J. theor. Biol. (1998) **194,** 391–407
Article No. jt980762

Unifying Genetic and Game Theoretic Models of Kin Selection for Continuous Traits

TROY DAY* AND PETER D. TAYLOR

Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

(*Received on* 21 *November* 1997, *Accepted in revised form on* 4 *June* 1998)

A framework is presented for unifying single locus genetic and game theoretic models of continuous traits under frequency-dependent selection when there are interactions among relatives. This framework serves two purposes. First, it is used to determine how "games between relatives" must be modeled to be genetically valid. There are two commonly employed phenotypic approaches used in this setting, and we demonstrate that, although some of their predictions are always genetically valid, others are invalid in general, and this is true for both haploid asexual and diploid sexual organisms. In particular, we show that both approaches obtain the correct equilibrium and convergence stability conditions, but neither obtains the correct condition for evolutionary stability. Unlike earlier results for discrete trait matrix games (Hines & Maynard Smith, 1979), there is no simple correspondence between phenotypic and genetic predictions, and we provide two examples to illustrate this point. It is possible however, to obtain these earlier results within the present setting by restricting attention to a particular class of fitness functions. These results demonstrate that, even when selection is weak, phenotypic models can fail if fitness is frequency-dependent. The second purpose is to determine when population mean inclusive fitness effect provides an adaptive topography in games between relatives. Our results show that the fitness function must have a special form for this to be true, and this form differs between haploid and diploid organisms.

1. Introduction

A simplistic classification scheme for evolutionary models might place them along a continuum of genetic detail. Explicitly genetic models of one or a few loci would be at one end, being most faithful to genetics. Optimization and phenotypic game-theoretic models would be at the other, presumably sacrificing genetic detail to include other biological factors.

At first glance, models of kin selection would appear to enjoy a rather unique status, occupying both ends of this hypothesized spectrum. Many kin selection models are explicitly genetic and track allele frequency changes over time (e.g. Gayley, 1993 and references therein). In particular, many of these are single locus, diallelic models and are used to model the evolution of discrete traits with two possible states (e.g. altruism or not). Also placed

Present address: Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z4.

* Author to whom correspondence should be addressed.

0022–5193/98/190391 + 17 $30.00/0

at this end of the spectrum might be the so-called inclusive fitness models for discrete traits based on Hamilton's Rule (Hamilton, 1964; Michod, 1982; Bulmer, 1994). Although these usually do not use population genetic arguments, they are founded on the realization that, under appropriate assumptions, the same results are obtained as those from more explicit genetic treatments (Taylor, 1996a and references therein). Nearer the other end of the spectrum however, might lie most kin selection models of continuous traits (e.g. sex ratio, dispersal probability; Bulmer, 1994). In most such cases the fitness of any individual depends on what other individuals in the population are doing (i.e. it is frequency-dependent) and therefore these models are often allied more closely with phenotypic, continuous-trait game theory than with single locus genetics.

As a result, the literature on kin selection is divided. It is not obvious how the findings of single locus genetic models for discrete traits tie in with kin selection models of continuous traits. In particular, for single locus genetic models of discrete characters, an adaptive topography can often be defined such that population mean inclusive fitness effect increases as allele frequencies change (Hamilton, 1964; Michod & Abugov, 1980; Gayley, 1993 and references therein). This provides an important unifying concept for such models, but its relationship to continuous trait models and their reliance on game theoretic notions such as evolutionary stability (ES) and convergence stability (CS) is unclear.

On the other hand, because many models of kin selection on continuous traits are purely phenotypic game theory models, it is not obvious when these are genetically valid. Progress towards answering this question has been made for discrete character, matrix games using haploid genetic models (Grafen, 1979; Hines & Maynard Smith, 1979, see also Queller, 1984), but the applicability of these results is somewhat restricted. Many fitness functions cannot be specified in matrix form, particularly fitness functions for continuous traits, and it would be useful to have results for diploid, sexual organisms as well.

This paper therefore has three aims. The first is to present a unified conceptual framework for single locus genetic and phenotypic models of kin selection on continuous traits, when fitness is frequency-dependent. There are two main approaches to constructing such phenotypic models of kin selection, both of which are founded in continuous-trait game theory (Maynard Smith, 1978; Oster *et al*., 1977; Mirmirani & Oster, 1978; Hines & Maynard Smith, 1979 and Taylor & Frank, 1996, respectively). We bring both approaches as well as a single locus genetic approach into a common, game theoretic context.

Our second objective is to use this common framework to determine when the phenotypic approaches are genetically valid. Although it is known that factors such as strong selection can cause phenotypic models of kin selection to fail (Charnov, 1977; Charlesworth, 1980; Michod & Hamilton, 1980; Seger, 1981; Bulmer, 1994; Taylor, 1996a), it is not clear that phenotypic models are valid even when selection is weak if fitness is frequency-dependent. In fact we show that, when fitness is frequency-dependent, some predictions of the phenotypic models are always valid but some are not. In particular, special care is needed when evaluating the game theoretic stability conditions, and we provide an example to illustrate this point.

The third aim is to use this common framework to illuminate the relationship between the adaptive topography results of single-locus genetic models for discrete traits and the game-theoretic results. In particular we ask, When does mean inclusive fitness effect provide an adaptive topography for frequency-dependent models with continuous traits?

2. Quantitative Trait Game Theory

Game theoretic models for quantitative traits usually begin by specifying a fitness function, $W(x, x^0)$, that gives the fitness of an individual using strategy x in a population monomorphic for the strategy x^0 (see Appendix A for notation). It is usually assumed that mutant individuals form a small enough part of the population that they do not affect the population-wide average strategy (Maynard Smith & Price, 1973; Maynard Smith, 1982; Bulmer, 1994 reviewed in Taylor, 1996a). An evolutionarily stable value of x (an ESS) say x^*, is then specified as a value

such that if all members of the population adopt that value, then any mutant value $x = x^* + \Delta x$ in that population has a lower fitness, i.e.

$$W(x^* + \Delta x, x^*) - W(x^*, x^*) \leqslant 0 \qquad (1)$$

for all Δx. Condition (1) says that fitness, W, is globally maximized in the individual's strategy value at $x = x^*$. This means that x^* is an ESS with respect to mutational jumps of any size (i.e. it is a global ESS). Using the subscripts 1 and 2 in reference to partial derivatives of W with respect to its first and second argument, local sufficient conditions corresponding to (1) are

$$W_1|_{x = x^0 = x^*} = 0 \qquad (2)$$

and

$$W_{1,1}|_{x = x^0 = x^*} < 0. \qquad (3)$$

Any x^* satisfying (2) and (3) is an ESS with respect to small mutational jumps because the use of local conditions only considers mutants that deviate from the population value by a small amount. Notice however, that if x^* is a global ESS than it is also a local ESS; therefore conditions (2) and (3) are usually employed to find values, x^*, that satisfy (1). We refer to condition (2) as the equilibrium condition and (3) as the ESS condition since (2) holds whether mutants have a higher or a lower fitness than x^* whereas (3) requires that they have a lower fitness.

