J. Math. Biol. (1996) 34: 511–532

Journal of Mathematical Biology

© Springer-Verlag 1996

# Darwinian adaptation, population genetics and the streetcar theory of evolution

**Peter Hammerstein**

Max-Planck-Institut für Verhaltensphysiologie, Abteilung Wickler, D-82319 Seewiesen, Germany

Received 22 April 1994; received in revised form 10 July 1995

**Abstract.** This paper investigates the problem of how to conceive a robust theory of phenotypic adaptation in non-trivial models of evolutionary biology. A particular effort is made to develop a foundation of this theory in the context of *n*-locus population genetics. Therefore, the evolution of phenotypic traits is considered that are coded for by more than one gene. The potential for epistatic gene interactions is not a priori excluded. Furthermore, emphasis is laid on the intricacies of frequency-dependent selection. It is first discussed how strongly the scope for phenotypic adaptation is restricted by the complex nature of 'reproduction mechanics' in sexually reproducing diploid populations. This discussion shows that one can easily lose the traces of Darwinism in *n*-locus models of population genetics. In order to retrieve these traces, the outline of a new theory is given that I call 'streetcar theory of evolution'. This theory is based on the same models that geneticists have used in order to demonstrate substantial problems with the 'adaptationist programme'. However, these models are now analyzed differently by including thoughts about the evolutionary removal of genetic constraints. This requires consideration of a sufficiently wide range of potential mutant alleles and careful examination of what to consider as a stable state of the evolutionary process. A particular notion of stability is introduced in order to describe population states that are phenotypically stable against the effects of all mutant alleles that are to be expected in the long-run. Surprisingly, a long-term stable state can be characterized at the phenotypic level as a fitness maximum, a Nash equilibrium or an ESS. The paper presents these mathematical results and discusses – at unusual length for a mathematical journal – their fundamental role in our current understanding of evolution.

**Key words:** Adaptation – Optimality – Nash equilibrium – ESS – N-locus genetics – Epistasis – Long-term evolution – Rationality paradox

## 1 Introduction

If Charles Darwin could ask us today about modern variants of his concept of evolutionary adaptation, what would we say? Population geneticists would perhaps report the struggle they have had during the second half of this century in order to make the idea of Darwinian adaptation precise in their own theoretical framework. Part of this report would be that in populations with sexual reproduction the mechanism of recombination acts as an undirected and potentially destructive evolutionary force which can strongly interfere with directed selective forces at the phenotypic level. Therefore, from a 'radical' population geneticist's point of view (e.g. Karlin 1975) it seems almost necessary to include genetics in models of how evolution has molded traits in animals and plants. Properties of the underlying genetics seem to have such a strong influence on organismic design that ecological conditions and phenotypic constraints appear to be insufficient to understand this design. Organismic biologists, however, would hold against this view the overwhelming empirical success of purely phenotypic analysis (e.g. Maynard Smith 1978) in their studies of evolutionary adaptation. They would emphasize that a 'reductionist' interest in genetic detail might draw the researcher's attention too far away from the primary scene of selection that matters most if one studies adaptive phenotypes.

Obviously, both views expressed in the preceeding paragraph make sense despite their apparent incompatibility. This calls for a theory that unifies phenotypic and genetic modelling approaches in evolutionary biology. Based on previous work by Eshel and Feldman (1984), Liberman (1988), Eshel (1991), and particularly on Hammerstein and Selten (1994), the present paper presents a draft of such a theory and aims at bridging the gap between population genetics and organismic biology. This reconciliatory mental exercise I call the 'streetcar theory' of evolution. A streetcar comes to a halt at various temporary stops before it reaches a final stop where it stays for much longer. It will be shown in this paper that an evolving population resembles a streetcar in the sense that it may reach several 'temporary stops' that depend strongly on genetic detail before it reaches a 'final stop' which has higher stability properties and is mainly determined by selective forces at the phenotypic level. At such a final stop the ecology has left its strongest traces in the genetic system. I argue that these are the traces of Darwinism that sometimes seemed to be so invisible in the world of modern evolutionary theory. It will be shown that population geneticists typically focus their attention on what happens before a final stop is reached. They have good reasons to defend this as their domain. In contrast, modest phenotypic modellers who only ask for 'final stops of the streetcar' have good reasons not to become explicit about genetics.

The streetcar theory is neither confined in scope to a single gene, nor is it based on the strong assumptions made in the theory of quantitative genetics. It deals with the rather untractable model world of explicit $n$-locus genetics

and analyses these models by asking questions that differ in a subtle way from those found in standard population genetics. As it often happened in theoretical physics, complicated systems may turn out to have simple properties if one learns to ask the appropriate questions. A major aim of the present paper is to show that this is also true for the biological systems under investigation. For the sake of a simple presentation of new concepts, theorems and proofs are formulated for the case $n = 2$. However, they can be generalized to more than two loci (see Weissing 1995).

The paper is organized as if – in a fictitious theatre play – a population geneticist and an organismic biologist tried, with some rivalry, to explain to Darwin their ideas about evolutionary adaptation. It begins with a simple one-locus example of why purely phenotypic considerations may not meet biological facts. It then explains how Sewall Wright would have dealt successfully with this particular one-locus problem and how his idea of an adaptive topography already breaks down in a similar two-locus model. This is the point where the population geneticist loses confidence in the idea of evolutionary optimization principles even with regard to models of frequency-independent selection (i.e., models that were thought to be most favourable to Darwinian adaptation).

If already the frequency-independent model does not 'obey the logic of Darwinian fitness' the introduction of frequency-dependence can only aggravate this problem. In the present study, the latter feature is added to the basic selection model by assuming that (a) phenotypic interaction can be described mathematically as a game and that (b) each genotype specifies a probability distribution over strategies of this game. Without the help of the streetcar approach, it would appear as if Maynard Smith's (1982) phenotypic notion of an evolutionarily stable strategy (ESS) was an inappropriate tool for analysing such systems. It looks indeed at this stage of the presentation very discouraging for any phenotypic concept of evolutionary adaptation. With some bitterness the organismic biologist might now recall a paper by Gould and Lewontin (1979) in which these authors ridiculed the 'adaptationist programme'.

Fortunately, the scientific story told in this paper has a more positive end. The organismic biologist demonstrates how a modern adptationist would define and successfully defend his programme. This adaptationist is not the strawman built up Gould and Lewontin. He is well familiar with all the problems mentioned so far and he knows the spandrels of San Marco. His philosophy will be commented at the end. Perhaps most important is the insight that the processes of non-trivial Darwinian evolution in $n$-locus models and that of rational decision making (as defined in economic theory) come to an important halt at very similar solutions. I call this the 'rationality paradox' because it makes animal behaviour looks as if it was based on rational decision making whereas human behaviour often fails (e.g. Tyszka 1983) to follow the logic of decision theory.

## 2 How 'well behaved' are models of frequency-independent selection?

Behavioural ecologists study organismic design most commonly under the assumption that evolution has a tendency to maximize expected fitness. At the theoretical level this either leads into phenotypic optimization theory or into phenotypic game theory. A lot can go wrong with this approach. To illustrate this, let us suppose that a fictitious behavioural ecologist did not know anything about the phenomenon of sickle cell anemia (in practice every biologists knows it) and that he would undertake a phenotypic analysis of this phenomenon at an African study site were Malaria is prevalent.

