Generic properties of evolutionary games and adaptationism[*]

Simon M. Huttegger

Department of Logic and Philosophy of Science

3151 Social Science Plaza A

University of California, Irvine, CA 92697-5100

shuttegg@uci.edu

August 26, 2009

Adaptationism maintains that natural selection is the principal factor causing evolutionary change. This statement is in need of qualification. Indeed, Peter Godfrey-Smith finds several possible usages of the term adaptationism.[1] Steven Orzack and Elliott Sober argue that for adaptationism to be an interesting research program it has to be given in a reasonably strong form.[2] Such a reasonably strong version describes adaptationism as committed to the claim that any factor other than natural selection can nearly always be ignored in the explanation of why a certain trait is found in a population. Thus, adaptationism allows one to ignore the role of non-selective factors such as mutation, drift, recombination, or epigenetic processes.

Evolutionary game theory as set forth by John Maynard Smith is often associated with adaptationism.[3] Evolutionary games are considered as nothing but

optimality models that take frequency-dependence into account. As a generalization of optimality models, evolutionary change in strategic settings is exclusively explained by the workings of natural selection. In contrast to this view, I shall put forward the claim that this conception rests on an impoverished view of what constitutes a game. When properly conceived, processes such as drift or mutation play a significant role in determining evolutionary outcomes in games, more specifically, in games with a non-trivial extensive form. Thus, evolutionary games provide a case for including non-selective factors into explanations of the evolution of behavioral traits. I start by briefly characterizing adaptationism and how some standard games conform well with models focusing on natural selection. After introducing extensive form games, which model sequences of moves, I will argue that games will typically have properties that make it necessary to include non-selective factors into evolutionary explanations. Finally, I describe the roles of mutation and drift in dynamic models of evolution in more detail, and I discuss possible criticisms as well as some implications this result has for different forms of adaptationism.

1 Adaptationism

Adaptationism remains a controversial topic to this day.[4] As a research program it is characterized by a strong focus on explaining the evolution of traits by natural selection. If most individuals in a population possess a certain morphological or behavioral trait, then it is exceedingly likely that past selection caused the predominance of the trait. Orzack and Sober propose three ways natural selection can be incorporated in a model of the evolution of a certain trait: natural selection could have some causal influence, an important causal influence, or natural selection could be the only important cause in the evolution of that trait.[5]

It is reasonable to assume that a statement like the third one should count as adaptationist; i.e., non-selective factors such as mutation, drift, epigenetic processes, or constraints can be ignored in a model that provides an adaptationist explanation. To capture this idea, Orzack and Sober introduce the notion of *sufficient explanation*. Natural selection is sufficient to explain a trait if a model which disregards the influence of non-selective factors—a so-called *censored* model—explains the evolution of the trait as well as any more complex model which includes other factors. There are a number of subtle issues involved in this.[6] But for the specific case of evolutionary games I think one can give a fairly appropriate characterization of this idea. In particular, I will focus my arguments on the role of natural selection and ignore the question in what sense possible outcomes of selection processes may be regarded as optimal. I believe this is justified, since adaptationists appear to accept at least the strong emphasis on natural selection in evolutionary explanations regardless of their take on the optimality of evolutionary outcomes.[7]

This specification of an adaptationist model makes it possible to define adaptationism. A reasonable formulation is offered by Orzack and Sober; according to them, the adaptationist thesis reads as follows: "Natural selection is a sufficient explanation for most nonmolecular traits."[8] In other words, adaptationism claims that it is almost always enough to consider censored models in order to explain or predict the prevalence of a phenotypic trait. As the most basic example of such a model consider first a one-locus two-allele population genetic model where the fitness of the A allele is 1 and the fitness of the a allele is $1 - s, s > 0$. If both alleles are present in the population, then selection will carry the A allele to fixation.[9] We can switch from this model to a phenotypic model where we are not talking about two alleles but about two phenotypic traits. It should be emphasized that adaptationism is concerned with nonmolecular, i.e., pheno-

typic traits. Adaptationists often admit that processes like drift may play an important role in molecular evolution; however, they deny that non-selective factors are important for phenotypic evolution.[10]

Models where one phenotype is superior to another in all circumstances are admittedly very simple. Their range of application is quite restricted and they should thus only serve as illustrations. One way to generalize them is given by game theory, where fitnesses, or payoffs, depend essentially on the phenotypes of other organisms. Game theory features prominently in adaptationist explanations of animal behavior. In the most simple game theoretic settings the adaptationist program appears to work perfectly fine. But, as we shall see, looking at slightly more complex (but still biologically meaningful) games changes this assessment considerably.

2 Evolutionary games

One of the paradigmatic games in evolutionary game theory is the Hawk-Dove game.[11] In this game, an individual can act as a hawk or as a dove regarding a resource when meeting another of its kind. If both act hawkishly, then the probability of serious injury is very high for both of them. If one of them acts as a hawk and the other as a dove, then the hawk wins the resource. Finally, if both act dovishly, then they do not fight and the resource is shared. The two individuals can be regarded as players in a two-strategy game, where the first strategy is choosing Hawk, H, and the second choosing Dove, D. The payoff table of the Hawk-Dove game looks like this:

	H	D
H	$0, 0$	$3, 1$
D	$1, 3$	$2, 2$

4

The most fundamental concept in game theory is the concept of a *Nash equilibrium*. Players are at a Nash equilibrium when each of them chooses a strategy that is optimal given the other players' choices; i.e., as long as the other players stay with their strategy choices, then no player has an incentive to switch strategies. In the Hawk-Dove game there are three Nash equilibria. At one the row player chooses H and the column player chooses D. At another these roles are reversed. And there is a third Nash equilibrium in mixed strategies where both players choose H with probability $1/2$ and D with probability $1/2$.

The game theoretic concept of a Nash equilibrium does by itself not tell us much about what might be an evolutionary outcome in a game. John Maynard Smith introduced the concept of an *evolutionarily stable strategy* for such considerations.[12] A strategy in a game is evolutionarily stable if a population who adopts this strategy cannot be invaded by a sufficiently small fraction of individuals playing a different strategy. This characterization of evolutionarily stable strategies can be stated in a mathematically precise way. The interpretation of evolutionarily stable strategies poses some interesting problems.[13] However, for our purposes this rather informal characterization of evolutionary stability is enough. Evolutionarily stable strategies and Nash equilibria are precisely related. Every evolutionarily stable strategy is a Nash equilibrium. The converse of this implication does not hold, however. Thus evolutionary stability is a refinement of the Nash equilibrium concept.