There is an additional stability criterion, sometimes termed convergence or continuous stability (CS; Eshel & Motro, 1981; Eshel, 1983; Taylor, 1989; Christiansen, 1991). This condition is meant to ensure the evolutionary attainability of x^*. In particular, it requires that when the population value is slightly perturbed from an ESS (i.e. $x^0 \neq x^*$), mutants with a strategy slightly closer to x^* than x^0 will have a larger than average fitness, whereas those slightly farther from x^* will have a smaller than average fitness. This is necessary for a population sufficiently close to x^* to actually evolve towards x^*. An ESS with this property is then also termed a CSS. Therefore for population values $x^0 < x^*$ we require

$$\text{if} \quad x^0 < x < x^* \quad \text{then} \quad W(x, x^0) - W(x^0, x^0) > 0 \qquad (4a)$$

$$\text{if} \quad x < x^0 < x^* \quad \text{then} \quad W(x, x^0) - W(x^0, x^0) < 0 \qquad (4b)$$

or, for small values of Δx,

$$\text{if} \quad x^0 < x^* \quad \text{then} \quad W_1|_{x = x^0} > 0. \qquad (5)$$

Analogous conditions must hold for $x^* < x^0$ which, for small values of Δx, requires that

$$\text{if} \quad x^* < x^0 \quad \text{then} \quad W_1|_{x = x^0} < 0. \qquad (6)$$

Conditions (5) and (6) are usually encapsulated in the single, local sufficient condition,

$$\left[\frac{\mathrm{d}W_1|_{x = x^0}}{\mathrm{d}x^0}\right]_{x^0 = x^*} < 0. \qquad (7)$$

In general there is no global version of (7) because the phenomenon captured by the CS condition is a local phenomenon. To produce a global condition, the evolutionary dynamics of the trait would need to be explicitly specified, and classical ESS models do not specify evolutionary dynamics (see Vincent & Brown, 1988; Iwasa *et al*., 1991; Abrams *et al*., 1993; Taylor, 1996b; Taylor & Day, 1997 for useful heuristic evolutionary dynamics in ESS models).

The above results hold for all fitness functions, W. To proceed further we now need to determine how W is specified. The first step in doing so is to specify the form of frequency-dependent interaction. In the remainder of this article we focus on pairwise interactions. Specifically, we suppose that the payoff to an actor using strategy x when interacting with an individual using strategy y is given by the payoff function $F(x, y)$. Although our results can be presented in a more general context, having a specific example at hand clarifies matters. Pairwise interactions have received a lot of attention in the literature and therefore this is probably a good choice. We also present results for a patch-structured population in which individuals play the field in Appendix E (Maynard Smith, 1982).

Given a function $F(x, y)$ we can then define W in terms of this payoff function and use this W

in the above stability conditions to generate predictions from the model. The crux of the issue at hand, however, is determining how W should be defined in terms of the payoff function under kin selection. When the interactants are *not* related the connection is very simple (Maynard Smith & Price, 1973; Maynard Smith, 1982); the payoff function *is* the fitness function (i.e. $W(x, x^0) \triangleq (x, x^0)$). When the interactants are genetically related however, this will no longer work. The genetic model and the two phenotypic models mentioned earlier each define W in different ways. The next section presents these different definitions of W, and calculates the stability conditions that result from each.

3. Genetic and Phenotypic Models of Kin Selection

First we examine how the genetic model defines fitness, W, because it provides the proper benchmark for evaluating the phenotypic approaches.

3.1. A SINGLE LOCUS GENETIC MODEL

Our genetic model is a diallelic, haploid model where one of the alleles (the mutant) is rare. The rarity assumption corresponds to the conceptual framework of the game theoretic approach, and the haploidy assumption is made for simplicity. A generalization to diploidy is presented in Appendix C.

The conceptual basis for using single locus genetics to model the evolution of a continuous trait is sometimes unfamiliar (see Taylor, 1996a for a review). The central idea in doing so is a separation of short and long-term evolutionary change (Eshel, 1996). Short-term evolutionary change is described by the standard population genetic equation for the allele frequency dynamics, with two alleles and no mutation. On a longer time scale, however, there is mutation. It is assumed that mutational events are infrequent so that the evolutionary dynamics of the two allele system reach an equilibrium before another mutational event occurs. Thus the overall evolutionary process is viewed as a cyclic process whereby a new mutation enters the population, the standard allele frequency equation describes the evolutionary dynamics until an equilibrium is reached, and then another mutation enters the population. This cycle repeats itself until some ultimate evolutionary equilibrium is reached at which presumably no new mutant can invade (Hammerstein, 1996). Notice that this process implicitly assumes that, during each selective sweep, one allele spreads to fixation so that there are only two alleles present at any time. Clearly not all mutants that invade will spread to fixation, particularly in diploid, sexual populations. Additionally, it is conceivable that no ultimate evolutionary equilibrium exists. Recent work on a general cyclic evolutionary process considers these phenomena (Hammerstein, 1996; Eshel *et al*., 1997).

Denote the mutant allele by "A" and the normal (resident) allele by "a". Thus, allele "a" codes for a particular value of the trait, x^0, while the mutant allele, "A", codes for a different value, $x = x^0 + \Delta x$ where Δx is the mutant's deviation from normal. The equation for allele frequency change is

$$\frac{\Delta \bar{g}}{\bar{g}} = \frac{1}{\bar{W}}(W^m - \bar{W}) \tag{8}$$

where W^m is the fitness of the mutant allele, and $\bar{W}$ is the mean fitness (Crow & Kimura, 1970). In general W^m will depend on x^0, x, and allele frequency, $\bar{g}$. We have divided the standard population genetic equation by $\bar{g}$ because we are interested in the limiting case where the mutant allele is rare. In the limit as $\bar{g} \to 0$, eqn (8) becomes

$$\frac{\Delta \bar{g}}{\bar{g}} = \frac{1}{W^a(x^0)}(W^A(x, x^0) - W^a(x^0)) \tag{9}$$

where the functions W^A and W^a are the genotypic fitnesses when the mutant is rare; i.e. $\lim_{\bar{g} \to 0} W^m = W^A(x, x^0)$ and $W^a(x^0) = W^A(x^0, x^0)$. Because we are assuming the mutant is rare, W^A is a function of both the mutant strategy and the normal strategy, whereas W^a is a function of the normal strategy only.

Equation (9) describes the invasion dynamics of a mutant allele, whatever its strategy, x, might be. If $\Delta \bar{g}/\bar{g} > 0$ then the allele will invade and if $\Delta \bar{g}/\bar{g} \leqslant 0$ it will not. Now if we define

$$W(x, x^0) \triangleq W^A(x, x^0), \tag{10}$$

then the game theoretic stability conditions (i.e. the ESS and CSS conditions) correspond to dynamic stability conditions of the genetic model. In particular, suppose that x^* is an ESS using the fitness function $W(x, x^0)$. From condition (1) and definition (10) this implies

$$W^A(x, x^*) - W^a(x^*) \leqslant 0, \tag{11}$$

for all $x \neq x^*$. From eqn (9), this implies that no mutant allele can invade a population in which the resident allele codes for x^*. Now suppose x^* is also a CSS using the fitness function $W(x, x^0)$. Then, for population values $x^0 < x^*$, conditions (4a), (4b), and definition (10) imply

$$\text{if} \quad x^0 < x < x^* \quad \text{then} \quad W^A(x, x^0) - W^a(x^0) > 0 \tag{12a}$$

$$\text{if} \quad x < x^0 < x^* \quad \text{then} \quad W^A(x, x^0) - W^a(x^0) < 0 \tag{12b}$$

with an analogous condition holding for $x^* < x^0$. From eqn (9), this implies that, when the resident allele codes for a value slightly different than x^*, mutant alleles that code for a value closer to x^* than x^0 can invade. These results clearly spell out the notion that, for game theoretic models to be genetically valid under haploidy, we must define the fitness function, $W(x, x^0)$ as in (10). While this result is probably obvious, it is often not realized that a similar approach can be used to determine the definition of W that makes game theoretic models genetically valid under diploidy as well [Appendix C, eqn (C.4)].

For the case of pairwise interactions and haploid genetics, the function W^A is very simple. Defining ρ as the probability that the recipient is a mutant, and noting that, when the mutant allele is rare, the relatedness of the recipient to the actor is $\bar{r} = \rho$, [Appendix B, eqn (B.1)] we have

$$W^A(x, x^0) = \bar{r}F(x, x) + (1 - \bar{r})F(x, x^0) \tag{13}$$

(Grafen, 1979).

