METHODOLOGY IN BIOLOGICAL GAME THEORY

SIMON M. HUTTEGGER AND KEVIN J.S. ZOLLMAN

Abstract. Game theory has a prominent role in evolutionary biology, in particular in the ecological study of various phenomena ranging from conflict behavior to altruism to signaling and beyond. The two central methodological tools in biological game theory are the concepts of *Nash equilibrium* and *Evolutionarily Stable Strategy* (ESS). While both were inspired by a dynamic conception of evolution, these concepts are essentially static – they only show that a population is uninvadable, but not that a population is likely to evolve. In this paper we argue that a static methodology can lead to misleading views about a dynamic evolutionary processes. Instead, we advocate a more pluralistic methodology, which includes both static and dynamic game theoretic tools. Such an approach provides a more complete picture of the evolution of strategic behavior.

1. Introduction

When an ecologist or an evolutionary biologist is confronted with an apparently maladaptive phenotype, she must answer two questions. Firstly, why is this phenotype stable? Since it is apparently maladaptive, why hasn't this phenotype been eliminated in favor of a more adaptive alternative? Secondly, what led to the evolution of that behavior in the first place? This second question is especially pressing if it seems likely that an ancestral population did not possess the apparently maladaptive phenotype.

These two questions seem quite similar. One might be inclined to think an answer to the first will provide an answer to the second. We here suggest that in the context of game theory these two questions are often conflated, and that this conflation leads to incorrect judgments about evolutionary processes. There is one case in particular – the case of signaling behavior – where the proffered answer to the first question has been regarded as satisfactory despite the fact that scholars have (unknowingly) introduced a yet more difficult to solve mystery in the form of the second question.

In biological game theory there is a prevailing methodology which we will call the *equilibrium methodology.* This methodology involves developing a model of evolution and considering potential endpoints of evolution utilizing so called equilibrium concepts. Most common among these are the concept of a (strict) Nash equilibrium and the concept of an Evolutionarily Stable Strategy (ESS). Since it looks primarily at end points of evolutionary processes, this methodology is most clearly aimed at answering the first question: why is a particular state stable? But it is also aimed at providing a partial answer to the second question. Biologists often claim that the equilibria they find are (at least) potential endpoints for an evolutionary process. So, the claim that a state is an equilibrium entails that the state is stable and also that it is reachable by evolution.

What is left out of the equilibrium methodology is any model of the dynamics of evolutionary processes. This is in contrast to what we will call the dynamic methodology, which models explicitly (to various degrees) the process of evolution. We will argue that recent results from the game theoretic study of signaling in biology demonstrates that the equilibrium methodology alone is inadequate to answer the second major question, and that it has in fact been misleading. While the theoretical possibility of such problems has been known for some time, those pitfalls have been regarded as either obvious or unrealistic. That there are biologically significant examples where the methodology has failed suggests that the limitations of the methodology may be endemic – a conclusion for which we shall indeed argue. We do not argue that the static methodology should be abandoned – equilibrium analysis is an important part of the process of understanding evolutionary games. Instead, we suggest that any full analysis of strategic interaction must proceed by utilizing both methodologies in tandem.

We will begin in Section 2 by describing the equilibrium methodology and discussing the already well known limitations of this method. In Sections 3 and 4 we describe two cases from the signaling games literature where we believe this methodology has misled investigators in their search for explanations of behavior. In contrast, we suggest that the dynamic methodology provides significant insight. Finally, in Section 5, we conclude.

2. THE EQUILIBRIUM METHODOLOGY

Game theory was initially developed in economics as a model for human strategic interaction. A *game in strategic form* is a mathematical object which includes a list of players, a set of strategies for each player, and a specification of a payoff for every combination of strategies by each player. Game theory was later introduced to biology by Maynard Smith and Price (1973) (although similar ways of approaching problems go back at least as far as Fisher (1915)). In a biological setting the strategies are interpreted as alternative phenotypes and the payoffs are interpreted as fitnesses.

In analyzing games in both economics and biology, it has become common to develop conditions required for a set of strategies to be in *equilibrium*. Most well known in game theory is the concept of Nash equilibrium, which merely requires that no player could improve her situation by unilaterally switching. Nash equilibria can be of two types. Pure strategy Nash equilibria represent situations where an entire population is monomorphic with respect to the equilibrium phenotype. Mixed strategy equilibria involve random distributions of strategies and require a more nuanced interpretation. In economics, these equilibrium strategies represent intentional randomization by individual players. In a biological setting, a mixed strategy might either represent a single organism whose phenotype is determined by a random process, or it might represent a population that is polymorphic – one that has several different phenotypes represented.

The Nash equilibrium criterion picks out a set of strategies as deserving special attention. Once there, no player has a positive incentive to leave, and so one might expect that this set of strategies (in biology, phenotypes) would be stable. More specifically, it's stable in a weak sense. A player might do equally well by switching.

