

The phenotypic gambit: selective pressures and ESS methodology in evolutionary game theory

Hannah Rubin¹

Received: 28 May 2015 / Accepted: 21 April 2016 / Published online: 11 May 2016
© Springer Science+Business Media Dordrecht 2016

Abstract The ‘phenotypic gambit,’ the assumption that we can ignore genetics and look at the fitness of phenotypes to determine the expected evolutionary dynamics of a population, is often used in evolutionary game theory. However, as this paper will show, an overlooked genotype to phenotype map can qualitatively affect evolution in ways the phenotypic approach cannot predict or explain. This gives us reason to believe that, even in the long-term, correspondences between phenotypic predictions and dynamical outcomes are not robust for all plausible assumptions regarding the underlying genetics of traits. This paper shows important ways in which the phenotypic gambit can fail and how to proceed with evolutionary game theoretic modeling when it does.

Keywords Evolutionary game theory · Philosophy of biology · Methodology · Evolutionary models

Uncertainty about the underlying genetics of a trait is often used in evolutionary game theory as a justification for the ‘phenotypic gambit’—the assumption that we can ignore genetics and look at the fitness of phenotypes, the observable traits of organisms determined by some mixture of genetics and environment, to determine the expected evolutionary dynamics of a population. In phenotypic models, it is assumed that reproduction is asexual and that offspring are clones of their parent. Genetic models on the other hand take into account that in sexually reproducing organisms, offspring get their genetic material from both parents. Offspring may have a different genetic make-up from either of their parents and thus a different phenotype as well. The genes encoding for different traits become the focus of

✉ Hannah Rubin
hannahr@uci.edu

¹ Department of Logic and Philosophy of Science, University of California, Irvine, 3151 Social Science Plaza A, Irvine, CA 92697-5100, USA

analysis rather than only the traits themselves. While the phenotypic approach both simplifies the mathematics and often generates accurate predictions of real world systems, it can also yield false predictions (Marrow et al. 1996; van Oers and Sinn 2011).

The emphasis on phenotypes rather than genotypes is so prevalent in evolutionary game theory that it is generally considered a phenotypic approach to evolutionary modeling (Hammerstein and Selten 1994, 935). However, there is no reason why evolutionary game theoretic models cannot be genetic in character. This paper will give concrete examples of genetic models in evolutionary game theory and demonstrate why the use of genetic models is important. In particular, it will show that varying which phenotype is associated with a particular genotype in genetic models can drastically affect the dynamics in a way that the phenotypic approach cannot predict or explain.

The organisms in the genetic models considered here are diploid; they have two sets of chromosomes. This means they can have two copies of a gene at a locus, or a place on a chromosome where a particular gene is found. Organisms may be homozygous, have two copies of the same allele (version of a gene), or heterozygous, have two different alleles, with respect to some locus. This paper will use evolutionary game theory to show that what assumptions are made about the heterozygote's phenotype can change both the type of selective pressure acting on the population and the expected outcome of evolution. As we will see, this is in contrast to the relatively unimportant role often played by heterozygotes when genetics are incorporated into evolutionary game theoretic models (“Preliminaries: genotype to phenotype maps and evolutionary stable strategies, Selective Pressures and Stability and ESS Methodology” sections).

This paper will analyze several models of well-known games—namely the Nash Demand game, Hawk–Dove, and the Prisoner’s Dilemma—to show how genetics can affect the evolutionary process. This will show some ways in which the phenotypic gambit can fail to predict evolutionary outcomes. This is important because the phenotypic gambit plays an important role in evolutionary modeling, especially in what are called ‘adoptionist programs’ or optimization approaches to evolutionary theory. These approaches are built on the expectation of fitness-maximization, or the expectation that evolution will lead to some optimal set of phenotypes. Genetic constraints which would prevent the population from reaching this optimal point are “a fly in [the] ointment” for the phenotypic gambit and optimization approaches (Grafen 2014, 157). While we know the assumptions of phenotypic gambit are false, we could be comfortable relying on it if we knew that it gave us *close* to the right answer in most cases and if we could categorize the exceptional cases in which predictions based on the phenotypic gambit would not be approximately true (Grafen 2014).

Part of the analysis presented here will demonstrate how the well-known phenomenon of heterozygote advantage, where the heterozygote is more fit than either homozygote, can transfer over from the non-social to the social context, where evolutionary game theory is used. Heterozygote advantage has been commonly used as a counter-example to adoptionist programs (see, for example Grafen 2014). If the heterozygote is most fit, then genetic constraints prevent the population from evolving to become composed entirely of the most fit organisms, as will be further explained (in the “ESS methodology” section). This paper will show that in the social context there

is an additional complication for adaptationist programs: the phenotype of the heterozygote can affect evolutionary outcomes even when its fitness is between the fitness of the two homozygotes.¹ Why this is important will be explained further (in the “[Selective pressures and stability](#) and [ESS methodology](#)” sections).

This demonstrates how the phenotype assigned to the heterozygote can importantly affect evolutionary processes and outcomes. However, it is often argued that in the long-term, the effects of these genetic constraints can be ignored and we can predict the evolutionary outcome based on phenotypic stability criteria, such as an evolutionary stable strategy (ESS) (see Marrow et al. 1996 and references therein). However, as we will see, connections between ESS methodology and long-term evolutionary outcomes rely on strong assumptions that may not be reasonable for every evolutionary process. So while the phenotypic gambit may pay off in many cases, we should be aware of its limitations and we should know how to proceed with evolutionary game theoretic modeling when the gambit fails.

The paper will proceed as follows. First, I will introduce genotype to phenotype maps and their use in evolutionary game theory. I will also review some relevant topics from game theory in a population setting and explain the concept of an ESS (“[Preliminaries: genotype to phenotype maps and evolutionary stable strategies](#)” section). Then, I will discuss dynamics commonly used in evolutionary game theory for phenotypic and genetic models (“[Dynamics for phenotypic and genetic models](#)” section). With all this in place, I will consider models of a number of well-known games where differing assumptions about the heterozygote phenotype make significant differences for evolutionary outcomes (“[Models](#)” section). Next, I will discuss some lessons about selective pressures and ESS methodology that can be drawn from this analysis (“[Selective pressures and stability](#) and [ESS methodology](#)” sections, respectively). Finally, I will discuss some ways to think about phenotypic modeling in evolutionary game theory (“[Concluding remarks](#)” section).