The two pure-strategy Nash equilibria in the Hawk-Dove game are not evolutionarily stable. In a population of Hawks, Doves do better, while Hawks outperform Doves when the population consists almost entirely of Doves. On the other hand, the mixed-strategy Nash equilibrium is evolutionarily stable. This means that if the population is at a state where $1/2$ of the population play H and $1/2$ play D, then small perturbations away from this state—which

correspond to D mutants and H mutants, respectively—will lead back to it. Alternatively, if all individuals in the population adopt the mixed strategy of $1/2$ H and $1/2$ D, then no mutant strategy employing a different mix can do better. The reason for the evolutionary stability of the mixed equilibrium is that each slight over-representation of H leaves D better off, while in a population with a D share of more than a half, H yields a higher payoff.

Games can be a little more complex than this. Another well studied game in evolutionary game theory is the Rock-Scissors-Paper game, where the strategies are denoted by R, S and P. In this game there is a cycle of best responses: R beats S, S beats P, and P beats R. There are two players, each of them choosing a strategy simultaneously. Moreover, there is only one Nash equilibrium where both players choose each strategy with probability $1/3$. This Nash equilibrium is not evolutionarily stable, however. The payoff table of the Rock-Scissors-Paper game looks as follows:

	R	P	S
R	$0,0$	$1,-1$	$-1,1$
P	$-1,1$	$0,0$	$1,-1$
S	$1,-1$	$-1,1$	$0,0$

The difference between a game like Hawk-Dove and Rock-Scissors-Paper is best brought out when taking into account evolutionary dynamics. This will also serve to illustrate a point that will be of some significance below. It is important to bear in mind that the concept of evolutionary stability is static, in the sense that it does not by itself answer the question if a population evolves to an evolutionarily stable strategy or state. Evolutionary stability only answers the question why a population stays at a certain state and not how it got there. The causal role of natural selection in the evolution of a trait is therefore insuffi-

6

ciently modeled by evolutionary stability alone. Considering some evolutionary dynamics explicitly is inevitable if one wants to get a thorough understanding of the role of natural selection.

The basic equations describing selection dynamics are given by the replicator dynamics.[14] Suppose we have a game like Hawk-Dove. In a population of Hawks and Doves, let x be the frequency of H and $y = 1 - x$ the frequency of D. Then the replicator dynamics of the Hawk-Dove game above can be given by one equation:

$$\dot{x} = xy(3 - 4x - 2y) = x(1 - x)(1 - 2x)$$

This follows from the general replicator dynamics. If there are types (strategies) $1, \ldots, n$ in the population and if x_i denotes the relative frequency of type i, then the replicator dynamics is given by

$$\dot{x}_i = x_i(u(i, x) - u(x, x)),$$

where $x = (x_1, \ldots, x_n)$ is the state of the population, $u(i, x)$ is the fitness of type i when the state of the population is x, and $u(x, x)$ is the average fitness of the population. The population is assumed to be (infinitely) large, so the fitness values can be identified with the expected payoffs coming from the payoff matrix of the underlying two-player game.[15]

In the Hawk-Dove game every population state except $x = 0$ or $x = 1$ evolves to the evolutionarily stable strategy under the replicator dynamics. The polymorphism of H and D is not only asymptotically stable—all nearby states stay close and converge to it; it is also globally stable—the set of states not converging to it is negligible. Thus, if one would like to explain a mixture of hawkish and dovish behavior in a population, then one could try to argue that natural selection explains the evolution of the mixture sufficiently well. The replicator

dynamics is a general model of selection where types with above average fitness increase in frequency and types with below average fitness decrease in frequency. Moreover, one just has to assume implicitly that there is some kind of mutation that pushes the population away from the pure states $x = 0$ and $x = 1$. The precise nature of the mutation does not matter, however. Other non-selective processes can be ignored. Thus, the Hawk-Dove game is an excellent example for an adaptationist explanation.[16]

The Rock-Scissors-Paper game provides more of a challenge. With the pay-offs as given above, the unique Nash equilibrium is not asymptotically stable. But it is stable in a weaker sense. All evolutionary trajectories are cycling around the unique Nash equilibrium. This implies that trajectories close to it remain close, although they do not converge to the Nash equilibrium. The example of the Rock-Scissors-Paper game shows that natural selection alone does not necessarily lead to an equilibrium. More complicated dynamical behavior is also possible.[17] We may find natural phenomena other than convergence to an equilibrium where natural selection alone could nevertheless explain the phenomenon sufficiently well.
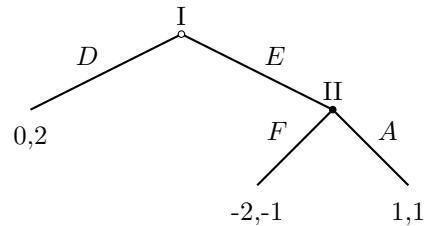
Adaptationist explanations work well for games like Hawk-Dove. They might also work well for more subtle games like Rock-Scissors-Paper. They work well in the sense that the replicator dynamics is an exemplary censored model in the sense of Orzack and Sober. Selection is the only factor causing change in the vector field that is generated by the replicator equations. But the question arises if we should expect the same nice fit between evolutionary games and adaptationist explanations in general. In order to answer this question we will have to look at the general structure of games that are important for evolutionary biology.
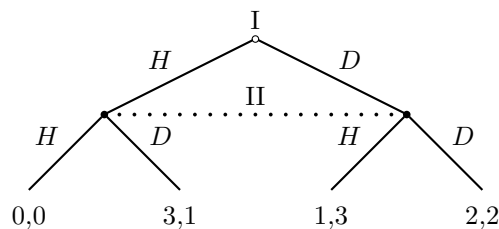
3   Normal form and extensive form

John von Neumann and Oskar Morgenstern started their development of the theory of games by considering games in *extensive form*.[18] A game in extensive form models the sequence of moves of players which determines the outcome of the game. This can be done by using the notion of a *game tree*. A game tree is a finite collection of ordered nodes where the ordering expresses a precedence relation. The precedence relation is assumed to be such that the game tree is connected, has one root, no cycles, and each node except the root is assumed to have exactly one predecessor. At each node, one of the players chooses between several actions, the number of which coincides with the number of nodes that are its successors. At each stage of the game nodes may be part of an *information set*. This means that the player does not know at which node in the information set she is. Random events are incorporated into extensive-form games by adding Nature as an additional player. At the nodes where Nature chooses, successor nodes are reached according to a fixed probability distribution.