Consider, for example, the game in Figure 1. Here the strategy profile (A, A) is a Nash equilibrium, because no one does strictly better by switching to B. But considered from an evolutionary perspective, this equilibrium seems suspect. Suppose one begins with a population of A-types. If a mutant B-type were introduced, she would not be eliminated by natural selection since she does as well as any other. Should another B-type arise and should they interact, their fitness will be enhanced and natural selection should favor the B-types. The reason for this is that A is weakly dominated by B. This means that B always gets at least as high a payoff as A and a higher payoff in at least one instance. As a result, the concept of Nash equilibrium is too general from an evolutionary point of view. It includes population-states which one would not expect to be stable.

	A	B
A	$0, 0$	$0, 0$
B	$0, 0$	$1, 1$

FIGURE 1. The strategy profile (A, A) is a Nash equilibrium, but A is weakly dominated by B.

One possibility is to restrict attention to *strict Nash equilibria.* To be a strict Nash equilibrium it must be the case that every individual will do strictly worse by switching. It is usually regarded that such a restriction is too strong. It seems clear that strict Nash equilibria should count as stable for any evolutionary process (at least in finite games). However, there are certain mixed Nash equilibria which should also be considered stable from an evolutionary point view. Mixed strategy

Nash equilibria cannot be strict. As a result, a concept that lies in between Nash equilibrium and strict Nash equilibrium must be considered.[1]

Maynard Smith and Price (1973) and later Maynard Smith (1982) suggested a notion of evolutionary stability that would coincide with the biological notion of uninvadability. Those phenotypes which cannot be invaded by small mutations are called Evolutionarily Stable Strategies (ESS). Formally, the definition of an ESS is,

Definition 1. *A strategy (i.e. phenotype)* s^* *is an ESS if and only if the following two conditions are met*

1. $u(s^*, s^*) \geq u(s, s^*)$ *for all alternative strategies* s *and*
2. *If* $u(s^*, s^*) = u(s, s^*)$*, then* $u(s^*, s) > u(s, s)$.

$u(x, y)$ represents the fitness (payoff) of strategy x against y. The first condition states that s^* is in Nash equilibrium with itself, i.e., there is no other strategy earning a higher payoff against s^*. The second condition guarantees stability in case of a mutant strategy s that earns the same payoff against s^* by requiring that s^* is doing better against s than the mutant strategy against itself.

The two conditions for the evolutionary stability of a strategy seem to be natural for a first approximation where we assume strategies to be distributed in a large, randomly interacting population, and where an incumbent strategy is confronted only with one mutant at a time. There is an alternative characterization of ESS that is particularly revealing in this context. It is easy to show that s^* is an ESS if and only if

$$u(s^*, \varepsilon s + (1 - \varepsilon)s^*) > u(s, \varepsilon s + (1 - \varepsilon)s^*).$$

for all ε that are less than some sufficiently small $\bar{\varepsilon}$. That is to say, in a population with a share of ε s strategies and $1 - \varepsilon$ s^* strategies, s^* gets a higher expected payoff than s.

The strategy s^* under consideration might be a pure strategy or a mixed strategy (a probability distribution over different pure strategies of the game). Like interpreting mixed strategy Nash equilibrium, interpreting mixed strategy ESS is a delicate matter (Bergstrom and Godfrey-Smith, 1998).

It should be emphasized that both Nash equilibria and ESS are static concepts. In a biological context, one considers a population at a given state and asks if this population would remain at that state. Therefore they cannot, necessarily, explain how a population arrived at that state. Maynard Smith did, in part, recognize this problem. He suggested that in two pathological cases, ecologists would have to consider "change as well as constancy" (1982, 8). His particular focus was on the issue of ESS, and we will limit our attention to that case here.

---

[1] The well-known evolutionary game Hawk-Dove is used as an illustration of this situation.

The first case is one where no strategy is an ESS. The simplest example of such a case is illustrated by the children's game Rock-Paper-Scissors, illustrated in Figure 2. In this game there is a single Nash equilibrium where each player plays each strategy with equal probability. However, this strategy is not evolutionarily stable. Because all strategies do equally well against this mixed strategy, we must consider the second condition of the ESS definition. This requires that the uniform mixture do better against any alternative than this alternative does against itself. However, the uniform mixture secures an expected payoff of 0 against all alternative strategies, which is precisely what an alternative strategy secures against itself. As a result, there is no ESS in this game.

	Rock	*Paper*	*Scissors*
Rock	$0, 0$	$-1, 1$	$1, -1$
Paper	$1, -1$	$0, 0$	$-1, 1$
Scissors	$-1, 1$	$1, -1$	$0, 0$

Figure 2. Rock-Paper-Scissors

Maynard Smith thought this represented no significant problem for his methodology for two reasons. First, the paucity of the ESS-methodology in this case was clear; there was no ESS. A theoretical ecologist confronting such a game would immediately see that the methodology was unhelpful. Second, Maynard Smith suggested he was aware of no biologically realistic situation where there is no ESS. However, it has since been discovered that the mating behavior of male side-blotched lizards (Sinervo and Lively, 1996) and toxin production in some bacteria (Kerr et al., 2002) both follow a Rock-Paper-Scissors structure, suggesting that this game is biologically plausible. One could, just possibly, still maintain that strategic interactions without an ESS are very rare.