The fictitious scientist would find organisms with three different blood types and considerable differences in the average fitness of the three types. This would look like a non-equilibrium phenomenon – one in which Darwinian selection is acting very strongly against the sickle cell disease. However, a long-term study would reveal that the proportion of sickle cell cases does not diminish between generations. The reason for this is well known and can be described in a one locus population genetics model with two alleles $A$ and $B$ and genotype fitnesses

$$w(AA) < w(BB) < w(AB)$$

The failure of a purely phenotypic modelling approach is here due to the heterozygote advantage and the fact that during meiosis this favourable type gets 'ripped into pieces'. Wright (1932) has shown us a way of how to find traces of Darwinian adaptation in this example. Suppose we calculated in the one-locus model mean population fitness as a function of the relative frequency of of allele $A$ (with the help of Hardy-Weinberg ratios). A plot of this function would give us a picture of how mean population fitness depends on the genetic composition of the population. This is Wright's fitness landscape and we know that the genotype frequency equilibrium is given for the frequency of $A$ at which the landscape has its peak.

Unfortunately, Wright's way of characterizing genotype frequency equilibrium as a maximum of the fitness landscape cannot be extended to selection models with two or more loci. Let us therefore take a short look at the two locus case of a standard model for viability selection.

### *2.1 Two-locus model for viability selection – the case with constant fitness*

The model describes an infinite population of diploid individuals with non-overlapping generations, sexual reproduction, and random mating. Consider in this population an evolving trait which is coded for by two genes $A$ and $B$ (note the shift in notation from alleles to genes!). Suppose that there are $n$ alleles $A_1, \ldots, A_n$ of gene $A$, and $m$ alleles $B_1, \ldots, B_m$ of gene $B$. A diploid

genotype then looks as follows:

$$\begin{array}{cc} A_i & A_k \\ | & | \\ B_j & B_l \end{array}$$

It is assumed that recombination takes place at a fixed rate $r$, with $0 < r \leqq \frac{1}{2}$.

Let us make a population census at each gametic state. Let $x_{ij}$ denote the relative frequency of a gamete with haploid genotype $A_iB_j$. At the moment of our census a population state $x = (x_{11}, \ldots, x_{ij}, \ldots, x_{nm})$ can be conceived as a frequency histogram for the population distribution of possible haploid genotypes $A_iB_j$. Due to the assumption of random mating, the following selection equation describes how this frequency histogram changes from one generation to the next if each diploid genotype $A_iB_j/A_kB_l$ has a constant fitness $w_{ijkl}$, with:

$$x_{ij}(t+1) = \frac{1}{\bar{w}(t)}\left[(1-r)\sum_{kl} w_{ijkl}x_{ij}(t)x_{kl}(t) + r\sum_{kl} w_{ilkj}x_{il}(t)x_{kj}(t)\right] \quad (1)$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Here, $\bar{w}$ is the mean fitness of the population:

$$\bar{w}(t) = \sum_{ijkl} w_{ijkl}x_{ij}(t)x_{kl}(t)$$

This equation is based on the classical Mendelian laws of inheritance with 'fair meiosis' and it should be added here that we assume $w_{ijkl} = w_{ilkj} = w_{kjil} = w_{klij}$. No exotic assumptions are made. Because of the constant fitness of genotypes one would naively think that there are very good chances for Darwinian adaptation to 'take place' in this model. However, it turns out that this equation has fairly exotic properties. Moran (1964) was the first to show that the trajectories of solutions of this equation may easily ignore Wright's fitness landscape in the sense that mean population fitness decreases over time. In colloquial terms this means 'survival of the less fit'. The genotype frequency equilibrium may be located at a peak, a slope or at a valley of the fitness landscape. In the latter two cases, Darwinism is facing a serious problem. Let us illustrate this with one of Moran's examples. Suppose there are two loci $A$ and $B$, and two alleles $A_1, A_2$ and $B_1, B_2$ at each locus. Let the fitness of diploid genotypes be determined by the following fitness matrix:

| | $B_1,B_1$ | $B_1,B_2$ | $B_2,B_2$ |
|---|---|---|---|
| $A_1,A_1$ | 1 | 1 | 0 |
| $A_1,A_2$ | 1 | 3 | 1 |
| $A_2,A_2$ | 0 | 1 | 1 |

This matrix describes a case with some similarity to the sickle cell example because the double heterozygote has the highest fitness. However, unlike the sickle cell example, mean population fitness is here not maximized at genotype

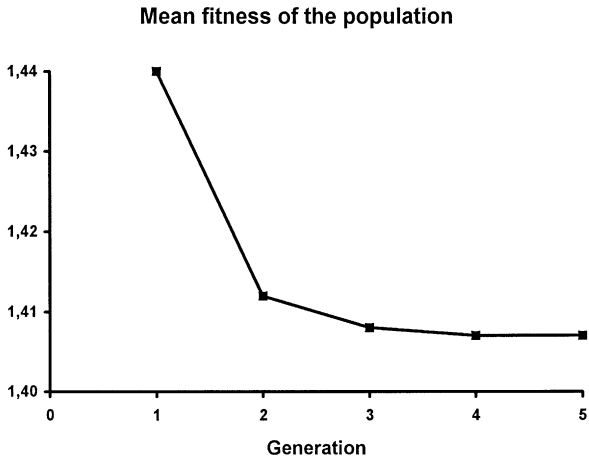


**Fig. 1.** Frequency-independent selection with a decrease in mean population fitness. Moran created this example in order to demonstrate that a population can move 'downhill' in Wright's fitness landscape if one considers a two-locus model

frequency equilibrium. Figure 1 demonstrates impressively how fitness can decrease from generation to generation in this model.

Moran's example may look somewhat artificial. Ewens (1968) and Karlim (1975) showed that the problem raised by Moran is not confined to very special cases. Even if the highest fitness could be achieved by a double homozygote, equation 1 may have stable equilibrium states at which there is no peak of the fitness landscape.

## 3 Evolutionary games and frequency-dependent selection

In the theory of frequency-dependent selection it is well known that mean population fitness is typically not maximized at evolutionary equilibrium. Therefore, at first glance the results reported in the previous section do not seem to contribute anything new in this context. However, this is not quite correct. If a frequency-independent selection model does not follow the simple logic of Darwinian fitness, then frequency-dependent variants of such a model can easily have stable equilibrium states that fail to satisfy phenotypic conditions for evolutionary stability, no matter how weak evolutionary stability is defined. In the light of the previous section this is – roughly speaking – for the following reason. If we were to take a 'snapshot' of an evolving population together with its co-evolving fitness-landscape, the picture of the landscape would not even reveal the instantaneous direction of evolutionary change at the moment of the snapshot.

Let us nevertheless describe a class of frequency-dependent selection models that are based on equation (1) and on a game-theoretic description of the phenotypic world. Suppose that phenotypes interact pairwise and that

pair formation is random with respect to genotypes. Let $S$ be a finite set of pure strategies and $E(s, t)$ be a function which defines the expected change of an individual's fitness if this individual plays strategy $s$ against an opponent who plays $t$. Using some technical jargon from mathematical economics we could then say that $S$ and $E$ define a symmetric 2-person game $G = (S, E)$ in strategic form. The Hawk-Dove game (Maynard Smith 1982) is the best known biological textbook example of such a game. A mixed strategy is a probability distribution over $S$. The payoff function $E$ can be extended in the usual way to pairs of mixed strategies $(p, q)$. The extended payoff function is bilinear.