Preliminaries: genotype to phenotype maps and evolutionary stable strategies

Evolutionary change depends on the fitness of organisms, which in turn depend on the organisms’ phenotypes. Thus changes in gene frequencies are crucially affected by what sort of genotype to phenotype mapping is assumed. A genotype to phenotype map simply describes what sort of phenotype is encoded for by each possible genotype, and so is responsible for connecting the genetic level with the phenotypic level. This mapping can be one-to-one, where each genotype encodes for a unique phenotype, or many-to-one, where multiple genotypes can encode for the same phenotype.² In the context of evolutionary game theory, the strategies of a game are

¹ In contrast, when selection is not frequency dependent, selection will eventually lead to a population composed of the most fit homozygote (although the heterozygote phenotype may affect the speed of this evolutionary process).

² In real life situations these maps are often one-to-many or many-to-many because environmental factors affect phenotypes as well as genetic factors. However, environmental factors are often ignored in models for simplicity, so it is assumed that the maps are one-to-one or many-to-one.

considered to be phenotypes of the organisms, so the genotype to phenotype map in this context determines what strategy an organism of a given genotype will play. Just as in the classic examples of heterozygotism, heterozygotes game theoretic models can play a strategy that is distinct from the strategies associated with homozygotes.³

When making the phenotypic gambit, one assumes a very simple genotype to phenotype map. That is, one assumes there is simply a gene which maps one-to-one onto a strategy. Often, genetic models also assume a fairly simple genotype to phenotype map: they assume that the heterozygote strategy is a convex combination of strategies played by the homozygotes (Maynard Smith 1982; Binmore and Samuelson 2011).⁴ A convex combination in this context is a linear combination of strategies where the coefficients are positive and sum to one. So for instance, if there are two strategies, s_1 and s_2 , a convex combination of the two would be $\alpha s_1 + \beta s_2$ where $\alpha + \beta = 1$. This means that the heterozygote could play the same strategy as one of the homozygotes (if α or β is zero), interpreted as one allele being dominant over the other. In this paper, I will provide examples where α and β aren't zero, meaning the heterozygote plays a mixed strategy.⁵

Genotype to phenotype maps, however, are not always straightforward, especially when it comes to behavioral traits.⁶ Besides mathematical convenience, there is no reason to assume that an intermediate heterozygote phenotype must be a convex combination of the homozygote strategies in the particular way commonly assumed by these biological models, or a convex combination of homozygote strategies at all. Examples of different assumptions will be discussed (in the “Models” section).

This paper uses evolutionary game theory to analyze the effects of the genotype to phenotype map. An evolutionary game theoretic analysis begins with a game that can be used as a model of some strategic interaction an organism might participate in. The payoffs of these interactions are then taken to affect the organism's fitness, so that a strategy which receives higher payoffs in the game can be generally expected to increase in frequency. In using evolutionary game theoretic analysis it is useful to relate the dynamical outcomes to equilibrium concepts in game theory, so I will introduce some of these concepts now.

Game theoretic equilibrium concepts are traditionally phrased in terms of a rational agent, but we can also apply these concepts in a biological setting where an organism's phenotype is the strategy it plays in a game. It will be useful to explain at the start how we can apply these concepts in the population setting without any assumptions about the rationality of organisms. To do this, we look at the strategy a

³ One classic example of heterozygotes exhibiting traits that differ from either homozygote is in primroses where homozygotes have either red or white flowers while the heterozygote's flowers are pink.

⁴ Hines and Bishop (1984a) explore cases where the heterozygote strategy is not a convex combination of mixed strategies, but their model is different from the ones presented here in that both the homozygotes and the heterozygotes employ mixed strategies, and mix from the same set of possible pure strategies. This is discussed in further detail in the “ESS methodology” section.

⁵ I assume for simplicity that homozygotes play pure strategies, but they could also play mixed strategies and similar analysis would apply.

⁶ Genotype to phenotype maps have been investigated with a variety of different aims. Within evolutionary game theory, for instance, Binmore and Samuelson (2011) evaluate the effect of second order forces (e.g. drift or mutation) in models with different underlying genetics.

population as a whole is playing, or the *effective strategy* of the population, by taking the weighted average of all the strategies played in the population (Cressman 1992, 61–2). In a population setting, a *Nash equilibrium* corresponds to an effective strategy of the population that is a best response to itself—no other effective strategy could do better if played against it. This could be a *pure Nash equilibrium*, where every organism plays the same pure strategy, or a *mixed Nash equilibrium*, where more than one strategy is played within the population.

A pure Nash equilibrium hashed out as a population state corresponds to a *phenotypic monomorphism*—a state where every organism in the population has the same strategy (phenotype). A mixed Nash equilibrium could either correspond to a phenotypic monomorphism, where every organism mixes over a number of pure strategies with the same probability, or a *phenotypic polymorphism* in which there are multiple strategies (phenotypes) present. A phenotypic polymorphism could then have organisms playing strategies with different probabilities with the effective strategy of the population corresponding to a mixed Nash equilibrium. How the individual strategies of organisms combine to determine the effective strategy of a population can have important effects on selective pressures and evolutionary outcomes (Bergstrom and Godfrey-Smith 1998), but for the purposes here these distinctions will not be important.

One of the most commonly used concepts of evolutionary stability in evolutionary game theory, an *evolutionary stable strategy* (ESS), is a refinement of the Nash equilibrium concept:

A strategy s^* is an ESS if and only if

1. $u(s^*, s^*) \geq u(s, s^*)$ for any s in the strategy set, and
2. if $u(s^*, s^*) = u(s, s^*)$, then $u(s^*, s) > u(s, s)$.

where $u(x, y)$ is the utility (or fitness benefit) an organism using strategy x gets from an interaction with an organism using strategy y . That is, (1) a strategy s^* is a best response to itself—it is a Nash equilibrium—and (2) if another strategy s is also a best response to s^* , then s^* does better against s than s does against itself. This is a way of making precise the idea that an ESS is an ‘uninvadable’ strategy, that when a population is at an ESS other strategies cannot arise through mutation and invade the population. For population games, it is helpful to view s and s^* as possible effective strategies of the population (Eshel 1996).

Dynamics for phenotypic and genetic models

In phenotypic evolutionary game theoretic models, the replicator dynamics are often used to model the evolutionary process. Under these dynamics, if the fitness of a strategy is greater than the average fitness of the population, the frequency of that strategy will increase. A strategy’s fitness is determined by its payoffs against other strategies in addition to the population composition, which determines how likely it is for organisms using that strategy to receive each of their payoffs. Let us call the allele in question A_1 , its frequency p , and its fitness w_p . The frequency of the other

allele, A_2 , is q where $q = 1 - p$ and its fitness is denoted by w_q . The replicator dynamic is then governed by the following equation:

$$\dot{p} = p[w_p - (pw_p + qw_q)]$$

where $(pw_p + qw_q)$ is the average fitness of the population.⁷ Although I have described this in terms of frequencies of alleles, phenotypes are determined by a single allele so tracking changes in allele frequencies is equivalent to tracking changes in phenotype frequencies.