Perhaps more importantly, Maynard Smith's second situation where one must consider "change as well as constancy" occurred when there was more than one ESS. A trivial version of such a game is illustrated in Figure 3. In this game both A and B are ESSs. Thus, the ESS theory cannot predict which should be expected to evolve.

Again, however, Maynard Smith regarded such situations as obvious. The presence of two ESSs will alert one to the presence of a troubling case. In both these cases, no ESS and more than one ESS, Maynard Smith cautions that one should not

	A	B
A	1, 1	0, 0
B	0, 0	1, 1

Figure 3. A coordination game

suppose that an ESS strategy is likely to evolve. Notice, though, that it is much harder to downplay the case of having more than one ESS, since in a strategic context such situations are presumably abundant.

After Maynard Smith, another potentially troubling case was suggested by Nowak (1990). Nowak presents a game with nonlinear payoffs where an ESS may not be the result of evolution. Again, however, these examples depend on fitness values which are not linear in the population proportions of the other strategies – another tipoff that the ESS analysis may be misleading.

Given the various limitations of the equilibrium methodology, why should Maynard Smith have championed it? In the first place, he seemed to think that the class of strategic interactions where it is applicable comprises nonetheless a large part of real-world interactions. Secondly, the equilibrium methodology appears very general. By focusing clearly on the stability question, it is hoped that the result of equilibrium analysis will apply to many different types of underlying evolutionary dynamics. One does not, for instance, need to make any assumptions about the role of mutation, drift, population size or structure, environmental heterogeneity, etc., in order to derive important conclusions from the model. We believe that it is this "dynamics agnosticism" which motivates many biologists to utilize the equilibrium methodology. A motivation for this might be the desire to derive conclusions that will apply to as many potential situations as possible and thus would require few idealized assumptions. If one is too specific, then one runs the risk of being inapplicable to many biologically realistic situations.

In the following two sections we will present two cases where the equilibrium analysis has been employed, but is nonetheless misleading. The first one (Section 3) is, in a sense, of Maynard Smith's second type. Here we illustrate a subtle slide in reasoning that has allowed scholars to ignore that the game is one of these cases, and thus has led to incorrect assumptions about the evolvability of certain strategies. The second case is more troubling. The Sir Philip Sidney game is a game invented by Maynard Smith and analyzed using the equilibrium methodology. This game does not conform to any of the above known pathological cases. In Section 4 we show how, even here, the equilibrium analysis is misleading in the answer it gives to the evolvability question.

3. Common interest signaling

Signaling has become a canonical example of the application of game theory to biological phenomena. One might wonder how two organisms evolved to use some arbitrary mechanism for the exchange of information.[2] Of special interest to biologists is how such a system of information transfer might come about in situations where the parties to the transfer might have diverging interests about what should be done with that information. We will return to these cases of signaling in the face of partial conflict of interest in Section 4. But for the time being we will focus on the case where all parties have common interests.

3.1. Lewis' signaling game.

Perhaps the simplest example of signaling was first discussed by Lewis (1969). Lewis described a game where there are two parties: a sender and a receiver. The sender observes some feature of the world that is relevant to both parties. She can send one of a set of messages to the receiver. The receiver can observe the message, but not the feature of the world, and take some action. Depending on the state of the world, different actions benefit both the sender and the receiver equally.

We can consider a class of finite versions of these games known as $N \times N \times N$ signaling games. In these games there are N states of the world and N actions, where there is exactly one action which is appropriate in each state and it differs from state to state. The sender has access to N messages which she can send to the receiver. We assume that both the sender and the receiver benefit when the receiver takes the appropriate action. This makes the Lewis signaling game a game of common interest.

3.2. Static analysis.

In the $N \times N \times N$ signaling game, there are many Nash equilibria. There are $N!$ combinations of strategies which are dubbed by Lewis as "signaling systems".[3] In these equilibria the sender chooses a different message for each state and the receiver chooses the appropriate act given the message. In these equilibria both the sender and the receiver do as well as possible. It is also easy to show that such states are only ESSs of this game (Wärneryd, 1993).[4]

The fact that there is no conflict of interest between the sender and the receiver in a Lewis signaling game has led some to conclude that common interest signaling

---

[2] The details of what should count as "signaling" are discussed in some detail in (Maynard Smith and Harper, 2003) and (Searcy and Nowicki, 2005). While an interested theoretical problem, we will leave it aside for the time being.

[3] The number of signaling systems corresponds to the number of all one-to-one functions from a set with N elements into a set with N elements.

[4] It should be noted that ESS as a concept is only applicable to symmetric games – those where all players have the same strategy set and the payoff of playing strategy x against y does not depend on the identity of the player playing strategy x. Lewis signaling games as defined are not symmetric, because the sender and receiver have different strategies. Wärneryd considers the symmetrized version of the game, where each player can be in the role of the sender and of the receiver. A player's strategy includes both a sender-strategy and a receiver-strategy. It is well known that the ESS of a symmetrized game correspond exactly to the strict Nash equilibria of the original game. In the discussion that follows we consider only the symmetrized Lewis signaling game.

games present no evolutionary mysteries. The following quote represents this stance fairly well:

> Honest signaling ... *would be expected* if the signaler and receiver have identical interests in an evolutionary sense... Communication between two such individuals would be akin to communication between two cells or two organs within an individual and one in general would not find reliability puzzling... (Searcy and Nowicki, 2005, 20, emphasis added)

While Maynard Smith suggested that cases with more than one ESS – like the Lewis signaling game – represent a situation where one must consider "change as well as constancy", Searcy and Nowicki appear to disagree. Although no explicit argument is given for Searcy and Nowicki's conclusion that signaling "would be expected" in this context, we understand why such a conclusion might seem appealing. If sender a receiver have identical interests, there appears to be no reason why they should not be able to communicate. While there is more than one ESS in the Lewis signaling game, they all feature perfect communication. If one is only interested in the property of signaling successfully without regard for how that signaling is achieved, then one expects signaling to evolve. We will see shortly that this argument is invalid because, even in a common interest game like this one, evolution need not take populations to an ESS.

3.3. **Dynamic analysis.** A more careful analysis of the dynamics of evolution in these games does not bear out Searcy and Nowicki's optimistic view of common interest signaling. First one should note that there are a variety of other Nash equilibria where less than perfect signaling is possible. In all these games there are equilibria known as pooling equilibria, where no information is transmitted. Suppose that the sender ignores the state of the world and sends the same signal regardless of state, and suppose that the receiver ignores the signal and takes the action which is most likely to be best (or chooses some action randomly among those who are most likely to be best). This set of strategies is an equilibrium; no player can do better by switching, but it is not an ESS.

When $N > 2$ there can also be states where some information is communicated, but less than perfect communication is achieved. Consider the strategies pictured in figure 4. Here x and y represent probabilities which lie strictly between 0 and 1. This strategy is an equilibrium; neither player can gain by switching. However, it features less than perfect communication. While state 3 is communicated perfectly, states 1 and 2 are pooled onto signal 1.

These equilibria are not ESSs, not because they can be invaded by a mutant, but instead because certain mutants will not be eliminated by evolution. Therefore in

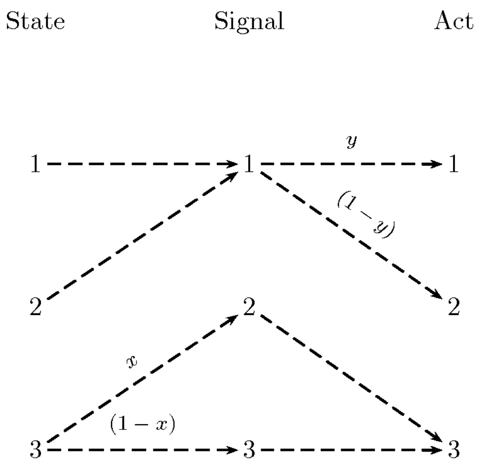

FIGURE 4. A partial pooling equilibrium

order to determine the evolutionary significance of total pooling and partial pooling equilibria we must turn to a model which explicitly considers how strategies change over time.

We will first consider perhaps the simplest model of evolution in games, the replicator dynamics (Taylor and Jonker, 1978; Hofbauer and Sigmund, 1998). This set of differential equations requires that individual strategies increase in frequency in a population only when they do better than the population average. In this model population proportions are treated as real numbers, and so the population is presumed to be infinitely divisible. Also, a strategy receives, as its payoff, the average it would receive against the population. As a result the replicator dynamics represents individuals as interacting at random. Using this underlying model for evolution we can return to the two questions we began with.

First, what populations are stable utilizing this model of evolution? A general fact about this dynamics is that all ESSs are asymptotically stable. If the population is at an ESS and there is a small perturbation in the strategy frequencies, the population will bounce back to the ESS. In other words, if the population's strategy frequencies start close to an ESS, the population will not only stay nearby but also converge to it.

However, other population states are Lyapunov stable. This means that small mutations do not snowball into large scale changes in the population; populations

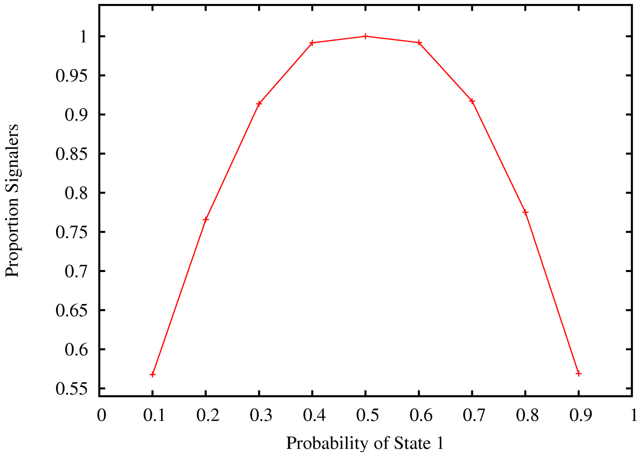

Figure 5. Basins of attraction for total pooling equilibria in the two-state, two-signal, two-act signaling game. The x-axis represents the probability of state 1.

starting close to a Liapunov stable population state will stay nearby.[5] Except in the special case where there are two states of the world, signals, and acts and the two states are equally probable, several of the pooling or partial pooling equilibria are Lyapunov stable (Huttegger, 2007; Pawlowitsch, 2008).