We can now use the definition of $W(x, x^0)$ from eqns (10) and (13) to calculate the stability conditions. The global condition (1) is

$$\bar{r}F(x, x) + (1 - \bar{r})F(x, x^*) - F(x^*, x^*) \leqslant 0. \tag{14}$$

Assuming Δx is small, the local conditions (2), (3) and (7) are

$$\text{EQUIL} : F_1 + \bar{r}F_2 = 0 \tag{15a}$$

$$\text{ESS} : F_{11} + 2\bar{r}F_{12} + \bar{r}F_{22} < 0 \tag{15b}$$

$$\text{CSS} : F_{11} + (1 + \bar{r})F_{12} + \bar{r}F_{22} < 0 \tag{15c}$$

where the subscripts 1 and 2 refer to partial derivatives with respect to the first and second arguments of F, respectively. These results assume that the altered behaviour of a mutant does not alter ρ and therefore does not alter relatedness. This is a common assumption for all game theoretic kin selection models because it then allows for simple pedigree definitions of relatedness to be used (Bulmer, 1994). The local conditions also assume that the alleles differ by a small amount, as has been assumed in many polygenic quantitative genetic models based on a continuum of alleles (Kimura, 1965). Here however, while there is a continuum of possible alleles, there are at most two alleles present at any given time.

Condition (14) is what we will call the correct global condition and (15a), (15b), and (15c) are what we call the correct local conditions. Under diploidy, conditions (15a) and (15c) remain the same (with the appropriate definition of $\bar{r}$) but conditions (14) and (15b) are somewhat more complicated (Appendix C).

3.2. THE PHENOTYPIC MODELS

The two main phenotypic approaches parallel the distinction between the inclusive and personal fitness perspectives (Orlove, 1979; Hines & Maynard Smith, 1979). The inclusive fitness perspective augments an individual's fitness with the effect of that individual's behaviour on its relatives, weighted by relatedness. In terms of mutant behaviour, this approach considers only one mutant, the actor, but uses a fitness measure that includes contributions from many individuals. On the

other hand, the personal fitness perspective augments an individual's fitness with the effect of its relatives' behaviours on itself. This approach allows several mutant individuals, but measures of the fitness of only one, and relatedness serves as a measure of the probability that the other individuals will be mutants. To bring both approaches into the common game theoretic framework presented above we simply need to identify the fitness function $W(x, x^0)$ in each case.

3.2.1. *The inclusive fitness approach*

The inclusive fitness approach begins by constructing a function which is defined as an individual's inclusive fitness when using strategy $x = x^0 + \Delta x$ in a population monomorphic at x^0 (Maynard Smith, 1978; Grafen, 1979; Hines & Maynard Smith, 1979; Oster *et al*., 1977; Mirmirani & Oster, 1978). This function is constructed in analogy with the inclusive fitness expression found in Hamilton's Rule (Hamilton, 1964). The actor has a payoff given by $F(x, x^0)$ and the actor's inclusive fitness is then obtained by augmenting this with the payoff to the recipient when interacting with the actor, weighted by the relatedness of the recipient to the actor, $\bar{r}$; i.e. $\bar{r}F(x^0, x)$ (Grafen, 1982). One then defines

$$W(x, x^0) \triangleq F(x, x^0) + \bar{r}F(x^0, x). \quad (16)$$

This approach has been used and discussed by several authors (Maynard Smith, 1978; Grafen, 1979; Hines & Maynard Smith, 1979; Oster *et al*., 1977; Mirmirani & Oster, 1978), but one additional feature of expression (16) is worth emphasizing. As given, expression (16) is the inclusive fitness of an individual using strategy x rather than its inclusive fitness *effect*. The distinction is that the inclusive fitness *effect* is the *change* in fitness due to a change in behaviour whereas expression (16) is not a change. The two become equivalent however, if we standardize normal fitness to be zero [i.e. $F(x^0, x^0) = 0$]. In either case the results given below remain the same.

Using fitness function (16) in condition (1) gives

$$\mathrm{F}(x, x^*) - F(x^*, x^*) + \bar{r}(F(x^*, x) - F(x^*, x^*)) \leqslant 0. \quad (17)$$

The left-hand side of (17) is the inclusive fitness *effect* of the actor altering its behaviour by an amount, Δx from x^*. The inclusive fitness approach requires that, at a globally stable equilibrium, x^*, the inclusive fitness effect of all mutants be non-positive. For small mutational jumps, Δx, the local conditions (2), (3), and (7) are

$$\text{EQUIL} : F_1 + \bar{r}F_2 = 0 \quad (18a)$$

$$\text{ESS} : F_{11} + \bar{r}F_{22} < 0 \quad (18b)$$

$$\text{CSS} : F_{11} + (1 + \bar{r})F_{12} + \bar{r}F_{22} < 0. \quad (18c)$$

The left-hand side of (18a) is the inclusive fitness effect of the actor increasing its strategy by a small amount.

3.2.2. *The personal fitness approach*

Taylor & Frank (1996) have introduced a technical device that generates the relatedness coefficient automatically as a derivative. In their approach, the phenotype of the recipient, y, is treated as a function of the actor's phenotype and the population-wide phenotype, $y(x, x^0)$, and then the fitness function is defined as

$$W(x, x^0) \triangleq F(x, y(x, x^0)). \quad (19)$$

Using definition (19), the global ESS condition is

$$F(x, y(x, x^*)) - F(x^*, y(x^*, x^*)) \leqslant 0 \quad (20)$$

and for small Δx, the equilibrium condition (2) is

$$F_1 + \frac{\mathrm{d}y}{\mathrm{d}x} F_2 = 0. \quad (21)$$

The next step is to notice that the derivative $\mathrm{d}y/\mathrm{d}x$ is analogous to the slope of a statistical regression of y on x (Taylor & Frank, 1996). In particular, the least-squares regression of y on x has a slope of $\beta = \mathrm{cov}(y, x)/\mathrm{var}(x)$. Assuming additive genetic effects, this slope can be identified as the expected relatedness of the

recipient to the actor (Appendix B). Thus (21) is written

$$\text{EQUIL} : F_1 + \bar{r}F_2 = 0. \tag{22}$$

Taylor & Frank (1996) show that this approach works for a wide variety of population structures and types of kin interactions. What their method did not specify clearly, however, is how the function $y(x, x^0)$ is defined. They identified only the derivative of y because they only treated the equilibrium condition. For now we will make the simplest, reasonable assumption; $y(x, x^0) = \bar{r}x + (1 - \bar{r})x^0$. This specifies the recipients' phenotype as a convex combination of the actor's phenotype and the population-wide phenotype where the weighting is the relatedness coefficient. This is intuitively sensible and it is similar to Grafen's geometric notion of relatedness (Grafen, 1985).

With this choice, the ESS and CSS conditions are calculated as

$$\text{ESS} : F_{11} + 2\bar{r}F_{12} + \bar{r}^2F_{22} < 0 \tag{23a}$$

$$\text{CSS} : F_{11} + (1 + \bar{r})F_{12} + \bar{r}F_{22} < 0, \tag{23b}$$

respectively.

There is actually one other approach that is commonly used to construct phenotypic models of kin selection. Mathematically, the model is built by starting with the derivative $W_1|_{x=x^0}$ rather than a fitness function, $W(x, x^0)$ (e.g. see Taylor, 1996a). This approach then obtains the same equilibrium condition by setting this derivative equal to zero. Because the CSS condition is calculated from this derivative [eqn (7)], it obtains the same CSS condition as well. It shortcoming, however, is that because there is no actual fitness function, it is not possible to calculate the ESS condition (3) and hence we do not discuss it further.

4. When do Phenotypic Models Fail?

We can now compare the above stability conditions to determine when the quantitative phenotypic models fail. First, given that the inclusive and personal fitness approaches differ in their definition of $W(x, x^0)$, and given that each differs from the genetic definition of $W(x, x^0)$, we do not expect either phenotypic method to obtain the correct global condition. With respect to the local stability conditions, we can see that both phenotypic approaches give the correct equilibrium and CSS conditions [compare (18a) and (22) with (15a), and (18c) and (23b) with (15c)]. In general however, neither phenotypic approach obtains the correct local ESS condition [compare (18b) and (23a) with (15b)]. In fact, it is possible to have a genetic ESS that is not a phenotypic ESS of either phenotypic approach and vice versa. All of these conclusions hold under both haploid and diploid genetics.