For genetic models of randomly mating, infinite populations, the Hardy–Weinberg dynamics is oftentimes employed. This is similar to the replicator dynamics but incorporates features of sexual reproduction. In this dynamic, offspring get their genetic material from two parents, one allele from each (unlike in the replicator dynamics where offspring are assumed to be clones of their parent). Selective pressure depends on the fitness of a strategy, but ultimately the fitness of each allele determines the evolution of a population. This is because the fitness of an organism is determined by its phenotype, but the frequency of alleles passed on through random mating will determine the phenotypes present in the next generation. The evolutionary trajectory is governed by the following equation:

$$p_{next} = \frac{w_{11}p^2 + w_{12}pq}{w_{11}p^2 + w_{12}2pq + w_{22}q^2}$$

where w_{xy} represents the fitness of the A_xA_y genotype. The numerator of this equation is the fitness of the A_1 allele⁸ and the denominator is the average fitness of the population.⁹ Since alleles recombine according to the laws of probability, p^2 is the frequency of individuals homozygous with respect to A_1 (denoted A_1A_1), q^2 is the frequency of A_2A_2 homozygotes and $2pq$ is the frequency of heterozygotes (A_1A_2).

In order to discuss the dynamical outcomes in the next section, it is important to first explain what I mean by the term ‘stable equilibrium’. First, a *steady state* is configuration of genotypes within a population where allele frequencies do not change over time in the absence of any drift or mutation. A *stable equilibrium* is then a steady state with the additional property of being *locally asymptotically stable*: when there is a small amount of mutation or drift that changes the gene frequencies, selective pressure will cause the population to return to the steady state.

The ESS concept is useful because it has been proven to match up with many common dynamics, such as the replicator dynamics and the Hardy–Weinberg dynamics, in such a way that if an effective strategy is an ESS of the game, then under certain assumptions it is also a stable equilibrium of the dynamic (Hofbauer and Sigmund 1998). The connection between the replicator dynamics and the ESS

⁷ This assumes there are only two alleles in the population, but of course the dynamics can be extended to account for more than two alleles. The same is true of the Hardy–Weinberg dynamics.

⁸ This is calculated from the fitness of the A_1A_1 homozygote times its frequency, plus the fitness of the heterozygote times pq . The frequency of these heterozygotes is $2pq$, but since they only pass on the A_1 allele half of the time, in calculating the fitness of A_1 , the heterozygote fitness is multiplied by half their frequency.

⁹ This is just the weighted average of the fitness of each genotype.

concept can be fairly straightforward: a strategy is a stable equilibrium of the replicator dynamics if it is an ESS of the game (Hammerstein and Selten 1994). However, even in phenotypic models, there are still many cases where an ESS of a game will fail to coincide with any steady states of the dynamics (Huttenegger and Zollman 2013). The connection between the ESS concept and stable equilibria in models which incorporate genetics is even less straightforward. The usefulness of ESS methodology for predicting outcomes in these models will be discussed (in the “ESS methodology” section).

In what follows I will look at the effects of the heterozygote strategy on the applicability of the standard phenotypic dynamics and ESS methodology in genetic models. This is, of course, only one particular de-idealization. The models here still make many assumptions in order to simplify analysis. In particular, it is assumed that the population is infinite and that interactions between organisms are random. Removing either of these assumptions can also affect evolutionary dynamics and the usefulness of ESS methodology (see, for example, Young and Foster 1991; Nowak et al. 2010; Fogel and Fogel 2011).

Models

In this section, I will present three examples using well-known games to illustrate the effects of different phenotype to genotype maps. Each example will begin with a description of a game. This will be followed by the analysis of a version of the game with two strategies using both replicator dynamics and a genetic model where the heterozygote plays a mixed strategy (this sort of genetic model will be referred to as the ‘mixed strategy case’). I will then provide an analysis of a version of the game with three strategies using replicator dynamics, followed by a genetic model in which the heterozygote plays one of the three pure strategies (the ‘pure strategy case’). In these models, the allele associated with the first strategy listed in the payoff table will be called A_1 and its frequency p . For ease of comparison, in the examples considered here I restrict attention to cases of intermediate inheritance, where the heterozygote phenotype is ‘between’ the homozygote phenotypes.

Example 1: Nash demand

The Nash Demand game represents a situation in which two actors try to divide a resource by demanding a certain fraction of the contested resource. If the two demands together are less than or equal to the total amount of the resource, each gets the share they demanded. If the demands add up to more than the whole, neither gets anything. Similar to Skyrms (1996), I assume that there is one allele for demanding $1/3$ of the resource and another for demanding $2/3$. We can get the fitness of each phenotype from the payoffs in the first and third rows of Table 1 (along with the population distribution).

Under the replicator dynamic, there is a stable polymorphic equilibrium at $p = 0.5$, where half the population is demanding $1/3$ and the other half is demanding $2/3$. In the genetic model with these same strategies, one homozygote

Table 1 Nash Demand game

	Demand 1/3	Demand 1/2	Demand 2/3
Demand 1/3	1/3	1/3	1/3
Demand 1/2	1/2	1/2	0
Demand 2/3	2/3	0	0

(A_1A_1) demands 1/3 and the other (A_2A_2) demands 2/3. The heterozygote (A_1A_2) will be modeled as mixing between these two demands with equal probability. How the heterozygote mixes between the two demands is arbitrary and does not qualitatively affect the dynamic in the sense that it does not alter the effective strategy, the weighted average of the strategies played in the population, at equilibrium. Why this is true of the mixed strategy case will be explained in the next section. There is again a stable polymorphic equilibrium at $p = 0.5$, where 25 % of the population demands 1/3, 50 % demands 1/3 half the time and 2/3 half of the time, and 25 % demands 2/3. So, half of the demands made in the population are for 1/3 and the other half are for 2/3; the effective strategy of the population is the Nash Equilibrium.

If the homozygotes demand 1/3 and 2/3 of a resource, the heterozygote could play the commonly assumed convex combination of these and mix between demanding these amounts. However, heterozygotes may also demand some portion of the resource between 1/3 and 2/3, say 1/2, which can also be viewed as convex combination of homozygote strategies since the amount demanded is a convex combination of the amounts demanded by each homozygote. These are both possible intermediate strategies and both might be of interest from a modeling standpoint.

