is a second noteworthy result in Huttegger and Zollman (2010). The hybrid equilibrium, when it exists, appears to have a significantly larger basin of attraction than the signaling ESS in similar cases of conflicting interests. It follows that the ESS methodology fails to a large extent in the SPS game. The signaling ESS, which is the only ESS of this game, fails to have a sufficiently large basin of attraction in the replicator dynamics. It therefore does not appear to be a good prediction of what the evolutionary outcome is in something similar to the SPS game. The two other kinds of Nash equilibria known to exist in the SPS game, the hybrid equilibrium and pooling equilibria, seem to be evolutionarily more significant than the signaling ESS despite their not being ESS.

These conclusions of course rest on using the replicator dynamics as a model for evolution. It is not clear at this stage what happens when one considers significantly different kinds of dynamics (finite populations, structured populations). To our knowledge there are no general results concerning the stability of ESSs in these models. Therefore, the ESS methodology is best suited when one has a model of evolution similar to the replicator dynamics in mind.

If we stay with the replicator dynamics, can we say a bit more about methodology other than that ESS methodology fails in the SPS game? There are two aspects of this question. Firstly, we chose the SPS game as a biologically relevant example of a game where an ESS methodology fails. Do we expect there to be other examples, or is the SPS game an odd exception? Secondly, is there an equilibrium concept other than ESS that leads to a more suitable methodology? We try to answer these questions in turn in the next section.

5 Extensive-form games

As to the first question we would like to emphasize that we expect the failure of the ESS methodology to be quite common. The SPS game is an example of a game with a non-trivial extensive-form structure. By this we mean that it is not a simultaneous-move game where all players have to make a choice without any (partial) information about aspects of the choices of the other players. In the SPS game, the receiver knows the action choice of the sender (she receives the sender's signal). In the Chain-Store game, player 2 knows if player 1 has entered the market. Thus, in games with a non-trivial extensive-form structure the sequence of decisions plays an essential role in determining the outcome of the game.

As has been shown elsewhere, games with a non-trivial extensive-form structure often lead to the kinds of equilibria observed in the SPS game (see Cressman (2003); see also Huttegger (2009, 2010)). To be more specific, the kind of behavioral equivalences mentioned above, such as "transfer" and "transfer only if signal" given the receiver strategy "don't signal", will be very common in such games. This is simply due to the fact that the sequence of decisions is important in games with a non-trivial extensive form. In many cases this will lead to several of a player's strategies being behaviorally equivalent given earlier choices of other players. This is important because it influences the topological structure of Nash equilibria. As in the case of pooling equilibria in the SPS game, behaviorally equivalent strategies create a Nash equilibrium component where there are other Nash equilibria arbitrarily close to any particular Nash equilibrium. But such a Nash equilibrium cannot correspond to an ESS since it is not a strict Nash equilibrium (by the existence of a behaviorally equivalent strategy).

As another example consider the Nash equilibrium in the chain store game where player 1 does not enter and player 2 chooses to fight (which is neither subgame perfect nor trembling-hand perfect). Given player 1's choice, both choices of player 2 are behaviorally equivalent. This implies that the Nash equilibrium is one element of a Nash equilibrium component. Each mixed strategy profile in this component is a Nash equilibrium as long as player 2 chooses to fight with sufficiently high probability.

Although these kinds of Nash equilibria do not correspond to ESSs, they do correspond to *neutrally stable strategies* (NSSs). For an NSS the second ESS condition is weakened to a non-strict inequality. This essentially means that there is no mutant strategy that can take over the population, but that there may be a mutant strategy that can exist alongside the strategy in question. A pooling equilibrium and the non-subgame perfect equilibrium of the Chain-Store game provide examples of NSSs. (There is again a qualification here. NSS, like ESS, only applies to symmetric games. When we talk about the NSSs of an asymmetric game like the SPS game, we also mean the corresponding NSS of the symmetrized game. There is a precise correspondence between the two, just as in the case of ESS (Cressman, 2003).)

In the replicator dynamics, NSS implies Liapunov stability, i.e., population frequencies starting close to an NSS will stay nearby (but don't necessarily converge to the NSS). Thus, one reason why the ESS methodology fails is because it does not take into account NSSs. Should we therefore replace ESS methodology by an NSS methodology? We think not. Taking NSSs into account is a step in the right direction, but one may still fail to capture the stability properties of evolutionary dynamics. NSS implies Liapunov stability, but the converse can fail (Bomze and Weibull, 1995). Thus there can be Liapunov stable states in the replicator dynamics that are not NSSs. Furthermore, there are many NSSs which are, in fact, evolutionarily insignificant as illustrated by the degenerate game where all strategies do equally well. For such a game, any evolutionary dynamics would just drift around and no state could be called stable. We shall not explore the significance of these results, but they should serve to caution us against positing NSS methodology in a strict sense.

The prevalence of Nash equilibria similar to the hybrid equilibrium is much harder to judge. The hybrid equilibrium is a so-called *Nash-Pareto pair* (Hofbauer and Sigmund, 1998). This essentially means that it is a mixed equilibrium that is very similar to a strict Nash equilibrium. But there are not many examples of such

Nash equilibria in the literature. We suspect that one reason for this has to do with the ESS methodology, which makes it difficult to discover these other kinds of equilibria. We conjecture that it plays an important role in interactions that have some zero-sum aspect to it (just as in the case of the SPS game).