Before we add genetics to the phenotypic game, let us see how economists would analyze it in the context of decision theory. If a strategy $r$ has the property that for a fixed opponent-strategy $q$ it maximizes the expected payoff $E(\cdot, q)$, then $r$ is called a best response to $q$. Furthermore, a pair of strategies $(p, q)$ is called a Nash equilibrium (Nash 1951) if $p$ and $q$ are best responses to each other. A Nash equilibrium is called symmetric if it is of the form $(p, p)$. The Nash equilibrium concept plays a central role in non-cooperative game theory (Fudenberg and Tirole 1991). Game theorists would solve a symmetric game $G$ by calculating its symmetric Nash equilibria. After this they would perhaps engage in long philosophical debates about which of them to choose (Harsanyi and Selten 1988). As we shall see in the streetcar theory (next section), biologists have good reasons to follow economists at the first step of this procedure. In order to show this we first have to connect the game with the genetic selection model depicted by equation (1).

### *3.1 The two-locus model with phenotypic game interaction*

Consider the two-locus model of Sect. 2. Suppose that a diploid genotype $A_iB_j/A_kB_l$ *causes individuals to 'play' a mixed or pure strategy* $u_{ijkl}$ of the phenotypic game $G$. For biological reasons we must assume that $u_{ijkl} = u_{ilkj} = u_{kjil} = u_{klij}$ because the corresponding genotypes are functionally equivalent. The function $u$ will be called the 'phenotype specification'. If one knows the relative frequencies of phenotypes in a population, one also knows the relative frequencies of strategies with the help of $u$. This makes it possible to calculate the mean population strategy at time $t$ as

$$\bar{q}_t(s) = \sum_{ijkl} x_{ij}(t)x_{kl}(t)u_{ijkl}(s) \quad \text{for all pure strategies } s$$

The expected fitness of an individual with genotype $A_iB_j/A_kB_l$ can now be defined as

$$w_{ijkl}(t) = w_0 + E(u_{ijkl}, \bar{q}_t) \tag{2}$$

This means that an individual has a basic fitness expectation $w_0$ which is changed by the game payoff $E$. We can now replace the constant fitness $w_{ijkl}$ in equation (1) by expression (2) and get the frequency-dependent version of this

model. For reasons outlined at the beginning of this section the new equation will in general not take a population to a Nash equilibrium. At this stage of the discussion economic reasoning and Maynard Smith's biological theory of evolutonarily stable strategies both seem to be fairly inconsistent with two-locus population genetics.

## 4 The streetcar theory of evolution

Evolutionary theorists who base their thoughts on the principle of optimality (in the case of frequency-independent selection) and on the game-theoretic notion of a Nash-equilibrium (in the case of frequency independent selection) are so successful in explaining facts that they have gained a considerable amount of attention by empiricists in organismic biology. If genetic models of evolution, such as the ones presented in Sects. 2 and 3, do not bear the phenotypic approach, this calls for the construction of a theoretical roof under which both phenotypic and genetic approaches to evolution can be housed in peaceful coexistence. I shall sketch in the remainder the construction plan of this roof – the streetcar theory of evolution.

Suppose that we are designing an evolutionary model in which we have a set of phenotypes, such as the strategies of the game in Sect. 3. Let a fitness function, e.g. the game payoff function $E$, define the 'Darwinian success' of phenotypes. So far the model can be based on extensive empirical information from organismic biology. However, when we try to introduce genetics in our model we are stepping on very unsafe grounds because in almost all interesting problems we simply do not know it well enough. The number of genes, alleles, and – most important – the specification of the genotype-phenotype relationship have to be made up artificially. Is it surprising that strange results can emerge from such an arbitrary specification?

But how to be more biological? A little thought experiment may help to answer this question. Suppose that in the spirit of the preceeding paragraph selection leads the evolving population to a stable genotype frequency equilibrium at which phenotypes fail to satisfy the phenotypic criteria (optimality or game equilibrium) of adaptation. Let us now take the idea serious that this reflects the weakness of the modeller's creativity more than the failure of Darwinism. The modeller then can start a post hoc search for features that he may have forgotten in the initial design of his model. In particular this leads to the question of whether important alleles may have been left out in the original setting. Perhaps a new allele can be added to the model so that the current population state would be destabilized if we were to add this allele at a low frequency to the equilibrium population. If one can mathematically show the existence of such an allele and an appropriate extension of the phenotype specification function, this allele can in theory be added to the original model and extends it.

The destabilized population would now evolve further and may reach another stable genotype frequency equilibrium. We may be able to go repeatedly

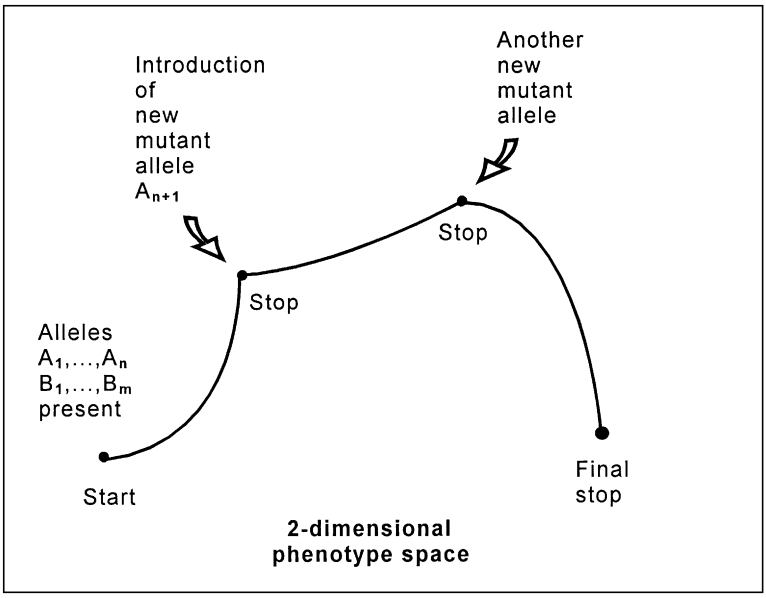


**Fig. 2.** Course of an evolving population in phenotype space. The population stops at a genotype frequency equilibrium. Temporary stops are then left after the population is perturbed by an appropriate new mutant allele. A final stop is phenotypically stable against genetic perturbation

through the same procedure. The evolutionary course of our population would then resemble that of a streetcar which repeatedly comes to a halt at temporary stops and starts moving after new passengers – new mutants – have entered it (note here that the tracks of the streetcar are not part of the paradigm).

A streetcar eventually comes to a halt at a final stop and so may our evolving population. A final stop of an evolving population is a phenotypically stable genotype frequency equilibrium which has the property that the above procedure cannot lead any more to phenotypic evolution because of the non-existence of a new allele with destabilizing properties. Extensions of the model via introduction of new alleles would in this case no longer help to keep the streetcar running.

This thought experiment was meant to show a way of how one might systematically think about innovative aspects of the evolutionary process – a process that does not only shape phenotypes but also the underlying genetics. The important question is now whether one can characterize final stops by phenotypic optimality conditions or the Nash equilibrium property. If this is indeed the case, we can answer Darwin's fictitious question about the modern theory of evolutionary adaptation by saying that it can best be conceived as the theory of final stops of the streetcar. Sickle cell anemia would not be such a stop because one could at least theoretically construct a new allele that produces the same effect as the 'classical' heterozygote.