Another standard assumption imposed on many extensive-form games is that they be games of *perfect recall*. This means that no player forgets any information she once had. In fact, the game considered throughout most of the rest of this paper will meet an even stronger requirement. In games of *perfect information* a player is never uncertain about the node she currently occupies. That is to say, all information sets are singletons. An important example of an extensive-form game of perfect information is provided by Reinhard Selten's Chain-Store game.[19] In this game, player I can decide to enter a market ,E, that is dominated by a chain store. If she does not enter, D, the market share of the chain store remains unchanged. If she decides to enter, player II (the chain store) can choose between fighting (dumping prices), F, and acquiescing, A. If she chooses to fight, both loose, but player I more heavily. If she decides

to acquiesce, I gets half of her share of the market. This decision situation is depicted in the following game tree:

$$
\begin{array}{c}
\text{I} \\
D \quad \quad E \\
\quad \quad \quad \text{II} \\
0,2 \quad \quad F \quad \quad A \\
\quad \quad \quad -2,-1 \quad \quad 1,1
\end{array}
$$

Notice the similarities between the Chain-Store game and the Hawk-Dove game. A biological interpretation of the Chain-Store game may describe the conflict between an owner and an intruder. In both games players have to make a decision between fighting and not fighting. The important difference between the two games is in the sequence of moves. In the Chain-Store, game player I chooses first, while in Hawk-Dove both players decide simultaneously. An extensive-form representation of the Hawk-Dove game would look differently. Player II does not know whether I is a hawk or a dove. Therefore, her information set includes two nodes:

$$
\begin{array}{c}
\text{I} \\
H \quad \quad D \\
\quad \text{II} \\
H \quad D \quad \quad H \quad D \\
0,0 \quad 3,1 \quad \quad 1,3 \quad 2,2
\end{array}
$$

As we have already seen in the previous section, games need not be given in extensive form. The representation used there was termed *normal form* by von Neumann and Morgenstern, and it is now often called the *strategic form* of a game. A game in strategic form consists of a set of players, a set of pure strategies for each player, and a utility function for each player which associates a real number—the player's payoff—with each *strategy profile*. A strategy profile

is a possible combination of the player's pure strategies. Two-player strategic-form games can be presented as payoff tables as in the case of Hawk-Dove and Rock-Scissors-Paper.

Analogously to the Hawk-Dove game, every game in strategic form can be represented by a game in extensive form. Von Neumann and Morgenstern have shown that the converse also holds. Every extensive-form game can be given as a game in strategic form. A pure strategy of a player simply determines the decisions of that player at each of her information sets. The set of all these strategies is the player's strategy set. Her utility function carries over from the extensive form game as well (where possible random moves can be taken into account by taking expected values).[20] The strategic-form of the Chain Store game looks therefore like this:

	F	A
N	$0,2$	$0,2$
E	$-2,-1$	$1,1$

It is important to notice that, from a game theoretic point of view, the transformation of extensive-form games into strategic-form games is more important than the converse transformation. The most basic game theoretic solution concepts, such as Nash equilibrium or the elimination of dominated strategies, are defined for the strategic form of a game. Moreover, the above considerations suggest that extensive-form games can always be given as strategic-form games without loosing any relevant information—in fact, this is what von Neumann and Morgenstern thought. This view has not gone unchallenged, though.[21] One reason for this will be given in the next section. As it turns out, considering certain properties of extensive-form games has also important consequences for adaptationism.

4 Payoff genericity and structural stability

We may ask ourselves what would happen if we changed the payoffs of a strategic-form game like Hawk-Dove slightly. After all, we never know the utility functions or fitness values with complete accuracy, so it is important to take arbitrarily small changes of the numbers that represent fitnesses and utilities into account. In doing this, we see that in the Hawk-Dove game the structure of Nash equilibria is invariant with respect to sufficiently small perturbations of the players' payoff entries. We still have two pure-strategy Nash equilibria where one player chooses H and the other chooses D. And we also still have a mixed Nash equilibrium that lies close to the mixed Nash equilibrium of the original game.

These facts do not just hold for the Hawk-Dove game but for strategic-form games in general. This follows from an important theorem in game theory.[22] Given the number of players and their sets of pure strategies, one may think of a game as completely specified by the players' payoffs. This means that we can parametrize games in terms of payoff entries. For instance, if we are looking at two-player games where each player has two strategies, then one specification of the payoffs will result in the Hawk-Dove game, while another specification will result in a different game such as, e.g., the Prisoner's Dilemma. This allows one to view a game as a subset of some finite-dimensional real space. A property is said to hold for *almost all games* if the set of payoff values for which the property does not hold has measure zero. Thus, the payoff values for which the property does not hold are confined to a subspace of the space of all games. To view games as points in a finite dimensional real space also allows one to talk about how close two games are. This can be done by introducing a metric on the real space of games. The exact definition of the metric is not so important, since all metrics induce the same topology on a finite-dimensional real space. A

Nash equilibrium is called *essential* if there exists a nearby Nash equilibrium for all nearby games. Nearby here means sufficiently close with respect to a metric. A *game* is called *essential* if all its Nash equilibria are essential. The theorem of Wu and Jiang states that *almost all finite strategic-form games are essential.*