Second, what populations are likely to evolve? In the special case discussed above where there are two states, signals, and acts and the states are equally probable, essentially every initial starting population evolves to a signaling system (Huttegger, 2007).[6] In this case, Searcy and Nowicki were right; signaling should be expected. However, in every signaling game where there are more signals or the states are not equally likely there is some significant set of initial populations that evolve to imperfect or no communication.

Figure 5 illustrates the situation for a $2 \times 2 \times 2$ signaling game. Here you can see that signaling is assured when the states are equally likely. However, signaling is far from guaranteed when one state is 9 times more likely than the other.[7]

Huttegger (2007), Pawlowitsch (2008), and Huttegger et al. (2010) have shown that the partial pooling equilibria have positive basins of attraction as well. Already when there are three states (which are equiprobable), signals, and acts, five percent of the initial starting proportions evolved to a state with partial communication.

The replicator dynamics does not include any method for representing mutation, where new strategies can be constantly introduced into the population. However,

---

[5] This is a weaker concept of stability than asymptotic stability. Liapunov stable population states are nonetheless significant as approximations.

[6] Formally, the set of states that evolve to a signaling system comprise a set of measure 1.

[7] This figure was generated by simulation using the related discrete time replicator dynamics.

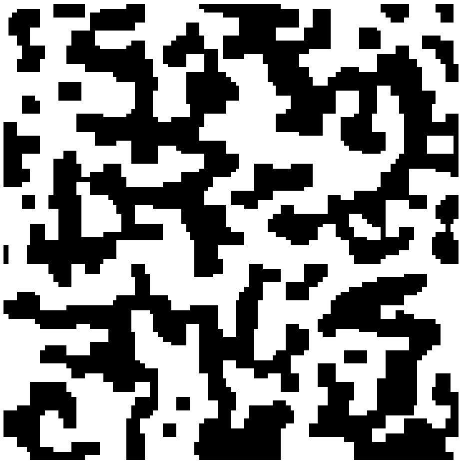

Figure 6. An end state for the spatial Lewis signaling game (from Zollman 2005).

one can modify this dynamics slightly in order to account for this possibility. One such modification is known as the selection-mutation dynamics (Hofbauer, 1985). Many things are possible under this dynamics. It might be the case that all initial states evolve to the signaling systems for certain amounts of mutation, but it need not be the case for others (Hofbauer and Huttegger, 2008).

One might also want to relax some of the other assumptions of the replicator dynamics. For instance, if one relaxes the assumption that the population is effectively infinite, one can use the Moran process (Moran, 1962). Again we get a variegated picture, where sometimes these non-signaling equilibria are avoided but not always (Huttegger et al., 2010; Pawlowitsch, 2007).

It is also important to consider a situation where individuals interact non-randomly. Perhaps they are constrained by physical space to only interact with people that are near them. Signaling games in this type of situation have been considered by Wagner (2009) and Zollman (2005). Zollman (2005) shows when individuals are mapped unto a plane and constrained to interact with others who are near them, new stable states emerge. Every simulation resulted in a state like the one pictured in Figure 6. Here there are two co-existing signaling systems (one pictured in white the other in black). This state is not an ESS, but is stable (in a weak sense) and, more to the point, is what will result from evolution. Wagner (2009) confirmed that these types of non-equilibrium states are possible even if one modifies the underlying structure of interaction to different types of social networks.

We believe that the results from evolutionary dynamics, taken together, indicate two important things. Firstly, the ESS analysis is incomplete. Considering the

Lewis signaling game only in terms of ESS will suggest that "signaling ought to be expected." An explicitly dynamical model of the process casts doubt on this strong conclusion. This is especially true for the replicator dynamics, which is strongly related to the concept of ESS in that both the replicator dynamics and ESS make similar assumptions about the population. What this illustrates is that the equilibrium methodology does not live up to its goal of being dynamics agnostic. Different dynamics yield different outcomes, some but not all of which coincide with the equilibrium analysis.

Secondly, considering evolutionary dynamics reveals a much more interesting and complicated picture of the relationship between the game, the process of evolution, and the expected outcomes which warrants investigation in order to develop a deeper understanding of the process by which signaling is to emerge.

For a variety of reasons the Lewis signaling game has not been extensively studied by biologists. Perhaps this is because of the simple facts about ESS or perhaps for some other reason such as the lack of conflicting interests and the resulting supposed triviality of signaling reliability. As a result, it is hard to say exactly how important the failure of the ESS methodology is in this case. We will now turn to a game that has been extensively studied and that (in some cases) has a unique strict Nash equilibrium and, therefore, a unique ESS. Again we will show how the static analysis has been misleading, and we will show why an explicitly dynamic analysis will make one doubt the general conclusion that results from finding a particular type of equilibrium in the game.

4. The Sir Philip Sidney game

The Lewis signaling game features common interest – in every state the sender and the receiver have the same ordering over the potential actions. This situation has not been extensively studied by ecologists. Instead they have focused on the apparently mysterious case where signaling exists, but where there are not perfectly overlapping interest between the sender and the receiver.

