Joint phenotypes, evolutionary conflict and the fundamental theorem of natural selection

David C. Queller

Department of Biology, Washington University in St Louis, St Louis, MO 63130, USA

Multiple organisms can sometimes affect a common phenotype. For example, the portion of a leaf eaten by an insect is a joint phenotype of the plant and insect and the amount of food obtained by an offspring can be a joint trait with its mother. Here, I describe the evolution of joint phenotypes in quantitative genetic terms. A joint phenotype for multiple species evolves as the sum of additive genetic variances in each species, weighted by the selection on each species. Selective conflict between the interactants occurs when selection takes opposite signs on the joint phenotype. The mean fitness of a population changes not just through its own genetic variance but also through the genetic variance for its fitness that resides in other species, an update of Fisher's fundamental theorem of natural selection. Some similar results, using inclusive fitness, apply to within-species interactions. The models provide a framework for understanding evolutionary conflicts at all levels.

1. Introduction

W. D. Hamilton’s inclusive fitness theory [1,2] has been important for many reasons [3–5]. It gave a way to calculate how selection would operate on social behaviours. It explained puzzling behaviours, such as altruism and spite. In outline, it was simple to understand and easy to apply because it was relatively independent of most genetic details. Perhaps most importantly, inclusive fitness provided a quantity that is maximized by natural selection and by extension identifies an agent—either the individual or the gene—that is adapted to behave as if it were maximizing it.

In this, Hamilton was following Darwin’s lead. Selection produces adaptations that perform as if they have been designed for survival and reproduction. One consequence of having the proper design criterion for social behaviour was a clarification of how conflict operates within a species. Just as a cheetah and gazelle can be selected differently for whether the former catches and eats the latter, so too can individuals within a species be in conflict. Even a mother and offspring may be selected differently, for example with respect to the amount of food the mother provides, when their inclusive fitnesses differ [6]. Though conflict is an important part of behavioural ecology, and of evolutionary biology in general, it has not been formalized to the same degree as selection in the absence of conflict. In this paper, I attempt such a formalization, roughly in the quantitative genetic tradition, by treating selection on a joint phenotype that is created by multiple parties.

Hamilton was also following Ronald Fisher’s lead. Fisher’s fundamental theorem of natural selection states that fitness increases at a rate equal to the additive genetic variance for fitness [7]. As such, it provided a formal foundation for the optimality notion that selection maximizes fitness. Hamilton’s theory can be viewed as an adjustment to this notion, not considered by Fisher, that is required when individuals affect neighbours who share their genes in non-random ways [3,4,8].

Fisher believed his fundamental theorem to be a very important contribution to biology, in some ways parallel to the second law of thermodynamics in physics [7]. But its reception was varied, in some ways parallel to the reception of inclusive fitness theory. Each was viewed by proponents as a spectacular
synthesis, yet each was viewed by others as at best an approximation that fails in many cases. The difference was that, for Fisher’s fundamental theorem, proponents were very scarce for over 40 years. The problem was that fitness does not always increase to the maximum. Dominance, epistasis and frequency-dependent selection often prevent fitness from reaching its highest possible value and can even cause average fitness to decline. Change of environment can do the same thing, particularly the change in the biotic environment. Fisher knew this and was not bothered by it, but he never explained his position clearly enough. The fundamental theorem was regarded as ‘entirely obscure’ [9] ‘recondite’ [10] or ‘very difficult’ [11]; it was suggested that it ‘mostly fails’ [12] and that attempts to save it ‘are quite pointless’ [13].

Fisher’s reasoning was eventually clarified by George Price [14]. In his view, Fisher was not talking about the total change in fitness but rather just the part of it that is owing to natural selection in the previous generation. Fitness might also change owing to changes in the environment but this was not his focus; he could ignore it and still capture the essence of Darwin’s insights about selection and adaptation. Fisher considered that the environment would often deteriorate, often because of competitors and enemies, so that total fitness would not always increase. Changes in the environment could also include dominance and epistasis changing the genetic environment of the next generation, a view that may have seemed odd at the time but which feels comfortable today given the genic view of selection elaborated by Dawkins [15,16]. Thus, Fisher’s result applies to the change in fitness owing to selection, keeping the average effects or breeding values constant.