An analogous result does not hold for extensive-form games, however. To see this, consider the extensive form of the Chain-Store game. If player I chooses N, then, no matter what strategy player II chooses, they always end up with the same payoffs of 0 and 2. Perturbing these payoffs will still result in a situation where the decision of player II has no influence on the outcome of the game whenever 1 chooses N. This results in a strategic-form game that is not essential. If I chooses N and it is sufficiently probable that II chooses F, then the players choose according to a Nash equilibrium of the Chain-Store game. To be more specific, if II chooses F with probability p and A with probability $1 - p$, then I is better off with playing N as long as

$$-2p + (1 - p) < 0,$$

i.e., as long as $p > 1/3$. Hence there is a *continuum* of Nash equilibria where player I chooses N with probability 1 and II chooses F with probability $p > 1/3$. This result continues to hold if the payoffs in the game tree are slightly perturbed. However, a continuum of Nash equilibria always contains a Nash equilibrium which is not essential: for some sufficiently close strategic-form game there is no nearby Nash equilibrium. This can, e.g., be seen by perturbing the payoff entries of the strategic-form representation of the Chain-Store game slightly. Therefore, for many game trees the payoffs are often constrained to be elements of a subspace of the space of corresponding strategic-form payoffs. Continua of Nash equilibria may always emerge when there are decision nodes of information sets that are unreached in a Nash equilibrium.[23]

Considering payoff perturbations is one way to assess the robustness of the properties of a game theoretic model.[24] When we move from a static analysis of games to a dynamic analysis, another type of robustness becomes very important. In the literature on dynamical systems this type of robustness is known as *structural stability*.[25] Consider the replicator dynamics of the Hawk-Dove game. We may ask the question what happens to the evolutionary trajectories when we perturb the replicator equations just slightly. To do this, we have to consider the replicator dynamics of the Hawk-Dove game as a point in a metric function space. The elements of this space are functions and the metric measures similarity according to whether the functions and its partial derivatives are close to each other.

Once we are able to consider dynamical systems close to the replicator equations, we can investigate the problem whether there is a qualitative change in the evolutionary trajectories of the replicator dynamics when we consider an arbitrarily close dynamics. If the evolutionary trajectories of the replicator dynamics and *all* sufficiently close dynamics are *topologically equivalent*, then the replicator dynamics is structurally stable. Topological equivalence means that there is a homeomorphism—a continuous, one-to-one transformation with continuous inverse—between the evolutionary trajectories of the replicator dynamics and another dynamics. To put it informally, the trajectories of topologically equivalent systems look the same up to various distortion operations such as stretching. From a qualitative perspective, topologically equivalent systems can be regarded as being the same. Topological equivalence defines equivalence classes of dynamical systems. Structural stability of a dynamical system states that a neighborhood of this system in function space is contained in its equivalence class. As an example of structural instability you may think of an asymptotically stable rest point of the dynamics becoming unstable for arbitrar-

ily small perturbations. A particularly exciting example is provided by a Hopf bifurcation, where an asymptotically stable rest point not only becomes unstable, but is also located within a periodic attractor which suddenly appears after the bifurcation. It should be emphasized that these transitions are not continuous; a new qualitative behavior appears suddenly due to continuous changes in the parameters.

Considering structural stability of the replicator equations (and of other dynamical systems as well) is important since we can never be sure whether the description of processes provided by the equations is completely accurate. If a plausible perturbation of the replicator dynamics leads to a different qualitative behavior of the system, then this calls the original results into question. It also indicates that a deeper theoretical and empirical analysis of the game is necessary.

The replicator dynamics of the Hawk-Dove game is structurally stable. This follows from Peixoto's theorem.[26] The replicator dynamics of the Rock-Scissors-Paper game is not structurally stable. This follows from the fact that the unique interior rest point of the dynamics (which coincides with the unique Nash equilibrium) is a center; i.e., all eigenvalues of the Jacobian matrix evaluated at this rest point are purely imaginary. This implies that arbitrarily small perturbations may change the zero real parts of the eigenvalues, yielding a different qualitative behavior of the evolutionary trajectories. A perturbation of the payoffs already indicates the same. Some payoff perturbations of the Rock-Scissors-Paper game turn the unique Nash equilibrium into a sink and some into a source.[27] It should be noted that the Rock-Scissors-Paper game is essential. This implies that we will always find a Nash equilibrium close to the original one after a perturbation of the payoffs. Qualitative changes of the replicator dynamics close to the Nash equilibrium will therefore change

its stability properties. This should not be too surprising, after all, since the replicator dynamics of Rock-Scissors-Paper with perturbed payoffs constitutes a possible perturbation of the replicator dynamics.

5   The roles of drift and mutation

What is the relation between essential Nash equilibria and structural stability on the one hand, and adaptationism on the other? This will become clear once we study the robustness properties of extensive-form games like the Chain-Store game. We have seen in the previous section that continua of Nash equilibria are unavoidable in the Chain-Store game; payoff perturbations of the extensive-form game will not destroy them. When we consider the replicator dynamics of the strategic-form game, there will thus exist a continuum of rest points since every Nash equilibrium is a rest point of the replicator dynamics. The figure below represents these rest points as black dots in the upper left part of the phase portrait:
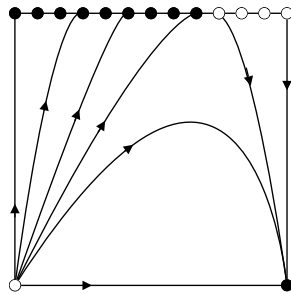
The evolutionary dynamics here does not arise from the replicator dynamics as described previously, but from the two-population replicator dynamics where we have one population for the role of player I and another population for the role of player II. In an evolutionary context one may think of inter-

actions between males and females, or of inter-species interactions. Moreover, the two-population replicator dynamics can be imbedded in a one-population replicator dynamics where each individual can be in both roles of the game.[28] For illustrative purposes the two-population model is much better suited. The two-population replicator dynamics is given by

$$\dot{x}_i = x_i(u(i,y) - u(x,y))$$

$$\dot{y}_j = x_j(u(j,x) - u(y,x)),$$

where x_i and y_j are the relative frequencies of types in the first and the second population, respectively; $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_m)$ describe the states of the two populations; and $u(i,y)$ and $u(j,x)$ are the payoffs to strategies i and j, while $u(x,y)$ and $u(y,x)$ are the population-specific average payoffs. As in the one-population replicator dynamics, the expressions in parentheses evaluate a type's payoff when interacting with the other population relative to the average payoff in its own population.