In situations where the sender an receiver no longer have common interest, there is a mystery about stability. Consider for example a canonical case – signaling between potential mates. Individuals of one sex (here we will say males) might differ in quality, and the other sex (here females) would prefer to mate with those of higher quality.[8] In such a situation high quality males would be selected to make their quality conspicuous. However, low quality males would also be selected to display whatever characteristic the high quality males use to signal their quality. But, if both high and low quality males appear the same in some respect, females

---

[8] This is the standard story which begs many questions, such as the exact nature of quality.

would be selected to ignore this trait, and the signal would cease to serve any evolutionary function.

Zahavi (1975) noted that despite this evolutionary story there appears to be a large number of cases where males successfully signal their quality to females. Zahavi developed the “handicap principle” where he suggested that only those signals which involved costs would be stable. This notion was formalized by Grafen (1990) and then simplified to a particular case by Maynard Smith (1991).

Maynard Smith considered a situation different from sexual signaling where individuals nonetheless have a conflict of interest. He considered the situation of a child begging for a resource from its parent. The child could be in one of two states, needy or healthy. While both needy and healthy children benefited from receiving the resource, the needy child benefited more. The child could communicate with the parent by sending a costly signal, which the parent can observe (the parent cannot directly observe the state of the child). Once observing the signal, the parent decides whether or not to transfer the resource, and reduce its own individual fitness to benefit its child.

4.1. **Static analysis.** If the two players are unrelated, the dominant strategy for the parent is to keep the resource. After all, what does it gain from reducing its own fitness? This remains true if the parent and the child are related to a low degree. But, once they are related to a sufficiently high degree, the parent wishes to donate the resource to the needy child but not to the healthy one, because the gain to the needy child is sufficiently high to warrant reducing her own fitness, whereas the gain for the healthy one is not. Both the healthy and the needy child would prefer to secure the resource, however. So, in this case there is a conflict of interest when the child is healthy (the child wants a resource that the parent would prefer not to give). When the parent and the child are related to sufficiently high degree, this game is similar to the common interest signaling game discussed in Section 3. The child only wants the resource if it is needy, and the parent only wants to transfer the resource if the child is needy.

As indicated before, the central mystery for ecologists has been the middle case where interests diverge. Why, when the child has an incentive to lie, would honest signaling persist? Maynard Smith showed that whenever one is in this situation one can impose a cost on the child for sending the signal which is sufficiently high that only the needy child is willing to pay the cost in order to secure the resource. When the game has partial conflict of interest and a single signal with sufficiently high cost, the unique ESS of the game is one where only the needy child signals and where the parent transfers the resource only if she observes the child signaling.

Maynard Smith's game has been taken to illustrate how cost (or handicaps) can explain the stability of signaling in the face of conflicts of interest. Signaling would not be an equilibrium, and would thus be unstable, if there were no cost. But when there is sufficiently high cost it is the only strict equilibrium.

4.2. **Other equilibria.**

Bergstrom and Lachmann (1997) first identified a potential problem with the explanation for signaling offered by Maynard Smith. They compare the signaling equilibrium to total pooling states. Like the Lewis signaling game, total pooling states are stable but in a weaker sense than ESS. However, they can have the property of being *Pareto superior* to the signaling state – i.e. the total pooling state is better for *both* the parent and child than the state of signaling. It would be odd, Bergstrom and Lachmann claim, for evolution to lead from a (weakly) stable superior state to an inferior one.

Huttegger and Zollman (2010) also show that another state of interest exists in this game, a *hybrid equilibrium.* When the cost of the signal is too low to support a signaling equilibrium, but nonetheless above a certain threshold, there exists an equilibrium where some communication occurs. In this equilibrium the needy child always sends the signal and the healthy child sometimes sends it, and the parent sometimes transfers the resource when it receives the signal but not always. This is an equilibrium of the game. Neither the parent nor the chick could do better by switching. But it is not an ESS.

In his initial paper, Maynard Smith was only concerned with offering an explanation of stability, and does not consider other equilibria. In later work (Maynard Smith and Harper, 2003) he seems sensitive to the existence of pooling equilibria and recognizes that one must offer an explanation for why the signaling equilibrium is arrived at rather than another. Even those who prefer an equilibrium based analysis should be concerned. By expanding our purview to consider equilibria beyond ESS, we have multiplied the number of possibilities. If one were to simply stop here, very little could be said about the evolvability of different types of signaling. One simply cannot answer that question now without considering dynamics explicitly.

4.3. **Dynamic analysis.**

When one turns to a dynamic analysis, one finds similar concerns that plagued the Lewis signaling game. In the Sir Philip Sidney game, pooling equilibria remain a problem. One can begin at the "best case" for signaling, when the parent and the child are clones. In such a case, a gain for the parent is a gain for the child and vice versa. In this situation there is no conflict of interest. The game is much like the Lewis signaling game. Just like in the Lewis signaling game there are two ESS where the parent and the chick perfectly coordinate, and like in the Lewis signaling game the dynamic analysis reveals similar concerns. Figure 7 (from Huttegger and Zollman (2010)) shows the basins of attraction for signaling