4.1. TWO EXAMPLES—ALTRUISM AND SELFISHNESS

To demonstrate how the failure of a phenotypic approach can affect the conclusions drawn from a particular model, we consider two examples. The first is a model of altruism in which we have an inclusive fitness ESS which is not a genetic ESS. The second is a model of selfishness in which we have a genetic ESS which is not an inclusive fitness ESS.

Suppose that x represents the level of altruism exhibited by an individual and the payoff to an actor with level x when interacting with a recipient with level y is $F(x, y) = -cx + b(y) + dxy$. Here c and d are positive constants and $b(y)$ is a function of y. Therefore, the cost of altruism to the actor is linear whereas the benefit to the recipient is not. In particular, we suppose that $b(y)$ exhibits diminishing returns. The third term represents a synergistic effect (Queller, 1985); when both individuals are altruistic there are additional benefits.

For example, consider the particular model

$$F(x, y) = -x + y(2 - y) + 2xy. \tag{24}$$

The equilibrium value of x in this model is easily calculated using either a phenotypic or genetic approach and is $\hat{x} = (1 - 2\bar{r})/2$. To examine the local ESS condition for the inclusive fitness approach [i.e. (18b)] we calculate $F_{11} + \bar{r}F_{22} = -2\bar{r}$ which is negative; therefore, $\hat{x}$ is an ESS of the inclusive fitness approach. To examine the local ESS condition for the genetic model [i.e. (15b)] we calculate $F_{11} + 2\bar{r}F_{12} + \bar{r}F_{22} = 2(2 - \bar{r})$ which is positive; therefore $\hat{x}$ is not an ESS of the genetic model. As a result, using the inclusive fitness phenotypic approach in this setting would lead to the

conclusion that $\hat{x}$ is an ESS when in fact it is not.

The easiest way to obtain an example in which the reverse holds is simply to change the sign of F. In particular, now suppose that x represents the level of selfishness of an individual, and take the payoff function to be

$$F(x, y) = x - y(2 - y) - 2xy. \qquad (25)$$

Here the third term represents a synergistic effect whereby when both individuals are selfish, there are additional costs. The equilibrium for this model is the same as the previous model, $\hat{x} = (1 - 2\bar{r})/2$, and it is easily checked that this equilibrium is now a genetic ESS but not an ESS of the inclusive fitness phenotypic approach. Therefore, the inclusive fitness phenotypic approach would lead to the conclusion that $\hat{x}$ is not an ESS when in fact it is.

4.2. ARE PHENOTYPIC APPROACHES USEFUL?

For the simple example above, of pairwise interactions under haploidy, it is probably easy enough to write down the correct definition of fitness directly from a genetic argument (e.g. Grafen, 1979), and therefore, one would never consider using either of the two phenotypic approaches. When the model becomes more complicated however, this is no longer true. For instance, it is probably not immediately obvious how to write down the correct fitness function for this simple example of pairwise interactions under diploidy. Also, in models where individuals play the field (e.g. sex ratio, dispersal; Appendix E), the exact genetic model is sometimes impossible to analyse without resort to numerical procedures and therefore the phenotypic approaches become invaluable (e.g. Bulmer, 1986; Frank, 1986; Taylor, 1988).

The strength of the phenotypic approaches are their simplicity. Neither approach requires any consideration of genetics and therefore their analysis is usually easier. Additionally, dispensing with the genetic details often makes the underlying biological factors at play more transparent. Since both phenotypic approaches obtain the correct equilibrium conditions, both are useful for characterizing the values of x that are candidates for an ESS. Additionally, both allow us to correctly analyse the convergence stability of all equilibria. This is true for other forms of interaction in addition to pairwise interactions (Appendix E) and applies regardless of whether the genetic system under study is haploid or diploid. The personal fitness approach has the additional feature that the fitness function evaluated in a monomorphic population takes on the correct value. In other words, expressions (10) and (19) both equal $F(x^0, x^0)$ when they are evaluated in a monomorphic population, whereas expression (16) from the inclusive fitness approach does not. This is particularly useful in models of dynamic characters where often one is required to know the value of this function in a monomorphic population (Day & Taylor, 1997).

The main shortcoming of the phenotypic approaches, however, is that they do not obtain the correct global or local ESS conditions. It is possible to characterize the relationship between the two phenotypic approaches and the genetic approach for various special cases of payoff function F. This allows a determination of when each phenotypic approach is completely valid (Table 1). Unfortunately, however, it is not possible to make any general statements, and if the form of the interaction among individuals is not pairwise (e.g. patch-structured interactions and playing the field; Appendix E), another tabulation of various special cases is required.

4.3. THE GRAFEN–HINES–MAYNARD SMITH RESULT FOR MATRIX GAMES

Our results also tie in with previous results from matrix games between relatives with discrete characters (Grafen, 1979). Grafen (1979) demonstrated that, in this setting, the inclusive

TABLE 1

The correspondence between the ESS condition from inclusive fitness, personal fitness, and genetic models of kin selection with pairwise interactions

A.	(1) $F_{12} = 0$	Genetic ESS $\Leftrightarrow$ Inclusive Fitness ESS
	(2) $F_{12} > 0$	Genetic ESS $\Rightarrow$ Inclusive Fitness ESS
	(3) $F_{12} < 0$	Inclusive Fitness ESS $\Rightarrow$ Genetic ESS
B.	(1) $F_{22} = 0$	Genetic ESS $\Leftrightarrow$ Personal Fitness ESS
	(2) $F_{22} > 0$	Genetic ESS $\Rightarrow$ Personal Fitness ESS
	(3) $F_{22} < 0$	Personal Fitness ESS $\Rightarrow$ Genetic ESS

fitness approach does not predict the same ESSs as a "correct" approach [note that Hines & Maynard Smith (1979) termed Grafen's correct approach, a personal fitness approach although it is really a haploid genetic approach. It is true however, that genetic models treat fitness in a personal rather than inclusive manner]. It was shown, however, that in this setting a genetic ESS (i.e. Grafen's method) is always an inclusive fitness ESS but the reverse is not true (Hines & Maynard Smith, 1979). These previous results can be related to the present framework as follows. First, in matrix games the fitness function has the special form, $\mathbf{x}^T\mathbf{A}\mathbf{y}$ where $\mathbf{x}$ and $\mathbf{y}$ are vectors specifying the frequency with which each of the possible discrete behaviours is used (Bulmer, 1994). The case of two discrete behaviours and mixed ESSs corresponds to our results. In particular, if we define $\mathbf{x}$ and $\mathbf{y}$ as $\mathbf{x}^T = [x \quad 1-x]$ and $\mathbf{y}^T = [y \quad 1-y]$, then the payoff function is

$$F(x, y) = \mathbf{x}^T\mathbf{A}\mathbf{y} \tag{26}$$

where $\mathbf{A}$ is the so-called 2×2 payoff matrix.

Now for a mixed ESS, we calculate $F_{11} = F_{22} = 0$ and $F_{12} > 0$. From Table 1 this corresponds to A2 and B1. Therefore, all genetic ESSs are inclusive fitness ESSs but not vice versa (from A2). Additionally, the personal fitness approach and the genetic approach correspond exactly (from B1).

Therefore, the results of Grafen (1979) and Hines & Maynard Smith (1979) for mixed ESSs are special cases within the setting of continuous trait game theory. Not all game theoretic models of interest will be matrix games, and not all interactions of interest will be pairwise. This is especially true for continuous traits, and therefore, the only general approach to obtaining the correct predictions is to define the fitness function as done in the genetic approach. In fact, the genetic approach presented here is a generalization of Grafen's (1979) approach that applies for any fitness function of continuous traits. Notice as well that, while the results for matrix games were formulated in terms of haploid, asexual genetics, the general results presented here apply equally well to diploid sexual populations.