6 ESS, equilibrium refinements and a dynamic perspective

As we have argued in the previous two sections, we think that the ESS methodology fails because it does not properly take into account the dynamic possibilities one already has in baseline dynamical models like the replicator dynamics (which is really tailored towards the ESS concept). We think that there is a deeper reason for this. As explained above, ESS shares with other Nash equilibrium refinements that one starts with what happens to an equilibrium when it is subject to some kind of small perturbation (mutants, trembles, and so on). This can at best capture part of the dynamics under consideration, be it evolutionary dynamics, learning dynamics, or the deliberational dynamics of a rational player, namely, the dynamics close to equilibrium. It does not answer the question how players get to an equilibrium if they are far from any equilibrium. (Far away may refer to population proportions or to beliefs of players, for example.) In certain special cases this question might not be relevant. For instance, if you have an interior ESS in the replicator dynamics, all interior population states will converge to it. But in many cases this question will be relevant. There could be more than one ESS. Or there could be additional Liapunov stable states that complicate the overall dynamical picture.

One way to address these questions is by determining the basins of attractions for each different equilibrium states. It should be noted that being an ESS does not guarantee a large basin of attraction or even a larger basin of attraction than non-ESS states. For instance, in (Huttegger and Zollman, 2010) we show that pooling equilibria have a larger basin of attraction than the signaling ESS (if there is conflict of interest between the sender and the receiver) by performing simulations.

While the initial motivations behind the ESS methodology (like the equilibrium refinements program in economics) were noble – theoretical simplicity combined with generality – we believe that this methodology is ultimately inappropriate. The ESS methodology is without a doubt theoretically more simple than an explicitly dynamic methodology, but one must pay too high a cost to secure this simplicity. Many non-ESS states are evolutionarily significant – sometimes more significant than ESS states. Instead we believe that in many cases one must turn to a dynamic methodology for analyzing strategic interactions in biology.

References

Bergstrom, C. T. and Lachmann, M. 1997. 'Signalling among relatives I: Is costly signalling too costly?', *Philosophical Transactions of the Royal Society London, Series B*, 352:609–617

Bomze, I. and Weibull, J. 1995. 'Does neutral stability imply Lyapunov stability?', *Games and Economic Behavior*, 11:173–192

Brandenburger, A. 2007. 'The power of paradox: some recent developments in interactive epistemology', *International Journal of Game Theory* 25:465–492.

Cressman, R. 2003. *Evolutionary Dynamics and Extensive Form Games.* MIT Press, Cambridge, MA

Grafen, A. 1990. 'Biological signals as handicaps.', *Journal of Theoretical Biology*, 144:517–546.

Hofbauer, J. and Sigmund, K. 1998. *Evolutionary Games and Population Dynamics.* Cambridge University Press

Huttegger, S.M. 2010. 'Generic properties of evolutionary games and adaptationism', *Journal of Philosophy*, 107:80–102

Huttegger, S. M. and Zollman, K. J. S. 2010. 'Dynamic stability and basins of attraction in the Sir Philip Sidney game', *Proceedings of the Royal Society London, B*, 277:1915–1922

Huttegger, S. M. 2009. 'On the relation between games in extensive form and games in strategic form', in A. Hiecke and H. Leitgeb (eds.) *Reduction and Elimination in Philosophy and the Sciences*, pages 375–385. Ontos Press, Frankfurt a. M.

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

Maynard Smith, J. 1991. 'Honest signalling: the Philip Sidney game', *Animal Behavior*, 42:1034–1035

Maynard Smith, J. and Price, G. 1973. 'The logic of animal conflict', *Nature*, 146:15–18

Moran, P. A. P. 1962. *The Statistical Processes of Evolutionary Theory.* Clarendon, Oxford

Searcy, W. and Nowicki, S. 2005. *The Evolution of Animal Communication* Princeton University Press

Selten, R. 1965. 'Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit', *Zeitschrift für die gesamte Staatswissenschaft*, 121:301–324, 667–689

Selten, R. 1975. 'Re-examination of the perfectness concept for equilibrium points in extensive games', *International Journal of Game Theory*, 4:25–55

Selten, R. 1978 ‘The chain store paradox’, *Theory and Decision*, 9:127–159.

Selten, R. 1980. ‘A note on evolutionary stable strategies in asymmetrical animal conflicts’, *Journal of Theoretical Biology*, 84:93–101

Selten, R. 1983. ‘Evolutionary stability in extensive two-person games’, *Mathematical Social Sciences*, 5: 269–363

Taylor, P. D. and Jonker, L. 1978. ‘Evolutionarily stable strategies and game dynamics’, *Mathematical Biosciences*, 40:145–156

van Damme, E. 1991. *Stability and Perfection of Nash Equilibria*. Springer, New York

Weibull, J. 1995. *Evolutionary Game Theory*. MIT Press, Cambridge, MA

Zahavi, A. 1975 ‘Mate selection – the selection of a handicap’, *Journal of Theoretical Biology*, 53:205–214