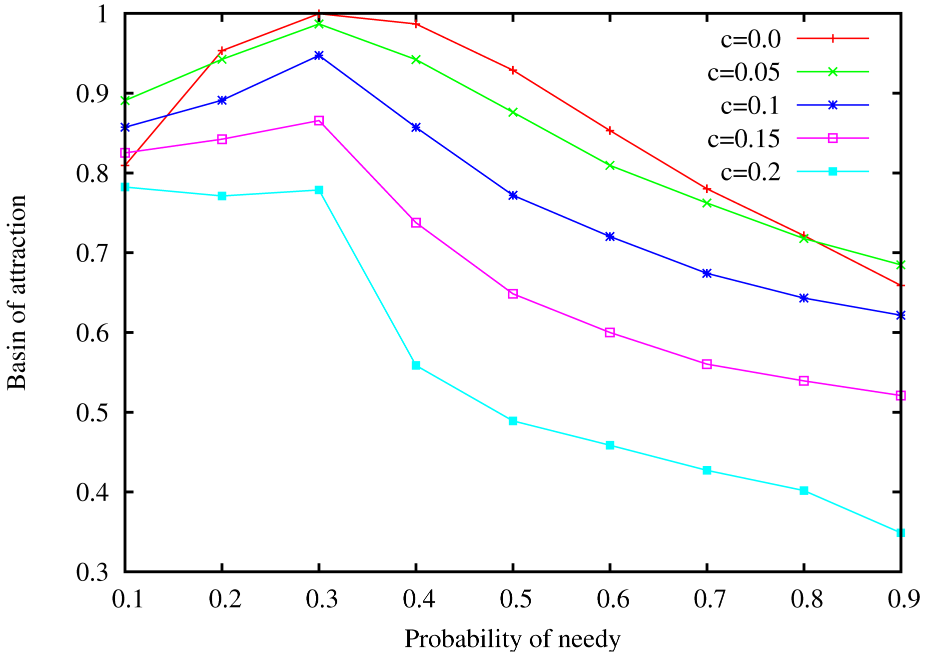

Figure 7. Basins of attraction for the signaling ESS in the Sir Philip Sidney game when the parent and the child are clones

for various amounts of cost. Again one sees that pooling equilibria pose a problem, especially when it is either very likely or very unlikely that the chick is needy.

What about the cases of primary interest to biologists, where there is conflict of interest between the parent and the child? Here too, Huttegger and Zollman find problems. In these situations, where there is a unique strict Nash equilibrium which features signaling, they find that signaling is less likely to evolve than pooling equilibria! In some cases they studied the probability of a random population evolving to signaling was less than 20%.

This presents a significant problem for the equilibrium methodology. In order to see why, let us revisit the explanation it attempts to offer. The question, which began much of biological investigation into signaling, is one about stability. Why do we observe many different instances of signaling in the face of conflict of interest, where there are apparently strong evolutionary forces that should drive us away from that state? The proffered explanation from Zahavi, Grafen, and Maynard Smith centers around cost: this state is stable because the signal carries a sufficiently high cost that it is not "profitable" for the signalers to lie.

In this very limited respect, the static analysis succeeds. That signaling is an ESS in the Sir Philip Sidney game does demonstrate that it is stable under most evolutionary dynamics. So, if our interest is the very narrow question of stability, then we have a legitimate answer. However, this explanation comes at a price. The very same model explains the stability of signaling and indicates that signaling is unlikely to evolve. An evolutionary mystery has been answered by substituting another. This new mystery might lead one to question the adequacy of the answer to the stability question, for if a behavior is made stable only by making it difficult to evolve, it seems unlikely that this is the correct explanation for its stability in the first place.

By restricting themselves only to an equilibrium based approach, those who have analyzed the Sir Philip Sidney game have obscured this fact. Thus, even if the equilibrium method succeeds at being general, this generality comes at a cost. One might produce explanations that, from a wider perspective, seem implausible.

5. Static and Dynamic Approaches

As we have mentioned repeatedly throughout the paper, the main advantage of the equilibrium methodology seems to be its claim of generality. We should expect equilibria like Maynard Smith's ESSs to be observable across a wide range of distinct ecological circumstances. When is this claim of generality justified? There are empirical aspects to this question. In this paper, we have focused on its theoretical aspects, however. We think that the generality claim of an equilibrium methodology also has implications for somewhat more specific models that include dynamical details of evolutionary processes. More precisely, if an equilibrium is claimed to be a very general outcome of evolution, then it should also be a significant state in many dynamic models of evolution. Otherwise, we have reason to doubt that our explanation of a real world state in terms of an equilibrium is correct. We have shown argued that in two biologically relevant situations this generality does not obtain.

In addition to the cases we discuss here, there are other situations that might cause one to worry about the generality of equilibrium based approaches. Some of the problems we discuss here are endemic of games that feature a non-trivial *extensive form* (Huttegger, 2010). Other important games have similar situations, like the widely studied ultimatum game which seeks to model situations of very simple economic exchange (Zollman, 2008). Furthermore, Wagner (2011) has recently provided another example which illustrates a problem with the equilibrium based approach. Wagner studies a zero-sum signaling game, where the sender and the receiver never agree about which action is best to perform in a given state of the world. One would expect when interests are so radically divergent that communication would never emerge, and this is exactly what the equilibrium based approach indicates. However, he shows in the replicator dynamics that one observes chaotic behavior, where some partial communication will emerge only to be later destroyed.