## 5 A streetcar named ESS

It is easy to imagine a streetcar but less easy to make the formal structure of the underlying theory precise for a class of models. This is particularly difficult in the context of frequency-dependent selection. Eshel and Feldman (1984) did the first analysis of frequency-dependent selection which in retrospect has the flavour of the streetcar theory. They had the ingenious idea to look at the external stability of genotype frequency equilibria by introducing *new* mutant alleles. They showed that genotype frequency equilibria can be destabilized by new mutant alleles if these equilibria are already in a neighbourhood of an evolutionarily stable strategy. But what can be said if these equilibria are not close to an ESS? Inspired by discussions with Ilan Eshel (see also Eshel 1991, 1995), Hammerstein and Selten (1994) dealt with this question and made an attempt to clarify the conceptual background of the theory under discussion. Also inspired by discussions with Franjo Weissing they introduced, in particular, a definition for 'phenotypic stability against genetic perturbation' that helps to capture the essence of final stops.

We now follow these authors on their streetcar journey. Like many absent-minded scientists they overlook all temporary stops, arrive at the final stop and ask themselves "where are we?". Instead of watching out of the window they theorize about properties of a final stop. This clearly requires a definition first. All following definitions relate to the frequency-dependent version of the selection model introduced in Sects. 2 and 3. Suppose therefore that there are two loci $A$ and $B$ with alleles $A_1, \ldots, A_n$ and $B_1, \ldots, B_m$. As explained before, the genetic population state can be described as a frequency histogram $x = (x_{ij})$, where $x_{ij}$ is the frequency of the haploid genotype, $A_iB_j$. It will be said that an initial genetic state $x(0)$ generates a sequence $p_0, p_1, \ldots$ of mean population strategies if strategy $p_t$ is the mean strategy of genetic state $x(t)$ in the solution sequence $x(0), x(t), \ldots$ of the selection equation.

If we extend the model by introducing a new allele $A_{n+1}$, we clearly have to consider extended population states $x = (x_{11}, \ldots, x_{n+1,1}, \ldots, x_{n+1,m})$. We also have to extend the phenotype specification function $u_{ijkl}$ that describes the genotype-phenotype relationship. Obviously, this can be done in many ways and we shall talk here of the phenotypic specifications of $A_{n+1}$. The following definition by Hammerstein and Selten (1994) helps to get closer to the idea of a final stop.

**Definition of phenotypic stability against genetic perturbation**
*Let $x$ be a population state with population mean strategy $p$. This state is called phenotypically stable against genetic perturbation if for every phenotypic specification of a new allele $A_{n+1}$* it has the following property: For every $\varepsilon > 0$ *a $\delta > 0$ can be found such that for every (extended) population state $y_0$ in the $\delta$-neighbourhood of $x$ the sequence of population mean strategies $q_0, q_1, \ldots$ generated by $y_0$ satisfies*

(i) $|q_i - p| < \varepsilon$ *for* $t = 0, 1, \ldots$

*and*

(ii) $\lim_{t \to \infty} q_t = p$

In the explicit formulation of this definition, a genetic perturbation is caused by a new mutant allele. However, implicitly this definition also captures perturbations that are caused by a slight change in the frequencies of genotypes that already exist (to see this, look at an $A_{n+1}$ that coincides with an existing allele as far as its phenotypic effects are concerned). In other words, internal and external aspects of stability are here combined. This led me to replace the term "phenotypic external stability" in Hammerstein and Selten (1994) by the expression "phenotypic stability against genetic perturbation". The new expression also makes it more transparent that even with regard to a 'truly new' allele we are not discussing the problem of uninvadability. At a phenotypically stable state, a new allele may indeed increase in frequency and the new genetic state may not be the same as the unperturbed one. Clearly, it then might occur in strange examples that the new genetic state is phenotypically unstable against genetic perturbation. This is not a problem for the streetcar theory as long as we are dealing with necessary conditions for a final stop and it makes it possible to state the following fundamental property of long-term evolution (Hammerstein and Selten 1994):

**Theorem 1**

*Consider a genetic population state $x$ in which alleles $A_1, \ldots, A_n$ and $B_1, \ldots, B_m$ are present at loci $A$, $B$, respectively. Let the fitness effects of phenotypic interaction be described by a finite game $G$ in strategic form, so that we consider selection equation* (1) *with frequency-dependent fitness* (2). *The following assertion holds*:

> *If state $x$ is phenotypically stable against genetic perturbation, then the population mean strategy $p$ of state $x$ is a best response to $p$ in the phenotypic game $G$.*

This theorem helps us to overcome Moran's and Karlin's criticism of adaptationist thought and it shows that there is no final stop without economically well behaved phenotypes. In order to underline the parallels with economic theory, the last part of Theorem 1 can be reformulated by saying: $p$ must have the property that $(p, p)$ is a Nash equilibrium of the phenotypic game $G$.

Let us now consider the special class of phenotypically monomorphic final stops in which the phenotype specification assigns the same strategy to all genotypes that do not contain the new mutant allele $A_{n+1}$. We are talking here about phenotypically monomorphic states without assuming that all alleles are fixed. This means that by 'throwing' $A_{n+1}$ into the population we may generate more than one new strategy. In this context we cannot reduce the problem of invasion by a new mutant allele to the one-locus case as it would be possible with fixed alleles. The following definition and Theorem 2 are based on Hammerstein and Selten (1994).

**Definition of phenotypic monomorphism and its stability properties**
*Consider the frequency-dependent selection model with alleles* $A_1, \ldots, A_n$ and $B_1, \ldots, B_m$ *for loci* $A$, $B$, *respectively. Suppose that all genotypes that can be formed with these alleles specify the same pure or mixed strategy* $p$ *of the finite game* $G$ *in strategic form. Let us call such a specification of the selection model a phenotypic monomorphism. We say that the monomorphism is phenotypically stable against genetic perturbation if all its genetic states have this stability property. We also need a partial version of this definition. The monomorphism is called phenotypically stable against a new mutant allele if it is phenotypically stable against perturbations caused by this particular allele. Furthermore, the monomorphism is called invasion stable against the new mutant allele if the frequency of this allele converges to zero in all solution sequences that start with a sufficiently small population frequency of this allel.*

**Definition of a mutant allele that cannot generate a given strategy**
*Let* p *be a strategy of the phenotypic game. Consider a new mutant allele* $A_{n+1}$ *and the set of strategies that are specified by genotypes containing one allele* $A_{n+1}$. *If* p *is not an element of the convex hull of this strategy set, we say that the new mutant allele cannot generate* p *in heterozygous condition. This has the following biological interpretation: Firstly, diploid genotypes with* $A_{n+1}$-heterozygotes do not code for p and, secondly, a mixed population of these genotypes cannot have p as its mean strategy.

**Theorem 2**
*Consider a phenotypic monomorphism in which all genotypes code for the same strategy* $p$ *of the finite phenotypic game* $G$. *The following two assertions hold:*

(i) *If the monomorphism is phenotypically stable against genetic perturbation, then* $p$ *is an evolutionarily stable strategy of* $G$.