The figure above expresses several basic facts about the selection dynamics of the Chain-Store game. In the lower left corner there is an asymptotically stable rest point which corresponds to the subgame perfect Nash equilibrium of the game. An open set of evolutionary trajectories converges to the component of rest points, however. This means that this component cannot be ignored in an analysis of the game. Along the component there is no selective pressure since there are no payoff differences. Thus, once it is reached by the replicator dynamics, *forces that remain unmodeled in the replicator dynamics will determine the further evolution of the system.* This implies that drift has a strong influence on the evolutionary outcome as long as we consider the replicator dynamics as the right evolutionary model. The replicator dynamics is a censored model in the sense of Orzack and Sober, however. Thus, even if we start with a

model that takes only natural selection into account, we end up in a situation where non-selective factors play a significant role in dtermining evolutionary outcomes.

Along the component of Nash equilibria, the population may be pushed to the left and end up in the state $(x, y) = (0, 1)$ due to stochastic events. Or it may be pushed to the right and eventually leave the component of rest points again when $p < 1/3$. In this case the population enters the basin of attraction of the subgame perfect equilibrium and selection takes over. If the population is very large the effects of drift will not be very pronounced. The population may drift along the component for a very long time. Whatever happens, we need to include processes other than natural selection in order to explain the evolutionary outcome. Selection alone does therefore not provide a sufficient explanation for the evolutionary outcomes of the model.

In a strategic-form game such reasoning could be countered by appealing to payoff robustness. As we have seen, continua of Nash equilibria are not essential. Moreover, almost all strategic-form games have a finite number of Nash equilibria.[29] Hence, for strategic-form games the argument given above has no teeth. It would depend on a degenerate specification of the payoff parameters. This counter-argument does not hold for games in extensive form, however. As we have seen, the continuum of Nash equilibria is robust regarding payoff perturbations of the extensive form. Thus, drift remains an important factor regardless of particular payoff choices.

There is another way to reach a similar conclusion which is based on structural stability. Like the Rock-Scissors-Paper game, Chain-Store-like games are not structurally stable. This follows from the existence of a continuum of rest points (which implies eigenvalues of the Jacobian matrix at those rest points with zero real part). Such a continuum will not persist under perturbations of

the dynamics. But unlike the Rock-Scissors-Paper game, the fact of structural instability persists under payoff perturbations in the Chain-Store game.

These results are important when we study a specific perturbation of the replicator dynamics, the *selection-mutation dynamics*.[30] Recall that the replicator dynamics does not model mutation explicitly. Rather, one uses implicit arguments to the effect that mutation provides enough variation such that every type will be present in the initial population. The selection-mutation dynamics includes mutation terms explicitly. For the two-population version it can be given by

$$\dot{x}_i = x_i(u(i,y) - u(x,y)) + \varepsilon(1 - nx_i)$$
$$\dot{y}_j = x_j(u(j,x) - u(y,x)) + \delta(1 - my_j),$$

where ε and δ are *uniform* mutation rates within each population, and n, m are the numbers of strategies in each population, respectively. The last term of each equation states that at each point in time the same share of each type mutates into any other type with equal probability. This is a simplification that allows one to derive analytical results in certain cases. (In the next section, I will remark on what happens when mutation rates are not assumed to be uniform.) Notice that we do not assume ε and δ to be equal. We will assume, however, that ε and δ are of the same order of magnitude as they go to zero.[31]

We are now in the position to derive the selection-mutation dynamics of the Chain-Store game, where $\bar{a} = 2x_E y_F - x_E(1 - y_F)$:

$$\dot{x}_E = x_E(1 - 3y_F - \bar{a}) + \varepsilon(1 - 2x_E)$$
$$\dot{y}_F = 2y_F x_E(y_F - 1) + \delta(1 - 2y_F)$$

Since $x_N = 1 - x_E$ and $y_A = 1 - y_F$, the two-population selection-mutation dy-

namics of the Chain-Store game is completely specified by these two differential equations. We are interested in what happens to the component of Nash equilibria under the perturbation. Will it disappear, or will there be rest points close to the component? To answer these questions, notice we look at the selection-mutation dynamics close to $x_E = 0$. A Taylor expansion in terms of ε based on the rest point condition

$$\dot{x}_E = x_E(1 - 3y_F - \bar{a}) + \varepsilon(1 - 2x_E) = 0$$

yields

$$x_E = \frac{\varepsilon}{3y_F - 1} + \text{ higher order terms in } \varepsilon.$$

Hence, up to higher order terms in ε,

$$\dot{y}_F = 2y_F \frac{\varepsilon}{3y_F - 1}(y_F - 1) + \delta(1 - 2y_F) = 0$$

or

$$\frac{\varepsilon}{\delta} = \frac{(2y_F - 1)(3y_F - 1)}{2y_F(y_F - 1)}.$$

Differentiating the right-hand side of this equation with respect to y_F shows that it has a unique maximum for $y_F \in [0, 1]$ at $y_F = -1 + \sqrt{2}$. The maximum value is $3/2 - \sqrt{2}$. Thus if

$$\frac{\varepsilon}{\delta} < \frac{3}{2} - \sqrt{2},$$

then there exists a rest point close to $x_E = 0, y_F > 1/3$. When we reverse the inequality sign, there exists no such rest point. Depending on the ratio ε/δ, all trajectories will either converge to the subgame perfect equilibrium, or there will be another rest point. In the latter case, index theory[32] establishes that there exists at least one further rest point. One of them is asymptotically stable and the other one is a saddle. This is illustrated in the following phase portrait:
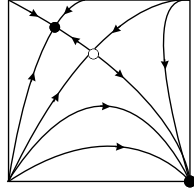
Thus, in this case there exists an asymptotically stable rest point which corresponds to a Nash equilibrium where player II is prepared to fight and player I does not enter. The following phase portrait depicts the situation where the subgame perfect equilibrium is globally stable (if ε/δ is sufficiently high):
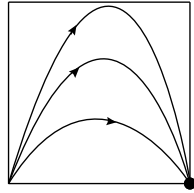
To summarize, for the replicator dynamics drift plays an essential role. Drift will play an equally essential role for all evolutionary dynamics that respect the equilibrium structure of the game, i.e. for which all Nash equilibria are rest points. In the selection-mutation dynamics, mutation plays an essential part in determining the range of evolutionary outcomes. Other biologically plausible perturbations will also have to include non-selective factors. This follows from the fact that selection results from payoff differences, which are absent along non-singleton Nash sets.