This revised and highly favourable view of the fundamental theorem seems to be the consensus opinion today [8,14,17–22]. I agree with this view but this paper will attempt a significant revision of the fundamental theorem, by trying to also capture part of the change in environment that has been ignored. This has been done to some degree for changes in the genetic environment but these effects are often small. Effects of changes in the physical environment would probably be hard to capture in a general way. Changes in the biotic environment are different, often being large, and typically (though not always) being deleterious. Moreover, certain changes in the biotic environment, specifically changes owing to natural selection on other parties, are heritable and can be easily captured. Indeed, Fisher’s result was not really about the full change owing to natural selection; it was about the change in a party’s fitness owing to natural selection on that party. This is very important, but I will explicitly incorporate the effects of selection on other parties in order to get closer to the total change in a party’s fitness owing to natural selection and to explicitly model why it often decreases.

I begin by introducing the Price equation that will be the basis of the models. Before coming to within-species interactions and inclusive fitness, I treat the case of between-species interactions, which is simpler in some respects. Hamilton was of course also interested in these, as exemplified by his host–parasite work. In each case, I will consider how selection operates on joint phenotypes that are affected by multiple parties. This will lead to a formalization and definition of selective conflict between the parties. Versions of the fundamental theorem can then be derived simply by considering fitness of one party as a joint phenotype that is also affected by other parties. The chief goal is not to analyse particular cases but to capture some general principles.

2. The Price equation

George Price is best known for his ‘Price equation’ [21,23], partly anticipated by Robertson [24]. It is a mathematical identity describing how selection operates that makes it easy to analyse selection through manipulation of high-level statistical parameters, such as means, variances and covariances. It states that average trait value \( z \) will increase in evolution as

\[
\Delta z = \text{Cov}(w, z) + E(w, \Delta z),
\]

where \( z \) is an individual’s trait value and \( w \) is its relative fitness [21,23]. Hamilton was the first to appreciate the importance of Price’s result and helped him to get it published in Nature.

Oddly, Price did not use his own equation in his exegesis of Fisher’s fundamental theorem, even though, as Steve Frank has shown, one can derive Fisher from Price in two simple steps [21]. The two have very similar structures. The first term of the Price equation captures the effect of selection, just as Fisher’s fundamental theorem did. The second term, the expected change in phenotype from a parent to its offspring \( \Delta z \) weighted by parental fitness, can include effects from dominance and epistasis that alter average effects in the next generation, and this is likely what allowed Price to see what Fisher had left out. I will follow the common practice of assuming the second term of Price’s equation is negligible (or of secondary interest), but with one very large exception: I will explicitly model changes owing to selection on other parties.

3. Joint phenotypes

Organisms often have extended phenotypes [16]. Traits outside of the organism’s conventional body, for example a beaver’s dam, are affected by the organism’s genes. When such traits affect the organism’s fitness, they can evolve under natural selection. Some extended phenotypes do not belong entirely to one organism and can be influenced by the genes of multiple parties [16]. The indirect-genetic-effect (IGE) approach in quantitative genetics uses this insight fruitfully [8,25–29] and my approach is fully in that spirit, but with a small shift in emphasis. Where IGE tends to speak of ‘interacting phenotypes’ that produce some combined result, I will focus on the combined result itself as the joint phenotype. Instead of viewing one party as the owner of the phenotype that happens to be affected by another, I treat both parties symmetrically as joint owners (though they usually contribute unequally). Examples of joint traits include the portion of a leaf eaten by an insect, health of an infected host, whether a peacock and peahen mate, blood flow to an embryo from its mother, and degree of meiotic drive during spermatogenesis. When individual \( i \) of one species interacts with individual \( j \) of another to produce a joint trait \( z \), we can write it as the sum of the two individuals’ breeding values for the joint trait plus an environmental deviation

\[
z = g_i + g_j + e_{ij}.
\]

The transmissible part of the joint trait is the sum of the two breeding values, each of which can be estimated through quantitative genetic methods [30,31].
For concreteness, let the two parties be gazelles (indexed by $i$) and cheetahs (indexed by $j$). They have many phenotypic traits such as sensory acuity, speed and agility that influence the interaction, but I will consider their summed effect on a joint phenotype: whether, when they encounter each other, the gazelle becomes dinner for the cheetah, which can be scored as zero or one for a single interaction.

4. Interactants of different species

We can modify the Price equation to accommodate a joint phenotype affected by two species. Two terms are needed because gazelle genes are passed only through gazelle fitness and cheetah genes only through cheetah fitness. Assuming no environmental change

$$
\Delta z = \Delta \bar{z}_i + \Delta \bar{z}_j = \text{Cov}(w_i, g_i) + \text{Cov}(w_j, g_j).
$$

(4.1)

The $w$s used here are relative fitnesses. If they are interpreted instead as absolute fitnesses, then the terms on the right-hand side need to be divided by $w_i$ and $w_j$ respectively, and these denominators would be carried through in the derivations that follow.