Under the replicator dynamic with these three strategies, there are two stable equilibria: one in which the entire population demands 1/2 and another in which half the population demands 1/3 while the other half demands 2/3 (Skyrms 1996). These effective strategies are both Nash equilibria of the game. On the other hand, in a genetic model with the heterozygote demanding 1/2, the population always evolves to a stable polymorphic equilibrium at $p \approx 0.63$. This means that roughly 36.7 % demand 1/3, 46.7 % demand 1/2, and 13.7 % demand 2/3.¹⁰ The reasons for ending up at this alternative equilibrium this will be discussed in the next two sections. For now, we can note that the polymorphism in the mixed strategy case, where the population demands are split evenly between demanding 1/3 and demanding 2/3, is not possible in the pure strategy case because, even if the population were to start at this effective strategy, mating would lead to the production of heterozygotes demanding 1/2 in the next generation.

Example 2: Hawk–Dove

Maynard Smith and Price (1973) introduced the Hawk–Dove game as a description of certain animal conflicts. ‘Hawk’ refers to an aggressive animal that fights for a

¹⁰ This case has also been considered by Tennant (1999), but the sorts of dynamic and stability concerns addressed here were not at issue.

Table 2 Hawk–Dove–Retaliator

	Dove	Retaliator	Hawk
Dove	1	1	0
Retaliator	1	1	−1
Hawk	2	−1	−1

resource and ‘dove’ refers to an animal that is unwilling to fight. A hawk then always gets the contested resource when encountering a dove, but will split the resource and cost of entering into a conflict (injuries incurred, expended energy, etc.) with another hawk. A dove will always surrender the resource to a hawk, but will split the resource peacefully with another dove. As in Maynard Smith (1982), when the value of the resource is 2 and the cost of conflict is 4, the payoffs are as in Table 2.

Under the replicator dynamics, there is one stable equilibrium at $p = 0.5$ where each allele is found in equal frequency and each strategy is played by half of the population. In the mixed strategy case, there is again one stable equilibrium at $p = 0.5$, but this corresponds to a population where 25 % of the population always plays dove, 25 % always plays hawk and 50 % mixes between the two strategies with equal probability. The effective strategy of the population is to play each strategy with equal probability, the Nash equilibrium of the game.

In the Hawk–Dove game, the heterozygote could be a ‘retaliator,’ which will respond to a hawk’s aggressive behavior with aggressive behavior, but acts like a dove otherwise. This amounts to a novel pure strategy which is not a convex combination of the homozygotes’ strategies.¹¹ With these three strategies, under the replicator dynamics the only stable equilibrium in which half of the organisms play hawk and half play dove, like in the game with two strategies (Maynard Smith 1982).¹² When heterozygotes are retaliators, the population ends up at a stable polymorphic equilibrium at $p \approx 0.89$. Roughly 69.4 % of the population plays dove, 27.8 % plays retaliator, and 2.8 % plays hawk. In the pure strategy case, there is much more dovish behavior than in the mixed strategy case. This is because the retaliator heterozygote acts as sort of a buffer for the doves against invading

¹¹ It is, however, still in some sense ‘intermediate’ between the two homozygote strategies because an organism with this type of strategy will sometimes play hawk and sometimes play dove, depending on what type of organism it interacts with. If we are thinking in terms of alleles encoding for a ‘dose’ of some gene product, we can think of the retaliator as having one ‘dose’ of whatever causes the disposition for hawkish behavior and one ‘dose’ of whatever causes dovish behavior—the retaliator has some disposition toward peaceful behavior, as seen when it interacts with a peaceful organism, but also has the disposition to act aggressively if faced with another aggressive organism. The same sort of explanation can be given in example 3 for why tit-for-tat can be considered an intermediate phenotype.

¹² Selective pressures can also lead toward a steady state of all retaliators or of some combination of retaliators and doves. This set of states can be evolutionarily important in that it can attract a large portion of the possible initial populations. However, states within this set are not stable equilibria in the sense defined above, but are instead what is called *Lyapunov stable*. That is, when there is some mutation, evolutionary pressures will not drive the population back to the state in which it started. Evolutionary pressures will instead drive the population to a state *near* where it started. A somewhat similar situation occurs with tit-for-tat in example 3. This sort of phenomenon is also taken to show the limitations of ESS methodology, but will not be addressed in detail here (see Huttegger and Zollman 2013, for a discussion).

hawks—as hawks and thus the A_2 allele increase in frequency so do the retaliators. The hawks do much less well against retaliators than do doves, so they are prevented from comprising a large part of the population.

Example 3: (Iterated) prisoner's dilemma

The Prisoner's Dilemma is another game frequently encountered in the biological literature. It is often used to explain the evolution of biological altruism as an organism who 'cooperates' can be interpreted as sacrificing its own fitness in order to increase the fitness of another.¹³ I assume that an individual can sacrifice 1 unit of fitness to increase the fitness of another by 3 units. Further, I assume the game is iterated, or played repeatedly against the same organism, 10 times. The payoffs are shown in Table 3.

In the game with only cooperators and defectors, there is only one stable equilibrium which corresponds to a population of all defectors ($p = 0$), the Nash equilibrium, for both the replicator dynamics and the mixed strategy case.

The fact that the game is iterated allows for a variety of possible additional strategies. One commonly used strategy is tit-for-tat, which cooperates on the first iteration and only defects if defected against in the previous iteration (Young and Foster 1991, Axelrod 2006, Nowak 2006). Applying the replicator dynamics to the game with these three strategies reveals that again there is only one stable equilibrium where the population is composed entirely of defectors.¹⁴ However, in the pure strategy case with the heterozygote playing tit-for-tat, there are two stable equilibria. Again, one of these is composed of all defectors, but there is another at $p \approx 0.81$ where of the population roughly 65.7 % cooperate, 30.7 % play tit-for-tat, and 3.6 % defect. This is similar to the phenomenon that occurs in the Hawk–Dove game. Here, there is more altruistic behavior in the pure strategy case than in the mixed strategy case because in the pure strategy case, tit-for-tat can act as a buffer protecting the altruists from invasion by defectors (Rubin 2015).