(ii) *If* $p$ *is an evolutionarily stable strategy of* $G$, *then the monomorphism is phenotypically stable and invasion stable against every new mutant allele that cannot generate* $p$ *in heterozygous condition.*

In this theorem we use the definition of evolutionary stability by Maynard Smith and Price (1973). A strategy $p$ is called evolutionarily stable if $(p, p)$ is a symmetric Nash equilibrium with the additional property that $E(p, q) > E(q, q)$ for all other strategies $q$ that are also best responses to $p$ (see Sect. 3). Theorem 2 shows that important properties of monomorphic final stops are surprisingly well described by the original ESS definition. Most likely one can get an even stronger result. Hammerstein and Selten (1994) had already formulated a version of this theorem, in which the ESS property was thought to be sufficient for phenotypic stability against unrestricted genetic perturbation. Very recently, however, Weissing (1995) has discovered a gap in our original proof. Probably, this gap can be bridged without resorting to strong additional assumptions.

## 6 Why final stops have Darwinian properties

This section contains proofs of the two theorems about final stops and a more precise description of the formal background of these theorems. The key idea that helps to reveal the Nash equilibrium property as a necessary condition for phenotypic stability against genetic perturbation should be stated before we 'dive into formal detail': if genetic constraints keep a population away from a phenotypically adaptive state, there is a possibility for a new mutant allele to code for phenotypes that perform better than the population mean, with a phenotype specification that is particularly resistant to the disruptive forces of recombination. This resistance can be achieved if the new allele codes for the better phenotype independent of the genetic context. Therefore, the proof of the important Theorem 1 is almost trivial. This is perhaps why it took evolutionary biologists so long to discover it, since nobody expected a truly fundamental theorem of natural selection to be so obvious. Theorem 2 requires considerably more thought.

Before we begin with the actual proofs it will be helpful to give a complete picture of the class of selection models that was gradually built up in Sects. 2 and 3. A model $M$ of this class is specified by the following components:

(a) the phenotypic game $G$ (as defined in Sect. 3)
(b) the number of alleles, $n, m$ for locus $A$, $B$, respectively
(c) the phenotype specification function $u$ which assigns a pure or mixed strategy $u_{ijkl}$ to every index combination with $1 \leqq i, k \leqq n$ and $1 \leqq j, l \leqq m$
(d) the recombination rate $r$ with $0 < r \leqq 1/2$

We shall speak here of the model $M = (G, n, m, u, r)$. During the thought experiment of a 'streetcar journey' we keep $G$ and $r$ constant. – not because $G$ and $r$ never change in evolution but because we do not need their change in order to overcome genetic constraints. If we ask for long-term stability against a particular new mutant allele $A_{n+1}$, we look at a modified model $\hat{M} = (G, n+1, m, \hat{u}, r)$, where $\hat{u}$ is an extension of $u$. We call $\hat{M}$ a one-allele extension of $M$. The 'specification of a new allele $A_{n+1}$' is a specification of a one-allele extension of $M$.

For a given selection model $M$ we conceive a genetic state as a matrix $x = (x_{ij})$ of relative population frequencies at which haploid genotypes occur in gametes. We obviously require that $0 \leqq x_{ij} \leqq 1$ for all index combinations. It is also obvious what it means to speak in sloppy language of the 'same genetic state' when we make a one step extension of a model $M$. A given selection model $M$ specifies the selection equation that results from linking Sects. 2 and 3 were the underlying assumptions are described:

$$x_{ij}(t+1) = \frac{1}{\bar{w}(t)}\left[(1-r)\sum_{kl} w_{ijkl}(t)x_{ij}(t)x_{kl}(t) + r\sum_{kl} w_{ilkj}(t)x_{il}(t)x_{kj}(t)\right] \tag{3}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

with
$$\bar{w}(t) = \sum_{ijkl} w_{ijkl}(t)x_{ij}(t)x_{kl}(t) \tag{4}$$

$$w_{ijkl}(t) = w_0 + E(u_{ijkl}, \bar{q}_t) \tag{5}$$

$$\bar{q}_t(s) = \sum_{ijkl} x_{ij}(t) x_{kl}(t) u_{ijkl}(s) \quad \text{for all pure strategies } s \tag{6}$$

**Proof of Theorem 1**

Assume that $(p, p)$ is not a Nash equilibrium of the phenotypic game $G$. This means that $p$ is not a best response to $p$. Another strategy $s$ must then exist which is a better response to $p$, so that $E(s, p) > E(p, p)$. Now, the central idea of the proof is to introduce a new mutant allele $A_{n+1}$which in all genotypic combinations specifies this strategy $s$ at the phenotypic level. The allele $A_{n+1}$ then dominates the rest of the genotype. $A_{n+1}$ will be used in order to perturb a genetic population state with mean strategy $p$. We shall see that after a small perturbation this allele will cause the population mean strategy to evolve away from $p$. For the biological interpretation of this proof it is important to state that there will always be an $s$ with the above property which differs only slightly from $p$ and, therefore, is no macro-mutation. In order to construct such an $s$, consider a pure strategy $\pi$ which is best response to $p$. We know from game theory (Fudenberg and Tirole, 1991) that such a $\pi$ exists. Let $s$ be that strategy that plays like $p$ with probability $1 - \varepsilon$ and $\pi$ with probability $\varepsilon$. In comparison with $p$ this strategy $s$ assigns a higher probability to $\pi$ and, therefore, must be a better reply to $p$ than $p$ itself.

Technically speaking we construct the above new allele $A_{n+1}$ by extending the phenotype specification function $u$ to an index range that includes $n + 1$:

$$u_{n+1,jkl} = s \quad \text{for } k = 1, \ldots, n+1, \quad j, l = 1, \ldots, m$$

For haploid genotypes with allele $A_{n+1}$ the equation (3) then looks as follows:

$$x_{n+1,j}(t+1) = \frac{w_0 + E(s, \bar{q}_t)}{w_0 + E(\bar{q}_t, \bar{q}_t)} \sum_{kl} [r x_{n+1,l}(t) x_{kj}(t) + (1-r) x_{n+1,j}(t) x_{kl}(t)] \tag{7}$$

for $j = 1, \ldots, m$.

When we make our population census at gametic stage we can look at two kinds of haploid genotypes, namely those with $A_{n+1}$ and those without it. Let $b$ denote the relative frequency of haploid genotypes with $A_{n+1}$. At generation $t$ this frequency is

$$b(t) = \sum_{j=1}^{m} x_{n+1,j}(t) \tag{8}$$

From (7) and (8) we get the following expression for how $b(t)$ changes between generations:

$$b(t+1) = \frac{w_0 + E(s, \bar{q}_t)}{w_0 + E(\bar{q}_t, \bar{q}_t)} b(t) \tag{9}$$

We finally use the property that $s$ is a better response to $p$ than $p$ is to itself. This means that $E(s, p) > E(p, p)$. In view of the continuity of $E(.,.)$ we can find an $\varepsilon > 0$ and a constant $c > 0$ such that $E(s, q) - E(q, q) > c$ holds for all strategies $q$ with $|p - q| < \varepsilon$. Consider now an arbitrary population state $x(0)$

with $b(0) > 0$ in the $\varepsilon$-neighbourhood of $p$. We take this state as the initial value for solution $x(t)$ of equation (3). This solution generates the sequence $\bar{q}_t$ of population mean strategies. It will now be shown that the sequence $\bar{q}_t$ cannot remain in the $\varepsilon$-neighbourhood of $p$.