5. Adaptive Topography and Population Mean Inclusive Fitness Effect

Now we look at the adaptive topography results for haploid, discrete genetic models. For all our results, a similar analysis can be conducted for diploid genetics unless otherwise stated. Allele frequency dynamics are described by eqn (8). Wright's (1969) notion of an adaptive topography is obtained by first expressing (8) as

$$\frac{\Delta\bar{g}}{\bar{g}} = \frac{(1-\bar{g})}{\bar{W}} \frac{\mathrm{d}\Omega(\bar{g})}{\mathrm{d}\bar{g}} \tag{27}$$

where the function $\Omega(\bar{g})$ is defined by

$$\Omega(\bar{g}) = \int \frac{(W^m - \bar{W})}{(1-\bar{g})} \mathrm{d}\bar{g} \tag{28}$$

with a condition $\Omega(0) = \Omega_0$. Equation (27) reveals that $\Delta\bar{g}$ and $\mathrm{d}\Omega/\mathrm{d}\bar{g}$ always have the same sign. This suggests that, as allele frequencies change, the function $\Omega(\bar{g})$ increases because the change in Ω in one generation is approximately $(\mathrm{d}\Omega/\mathrm{d}\bar{g})\Delta\bar{g}$ which is always positive. Thus, as allele frequencies change, the population climbs to higher values of $\Omega(\bar{g})$.

For this adaptive topography to be of use however, the equation defining $\Omega(\bar{g})$ must be given a biological interpretation. For classical, constant selection models it turns out to be population mean fitness (Wright, 1969). For certain models of kin selection on discrete traits, it has been shown that $\Omega(\bar{g})$ can be interpreted as population mean inclusive fitness effect (Hamilton, 1964; Michod & Abugov, 1980; Gayley, 1993 and references therein). The question of interest here is, When can this be done for models of continuous traits and frequency-dependent selection?

To answer this question we require the limiting case of the above equations when one allele is rare. In this case (8) becomes eqn (9) and eqn (27) becomes,

$$\frac{\Delta\bar{g}}{\bar{g}} = \frac{1}{W^a(x^0)} \frac{\mathrm{d}\Omega(\bar{g})}{\mathrm{d}\bar{g}}. \tag{29}$$

A comparison of (29) with (9) reveals that, under rarity, we can define

$$\Omega(\bar{g}) = (W^A(x, x^0) - W^a(x^0))\bar{g} \tag{30}$$

where we have taken the constant of integration to be zero. Of course as allele frequencies change, again Ω increases, but now this result is valid only when the mutant is rare.

When is $\Omega(\bar{g})$ the population mean inclusive fitness effect? For the case of pairwise interactions, putting eqn (13) into (30) gives

$$\Omega(\bar{g}) = (F(x, x^0) - F(x^0, x^0) + \bar{r}(F(x, x) - F(x, x^0)))\bar{g}. \quad (31)$$

Now, the population mean inclusive fitness effect is

$$[F(x, x^0) - F(x^0, x^0) + \bar{r}(F(x^0, x) - F(x^0, x^0))]\bar{g} \quad (32)$$

since the inclusive fitness effect of a mutant is the effect of altering an individual's strategy from x^0 to x and inclusive fitness effect of the resident allele is zero by definition. Thus when the difference between (32) and (31) is zero, population mean inclusive fitness effect will provide an adaptive topography. This difference is zero whenever

$$F(x, x^0) + F(x^0, x) = F(x, x) + F(x^0, x^0) \quad (33)$$

for all x^0 and x. It is demonstrated in Appendix D that the payoff function $F(x, y)$ satisfies condition (33) if and only if it is the sum of a function that is skew-symmetric plus a function of x and a function of y [S is skew-symmetric if $S(x, y) = -S(y, x)$ for all x, y]. That is, $F(x, y)$ must have the form

$$F(x, y) = S(x, y) + A(x) + B(y). \quad (34)$$

This has an interesting biological interpretation. If the payoff function is skew-symmetric, then the two individuals are playing a constant-sum game; a benefit to one player is a cost to the other (Weibull, 1995). Put another way, the interaction involves a net one-way transfer of some quantity. If the payoff function consists of a function of x plus a function of y, then the effect of one player altering its strategy does not depend on the other player's strategy. This corresponds to the familiar case of additive costs and benefits in an interaction. Therefore, population mean inclusive fitness effect provides an adaptive topography for quantitative trait models of kin selection if and only if the payoff function is a sum of these two types of interaction. This result is valid for haploid organisms; however, it seems that population mean inclusive fitness effect provides an adaptive topography for diploid organisms only when the payoff function is of the form $F(x, y) = ax + by$, which is a special case of (34). Interestingly, it also provides an adaptive topography for any payoff function, $F(x, y)$, when Δx is small because both (31) and (32) become

$$(F_1 + \bar{r}F_2)\Delta x\bar{g} \quad (35)$$

which is the population mean inclusive fitness effect of a small change in strategy when the mutant allele is rare. This is true for both haploid and diploid organisms.

6. Summary

The framework presented here helps to illuminate the interrelationship between single locus genetic and phenotypic models of kin selection on continuous traits. In particular, we have used this framework to address two broad issues. First, When are phenotypic game theory models of kin selection on continuous traits genetically valid?, and second, How do the adaptive topography results from single locus genetics relate to the game theoretic stability notions of these phenotypic models? The answer to these questions generalize and tie together previously disparate results and demonstrate how "games between relatives" must be modeled.

Although phenotypic models often lend more insight to a problem than the corresponding genetic model, they are only useful to the extent that they do not provide erroneous conclusions. It is well documented that phenotypic models of kin selection can fail when selection is strong; however, even when selection is weak, phenotypic models can still fail when fitness is frequency-dependent. In particular, we have shown that, in general, phenotypic models fail to obtain the correct ESS stability condition. Previous results for matrix games between relatives have shown that all genetic ESSs are also ESSs of the corresponding inclusive fitness phenotypic model, but not vice versa. Our results generalize these findings by demonstrating how

they can be treated as a special case within a broader setting. This reveals that, in general, no simple correspondence between genetic and phenotypic ESSs exists. Although it is possible to tabulate the conditions required to have correspondence between various models (Table 1), different tabulations are required for different forms of frequency-dependent interaction. As a result, the only sure method for predicting the correct ESS is to construct a genetic model. The framework outlined here also allows an easy extension to diploidy (Appendices C and E). This has shown that the same conclusions hold under both haploid and diploid genetics and it provides a generalization of Grafen's (1979) approach for obtaining the correct ESS stability condition to models of continuous traits in diploid organisms.

Lastly, the framework presented here allows a better understanding of the correspondence between previous results for single locus genetic models of discrete characters (in particular, the adaptive topography results) and the game theoretic approach of continuous trait kin selection. An important conclusion is that population mean inclusive fitness effect provides an adaptive topography in continuous trait kin selection models only if the payoff function $F(x, y)$ has a certain form. Additionally, population mean inclusive fitness effect provides an adaptive topography for any payoff function $F(x, y)$ under either haploid or diploid genetics, provided that change in behaviour exhibited by the actor is small.

We thank Sally Otto for a critical review of the manuscript. This research was funded by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Abrams, P. A., Harada, Y. & Matsuda, H. (1993). On the relationship between quantitative genetic and ess models. *Evolution* **47,** 982–985.

Bulmer, M. G. (1986). Sex ratio theory in geographically structured populations. *Heredity* **56,** 69–73.

Bulmer, M. (1994). *Theoretical Evolutionary Ecology*. Sunderland, MA: Sinauer Associates.

Charlesworth, B. (1980). Models of kin selection. In: *Evolution of Social Behaviour*: *Hypotheses and Empirical Tests* (Mark, H., ed.) pp. 11–26. Dahlem Konferenzen.

Charnov, E. L. (1977). An elementary treatment of the genetical theory of kin-selection. *J. theor. Biol.* **66,** 541–550.

Christiansen, F. B. (1991). On conditions for evolutionary stability for a continuously varying character. *Am. Nat.* **138,** 37–50.

Crow, J. F. & Kimura M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.

Day, T. & Taylor, P. D. (1997). Hamilton's rule meets the Hamiltonian—kin selection on dynamic characters. *Proc. R. Soc. London B* **264,** 639–644.

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

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. & Motro U. (1981). Kin selection and strong evolutionary stability of mutual help. *Theor. Pop. Biol.* **19,** 420–433.