It should be noted that the Chain-Store game is by no means an exception.[33]

Non-singleton Nash sets may appear whenever decision nodes or information sets are unreached in a Nash equilibrium. Therefore, when considering games in extensive form, evolutionary game theory ceases to fit into a strictly adaptationist framework. In general, natural selection alone does not explain evolutionary outcomes in game theoretic models.

6   Discussion

I argued in the previous section that if we consider the replicator dynamics as the appropriate evolutionary dynamics, then drift enters as an inevitable factor for many extensive-form games such as the Chain-Store game. One could object to that conclusion with an argument that is based on the logic of subgame perfectness.[34] Reinhard Selten developed a trembling-hand approach to evolutionary stability.[35] The logic behind this reasoning is quite simple. Consider the line of Nash equilibria in the Chain-Store game. In all these equilibria player II does not reach her decision node. But what if player I "trembles", i.e. makes a mistake and chooses E? Then player II would have to play F with positive probability, which gives her a suboptimal payoff once she is at the subgame which starts at her decision node. Hence, any Nash equilibrium other than (E, A) is not subgame perfect.

This is a potential criticism to what I have stated above. A trembling-hand approach would force us to consider a game without a non-singleton Nash set instead of our original game. The original game would be a degenerate limiting case when we let the probability of trembling go to zero. But, as Ken Binmore has convincingly argued with respect to the rational choice concept of subgame perfectness, trembling-hand considerations are not the only or the most reasonable way to deal with unreached information sets.[36] The trembling-hand paradigm effectively assumes that the mistakes of players are probabilistically

independent noise. But without considering equilibrating processes, this assumption is quite arbitrary. A similar reasoning applies to the trembling-hand approach for evolutionary stability in extensive-form games. Without considering what happens to the evolutionary dynamics close to a component of Nash equilibria when players tremble, not much can be said about the evolutionary significance of the component. If player II were called upon to choose at her decision node in the Chain-Store game, then choosing A is better than choosing F. But how often is her decision node reached compared to when it is not reached? How does player I act and in what way are her actions stochastic? In order to analyze these problems, one has to take account of the dynamics close to Nash components.

Studying the selection-mutation dynamics close to a component reveals some of this information. One should bear in mind, however, that the selection-mutation equations as given above have an extremely simple structure. The relationships of mutation between the strategies are very symmetric and idealized. This need not be the case. Such relations may be highly non-linear and asymmetric. In this case, many different scenarios would obtain close to the component of Nash equilibria, e.g. any number of equilibria with all kinds of stability properties.[37] The significance of the result above derives from the fact that even in this most benign case there is a non-trivial qualitative change in the evolutionary dynamics which depends essentially on the mutation rates.

Ken Binmore and Larry Samuelson studied learning dynamics close to Nash components quite generally.[38] Their learning dynamics does have a very general functional form and can straightforwardly be considered as a perturbed selection dynamics. The effects of the selection part of the dynamics become very small when populations come close to a component of Nash equilibria, because the payoff differences go to zero. Thus, close to Nash components factors Binmore

and Samuelson collectively call "drift" will play a crucial role in determining the properties of the dynamical system. They prove under quite general conditions that drift—i.e. originally unmodeled factors—can stabilize any Nash component such as the one that we have encountered in the Chain-Store game. Hofbauer and Huttegger provide similar results to the one given in the previous section for signaling games, which are another biologically relevant example.[39] Games of indirect and direct reciprocity provide further examples where continua of Nash equilibria are relevant.[40]

Another possible criticism may be based on the following reasoning. In the selection-mutation dynamics the ratio of mutation rates does indeed influence the evolutionary outcome. But may selection not also adjust the mutation rates such that a population will reach one specific evolutionary outcome? In other words, would natural selection not favor individuals which have the "right" mutation rate given the mutation rate in the other population?

To answer this question, notice first that for the Chain-Store game we have to assume that the subgame perfect Nash equilibrium is in some way better than the other Nash equilibria, since it is a rest point of the selection-mutation dynamics for any sufficiently small ε, δ. This is not an innocuous statement, since the concept of subgame perfectness is not vindicated from an evolutionary point of view.[41] But let us grant it for the moment. In this case, drift enters the picture again. ε and δ may change continuously, but the only relation that matters regarding which evolutionary outcomes are possible is whether the ratio ε/δ is big enough. Thus, small changes to a mutation rate will result in the same possible evolutionary outcomes. The phenotype of having a certain mutation rate thus can have only two consequences. I would like to emphasize, though, that the fitness consequences of having a certain mutation rate are far from clear in the case of the Chain-Store game.

The ubiquity of non-singleton Nash sets similar to the one of the Chain-Store game has profound consequences for adaptationism. Godfrey-Smith distinguishes between three kinds of adaptationism: empirical, explanatory and methodological adaptationism.[42] Empirical adaptationism claims that natural selection is in fact the only important causal force in determining evolutionary trajectories. This form of adaptationism is prone to well known criticisms from population genetics. This appears not to be the case for explanatory adaptationism, which holds that though other processes may be important at, e.g., the molecular level (such as genetic drift or details of mechanisms underlying mutation), natural selection is the only important factor in explaining phenotypic adaptations of organisms (i.e. their apparent design). Methodological adaptationism, finally, states that looking for evolutionary explanations in terms of natural selection is a good methodological maxim.[43]

Concerning empirical adaptationism, the conclusions to be drawn from my arguments are not conclusive in themselves, since arguments solely based on theoretical models cannot decide empirical issues. But the explanatory non-sufficiency of selection in extensive-form games indicates that we should not expect empirical adaptationism to be adequate in general. Interactions that involve sequences of moves of several players are presumably abundant between all kinds of organisms. In games modeling these interactions, non-singleton Nash sets where selection ceases to work are inevitable. From a theoretical viewpoint it seems thus to be likely that in order to explain certain behaviors one must also include non-selective factors in an essential way.

Explanatory adaptationism seems to be quite suspect given the arguments from extensive-form games. Consider again the Chain-Store game. Both the subgame perfect equilibrium and the equilibria in the other Nash set are evolutionarily optimal in the minimal sense that selection stops to work once one

of these states is reached. Hence both kinds of equilibria describe a possible behavioral adaptation in a very basic sense. Even if we grant the point that the apparent adaptations of organisms constitute the most important problem of biology (which is not an obvious point), explaining which behavioral adaptation evolves in the Chain-Store game requires something more than just appealing to selection.