Define a new constant as $c' = \dfrac{c}{\max[w_0 + E(.,.)]}$. Then $c' > 0$ and

$$\frac{w_0 + E(s, \bar{q}_t)}{w_0 + E(\bar{q}_t, \bar{q}_t)} > 1 + c' \tag{10}$$

holds for $|q_t - p| < \varepsilon$. Therefore, as long as the population mean strategy $q_t$ remains in this $\varepsilon$-neighbourhood of $p$ the between generation change in relative frequency of haploid genotypes with allele $A_{n+1}$ must satisfy the following inequality:

$$b(t+1) > (1 + c')b(t) \tag{11}$$

We conclude from (11) that eventually $q_t$ must leave the $\varepsilon$-neighbourhood of $p$. This shows that for the $\varepsilon$ under consideration no $\delta$ can be found which satisfies the requirement of the definition of phenotypic stability against genetic perturbation.

**Proof of Theorem 2**

Consider alleles $A_1, \ldots, A_n$ and $B_1, \ldots, B_m$. Assume that all genotypes that can be formed with these alleles specify the same phenotypic strategy $p$. Technically speaking this means that we consider a selection model $M = (G, n, m, u, r)$ with the phenotype specification function $u_{ijkl} = p$ for $i, k = 1, \ldots, n$, $j, l = 1, \ldots, m$. We call such a model '$p$-monomorphic' or a '$p$-monomorphic specification of equation (3)'. Theorem 2 addresses $p$-monomorphic models and their one-allele extensions. Although only one allele is added, these extensions can be highly polymorphic!

We first prove the following **assertion 1:**

If strategy $p$ is an ESS of the phenotypic game $G$, then every population state of the $p$-monomorphic specification of equation (3) is phenotypically stable against every new mutant allele for which $p$ is outside the convex hull of the set containing strategies $u_{n+1,jkl}$, $k \neq n+1$. Furthermore, if the initial frequency of the new mutant allele is small enough, its frequency will converge to zero in solution sequences of the selection equation.

For generation $t$ let $a(t)$ denote the relative population frequency of haploid genotypes that do not contain the new allele $A_{n+1}$. Let $b(t)$ denote the complementary frequency of those with $A_{n+1}$. Our aim is now to find a suitable expression for the difference equation that describes how $a(t+1)$ depends on $a(t)$. This requires some thought. We shall first calculate various genotype frequencies and average strategies. For the haploid genotype classes with and without allele $A_{n+1}$ the genotype frequencies $(t)$ and $a(t)$ are:

$$a(t) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}(t) \quad \text{and} \quad b(t) = \sum_{j=1}^{m} x_{n+1,j}(t) \tag{12}$$

**Table 1.** Genotype classes and their average phenotypes

| Class of diploid genotypes of generation $t$ | Relative frequency of class genotypes | Average strategy of a class |
|---|---|---|
| Without allele $A_{n+1}$ | $a^2(t)$ | $p$ |
| With one allele $A_{n+1}$ | $2a(t)b(t)$ | $\bar{u}_t = \frac{1}{a(t)b(t)} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{m} x_{n+1,j}(t) x_{kl}(t) u_{n+1,jkl}$ |
| With two alleles $A_{n+1}$ | $b^2(t)$ | $\bar{v}_t = \frac{1}{b^2(t)} \sum_{j=1}^{m} \sum_{l=1}^{m} x_{n+1,j}(t) x_{n+1,l}(t) u_{n+1,j,n+1,l}$ |
| All genotypes | 1 | $\bar{q}_t = a^2(t)p + 2a(t)b(t)\bar{u}_t + b^2(t)\bar{v}_t$ |

We now consider diploid genotypes that are produced by gametes of generation $t$. With the help of (12) and using Hardy-Weinberg ratios we can calculate for different classes of diploid genotypes their relative frequency in the population. This in turn makes it possible to calculate the average strategy of a genotype class with the help of the genotype specification function $u$. Table 1 contains the results of these calculations and the definition of the important expressions $\bar{u}_t$ and $\bar{v}_t$. The way to read Table 1 is column by column in a 'top down' fashion because cell entries depend on each other.

It is useful to also consider for generation $t$ the average strategy of classes of alleles. Let $\bar{\alpha}_t$ denote the average strategy of all alleles $A_1, \ldots, A_n$. Let $\bar{\beta}_t$ denote the average strategy of $A_{n+1}$. With the help of Table I we get

$$\bar{\alpha}_t = a(t)p + b(t)\bar{u}_t \quad \text{and} \quad \bar{\beta}_t = a(t)\bar{u}_t + b(t)\bar{v}_t \tag{13}$$

An allele $A_i$ with $i < n+1$ occurs in genotypes without $A_{n+1}$ at frequency $a(t)$ and in those with $A_{n+1}$ at frequency $b(t)$. Therefore, the mean strategy $\bar{q}_t$ of the population satisfies the following equation:

$$\bar{q}_t = a(t)\bar{\alpha}_t + b(t)\bar{\beta}_t \tag{14}$$

Aiming now at the relationship between $a(t)$ and $a(t+1)$, we first calculate $a(t+1)$ with the help of (3) and (12):

$$\begin{aligned} a(t+1) = {} & \frac{w_0 + E(p, \bar{q}_t)}{\bar{w}(t)} \sum_{i,k=1}^{n} \sum_{j,l=1}^{m} ((r x_{il}(t)\alpha_{kj}(t) + (1-r)x_{ij}(t)\alpha_{kl}(t)) \\ & + \frac{1}{\bar{w}(t)} \sum_{i=1}^{n} \sum_{j,l=1}^{m} [r[w_0 + E(u_{il,n+1,j}, \bar{q}_t)] x_{il}(t) x_{n+1,j}(t) + (1-r) \\ & \times [w_0 + E(u_{ij,n+1,l}, \bar{q}_t)] x_{ij}(t) x_{n+1,l}(t)] \end{aligned}$$

Using Table I, it is easy to see that this equation is equivalent with

$$a(t+1) = \frac{1}{w_0 + E(\bar{q}_t, \bar{q}_t)} \{a^2(t)[w_0 + E(p, \bar{q}_t)] + a(t)b(t)[w_0 + E(\bar{u}_t, \bar{q}_t)]\}$$

In view of the definition (13) of $\bar{\alpha}_t$ this yields:

$$a(t+1) = \frac{w_0 + E(\bar{\alpha}_t, \bar{q}_t)}{w_0 + E(\bar{q}_t, \bar{q}_t)} a(t)$$

It follows by (14) that we have

$$\frac{w_0 + E(\bar{\alpha}_t, \bar{q}_t)}{w_0 + E(\bar{q}_t, \bar{q}_t)} = 1 + \frac{E(\bar{\alpha}_t, \bar{q}_t) - E(\bar{q}_t, \bar{q}_t)}{w_0 + E(\bar{q}_t, \bar{q}_t)} = 1 + \frac{b(t)E(\bar{\alpha}_t, \bar{q}_t) - b(t)E(\bar{\beta}_t, \bar{q}_t)}{w_0 + E(\bar{q}_t, \bar{q}_t)}$$

This yields

$$a(t+1) = \left[1 + (1 - a(t)) \frac{E(\bar{\alpha}_t, \bar{q}_t) - E(\bar{\beta}_t, \bar{q}_t)}{w_0 + E(\bar{q}_t, \bar{q}_t)}\right] a(t) \tag{15}$$