Although equation (4.1) is based on interaction of two parties, it appears to leave no room for actual non-additive interaction between $g_i$ and $g_j$, i.e. a between-individual epistasis. But as with normal epistasis, the non-additive component part of this interaction—the part not captured by breeding values—is relegated to the second term of Price’s equation (2.1) and the uncaptured ‘change of environment’ term of the fundamental theorem.

Equation (4.1) is modified using methods that involve choosing appropriate components or predictors of fitness [32–34]. A set of familiar statistical identities are used: if $x$, $y$ and $z$ are variables and $k$ is a constant, $\text{Cov}(x, x) = \text{Var}(x)$, $\text{Cov}(x,y) = k \text{Cov}(x,y) + \text{Cov}(x,y + z) = \text{Cov}(x,y) + \text{Cov}(x,z)$ and $\beta_x = \text{Cov}(x,z)/\text{Var}(x)$, where $\beta_x$ is a simple regression coefficient of $y$ on $x$ [35].

First, we can model the effects of an interaction, here assumed to be linear, on fitnesses of gazelles and cheetahs as

$$
w_i = a_i + n_i \bar{z}_i b_{i \bar{z}_i}
$$

(4.2)

and

$$
w_j = a_j + n_j \bar{z}_j b_{j \bar{z}_j}.
$$

(4.3)

The $a$s are the fitnesses in the absence of the interaction; $n_i$ and $n_j$ are the numbers of interactions experienced by the $i$th gazelle and $j$th cheetah (i.e. the number of encounters, which may be zero for some individuals), respectively. If the joint phenotype involves one individual of each species, the sums of $n_i$ and $n_j$ would be equal in the two species. $\bar{z}_i$ and $\bar{z}_j$ are the mean joint phenotypes experienced by gazelle $i$ and cheetah $j$, across all their interactions, respectively. $b_{i \bar{z}_i}$ and $b_{j \bar{z}_j}$ represent the expected fitness change to the gazelle and cheetah, respectively, in a single interaction, per unit change in the joint phenotype $z$. The $b$s are considered to be constants and they can be estimated by regression.

As noted earlier, this is essentially an IGE approach [8,25–29], extended to multiple species, but with some minor differences from conventional usage. Instead of treating each party’s individual traits, such as speed and agility, as the phenotypes, I use the joint phenotype caused by their interaction. As such, I do not view one party as the owner of the trait but instead treat both parties’ effects on the trait symmetrically. Each simply makes its contribution to the trait, with the breeding value representing the heritable component, with neither necessarily considered less direct than the other. It seems probable that similar results could be obtained with the standard IGE model, and this might lead to more insights on the individual traits that lead to the joint phenotype, but my goal is to highlight conflict over joint phenotypes.

Substituting equations (4.2) and (4.3) into (4.1) and assuming that the baseline fitnesses ($a$) are uncorrelated with breeding values $g_i$ and $g_j$ yields

$$
\Delta z = \text{Cov}(g_i, n_i \bar{z}_i b_{i \bar{z}_i}) + \text{Cov}(g_j, n_j \bar{z}_j b_{j \bar{z}_j}).
$$

(4.4)

We can extract the constant $b$s from the covariances and also the means of the $n$s, provided that these are independent of the breeding values ($g$) in their terms. This means that the genes determining the outcome of an interaction are independent of the number of interactions experienced, as when the number of interactions is determined by environmental factors

$$
\Delta z = n_i b_{i \bar{z}_i} \text{Cov}(g_i, \bar{z}_i) + n_j b_{j \bar{z}_j} \text{Cov}(g_j, \bar{z}_j).
$$

(4.5)

The mean phenotypes experienced by gazelle $i$ and cheetah $j$ can be written as

$$
\bar{z}_i = g_i + \bar{g}_j + e
$$

(4.6)

and

$$
\bar{z}_j = g_j + \bar{g}_i + e,
$$

(4.7)

where $\bar{g}_i$ is the mean breeding value for the joint phenotype of a gazelle’s cheetah interactants and $\bar{g}_j$ is the mean breeding value of the cheetah’s gazelle partners, each being weighted by the number of interactions with that partner. Note that in equations (4.5)–(4.7) the means of $n$s are over all individuals of a species, while the means of $z$s and $g$s are taken over partners of one individual. Substituting equations (4.6) and (4.7) into (4.5), assuming no gene–environment correlation, yields

$$
\Delta z = n_i b_{i \bar{z}_i} \text{Var}(g_i) + \text{Cov}(g_i, \bar{g}_i) + n_j b_{j \bar{z}_j} \text{Var}(g_j) + \text{Cov}(g_j, \bar{g}_j).
$$

(4.8)