Finally, the prospects of methodological adaptationism also do not appear to be enhanced by what we know from extensive-form games. As long as one is working with games which have a trivial extensive-form structure (i.e. simultaneous move games), one need not worry too much about non-selective factors. This does not hold for games with a non-trivial extensive-form structure. In this case one should not proceed as if selection were the only important causal factor in the evolution of a trait. Once possible obstacles like the Nash component in the Chain-Store game are identified, a more careful analysis has to be undertaken, which includes formal analysis as demonstrated here for the Chain-Store game. Understanding the results from different kinds of perturbations may then allow one to make predictions (such as the ratio of mutation rates) which can in principle be tested.[44]

7   Conclusion

We have seen that for games in extensive form, the adaptationist program is not feasible from the point of view of evolutionary game theory. Information sets that are unreached in a Nash equilibrium play an important role in determining evolutionary outcomes. They allow non-selective processes to become significant causal factors for influencing evolutionary trajectories. This conclusion is about behavioral traits that arise from frequency-dependent interactions. Perhaps a similar argument regarding morphological traits can be put forward. An

analysis in terms of other evolutionary formalisms, such as the Price equation or other descriptions of evolution by natural selection would also be desirable.[45] It remains to be seen whether my argument will carry over to these other contexts.

Notes

[1] Godfrey-Smith, "Three Kinds of Adaptationism," in S. H. Orzack and E. Sober, eds., *Adaptationism and Optimality* (New York: Cambridge, 2001), pp. 353-357.

[2] Orzack and Sober, "Optimality Models and the Test of Adaptationism," *American Naturalist*, CXLIII (1994): 361-380.

[3] See Maynard Smith and G. Price, "The Logic of Animal Conflict," *Nature*, CXLVI (1973): 15-18; Maynard Smith, *Evolution and the Theory of Games* (New York: Cambridge, 1982). For the relation between adaptationism, optimality models and evolutionary game theory see in particular Maynard Smith, "Optimization Theory in Evolution," *Annual Review of Ecological Systems*, IX (1978), 31-56, and G. A. Parker and Maynard Smith, "Optimality Theory in Evolutionary Biology," *Nature*, CCCXLVIII (1990): 27-33.

[4] Some important early papers are R. C. Lewontin, "Adaptation," *Scientific American*, CCIIIIX (1978): 212-228, S. J. Gould and Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," *Proceedings of the Royal Society B*, CCV (1979): 581-598, and Maynard Smith, "Optimization Theory in Evolution." An overview of the discussion is given by M. Pigliucci and J. Kaplan, "The Fall and Rise of Dr. Pangloss: Adaptationism and the Spandrels Paper 20 Years Later," *Trends in Ecology and Evolution*, XV: 66-70; see also Orzack and Sober, eds., '*Adaptationism and Optimality* (New York: Cambridge, 2001); and P. Forber, ed., *Spandrels and Adaptationism: Marking Thirty Years of Controversy*, Special Issue of *Biology and Philosophy*, XXIV (2009): 155-280.

[5] See Orzack and Sober, "Optimality Models and the Test of Adaptationism."

[6] See the exchange between Orzack and Sober on the one hand and Robert Brandon and Mark Rausher on the other. Orzack and Sober, "Optimality Models and the Test of Adaptationism"; and "How to Formulate and Test Adaptationism," *American Naturalist*, CXLVIII (1996): 202-210; and R. N. Brandon and M. D. Rausher, "Testing Adaptationism: A Comment on Orzack and Sober," *American Naturalist*, CXLVIII (1996): 189-201. Later on, Sober prefers to specify censored models in terms of predictions instead of explanations; see Sober, *Philosophy of Biology* (Boulder: Westview, 2000). This does not alter my arguments, though.

[7] Cf. Parker and Maynard Smith, who shift the emphasis from the optimality of adaptations

to the claim that natural selection is the main causal factor in the evolution of adaptations—be they perfect or not. Sober does not include the optimality of phenotypes in his formulation of adaptationism as well and focuses on selection instead; see Sober, *Philosophy of Biology* (Boulder: Westview, 2000).

[8] Orzack and Sober, "Optimality Models and the Test of Adaptationism," p. 364.

[9] This holds across different dynamical models; see R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation* (New York: Wiley, 2000) for a comprehensive overview.

[10] E.g., Maynard Smith, "Optimization Theory in Evolution."

[11] See Maynard Smith and Price.

[12] See Maynard Smith and G. Price; Maynard Smith, *Evolution and the Theory of Games.*

[13] For instance, evolutionarily stable strategies might not exist, such as in the Rock-Scissors-Paper game that we consider below. Moreover, there is an ambiguity in interpreting evolutionarily stable strategies. They can be interpreted either as population states or as a population adopting a certain strategy. In the case of mixed strategies this leads to ambiguities. On this point, see Orzack and Sober, "Optimality Models and the Test of Adaptationism"; Bergstrom and Godfrey-Smith "On the Evolution of Behavioral Complexity in Individuals and Populations," *Biology and Philosophy*, XIII (1998): 205-231. The classical definition of evolutionary stability is also not applicable to finite populations; see Maynard Smith, "Can a Mixed Strategy be Stable in a Finite Population," *Journal of Theoretical Biology*, CXXX (1988): 247-251; M. E. Schaffer, "Evolutionarily Stable Strategies for a Finite Population and a Variable Contest Size," *Journal of Theoretical Biology*, CXXXII (1988): 469-478; and M. A. Nowak, A. Sasaaki, C. Taylor, and D. Fudenberg, "Emergence of Cooperation and Evolutionary Stability in Finite Populations," *Nature*, CCCCIIVIII (2004): 646-650.

[14] For a comprehensive treatment see J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (New York: Cambridge, 1998).

[15] If there are n types, then one equation can be ignored since $x_i = \sum_{j \neq i} x_j$.

[16] The Hawk-Dove game is also a nice counterexample to a crude kind of adaptationism that emphasizes the optimality of evolutionary outcomes, since the mixed equilibrium may not be very good. For this reason Hawk-Dove-like situations go together well with a more subtle version with a gene-eye view of evolution such as in K. Sterelny and P. Kitcher, "The Return of the Gene," this JOURNAL, LIIIV (1988): 339-361.