These concerns have shown that the ESS methodology does not always achieve its primary aim: generality. But, even if it were to achieve this, we believe that there are other concerns. In the Sir Philip Sidney game, we have shown how the equilibrium methodology's focus on the question of stability obscured other relevant considerations. The purported explanation for the stability of signaling in the face of partial conflict of interest succeeded, but only by introducing another mystery: how could such a behavior have evolved? If we explain the stability of a behavior only by introducing a model that also suggests this behavior is very difficult to evolve, one might want to question the adequacy of the explanation for stability as well. One would not have realized that such a mystery had been introduced without considering an explicitly dynamic model. This is an independent reason, beyond concerns of generality, to eschew any methodology which focuses exclusively on the question of stability.

Although we have been critical of ESS methodology or, more general, of a static equilibrium methodology, we would like to emphasize that we don't think that ESS and other equilibrium concepts are useless. Quite to the contrary, we think that they are indispensable tools that allow us to get a basic understanding of evolutionary processes without getting tangled up in subtle dynamical considerations right in the beginning of an investigation. What we deny is that one can conclude that a state is a significant evolutionary outcome from the fact that it has been shown to be an equilibrium. We propose that a result like this one is an ingredient in an evolutionary explanation that needs to be supplemented by other results, empirical as well as theoretical. Other theoretical results are particularly important, first of all in order to guide empirical research, and secondly to explore the equilibrium in terms of more specific dynamical assumptions. We thus argue for a pluralistic approach to study evolutionary outcomes that takes advantage of the plethora of methods that are available in evolutionary and mathematical biology.

References

Bergstrom, C. T. and M. Lachmann (1997). Signalling among relatives. I. Is costly signalling too costly? *Philosophical Transactions of the royal Society of London B 352*, 609–617.

Bergstrom, C. T. and P. Godfrey-Smith (1998). On The Evolution of Behavioral Heterogeneity in Individuals and Populations. *Biology and Philosophy 13*, 205–231.

Fisher, R. A. (1915). The evolution of sexual preference. *Eugenics Review 7*, 184–192.

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

Hofbauer, J. (1985). The selection mutation equation. *Journal of Mathematical Biology 23*, 41–53.

Hofbauer, J. and S. Huttegger (2008). Feasibility of communication in binary signaling games. *Journal of Theoretical Biology 254*(4), 843–849.

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

Huttegger, S. (2007, January). Evolution and explanation of meaning. *Philosophy of Science 74*(1), 1–27.

Huttegger, S. (2010). Generic Properties of Evolutionary Games and Adaptationism *Journal of Philosophy 107*(2), 80–102.

Huttegger, S., B. Skyrms, R. Smead, and K. Zollman (2010). Evolutionary dynamics of Lewis signaling games: Signaling systems vs. partial pooling. *Synthese 172*(1), 177–191.

Huttegger, S. and K. J. Zollman (2009). Signaling games: The dynamics of evolution and learning. In A. Benz (Ed.), *Language, Games, and Evolution.*

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

Kerr, B., M. A. Riley, M. W. Feldman, and B. J. M. Bohannan (2002, 11 July). Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature 418*, 171–173.

Lewis, D. (1969). *Convention: A Philosophical Study.* Cambridge: Harvard University Press.

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

Maynard Smith, J. (1991). Honest signaling, the Philip Sidney game. *Animal Behavior 42*, 1034–1035.

Maynard Smith, J. and D. Harper (2003). *Animal signals.* Oxford: Oxford University Press.

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

Moran, P. (1962). *The Stastical Process of Evolutionary Theory.* Oxford: Clarendon Press.

Nowak, M. (1990, January). An evolutionary stable strategy may be inaccessible. *Journal of Theoretical Biology 142*(2), 237–241.

Pawlowitsch, C. (2007). Finite populations choose an efficient language. *Journal of Theoretical Biology 249*, 606–617.

Pawlowitsch, C. (2008). Why does evolution not always lead to an optimal signaling system. *Games and Economic Behavior 63*, 203–226.

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

Sinervo, B. and M. Lively (1996, March 21). The rock-paper-scissors game and the evolution of alternative male strategies. *Nature 380*, 240–243.

Taylor, P. and L. Jonker (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences 40*, 145–156.

Wagner, E. (2009). Communication and structured correlation. *Forthcoming in Erkenntnis*.

Wagner, E. (2011). Deterministic Chaos and the Evolution of Meaning. *British Journal for the Philosophy of Science* forthcoming.

Wärneryd, K. (1993). Cheap talk, coordination, and evolutionary stability. *Games and Economic Behavior 5*, 532–546.

Zahavi, A. (1975). Mate selection – a selection for a handicap. *Journal of Theoretical Biology 53*, 205–214.

Zollman, K. J. (2005). Talking to neighbors: The evolution of regional meaning. *Philosophy of Science 72*, 69–85.

Zollman, K. J. (2008). Explaining Fairness in Complex Environments. *Politics, Philosophy, and Economics 7*(1), 81 – 97.