The variance terms are for the direct effect of an individual’s genes on its own fitness and the covariance terms are indirect selection. The indirect terms will be relevant when there are genetic correlations between interactants, for example, owing to partner choice or selection favouring particular combinations of interactants in the same way that epistasis can favour correlation among genes (linkage disequilibrium) [36,37]. But these terms will often be at or near zero, in which case (4.8) becomes simply

$$
\Delta z = n_i b_{i \bar{z}_i} \text{Var}(g_i) + n_j b_{j \bar{z}_j} \text{Var}(g_j).
$$

(4.9)

The joint trait evolves according to the additive genetic variance in the two parties, each multiplied by the mean selective effect on that party. Considering selection on a joint phenotype provides a clear definition of conflict: current selective conflict exists when the two effects of the joint phenotype on fitness, $b_{i \bar{z}_i}$ and $b_{j \bar{z}_j}$, are of different signs. Then the two parties push the joint phenotype in opposite directions, as when gazelle genes are selected to decrease the gazelle-as-cheetah-meal trait, while cheetah genes are selected to increase it.
For pairwise interactions \( \bar{N}_i \bar{N}_j = \bar{n}_i \bar{N}_j \), where \( \bar{N}_i \) and \( \bar{N}_j \)
are the population sizes of gazelles and cheetahs, respectively. Substituting \( \bar{n}_j = \bar{n}_i \bar{N}_j / \bar{N}_i \) into (4.9) makes the role
of population size explicit. For example, if there are more
gazelles than cheetahs, then the second term is elevated relative
to the first, reflecting that the average cheetah must experience
more interactions than the average gazelle.

Now assume that the trait of interest, \( z_i \), is the fitness of
gazelles, \( w_i \). Equations (4.8) and (4.9) become

\[
\Delta \bar{w}_i = \{ \text{Var}(g_i) + \text{Cov}(g_i, g'_j) \} + \bar{n}_i b_{i\text{un}} \{ \text{Var}(g_j) + \text{Cov}(g_j, g'_j) \},
\]

(4.10)

\[
\Delta \bar{w}_j = \text{Var}(g_j) + \bar{n}_j b_{j\text{un}} \text{Var}(g_j).
\]

(4.11)

The breeding values \( g_i \) and \( g_j \) are now interpreted as breeding
values of gazelle and cheetah genes for gazelle fitness. The \( b \)
in the first term of each equation disappears because \( b_{i\text{un}} \) must
equal 1 (as the phenotype—now fitness—of a gazelle changes, it
changes gazelle fitness by 1). \( \bar{n}_i \) also disappears because each
gazelle experiences only one instance of the phenotype gazelle
fitness and \( \bar{n}_j \) is now interpreted as the number of cheetahs
affecting the gazelle’s fitness.

Note that if the fitness model (equations (4.2) and (4.3)) is
more appropriate for absolute rather than relative fitness, as
will often be the case, then the two terms of the right-hand
sides of equations (4.8)–(4.11) should be divided by \( \bar{w}_i \) and \( \bar{w}_j \).

Equations (4.10) and (4.11) represent extensions of the
fundamental theorem for gazelle fitness when affected by
cheetahs, with (4.11) being the simpler form that applies in
the usual case when partners’ breeding values are uncorre-
lated. If cheetahs have no heritable effect on gazelle fitness,
then the second term of (4.11) is zero, yielding Fisher’s origi-
nal fundamental theorem. If there are cheetah genes that
affect gazelle fitness (\( \text{Var}(g_j) \neq 0 \)) and this in turn affects
cheetah fitness (\( b_{i\text{un}} \neq 0 \)), there will be selection in cheetahs
for genes that change mean fitness in gazelles. More simply, if
cheetahs evolve to be better at catching gazelles, it will reduce
gazelle fitness. This is hardly a novel concept, but it is not the
one that has been formally incorporated into the fundamental
theorem. Mean gazelle fitness can decline if cheetah genes
(for gazelle fitness) have a larger variance or have a larger
selection gradient. Exactly parallel expressions for change in
cheetah fitness can be written with \( g_j \) and \( g_i \) now being the
breeding values for cheetah fitness, by switching the
subscripts \( i \) and \( j \).