One might be concerned that, at least with Hawk–Dove and the iterated Prisoner's Dilemma, the heterozygote strategies chosen in the pure strategy case are more cognitively demanding than the heterozygote strategies in the mixed strategy case. These heterozygote strategies were chosen because they are salient in the literature. A reader familiar with models of these games will immediately see how the results differ from those produced in other models of the same games. However, the results are not dependent on the cognitively complexity of the heterozygote strategy; the point can be made with models using less cognitively complex heterozygote strategies. For an example of such models for Hawk–Dove, see the "[Appendix](#)".

¹³ An altruistic act in this sense is just one that decreases the fitness of the actor and increases the fitness of another.

¹⁴ There are also steady states composed of mixtures of altruists and organisms playing tit-for-tat, but these are not stable equilibria in the sense described above.

Table 3 Iterated Prisoner's Dilemma

	Cooperate	Tit-for-tat	Defect
Cooperate	30	30	-10
Tit-for-tat	30	30	-1
Defect	40	4	0

Selective pressures and stability

Before investigating stability criteria, there are a couple of significant observations to be made about selective pressures in the examples above. The selective pressure acting on the population at the phenotypic level is best visualized by graphing the fitness of each phenotype in the population for different proportions of strategies, as in Fig. 1A–C. The first thing to note is that in the pure strategy case there are phenotypic fitness differentials at a stable equilibrium. It might seem odd to call such a population state an equilibrium, as equilibria are generally thought of as population states in which there are no fitness differences. However, each *allele* has the same fitness at equilibrium, so the allele frequencies, and thus the genotype and phenotype frequencies, do not change over time. Second, this means that different types of selective pressure are at work in pure strategy cases. For example, in Fig. 1Ab, the heterozygote strategy is the most fit around the equilibrium. This points to an important aspect of the dynamics of the pure strategy case: heterozygote advantage can play a vital role in maintaining a stable equilibrium of the population.¹⁵ While not generally considered as part of evolutionary game theoretic analysis, there are noteworthy examples of heterozygote advantage. One classic example is the sickle cell mutation, which causes sickle cell anemia in people with two mutated genes, but in heterozygotes produces malaria resistance with little to no symptoms of sickle cell anemia.

Neither of these features are found in the mixed strategy case. When the effective strategy is a pure Nash equilibrium, there is only one strategy present in the population, so all strategies present will have the same fitness by default (because there is only one strategy present in the population). As seen in Fig. 1Aa, Ba, when the effective strategy of the population is a mixed Nash equilibrium, all strategies present have the same fitness so the population composition is not changing. In fact, in the mixed strategy case (with intermediate inheritance), there can *never* be fitness differentials at equilibrium because the heterozygote's fitness is always between the fitness of the two homozygotes' (except at an equilibrium where all the phenotypes present have the same fitness). Since the heterozygote plays the same strategy as one homozygote with some probability and the strategy of the other homozygote with some probability, it will receive the same payoff as one or the other homozygote with some probability, meaning that its fitness is a weighted average of the

¹⁵ I should make it clear that I will use the term heterozygote advantage in the loose sense just to mean that the heterozygote tends to have the highest fitness. Cases of heterozygote advantage, while seemingly rare (Bubb et al. 2006), are of interest in evolutionary biology. For one thing, they are a way of maintaining genetic variability in diploid populations. However, the significance of genetics in the pure strategy case is not dependent on the heterozygote being the most fit. As can be seen in Fig. 1Ba, the heterozygote may simply shift the location of the rest point away from the population ESS.

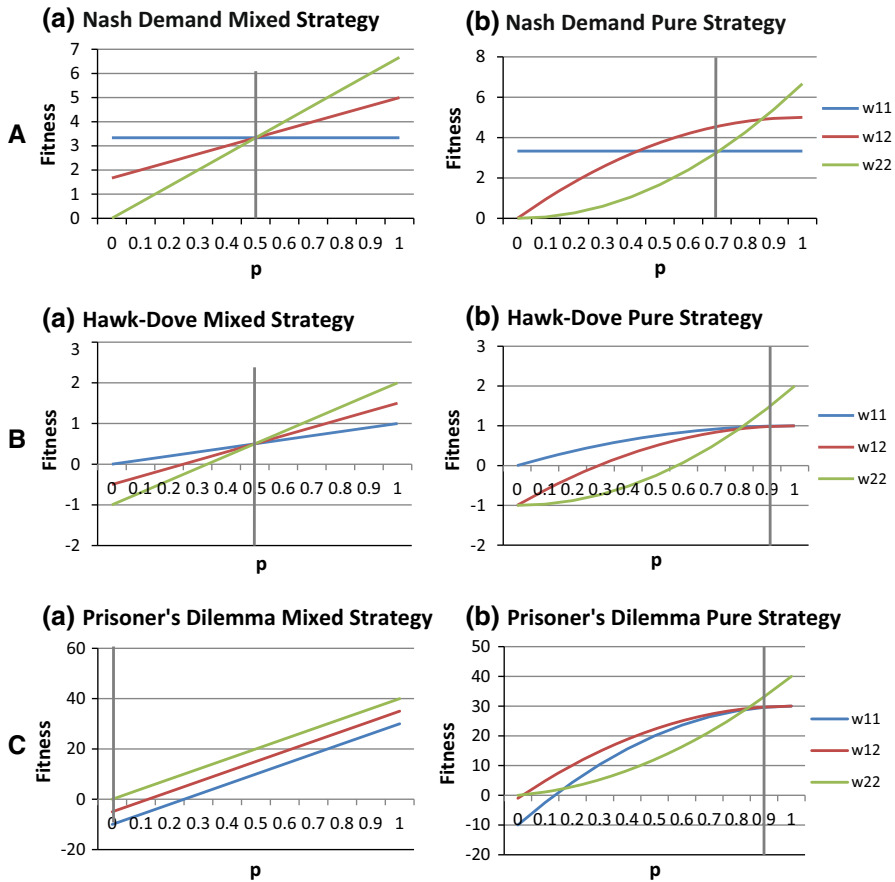


Fig. 1 Graphs of the fitness of each phenotype over possible gene frequencies: w_{11} represents the fitness of the strategy associated with the A_1A_1 homozygote, w_{12} the A_1A_2 heterozygote, and w_{22} the A_2A_2 homozygote. The stable polymorphic equilibrium is indicated by a vertical line.

homozygotes' fitnesses. That is, the heterozygote's fitness is a convex combination of the homozygote fitnesses because of the way the heterozygote strategy is a convex combination of homozygote strategies.