Since $0 \leqq a(t) \leqq 1$, the difference $E(\bar{\alpha}_t, \bar{q}_t) - E(\bar{\beta}_t, \bar{q}_t)$ is decisive for the movement of $a(t)$. It is easy to show that this difference can be rewritten in the following way (dropping $t$):

$$\begin{aligned} E(\bar{\alpha}, \bar{q}) - E(\bar{\beta}, \bar{q}) = {} & a^3[E(p,p) - E(\bar{u}, p)] \\ & + a^2 b\{2[E(p, \bar{u}) - E(\bar{u}, \bar{u})] + E(\bar{u}, p) - E(\bar{v}, p)\} \\ & + ab^2\{2[E(\bar{u}, \bar{u}) - E(\bar{v}, \bar{u})] + E(p, \bar{v}) - E(\bar{u}, \bar{v})\} \\ & + b^3[E(\bar{u}, \bar{v}) - E(\bar{v}, \bar{v})] \end{aligned} \tag{16}$$

Let us for a moment replace $\bar{u}$ and $\bar{v}$ in (16) by time-independent strategies $u$ and $v$. Consider the set $S_1$ of strategies that are specified by genotypes containing exactly one allele $A_{n+1}$. The convex hull of $S_1$ will be denoted by $h(S_1)$. We now show that for every strategy $u \in h(S_1)$, a value $\delta_u > 0$ can be found so that for all $0 < b < \delta_u$ and for every strategy $v$ the expression (16) is positive. Clearly, the latter expression would vanish for $u = v = p$. However, we know that $u \neq p$ because $A_{n+1}$ cannot generate $p$ in heterozygous condition. This statement could not be made in the proof by Hammerstein and Selten (1994) and it helps here to avoid the gap in our original proof described by Weissing (1995). If $u$ is not a best reply to $p$, then the first term in (16) is positive. If $u$ is a best reply to $p$ with $u \neq p$, then the first term in (16) is zero and the second one is positive. With the help of these results, it is easy to show that a continuous $\delta_u > 0$ with the required property can be found. Let $\delta$ be the minimum of $\delta_u$ with respect to the compact set $h(S_1)$. Obviously, $\delta > 0$.

In order to complete the proof of assertion 1, consider a sequence $x(0), x(1), \ldots$ generated in the one-allele extension model. Assume that the quantity $b(0)$ associated with $x(0)$ is smaller than $\delta$, then $b(t) = 1 - a(t)$ is decreasing. The sequence of population mean strategies $\bar{q}_0, \bar{q}_1, \ldots$ therefore remains in an $\varepsilon$-neighbourhood of $p$ if $b(0)$ is sufficiently small. Furthermore, the sequence $a(t)$ must converge to a value $a_0$ because it is increasing and has the upper bound 1. The corresponding sequence $\bar{q}_0, \bar{q}_1, \ldots$ must have an accumulation point $q$. We now show that $a_0 = 1$ and therefore $q = p$. Suppose that $a_0 < 1$. In view of (15) and the results about (16), a positive constant $c$ must then exist so that for all $\bar{q}_t$ that are sufficiently close to $q$ we get

$a(t+1) > (1+c)a(t)$. This means an increase of the underlying frequency $a(t)$ of haploid genotypes without the new mutant allele: Clearly, this is impossible since the sequence $a(t)$ cannot decrease otherwise for any $t$ and, therefore, would grow beyond one. We now have shown assertion 1.

In the remainder of the proof we show that the following **assertion 2** holds: If every population state of the $p$-monomorphic specification of equation (3) is phenotypically stable against genetic perturbation, then $p$ is an ESS of the phenotypic game $G$.

We know from Theorem 1 that if a population state $x$ of a $p$-monomorphic specification of equation (3) is phenotypically stable, then $(p, p)$ is a Nash equilibrium of the phenotypic game $G$, i.e. $E(p, p) \geqq E(s, p)$ for all strategies $s$ of $G$. It remains to show that $p$ also satisfies the second ESS condition, namely that for all other strategies $s$ with $E(s, p) \geqq E(p, p)$ we have $E(p, s) \geqq E(s, s)$.

Going the indirect way, let us assume that $p$ is a strategy that satisfies the Nash equilibrium condition but fails to satisfy the second ESS condition. A strategy $v$ must then exist for which the following statements (a) and (b) are true:

(a) $v$ is a best response to $p$

(b) $E(p, v) \leqq E(v, v)$ .

Let us now consider the following one-allele extension of the $p$-monomorphic selection model.

$$u_{n+1,jkl} = p \quad \text{for } j, l = 1, \ldots, m \text{ and } k = 1, \ldots, n$$

$$u_{n+1,j,n+1,l} = v \quad \text{for } j, l = 1, \ldots, m$$

We examine a sequence $x(0), x(1), \ldots$ generated by this selection model. The mean strategies $\bar{u}_t$ and $\bar{v}_t$ (defined in Table 1) do not depend on $t$ and we always have $\bar{u}_t = p$ and $\bar{v}_t = v$. Now, if we study the case $E(p, v) < E(v, v)$, expression (15) shows that $E(\bar{\alpha}_t, \bar{q}_t) - E(\bar{\beta}_t, \bar{q}_t)$ is always negative. From (15) we then conclude that the sequence $a(t)$ always converges to 0 as long as we start with an $a(0) < 1$. In the remaining case $E(p, v) = E(v, v)$ the difference $E(\bar{\alpha}_t, \bar{q}_t) - E(\bar{\beta}_t, \bar{q}_t)$ always vanishes and we have $\bar{q}_t = [1 - b^2(0)]p + b^2(0)v$ for $t = 0, 1, \ldots$ In both cases the sequence $\bar{q}_0, \bar{q}_1, \ldots$ of the population mean strategies does not converge to $p$ if we start with $b(0) > 0$. Therefore, a $p$-monomorphic state is not phenotypically stable against genetic perturbation.

## 7 Concluding remarks on the 'streetcar' philosophy'

In biology it is often the case that phenotypic aspects of an evolutionary problem are empirically far more accessible than the underlying genetics. Even if one uses the most modern technical equipment in order to sequence strands of DNA completely for their molecular structure, this does not yet reveal how genotypes and phenotypes relate to each other. For a theoretician

it is therefore tempting to study the evolution of phenotypic traits in purely phenotypic models. Admittedly, this approach has the important drawback that it ignores how severely the genetic system can constrain phenotypic evolution (as shown in Sects. 2, 3). However, it also seems inappropriate to place too much emphasis on the role of genetic constraints in evolution because the genetic system as such is subject to evolution. The streetcar theory hinges on the idea that evolution has a tendency to change genetic arrangements when genetic constraints become a strong impediment to phenotypic adaptation.

In our models, the evolution of genetics is made possible by taking a particularly wide range of potential mutant alleles into account. This is indeed what makes the course of the evolving population look like the journey of a streetcar. Sooner or later, an appropriate allele will destabilize a phenotypically maladaptive state. This allele enters the streetcar like a new passenger and the streetcar starts moving. In other words, the population begins to evolve away from its phenotypic rest point. The biological streetcar may again come to a halt where there is another passenger, etc. At a final stop, no such passenger exists. Here, it must be emphasized that 'final' means the end of the process of successive removal of genetic constraints in our theory about adaptation but it does not mean that in reality evolution would stand still forever. A change of the environment might lead to new evolution as well as a number of molecular processes (e.g. meiotic drive, segregation distortion, molecular drive) that disturb the picture of Mendelian genetics. At present, the streetcar theory only analyzes how this 'clean' Mendelian genetics affects phenotypic evolution. It would be a desirable extension of this theory to answer the question of what keeps genetics so clean and why the scope for selfish genetic elements is so limited.