Eshel, I., Motro, U. & Sansone, E. (1997). Continuous stability and evolutionary convergence. *J. theor. Biol.* **185,** 333–343.

Frank, S. A. (1986). Dispersal polymorphisms in subdivided populations. *J. theor. Biol.* **122,** 303–309.

Gayley, T. (1993). Genetics of kin selection: the role of behavioral inclusive fitness. *Am. Nat.* **141,** 928–953.

Grafen, A. (1979). The hawk–dove game played between relatives. *Anim. Behav.* **27,** 905–907.

Grafen, A. (1982). How not to measure inclusive fitness. *Nature* **298,** 425–426.

Grafen, A. (1985). A geometric view of relatedness. *Oxford Surveys Evol. Biol.* **2,** 28–89.

Hamilton, W. D. (1964). The genetical evolution of social behaviour, I. *J. theor. Biol.* **7,** 1–16.

Hammerstein, P. (1996). Darwinian adaptation, population genetics and the streetcar theory of evolution. *J. Math. Biol.* **34,** 511–532.

Hines, W. G. S. & Maynard Smith, J. (1979). Games between relatives. *J. theor. Biol.* **79,** 19–30.

Iwasa, Y., Pomiankowski, A. & Nee, S. (1991). The evolution of costly mate preferences. ii. The "handicap" principle. *Evolution* **45,** 1431–1442.

Kimura, M. (1965). A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proc. Nat. Acad. Sci.* **54,** 731–736.

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

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

Maynard Smith, J. & Price, G. R. (1973). The logic of animal conflict. *Nature* **246,** 15–18.

Michod, R. E. (1982). The theory of kin selection. *Annu. Rev. Ecol. System.* **13,** 23–55.

Michod, R. E. & Abugov, R. (1980). Adaptive topography in family-structured models of kin selection. *Science* **210,** 667–669.

Michod, R. E. & Hamilton, W. D. (1980). Coefficients of relatedness in sociobiology. *Nature* **288,** 694–697.

Mirmirani, M. & Oster, G. (1978). Competition, kin selection, and evolutionary stable strategies. *Theor. Pop. Biol.* **13,** 304–339.

Orlove, M. (1979). Putting the diluting effect into inclusive fitness. *J. theor. Biol.* **78,** 449–450.

Oster, G., Eshel, I. & Cohen, D. (1977). Worker-queen conflict and the evolution of social insects. *Theor. Pop. Biol.* **12,** 49–85.

QUELLER, D. C. (1984). Kin selection and frequency dependence: a game theoretic approach. *Biol. J. Linn. Soc.* **23,** 133–143.
QUELLER, D. C. (1985). Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* **318,** 366–367.
SEGER, J. (1981). Kinship and covariance. *J. theor. Biol.* **91,** 191–213.
TAYLOR, P. D. (1988). Inclusive fitness models with two sexes. *Theor. Pop. Biol.* **34,** 145–168.
TAYLOR, P. D. (1989). Evolutionary stability in one-parameter models under weak selection. *Theor. Pop. Biol.* **36,** 125–143.
TAYLOR, P. D. (1996a). Inclusive fitness arguments in genetic models of behaviour. *J. Math. Biol.* **34,** 654–674.
TAYLOR, P. D. (1996b). The selection differential in quantitative genetics and ess models. *Evolution* **50,** 2106–2110.
TAYLOR, P. D. & DAY, T. (1997). Evolutionary stability under the replicator and the gradient dynamics. *Evol. Ecol.* **11,** 579–590.
TAYLOR, P. D. & FRANK, S. A. (1996). How to make a kin selection model. *J. theor. Biol.* **180,** 27–37.
VINCENT, T. L. & BROWN, J. S. (1988). The evolution of ess theory. *Annu. Rev. Ecol. System.* **19,** 423–443.
WEIBULL, J. W. (1995). *Evolutionary Game Theory*. Cambridge, MA: MIT Press.
WRIGHT, S. (1969). *Evolution and the Genetics of Populations, Vol.* 2, *The Theory of Gene Frequencies*. Chicago, IL: University of Chicago Press.

APPENDIX A

Notation

General notation

x^0: the population-wide strategy
x; $x = x^0 + \Delta x$: the mutant strategy; Δx is the mutant's deviation from "normal"
x^*: an evolutionarily stable strategy value
$W(x, x^0)$: the fitness of a mutant using x in a population using x^0
$F(a, b)$: the payoff to an individual using a when interacting with a recipient using b
$\bar{r}$: the expected relatedness of the recipient to the actor
$y(x, x^0)$: the recipients phenotype as a function of x and x^0; for the PF approach

Haploid model

A, a: the two alleles; "A" is the mutant and "a" is the resident
W^A, W^a: the fitness of mutant and resident alleles when the mutants are rare
W^m: the fitness of the mutant allele
$\bar{W}$: the population mean fitness
$\bar{g}$: the frequency of the mutant allele
ρ: the probability that the recipient is a mutant
g_x: the genotype of the actor
g_y: the genotype of the recipient
$\Omega(\bar{g})$: the adaptive topography function

Diploid model

W^{AA}, W^{Aa}, W^{aa}: the three genotypic fitness functions when the mutant allele is rare
f: Wright's inbreeding coefficient
P_k: the probability that a random mutant homozygous actor interacts with a recipient having k copies of the mutant allele. $k = 0$, 1 or 2
p_k: the probability that a random mutant heterozygous actor interacts with a recipient having k copies of the mutant allele. $k = 0$, 1 or 2

APPENDIX B

Haploid Model

Here we show that

$$\bar{r} = \rho \tag{B.1}$$

is the correct definition of relatedness when the mutant allele is rare. We start with the definition

$$\text{Relatedness} = \frac{\text{cov}(g_x, g_y)}{\text{var}(g_x)} \tag{B.2}$$

where g_x is the genotype of the actor and g_y is the genotype of the recipient (Michod & Hamilton, 1980). The covariance is

$$\begin{aligned}
\text{cov}(g_x, g_y) &= \text{E}[g_x g_y] - \text{E}[g_x]\text{E}[g_y] \\
&= \text{E}[g_x g_y | g_x = 0]\Pr\{g_x = 0\} \\
&\quad + \text{E}[g_x g_y | g_x = 1]\Pr\{g_x = 1\} - \bar{g}^2 \\
&= \bar{r}\Pr\{g_x = 1\} - \bar{g}^2 \\
&= \bar{r}\bar{g} - \bar{g}^2 \\
&= \bar{g}[\bar{r} - \bar{g}]
\end{aligned} \tag{B.3}$$

and the variance is

$$\mathrm{var}(g_x) = \mathrm{E}[g_x^2] - \mathrm{E}[g_x]^2 = \bar{g} - \bar{g}^2 = \bar{g}[1 - \bar{g}]. \tag{B.4}$$

Therefore we have

$$\text{Relatedness} = \frac{\bar{r} - \bar{g}}{1 - \bar{g}} \tag{B.5}$$

and for rare mutants,

$$\lim_{\bar{g} \to 0} \text{Relatedness} = \bar{r}. \tag{B.6}$$

For rare mutants, relatedness is simply the ratio of the expected value of the recipient's genotype to the expected value of the actor's genotype *from the perspective of a mutant allele*.