Heterozygote advantage has been traditionally presented as a problem for adaptationist thinking when there is no frequency dependent selection, so it should be unsurprising that remains a problem when there is frequency dependent selection. However, with frequency dependent selection there's an additional complication: the phenotype of the heterozygote can affect the location of the rest point even when it is not the most fit. For example, Fig. 1Bb shows that the heterozygote retaliator strategy is never the most fit in the Hawk–Dove game. Yet, if we compare the location of the rest point in the pure strategy case to the mixed strategy case, we see that there is much more dovish behavior in the pure strategy case. The fact that heterozygotes play the intermediate retaliator strategy rather than an intermediate mixed strategy

means that doves have a fitness advantage for a larger portion of the state space. Since the fitness of a strategy depends on the frequency of the other strategies in the population, the strategy assigned to the heterozygote can affect the location of the rest point by affecting the fitness of the homozygotes. In this case, the presence of the heterozygote retaliators increases the fitness of doves and decreases the fitness of hawks, because doves do better against retaliators than hawks do.

ESS methodology

These observations show that the intermediate strategy assigned to the heterozygote can qualitatively affect the evolutionary process. The effect can be highlighted by considering how ESS methodology fails to predict these evolutionary outcomes. We will first look at this stability concept for short-term evolution, evolution with a given set of genotypes. It is well-known that the heterozygote strategy, or other constraints, can make ESS predictions and dynamical outcomes fail to coincide in the short-term.

In the mixed strategy case, if the ESS can be reached by some combination of the strategies in the population, as is the case in the examples here, then the ESS is a stable equilibrium of a genetic model (Maynard Smith 1982).¹⁶ This means that the evolutionary outcome, in terms of the effective strategy of the population, can be predicted based solely on the payoff table of the game. This calls attention to the fact that the role the heterozygote usually plays in genetic models is relatively unimportant. The heterozygote may be assigned to play the same strategy as either one of the homozygotes or it may be assigned to play some mix of the two strategies, but in any of these cases the same phenotypic outcome is predicted by the model. The same phenotypic outcome will also be predicted by a model using the replicator dynamics. That is, the effective strategy of the population at equilibrium is not affected by a choice among one of these modeling assumptions.

This is in contrast to the pure strategy case, where we have seen that the choice of strategy assigned to the heterozygote can drastically affect the population composition at equilibrium and the population is prevented from reaching an ESS due to the probabilistic recombination of genes. In this case, the underlying genotype to phenotype map makes ESS methodology fail to be useful in describing the dynamic process, at least in the short term. This phenomenon is similar to that described in Hines and Bishop (1984a). They show, using the Rock-Paper-Scissors game, how an ESS can fail to be a stable equilibrium. In Hines and Bishop's examples, both homozygotes and the heterozygote play different mixtures of rock, paper, and scissors. Notably, they consider cases where the heterozygote strategy is not a convex combination of the homozygotes' strategies. They show how the ESS can be 'trapped': if we represent geometrically the set of all possible population effective strategies, the effective strategy corresponding to the ESS can fall in an empty space or 'hole' in this set.

¹⁶ This might not be the case if, for example, in the Hawk–Dove game each phenotype played hawk more than half the time.

Their examples illustrate how an ESS must be attainable and maintainable in order to be a stable equilibrium. It must be attainable in that some combination of the genes present in the population can yield a population effective strategy that is the ESS. For instance, if in the prisoners' dilemma there is an allele for cooperating 50 % of the time and an allele for cooperating 100 % of the time, the ESS of playing defect 100 % of the time would not be attainable. The ESS also must be maintainable in that once the population's effective strategy is the ESS it will remain that way. For instance, in the pure strategy case of the Nash Demand game, the ESS of demand 1/2 is not maintainable because even if the population were to start out entirely composed of the heterozygous demand 1/2 organisms, there would be homozygotes introduced after mating. This paper makes a point similar to Hines and Bishop (1984a), but differs in that it explores how the phenotypic gambit fails in simpler cases with a more intuitive genotype-to-phenotype map. These differences motivate concerns about the use of ESS methodology for long-term evolutionary processes, which we turn to next.

If the ESS is both attainable and maintainable then, under certain assumptions, the effective strategy of a diploid population at equilibrium is an ESS (Hines and Bishop 1983) and the ESS will be a stable equilibrium of the dynamics (Hines and Bishop 1984b). However, these results depend on assumptions about the underlying genetics and genetic considerations, such as linkage¹⁷ and epistasis,¹⁸ have led some to argue that ESS methodology is best suited for analysis of long-term evolutionary processes. (Eshel 1982, 1996; Eshel and Feldman 2001). For long-term evolution, we consider that any possible allele, from some large enough set of 'relevant' alleles, could be introduced through mutation. The process is one of successively introducing potentially beneficial mutations into a population. Each time a mutant allele successfully invades, this starts a new process of short-term evolution which evolves to its end point before another potentially beneficial allele is introduced. This is often described as the 'streetcar theory of evolution' as the end points of short-term evolutionary processes can be thought of as stops along the way to the final destination reached by long-term evolution (Hammerstein 1996). While short-term and long-term evolutionary processes are connected, they are qualitatively different.

It has been suggested that in the context of long-term evolution the phenotypic gambit is a safe assumption, that predictions based on genetic models and ESS methodology will "meet at the terminus of the evolutionary streetcar" (Marrow et al. 1996). There have been many results connecting the outcomes of long-term evolutionary processes with ESS methodology, which are proven on the assumption that *any* possible allele can invade the population (see Hammerstein and Selten 1994; Weissing 1996; Eshel and Feldman 2001 and references therein). Note that the set of possible strategies may be still be restricted to those considered in a model, so 'any possible allele' invading means that the mutant alleles can code for

¹⁷ Linkage describes the fact that genes located close together on the same chromosome tend to be inherited together.

¹⁸ Epistasis occurs when a gene or genes at one locus affect or 'modify' the effects of a gene or genes at another locus.

any propensity for playing some combination of these strategies. When we generalize from simple models to more abstract scenarios, with many possible strategies and alleles and no specific game to ground our intuitions, it might be hard to tell how strong this assumption is. The models provided above can serve to make clear the sorts of intuitions we might have when thinking about the mixed strategy case or the pure strategy case and how these intuitions transfer over into thinking against a more abstract backdrop.

In the mixed strategy case, invasion by other possible mutant alleles means that the population's genetic composition can change substantially over time. However, since the population is already at an ESS, the effective strategy of the population does not change over time. This means that any new alleles that persist in the population will have the same average fitness as the existing alleles and will only increase in frequency through drift. For instance, in example 2, the effective strategy of the population is to play hawk half of the time and dove half of the time. If a new allele were to arise within the population which caused any individual with that allele to play each strategy with equal probability, then this allele would have the same fitness as the population average fitness (since organisms with this allele play the same strategy as the population plays on average). What the underlying genetics of the end product of long-term evolution will look like cannot be predicted by ESS methodology but the effective strategy of the population can be (Eshel 1996).