[17] For chaotic behavior in game dynamics see W. Schnabl, P. F. Stadler, C. V. Forst, and P. Schuster, "Full Characterization of a Strange Attractor: Chaotic Dynamics in Low Dimensional Replicator Systems," *Physica D*, XLVIII (1991): 65-90.

[18]See von Neumann and Morgenstern, *Theory of Games and Economic Behavior* (Princeton: Princeton, 1944). For an introduction to game theory see, e.g., D. Fudenberg and J. Tirole, *Game Theory* (Cambridge: MIT, 1991).

[19]Selten, "The Chain Store Paradox," *Theory and Decision*, IX (1978): 127-159.

[20]To be more specific, there is more than one strategic form associated with a game in extensive form. Other strategic forms eliminate certain equivalent strategies from the strategic-form representation, e.g. payoff equivalent strategies. Moreover, a game in extensive form can also be described by behavior strategies. A behavior strategy is a probability distribution over actions at an information set in a game tree. For our chief example—the Chain-Store game— all these representations yield the same payoff table. For more information on strategic-form representations, also in relation to the replicator dynamics, see R. Cressman, *Evolutionary Dynamics and Extensive Form Games* (Cambridge: MIT, 2003).

[21]See, in particular, Selten, "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit I und II," *Zeitschrift für die gesamte Staatswissenschaft*, CXXI (1965): 301-324, 667-689; "Re-Examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, IV (1975): 25-55; and Harsanyi and Selten, *A General Theory of Equilibrium Selection in Games* (Cambridge: MIT, 1988). See also K. G. Binmore, "Modeling Rational Players I," *Economics and Philosophy*, III (1987): 179-214, and "Modeling Rational Players II," *Economics and Philosophy*, IV (1988): 9-55.

[22]See W.-T. Wu and J.-H. Jiang, "Essential Equilibrium Points of n-Person Non-Cooperative Games," *Scientia Sinica*, XI (1962): 1307-1322.

[23]For more on this topic see Huttegger, "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* (New Brunswick: Ontos, 2009), pp. 375-385.

[24]The issue of robustness of models is of obvious importance to philosophy of science. See, e.g., R. Levins, "The Strategy of Model Building in Population Biology," in Sober, ed., *Conceptual Issues in Evolutionary Biology* (Cambrdige: MIT, 1966): pp 18-2; W. C. Wimsatt, "Robustness, Reliability, and Overdetermination," in M. Brewe and B. Collins, *Scientific Inquiry and the Social Sciences* (San Francisco: Jossey-Bass, 1981), pp. 124-163; and M. Weisberg, "Robustness Analysis," *Philosophy of Science*, LXXIII (2006): 730-742.

[25]See, e.g., J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (New York: Springer, 1983). The form of robustness we are looking at is not the only one possible. One might, for instance, also consider qualitative robustness of dynamical systems, where people are studying dynamical systems which meet some qualitative axioms. For examples in evolutionary game theory see Hofbauer and

Sigmund (*op. cit.*); B. Skyrms, "Stability and Explanatory Significance of Some Simple Evolutionary Models," *Philosophy of Science*, LXVII (2000): 94-113; and Huttegger, "Robustness in Signaling Games," *Philosophy of Science*, LXXIV (2007): 839-847.

[26]M. M. Peixoto, "Structural Stability on Two-Dimensional Manifold," *Topology*, I (1962): 101-120.

[27]See Hofbauer and Sigmund.

[28]This is the so-called Wright manifold. In population genetics, the Wright manifold refers to the manifold of genotype frequencies where the frequencies of alleles from different loci are probabilistically independent. For details and applications to game theory see Cressman.

[29]Harsanyi, "Oddness of the Number of Equilibrium Points: A new Proof," *International Journal of Game Theory*, II (1973): 235-250.

[30]See Hofbauer, "The Selection Mutation Equation," *Journal of Mathematical Biology*, XXIII (1985): 41-53; Hofbauer and Sigmund; and Bürger.

[31]The selection-mutation dynamics of the Rock-Scissors-Paper game may be instructive in this context. If we consider the one-population selection-mutation dynamics with a uniform mutation rate ε, then the mutation term causes the unique Nash equilibrium to become asymptotically stable. Thus adding mutation has an effect similar to certain payoff perturbations of the Rock-Scissors-Paper game.

[32]See Hofbauer and Sigmund, Section 13.2.

[33]For other examples see Gale, Binmore, and Samuelson, "Learning to be Imperfect: The Ultimatum Game," *Games and Economic Behavior*, VIII (1995): 56-90; Binmore and Samuelson, "Evolutionary Drift and Equilibrium Selection," *Review of Economic Studies*, LXVI (1999): 363-393; and Cressman.

[34]Selten, "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit I und II," and "Re-Examination of the Perfectness Concept for Equilibrium Points in Extensive Games."

[35]Selten, "Evolutionary Stability in Extensive Two-Person Games," *Mathematical Social Sciences*, V (1983): 269-363.

[36]Binmore, "Modeling Rational Players I" and "Modeling Rational Players II."

[37]Hofbauer; and Hofbauer and Sigmund.

[38]See in particular Binmore and Samuelson. Their analysis of the Chain-Store game parallels the analysis in the previous section.

[39]Hofbauer and Huttegger, "Feasibility of Communication in Binary Signaling Games," *Journal of Theoretical Biology*, CCLIV (2008): 843-849.

[40]The corresponding games are analyzed in Brandt and Sigmund, "The Good, the Bad and

the Discriminator: Errors in Direct and Indirect Reciprocity," *Journal of Theoretical Biology*, CCXXXIX (2006): 183-194.

[41] See the results in the previous section and Gale, Binmore and Samuelson; and Binmore and Samuelson.

[42] Godfrey-Smith, "Three Kinds of Adaptationism."

[43] Sober also emphasizes the distinction between adaptationism as a theory about evolutionary change and adaptationism as a methodological recommendation on how to study evolution; see Sober, *Philosophy of Biology*, (Boulder: Westview, 2000).

[44] See Binmore and Samuelson for an application to laboratory experiments in economics.

[45] See Godfrey-Smith, "Conditions of Evolution by Natural Selection," this JOURNAL, CIV (2007): 489-516.