APPENDIX C

Diploid Model

In a diploid model of kin selection we need to allow for the possibility of inbreeding and therefore, the three genotypes occur with frequencies

$$\begin{aligned}
&\text{AA } (1-f)\bar{g}^2 + \mathrm{f}\bar{g} \\
&\text{Aa2 } (1-f)\bar{g}(1-\bar{g}) \\
&\text{aa } (1-f)(1-\bar{g})^2 + f(1-\bar{g})
\end{aligned} \tag{C.1}$$

(Crow & Kimura, 1970). The equation for allele frequency change is

$$\frac{\Delta\bar{g}}{\bar{g}} = \frac{1}{\bar{W}}(W^m - \bar{W}) \tag{C.2}$$

where W^m is now the marginal fitness of the mutant allele (Crow & Kimura, 1970). We assume that genetic effects are additive and taking the limit as $\bar{g} \to 0$, eqn (C.2) becomes

$$\begin{aligned}
\frac{\Delta\bar{g}}{\bar{g}} = \frac{1}{W^{aa}(x^0)}&(fW^{AA}(x, x^0) \\
&+ (1-f)W^{Aa}(x, x^0) - W^{aa}(x^0))
\end{aligned} \tag{C.3}$$

where the functions W^{AA}, W^{Aa}, and W^{aa} are the genotype fitnesses. As done for the haploid genetic model, we can now see that to obtain a genetically valid game theoretic model under diploidy, it suffices to define the fitness function, $W(x, x^0)$, as

$$W(x, x^0) \triangleq fW^{AA}(x, x^0) + (1-f)W^{Aa}(x, x^0). \tag{C.4}$$

We can write the genotypic fitnesses as

$$W^{AA}(x, x^0) = \sum_{k=0}^{2} P_k F\left(x, \frac{kx + (2-k)x^0}{2}\right) \tag{C.5a}$$

$$W^{Aa}(x, x^0) = \sum_{k=0}^{2} p_k F\left(\frac{x + x^0}{2}, \frac{kx + (2-k)x^0}{2}\right) \tag{C.5b}$$

where P_k and p_k are defined in Appendix A. Using these in the fitness function (C.4) and assuming that the altered behaviour of a mutant does not alter the probabilities P_k and p_k, the equilibrium condition (2) is calculated as

$$\begin{aligned}
W_1(x, x^0)|_{x = x^0 = x^*} &= f\sum_{k=0}^{2} P_k\left(F_1 + \frac{k}{2}F_2\right) \\
&\quad + (1-f)\sum_{k=0}^{2} p_k\left(\frac{1}{2}F_1 + \frac{k}{2}F_2\right)
\end{aligned} \tag{C.6a}$$

$$\begin{aligned}
&= \left(\frac{1+f}{2}\right)F_1 + \left[f\left(\sum_{k=0}^{2} P_k\frac{k}{2}\right)\right. \\
&\quad \left. + (1-f)\left(\sum_{k=0}^{2} p_k\frac{k}{2}\right)\right]F_2
\end{aligned} \tag{C.6b}$$

which has the same sign as

$$F_1 + \bar{r}F_2 \tag{C.7}$$

where

$$\bar{r} = \frac{f\left(\Sigma_k P_k\frac{k}{2}\right) + (1-f)\left(\Sigma_k p_k\frac{k}{2}\right)}{(1+f)/2} \tag{C.8}$$

is the relatedness of the recipient to the actor (see overleaf).

The ESS condition (3) is calculated as

$$W_{11}(x, x^0)|_{x=x^0=x^*} = f \sum_{k=0}^{2} P_k\left(F_{11} + 2\frac{k}{2}F_{12} + \left(\frac{k}{2}\right)^2 F_{22}\right) + (1-f)\sum_{k=0}^{2} p_k\left(\frac{1}{4}F_{11} + \frac{k}{2}F_{12} + \left(\frac{k}{2}\right)^2 F_{22}\right) \tag{C.9a}$$

$$= \left(\frac{1+3f}{4}\right)F_{11} + \left(2f\sum_{k=0}^{2} P_k\frac{k}{2} + (1-f)\sum_{k=0}^{2} p_k\frac{k}{2}\right)F_{12} + \left(f\sum_{k=0}^{2} P_k\left(\frac{k}{2}\right)^2 + (1-f)\sum_{k=0}^{2} p_k\left(\frac{k}{2}\right)^2\right)F_{22}. \tag{C.9b}$$

The CSS condition (7) is calculated as

$$\frac{\mathrm{d}}{\mathrm{d}x^0}[W_1(x, x^0)|_{x=x^0}]\Big|_{x^0=x^*} = \left(\frac{1+f}{2}\right)(F_{11}+F_{12}) + \left[f\left(\sum_{k=0}^{2} P_k\frac{k}{2}\right) + (1-f)\left(\sum_{k=0}^{2} p_k\frac{k}{2}\right)\right](F_{12}+F_{22})$$

which has the same sign as

$$F_{11} + (1+\bar{r})F_{12} + \bar{r}F_{22} \tag{C.10}$$

where $\bar{r}$ is defined in (C.8).

To see that (C.8) is the correct definition of relatedness, we proceed as in the haploid case. Under additive gene action relatedness is defined as

$$\text{Relatedness} = \frac{\mathrm{cov}(g_x, g_y)}{\mathrm{var}(g_x)} \tag{C.11}$$

where g_x is the genotype of the actor and g_y is the genotype of the recipient (Michod & Hamilton, 1980). Using the shorthand

$$\hat{P} = \sum_{k=0}^{2} P_k\frac{k}{2} \tag{C.12a}$$

$$\hat{p} = \sum_{k=0}^{2} p_k\frac{k}{2} \tag{C.12b}$$

the covariance is

$$\begin{aligned}\mathrm{cov}(g_x, g_y) &= \mathrm{E}[g_x g_y] - \mathrm{E}[g_x]\mathrm{E}[g_y] \\ &= \mathrm{E}[g_x g_y|g_x = 1/2]\Pr\{g_x = 1/2\} \\ &\quad + \mathrm{E}[g_x g_y|g_x = 1]\Pr\{g_x = 1\} - \bar{g}^2 \\ &= \frac{1}{2}\hat{p}\Pr\{g_x = 1/2\} + \hat{P}\Pr\{g_x = 1\} - \bar{g}^2 \\ &= \hat{p}(1-f)\bar{g}(1-\bar{g}) + \hat{P}((1-f)\bar{g}^2 + \bar{g}f) - \bar{g}^2 \\ &= \bar{g}[\hat{p}(1-f)(1-\bar{g}) + \hat{P}((1-f)\bar{g} + f) - \bar{g}]\end{aligned} \tag{C.13}$$

and the variance is

$$\begin{aligned}\mathrm{var}(g_x) &= \mathrm{E}[g_x^2] - \mathrm{E}[g_x]^2 \\ &= \frac{1}{4}2\bar{g}(1-\bar{g})(1-f) + (1-f)\bar{g}^2 + \bar{g}f - \bar{g}^2 \\ &= \bar{g}\left[\frac{1}{2}(1-\bar{g})(1-f) + f + (1-f)\bar{g} - \bar{g}\right].\end{aligned} \tag{C.14}$$

Therefore we have

$$\text{Relatedness} = \frac{\hat{p}(1-f)(1-\bar{g}) + \hat{P}(1-f)\bar{g} + f) - \bar{g}}{\frac{1}{2}(1-\bar{g})(1-f) + f + (1-f)\bar{g} - \bar{g}} \tag{C.15}$$

and for rare mutants,

$$\lim_{\bar{g}\to 0}\text{Relatedness} = \frac{\hat{p}(1-f) + \hat{P}f}{\frac{1}{2}(1+f)} \tag{C.16}$$

which is eqn (C.8). Again, for rare mutants, relatedness is the ratio of the expected value of the recipient's genotype to the expected value of the actor's genotype *from the perspective of a mutant allele*.

APPENDIX D

The Form of $F(x, y)$ Satisfying Condition (33)

Here we demonstrate that the payoff function $F(x, y)$ satisfies

$$F(a, b) + F(b, a) = F(a, a) + F(b, b) \tag{D.1}$$

for all a, b [i.e. condition (33) of the text] if and only if it is of the form

$$F(a, b) = S(a, b) + A(a) + B(b) \quad \text{(D.2)}$$

where $S(a, b)$ is skew-symmetric [i.e. $S(a, b) = -S(b, a)$ for all a, b]. If $F(x, y)$ satisfies (D.1) then

$$2F(a, b) = (F(a, b) - F(b, a)) + (F(a, b) + F(b, a)). \quad \text{(D.3)}$$

The first set of terms in (D.3) is a function that is skew-symmetric and the second set of terms is a function of a plus a function of b from (D.1). Therefore, if $F(x, y)$ satisfies (D.1), then it has the form (D.2).