When we generalize to abstract scenarios with many possible strategies and alleles, where the ESS strategy might not initially be reachable by some combination of available strategies in the population, having the mixed strategy case in mind makes considering any possible invading allele seem unproblematic: if the alleles code for different mixtures over the strategies it is reasonable to suppose that any mixture is possible in the long run. Similarly, if we think of the heterozygote as playing a different mix of the same pure strategies as the homozygotes, as in Hines and Bishop (1984a), considering mutations that allow different genotypes to play these pure strategies with different probabilities likely seems uncontroversial.

However, there is no reason to assume that a genotype-to-phenotype map needs to assign genotypes different probabilities of playing certain strategies. Genotypes might instead map to different pure strategies, as in the Nash Demand game where they mapped to different amounts an organism could demand. This leads us to consider the pure strategy case, in which the strategy with the highest fitness at the equilibrium may be played by the heterozygote and is only prevented from taking over the population by probabilistic recombination of genes. If this strategy is an ESS, as in example 1, any mutant gene that allows a homozygote to play this strategy would take over the population. Thus, if we allow that any possible mutant allele can invade, the average phenotype of the population will be different when looking at short-term versus long-term evolution.

However, in some cases there might be reason to believe that certain strategies may only be played by heterozygotes and that assigning a certain strategy to a heterozygote is the most biologically realistic modeling assumption, even in the long-term. Then, the set of relevant mutations to consider would *not* include those which would allow the population effective strategy to be the ESS. One might again think of sickle cell

and malaria resistance as an example: in heterozygotes one allele codes for normal hemoglobin and the other for sickle cells. It is not clear that a homozygote could have both the malaria resistance associated with sickle cells, and the lack of sickle cell anemia from also having normal cells.¹⁹ While things are obviously more complex in the behavioral setting, we might want to similarly restrict the set of relevant mutations to consider. This is not to argue that the pure strategy case is more biologically plausible. It is difficult to see how we could know a priori which genotype to phenotype map is more plausible in any particular case, in which case we should not ignore the possibility that the underlying genetics of certain behaviors could be more like the pure strategy case than the mixed strategy case.

There are a couple ways to think of this in relation to the argument that we can use ESS methodology to make predictions about the outcome of long-term evolution. First, and most obviously, it shows that genetic considerations can prevent ESS methodology from being used reliably to predict long term evolutionary outcomes. Heterozygote advantage is a well-known phenomenon but examples of how it can come into play in evolutionary game theory are hard to come by and it is not often discussed as a reason why evolution might not lead to an ESS. Second, you might think of this as demonstrating how strong of an assumption it is to allow that any possible allele can invade. The assumption that any allele can invade effectively assumes away the possibility that the importance of the heterozygote strategy will be maintained in the long term by assuming it is always possible that a mutation could arise which lets the heterozygote strategy be played by a homozygote. In other words, it assumes that in the long-run the ESS will always end up being genetically attainable and maintainable.

When talking about long-term evolution, the set of possible mutations is restricted to those which encode for propensities to play strategies from some predetermined set, so it is not unreasonable to suggest that we might restrict the set of possible mutations in some other way. If there is reason for a restriction in some context, then in that context there is reason to expect that the connection between ESS methodology and long-term evolutionary stability will break down. This goes to show that one should be wary of considering ESS predictions to be robust for all reasonable genotype spaces.

Concluding remarks

This paper has shown important ways in which the phenotypic gambit can fail and, when it does fail, how one can build simple genetic models which illuminate the possibilities of evolution. The models provided here demonstrate that genetic

¹⁹ For instance, one might argue that a translocation could allow for a copy of each allele to be on one chromosome. Then, an organism could be homozygous with respect to each locus (one locus is homozygous for the allele encoding sickle cells, the other locus homozygous for the allele encoding normal blood cells). However, these mutations are generally not considered in a long-term evolutionary models, which assume a specific number of loci. One way of thinking about this point is that these sort of mutations change the genetic system— a trait with one-locus inheritance, for example, would become a trait with two-locus inheritance (Grafen 1984, 65).

models can lead to interesting new equilibria in several well-known games. The pure strategy case shows how the heterozygote strategy can play an important role in evolutionary game theoretic dynamics which, for obvious reasons, is not discussed in phenotypic modeling. Relatedly, the pure strategy case shows an important area where an ESS methodology can fail to apply reliably as a stability concept for both short-term and long-term evolution. The more commonly assumed mixed strategy case obscures the fact that heterozygote strategies can play an important role in evolution by assuming the heterozygote strategy is a particular kind of linear combination of the homozygotes' strategies that makes ESS methodology easy to apply.

This is not to argue that we should be wary of models employing the phenotypic gambit. These models are a valuable way of investigating the evolutionary impacts of traits of interest. Nor is it to argue that none of the authors cited here are aware of the limitations discussed above. Maynard Smith (1982) mentioned heterozygote advantage as a problematic case for ESS methodology from the beginning. That said, models that fall under the pure strategy case are not used often because they add mathematical complexity and many times connections between ESS methodology and long-term evolutionary outcomes are taken to be robust for any type of genotype to phenotype map. The simple models here demonstrate concretely that there are plausible genotype to phenotype maps that may be of interest to modelers for which ESS methodology will not apply even in long-term evolution. Further, they show that heterozygote advantage is not the only problematic case for the phenotypic gambit and adaptationist programs; the heterozygote can affect the evolutionary outcome even when it is never the most fit.

Evolutionary game theory is useful because it provides tools for exploring the possibilities of evolutionary change, and genetic models are a necessary component of a complete exploration. Assuming we know the underlying genetics of a trait and making predictions given these assumptions provides us with important theoretical knowledge about evolutionary processes. This theoretical knowledge in turn allows us to draw conclusions about the nature of evolutionary change (Eshel 1996, 490–493). However, in drawing these conclusions it is important to consider and reconsider the assumptions made in models. When making a gambit, one should be aware of the ways in which the gambit can fail and how to proceed with modeling when it does.

Acknowledgments I would like to thank Justin Bruner, Simon Huttegger, Cailin O'Connor, Brian Skyrms, Kyle Stanford, two anonymous referees, and audiences at the Philosophy of Science 2014 meeting for helpful comments.