What are the properties of final stops? If the phenotypic scenario is depicted by a game in strategic form we learn from Theorem 1 that the population mean strategy $p$ corresponds to a Nash equilibrium. This property happens to be the most important criterion for a solution of such a game in economic decision theory. Thus we are dealing here with an economic property of the final stop. The important interdisciplinary message behind this result is that the evolutionary process and the process of rational economic decision making both come to an important halt at the same point. This can be taken as an indication that major parts of economic theory are relevant to biological evolution and perhaps even more relevant to biology than to human economics.

**Rationality paradox:** Humans typically claim a 'monopoly' on rational decision making when they compare themselves with the animal world. However, from recent literature in social psychology and experimental economics we know that actual human decision processes deviate systematically from the principles of economic rationality. Paradoxically we learn from contemporary biology that animals often behave as if they had taken a course in rational economic decision making.

'Streetcar Theorem l' helps us indeed to understand this paradoxical aspect of the animal world. The empirically documented 'quasi-rationality' of animal behaviour creates some confidence in the importance of Theorem 1 for evolutionary biology. Despite its simple proof, this theorem seems to support more biological facts than Fisher's fundamental theorem of natural selection or Wright's adaptive topography in its classical form. We have seen in Sects. 2 and 3 how the latter principles fail to operate if one only changes from one-locus to two-locus population genetics. Theorem 1 also shows that a naive interpretation of the idea of the 'selfish gene' (Dawkins 1976) can easily direct our attention to an inappropriate level of biological organisation (genes instead of phenotypes). This is so because in the $n$-locus case of our selection model the genetic scene can only be described as an 'incredible mess' although very clear economic principles hold – in the long run – at the phenotypic level.

Are there results stronger than Theorem 1? If we try to characterize in general all those final stops that are phenotypically polymorphic (i.e. stops in which different genotypes occur that have different phenotypic specifications), then Theorem 1 is the best we can hope for. A stronger result can be achieved if one focuses attention on phenotypically monomorphic stops in which all genotypes play the same strategy. From 'streetcar Theorem 2' we learn the following. If these stops are final stops, the population mean strategy is an evolutionarily stable strategy as defined by Maynard Smith and Price (1973). Conversely, if the population mean strategy is an ESS, a monomorphic stop is long-term stable against mutant alleles that cannot generate the ESS in heterozygous condition. It is not astonishing that the streetcar theory lends its best support to the monomorphic case since the ESS-concept was originally developed with the intuitive picture of a monomorphic population in mind. Note here once again that 'monomorphic' does not mean fixed alleles for all relevant loci, nor does it mean that all animals act alike as if they were chorus girls on a Broadway stage. A strategy is after all a programme that specifies how animals 'translate' the information about their individual situation into action patterns.

One further theorem about streetcars should be mentioned here which was not stated explicitly in the present paper. If selection is frequency-independent, the final stops of the streetcar are those at which expected fitness of phenotypes is maximized. This criterion is much simpler than the complex optimization criterion of the Nash equilibrium. Altogether the streetcar theory follows the conceptual recipe often used in theoretical physics: 'define stable equilibrium in an appropriate way and characterize it afterwards by appropriate optimality conditions'.

The discussion of final stops would be incomplete without mentioning the possibility that there may not exist a final stop for an evolving population. Furthermore, there may not be enough time for evolution to reach a final stop before dramatic changes occur in the environment. Needless to say that the latter problem is very likely to occur in the evolution of human behaviour because of a quickly changing cultural background. Therefore, the streetcar theory reveals the fragile theoretical ground on which human sociobiology is

built rather than providing strong support for this biological excursion into human affairs. Similarly, one would expect in host-parasite coevolution that parasites evolve too quickly for their hosts to reach a final stop. Probably it is more than just a coincidence that the best known example of a temporary stop with heterozygote advantage is that of sickle cell anemia.

It would indeed mean to throw out the baby with the bathwater if we ignored the biological relevance of temporary stops. These are the halting points of evolutionary processes that definitely require proper treatment in the explicit framework of population genetics. However, unless there are strong hints in favour of temporary stops or of non-Mendelian genetic effects, researchers in evolutionary biology are well advised to use the modern theory of phenotypic adaptation when they build their first working hypotheses. Charles Darwin would be pleased to see how well this modern theory has maintained the flavour of Darwinism.

*Acknowledgements.* I wish to thank Franjo Weissing for the long discussions we had about Darwinian streetcars. I am grateful to Odo Diekmann, Josef Hofbauer, Eörs Szathmáry, and Rolf Weinzierl for a number of very helpful comments. I also thank the Collegium Budapest for supporting the final revision of this paper.

## References

Dawkins, R. (1976). The selfish gene. Oxford: Oxford University Press

Eshel, I. (1991). Game theory and population dynamics in complex genetical systems: the role of sex in short term and in long term evolution. In: R. Selten (Eds), Game Equilibrium Models I: Evolution and Game Dynamics (pp. 6–28). Berlin: Springer–Verlag

Eshel, I. (1983). Evolutionary and continuous stability. J. Theor. Biol. **103**, 99–111

Eshel, I. (1995). On the changing concept of population stability as a reflection of changing problematics in the quantitative theory of evolution. J. Math. Biol. **34**, 485–510

Eshel, I. and Feldman, M. W. (1984). Initial increase of new mutants and some continuity properties of ESS in two locus systems. Am. Nat., **124**, 631–640

Ewens, W. J. (1968). A genetic model having complex linkage behavior. Theor. and Appl. Genet., **38**, 140–143

Fudenberg, D. and Tirole, J. (1991). Game Theory. Cambridge, Massachusetts: The MIT Press

Gould, S. J. and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme: Proc. R. Soc. Lond. B, **205**, 581–598

Hammerstein, P. and Selten, R. (1994). Game theory and evolutionary biology. In: R. J. Aumann and S. Hart (Eds), Handbook of Game Theory with Economic Applications. Volume 2, pp. 929–993. Amsterdam: Elsevier

Harsanyi, J. C. and Selten, R. (1988). A general theory of equilibrium selection in games. Cambridge, Massachusetts: The MIT Press

Karlin, S. (1975). General two-locus selection models: some objectives, results and interpretations. Theor. Pop. Biol., **7**, 364–398

Lessard, S. (1984). Evolutionary dynamics in frequency-dependent two phenotype models. Theor. Pop. Biol., **25**, 210–234

Liberman, U. (1988). External stability and ESS: criteria for initial increase of a new mutant allele. J. Math. Biol., **26**, 477–485

Maynard Smith, J. (1978). Optimisation theory in evolution. Am. Rev. Ecol. Syst., **9**, 31–56

Maynard Smith, J. (1982). Evolution and the Theory of Games. Cambridge: Cambridge University Press
Maynard Smith, J. and Price, G. R. (1973). The logic of animal conflict. Nature, **246**, 15–18
Moran, P. A. P. (1964). On the nonexistence of adaptive topographies. Am. Human Genet., **27**, 338–343
Nash, J. F. (1951). Non-cooperative games. Ann. Math. **54**, 286–295
Tyszka, T. (1983). Contextual multiattribute decision rules. In: L. Sjöberg, T. Tyszka and J. A. Wise (Eds), Human Decision Making (pp. 243–256)
Weissing, F. J. (1995). Genetic versus phenotypic models of selection: can genetics be neglected in a long-term perspective? J. Math. Biol. **34**, 533–555
Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proc. XI. Internat. Congr. Genetics, **1**, 356–366