Now, if $F(x, y)$ has the form (D.2), then $F(b, a) = S(b, a) + A(b) + B(a)$, and adding this to (D.2) gives

$$F(a, b) + F(b, a) = A(a) + B(b) + A(b) + B(a) \quad \text{(D.4)}$$

since S is skew-symmetric. Now, from (D.2) we can also see that

$$F(a, a) = A(a) + B(a) \quad \text{(D.5)}$$

and

$$F(b, b) = A(b) + B(b) \quad \text{(D.6)}$$

and therefore

$$F(a, a) + F(b, b) = A(a) + B(b) + A(b) + B(a). \quad \text{(D.7)}$$

Substituting (D.7) into (D.4) then shows that if $F(x, y)$ has the form (D.2), then it satisfies (D.1).

APPENDIX E

Results for Path-structured Populations—Playing the Field

Here we present the analogous results for the genetic model and the two phenotypic models assuming a patch-structured population in which individuals play the field. Specifically, we assume that the payoff to an individual using strategy x in a patch where the average strategy is z is given by the function, $G(x, z)$.

THE GENETIC MODEL

Haploidy

Let n denote the number of individuals per patch, and ρ_k the probability that the patch has a *total* of k copies of the mutant allele where k runs from 1 through n. Therefore we have

$$W^A(x, x^0) = \sum_k \rho_k G\left(x, \frac{k}{n}x + \frac{n-k}{n}x^0\right). \quad \text{(E.1)}$$

Using (E.1) in definition (10) we can calculate the equilibrium, ESS, and CSS conditions. Equilibrium condition (2) is

$$W_1(x, x^0)|_{x = x^0 = x^*} = \sum_k \rho_k \left(G_1 + \frac{k}{n}F_2\right) \quad \text{(E.2a)}$$

$$= G_1 + \bar{r}G_2 \quad \text{(E.2b)}$$

where we have defined $r = k/n$ and

$$\bar{r} = \sum_k \rho_k \frac{k}{n} \quad \text{(E.3)}$$

is the relatedness of two randomly chosen patch members with replacement. That this is the correct definition of relatedness when the mutant allele is rare can be demonstrated using calculations similar to those in Appendix B. ESS condition (3) is

$$W_{11}(x, x^0)|_{x = x^0 = x^*} = \sum_k \rho_k \left(G_{11} + 2\frac{k}{n}G_{12} + \left(\frac{k}{n}\right)^2 G_{22}\right) \quad \text{(E.4a)}$$

$$= G_{11} + 2\bar{r}G_{12} + \bar{r}^2 G_{22}, \quad \text{(E.4b)}$$

and CSS condition (7) is

$$\frac{\mathrm{d}}{\mathrm{d}x^0}\left[W_1(x, x^0)|_{x = x^0}\right]\Big|_{x^0 = x^*} = G_{11} + (1 + \bar{r})G_{12} + \bar{r}G_{22}. \quad \text{(E.5)}$$

Again, these results all assume that the altered behaviour of the mutant does not affect the probability distribution ρ_k and consequently does not affect the relatedness, $\bar{r}$.

Diploidy

Under diploidy we have

$$W^{AA}(x, x^0) = \sum_k P_k G\left(x, \frac{k}{2}x + \frac{2n-k}{2n}x^0\right) \quad \text{(E.6a)}$$

$$W^{Aa}(x, x^0) = \sum_k p_k G\left(\frac{x+x^0}{2}, \frac{k}{2n}x + \frac{2n-k}{2n}x^0\right) \quad \text{(E.6b)}$$

where P_k and p_k are the probabilities that a random mutant homozygote and heterozygote, respectively, is in a patch with a total of k mutant alleles where k runs from 1 to $2n$. Using these in the fitness function (C.4), and assuming that the altered behaviour of a mutant does not alter the probability distributions P_k and p_k, the equilibrium condition (2) is calculated as

$$W_1(x, x^0)|_{x=x^0=x^*} = f\sum_k P_k\left(G_1 + \frac{k}{2n}G_2\right) + (1-f)\sum_k p_k\left(\frac{1}{2}G_1 + \frac{k}{2n}G_2\right) \quad \text{(E.7a)}$$

$$= \left(\frac{1+f}{2}\right)G_1 + \left[f\left(\sum_k P_k\frac{k}{2n}\right) + (1-f)\left(\sum_k p_k\frac{k}{2n}\right)\right]G_2 \quad \text{(E.7b)}$$

which has the same sign as

$$G_1 + \bar{r}G_2 \quad \text{(E.8)}$$

where

$$\bar{r} = \frac{f\left(\Sigma_k P_k\frac{k}{2n}\right) + (1-f)\left(\Sigma_k p_k\frac{k}{2n}\right)}{(1+f)/2} \quad \text{(E.9)}$$

is the relatedness of the recipient to the actor. This can be demonstrated using the approach from Appendix C.

ESS condition (3) is calculated as

$$W_{11}(x, x^0)|_{x=x^0=x^*} = f\sum_k P_k\left(G_{11} + 2\frac{k}{2n}G_{12} + \left(\frac{k}{2n}\right)^2 G_{22}\right) + (1-f)\sum_k p_k\left(\frac{1}{4}G_{11} + \frac{k}{2n}G_{12} + \left(\frac{k}{2n}\right)^2 G_{22}\right) \quad \text{(E.10a)}$$

$$= \left(\frac{1+3f}{4}\right)G_{11} + \left(2f\sum_k P_k\frac{k}{2n} + (1-f)\sum_k p_k\frac{k}{2n}\right)G_{12} + \left(f\sum_k P_k\left(\frac{k}{2n}\right)^2 + (1-f)\sum_k p_k\left(\frac{k}{2n}\right)^2\right)G_{22} \quad \text{(E.10b)}$$

CSS condition (7) is calculated as

$$\frac{\mathrm{d}}{\mathrm{d}x^0}[W_1(x, x^0)|_{x=x^0}]\bigg|_{x^0=x^*} = \left(\frac{1+f}{2}\right)(G_{11}+G_{12}) + \left[f\left(\sum_k P_k\frac{k}{2n}\right) - (1-f)\left(\sum_k p_k\frac{k}{2n}\right)\right](G_{12}+G_{22}) \quad \text{(E.11)}$$

which has the same sign as

$$G_{11} + (1-\bar{r})G_{12} + \bar{r}G_{22} \quad \text{(E.12)}$$

where $\bar{r}$ is defined in (E.9).

THE INCLUSIVE FITNESS APPROACH

The inclusive fitness approach makes the definition

$$W(x, x^0) \triangleq G(x, y) + (n-1)RG(x^0, y), \quad \text{(E.13)}$$

where R is the expected relatedness of a randomly chosen one of the $n-1$ remaining

group members to the actor. Conditions (2), (3), and (7) are

$$\text{EQUIL} : G_1 + \bar{r}G_2 = 0 \tag{E.14a}$$

$$\text{ESS} : G_{11} + \frac{2}{n} G_{12} + \frac{\bar{r}}{n} G_{22} < 0 \tag{E.14b}$$

$$\text{CSS} : G_{11} + (1 - \bar{r})G_{12} + \bar{r}G_{22} < 0 \tag{E.14c}$$

where $\bar{r}$ is now the expected relatedness between two randomly chosen patch mates with replacement.

THE PERSONAL FITNESS APPROACH

The personal fitness approach uses the definition

$$W(x, x^0) \triangleq G(x, z(x, x^0)) \tag{E.15}$$

where we define $z(x, x^0) = \bar{r}x + (1 - \bar{r})x^0$, and again $\bar{r}$ is the expected relatedness between two randomly chosen patch mates with replacement. Conditions (2), (3), and (7) are then

$$\text{EQUIL} : G_1 + \bar{r}G_2 = 0 \tag{E.16}$$

$$\text{ESS} : G_1 + 2\bar{r}G_{12} + \bar{r}^2G_{22} < 0 \tag{E.17}$$

$$\text{CSS} : G_{11} + (1 + \bar{r})G_{12} + \bar{r}G_{22} < 0. \tag{E.18}$$

Again we can see both phenotypic models obtain the correct equilibrium and CSS conditions but they fail to obtain the correct ESS condition.