Appendix 1: Hawk–Dove–Goose

One might be concerned that the analysis in this paper is somehow affected by the fact that the heterozygote strategies described in the pure strategy case for the Hawk–Dove game and the iterated Prisoner's Dilemma are more cognitively complex than in the mixed strategy case. To ease this concern, I consider an alternative approach to modeling the heterozygote pure strategy in Hawk–Dove. A

Table 4 Hawk–Dove–Goose option 1

	Dove	Goose	Hawk
Dove	1	0	0
Goose	2	0	0
Hawk	2	2	−1

Table 5 Hawk–Dove–Goose option 2

	Dove	Goose	Hawk
Dove	1	0	0
Goose	2	1	0
Hawk	2	2	−1

natural suggestion is that the heterozygote does not condition their behavior on the type of their counterpart, but the level of aggression displayed is somewhere between hawkish and dovish. We might call such an individual a ‘goose’.

When a goose interacts with a hawk, the hawk gets the resource because it is more aggressive. Likewise, when a dove interacts with a goose, the goose gets the resource because it is more aggressive. What happens when two geese meet? There are two plausible options. First, the two geese may get enter a conflict and end up splitting the resource and the cost of conflict. In this case we would reasonably assume that the cost of conflict for two geese is less than the cost for two hawks, as the level of aggression displayed is less. The payoffs for the interaction among these three types is shown in Table 4.

For option 1, the replicator dynamics leads to a polymorphic equilibrium where a third of the population are geese and two-thirds of the population are hawks. By contrast, a genetic model with heterozygous geese leads to one stable polymorphic equilibrium at $p \approx 0.20$, where the population is roughly 3.9 % doves, 31.7 % geese, and 64.4 % hawks.

The second option is that the two geese end up peacefully splitting the resource. The idea behind this is that geese are aggressive enough to signal aggression but not aggressive enough to actually engage in conflict. So both will signal aggression, but back down once they see the aggressive signal. The payoffs for the second option are shown in Table 5.

In option 2, for the replicator dynamics there is one stable equilibrium, composed of 50 % geese and 50 % hawks. For the genetic model there is one polymorphic equilibrium at $p \approx 0.26$, where about 6.8 % are doves, 38.6 % are geese, and about 54.6 % are hawks. In either case, we can see that the genetic models differ from the phenotypic models and, more importantly, differ from the genetic mixed strategy case discussed in the main text.

References

- Axelrod R (2006) The evolution of cooperation, 2nd edn. Basic Books, New York, pp 27–69
 Bergstrom C, Godfrey-Smith P (1998) On the evolution of behavioral heterogeneity in individuals and populations. *Biol Philos* 13:205–231

- Binmore K, Samuelson L (2011) Sex and evolutionary stability. *J Theor Biol* 278:63–73
- Bubb KL, Bovee D, Buckley D, Haugen E, Kibukawa M, Paddock M, Palmieri A, Subramanian S, Zhou Y, Kaul R, Green P, Olson MV (2006) Scan of human genome reveals no new loci under ancient balancing selection. *Genetics* 173:2165–2177
- Cressman R (1992) The stability concept of evolutionary game theory, Lecture Notes in Biomathematics. Ser. 94. Springer, Berlin, pp 58–59
- Eshel I (1982) Evolutionarily stable strategies and viability selection in mendelian populations. *Theor Popul Biol* 22:204–217
- Eshel I (1996) On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *J Math Biol* 34:485–510
- Eshel I, Feldman M (2001) Optimisation and evolutionary stability under short-term and long-term selection. In: Sober E, Steven Orzack NP (eds) *Adaptationism and optimality*. Cambridge University Press, Cambridge, pp 161–190
- Fogel GB, Fogel DB (2011) Simulating natural selection as a culling mechanism on finite populations with the hawk–dove game. *BioSyst* 104:57–62
- Grafen A (1984) Natural selection, kin selection and group selection. In: Krebs J, Davies N (eds) *Behav Ecol*. Blackwell, Oxford, pp 62–84
- Grafen A (2014) The formal darwinism project in outline. *Biol Philos* 29:155–174
- Hammerstein P (1996) Darwinian adaptation, population genetics and the streetcar theory of evolution. *J Math Biol* 34:511–532
- Hammerstein P, Selten R (1994) Game theory and evolutionary biology. In Aumann R, Hart S (eds) *Handbook of game theory with economic applications*, vol 2. Elsevier, Amsterdam, pp 929–993
- Hines WGS, Bishop DT (1983) Evolutionarily stable strategies in diploid populations with general inheritance patterns. *J Appl Probab* 20(2):395–399
- Hines WGS, Bishop DT (1984a) Can and will a sexual diploid population attain an evolutionary stable strategy? *J Theor Biol* 111:667–686
- Hines WGS, Bishop DT (1984b) On the local stability of an evolutionarily stable strategy in a diploid population. *J Appl Probab* 21(2):215–224
- Hofbauer J, Sigmund K (1998) *Evolutionary games and replicator dynamics*. Cambridge University Press, Cambridge
- Huttegger SM, Zollman KJ (2013) Methodology in biological game theory. *Br J Philos Sci* 64(3):637–658
- Marrow P, Johnstone RA, Hurst LD (1996) Riding the evolutionary streetcar: where population genetics and game theory meet. *Trends Ecol Evol* 11(11):445–446
- Maynard Smith J (1982) *Evolution and the theory of games*. Cambridge University Press, Cambridge
- Maynard Smith J, Price GR (1973) The logic of animal conflict. *Nature* 246(1973):15–18
- Nowak M (2006) Evolutionary dynamics of cooperation. *Proc Int Congress Math* 3(2006):1523–1540
- Nowak MA, Tarnita CE, Antal T (2010) Evolutionary dynamics in structured populations. *Philos Trans R Soc Lond B* 365:19–30
- Rubin H (2015) Genetic models in evolutionary game theory: the evolution of altruism. *Erkenntnis* 80(6):1175–1189
- Skyrms B (1996) *Evolution of the social contract*. Cambridge University Press, New York
- Tennant N (1999) Sex and the evolution of fair dealing. *Philos Sci* 66(3):391–414
- van Oers K, Sinn DL (2011) Toward a basis for the phenotypic gambit: advances in the evolutionary genetics of animal personality. In: Inoue-Murayama M, Kawamura S, Weiss A (eds) *From genes to animal behavior*. Springer, Japan
- Weissing FJ (1996) Genetic versus phenotypic models of selection: can genetics be neglected in a long-term perspective? *J Math Biol* 32(1996):533–555
- Young HP, Foster D (1991) Cooperation in the long-run. *Games Econ Behav* 3:145–156