More generally, it is easy to show that if the fitness of
gazelles is affected by multiple species indexed by \( S = 2 \ldots \)
\( S_{\text{max}} \), (4.11) becomes

\[
\Delta \bar{w}_i = \text{Var}(g_i) + \sum_{S=2}^{S_{\text{max}}} \bar{n}_j b_{i\text{un}} \text{Var}(g_j),
\]

(4.12)

where the \( j \)s now index separately for each partner species
(and again, if \( w \) is absolute fitness each term would be
divided by the species’ mean fitness). Covariance terms
such as those in (4.10) can be added if necessary. The sum-
mation, when negative, describes much of Fisher’s ‘deterioration
of the environment’ [7]. It is also a way of representing van Valen’s Red Queen effect [38], that because
of other species it is necessary to keep evolving just to stay in
the same place. However, for some species the \( b_{i\text{un}} \) effects
may be positive (e.g. mutualism) and will enhance the fitness
of their partner species.

5. Conspecific interactions

In this section, I address the question of joint phenotypes and
conflict between individuals of the same species playing
different roles. Interactions within a species give a similar
result to those between species, but can be complicated by
several factors. First, when partners are related there are
inclusive fitness effects, and the fundamental theorem
should take an inclusive fitness form [8]. In addition, where
conflicts involve individuals in two roles, such as male and
female, owner and intruder, or mother and offspring, each
individual carries genes for both roles, even if it does not
express both. In some cases, each individual might play
either role at different times. To fix ideas and to draw a
close parallel to the first model, consider small tadpoles
(potential victims) that may be cannibalized by large tad-
poles (potential cannibals) of the same species. Unlike the
gazelle–cheetah case, there is only one fitness, so the equation
parallel to (4.1) is

\[
\Delta \bar{z} = \Delta \bar{g}_i + \Delta \bar{g}_j = \text{Cov}(w, g_i) + \text{Cov}(w, g_j),
\]

(5.1)

where \( g_i \) and \( g_j \) are now the breeding values for joint trait \( z \)
for genes expressed in victims and cannibals (note that
Bijma’s derivation [8] uses different meanings of \( gw \) as
direct and indirect effects, with roles not explicitly treated).

Interaction changes an individual’s fitness by the sum of
what happens when it is a victim and a cannibal. Letting
primes designate genes of partners, the model is

\[
w = \bar{n}_i \bar{w}_i + \bar{n}_j \bar{w}_j + \bar{n}_j \bar{w}_j b_{i\text{un}},
\]

(5.2)

where \( \bar{n}_i \) and \( \bar{n}_j \) are the number of times an individual interacts
in the roles of potential victim and potential cannibal, respecti-
vately, \( \bar{z}_i \) and \( \bar{z}_j \) are the mean joint phenotypes it experiences in
the two roles, and \( b_{i\text{un}} \) and \( b_{j\text{un}} \) are the effects of phenotype \( z \)
fitness of individuals playing potential victim and potential
cannibal. Substituting into (5.1) and carrying out steps directly
parallel to the multi-species derivation yields the following
results parallel to (4.8) and (4.9), respectively:

\[
\Delta \bar{z} = \bar{n}_i \bar{w}_i b_{i\text{un}} [\text{Var}(g_i) + \text{Cov}(g_i, g'_j) + \text{Cov}(g_j, g_i)] + \bar{n}_j \bar{w}_j b_{j\text{un}} [\text{Var}(g_j) + \text{Cov}(g_i, g'_j) + \text{Cov}(g_j, g_i)],
\]

(5.3)

\[
\Delta \bar{z} = \bar{n}_j \bar{w}_j [\text{Var}(g_i) + \text{Cov}(g_i, g'_j)] + \bar{n}_i \bar{w}_i [\text{Var}(g_j) + \text{Cov}(g_i, g'_j) - \text{Cov}(g_j, g_i)].
\]

(5.4)

As before, we move from equation (5.3) to (5.4) by omitting
terms owing to correlation between genes for being a victim
and a cannibal (\( g_i \) and \( g_j \), respectively). In contrast to the
two-species case, such correlations are here easily caused by pleio-
tropic genes affecting both roles [39] but we neglect this to
highlight the role of social selection. Grouping term 1 with 4,
2 with 3 and then factoring out the variances yields an inclusive
fitness form

\[
\Delta \bar{z} = \text{Var}(g_i) [\bar{n}_i b_{i\text{un}} + \bar{n}_i b_{i\text{un}}] \beta_{i\text{un}} + \text{Var}(g_j) [\bar{n}_j b_{i\text{un}} + \bar{n}_j b_{i\text{un}}] \beta_{i\text{un}}.
\]

(5.5)

The first bracket term is the inclusive fitness effect of genes
expressed in the victim role \( g_i \), with the effect on self, \( \bar{n}_i b_{i\text{un}} \),
added to the effect on cannibals, \( \bar{n}_j b_{i\text{un}} \), multiplied by the
regression relatedness of victims to the cannibal interactants,
\( \beta_{i\text{un}} \). The second bracketed term is a similarly constructed
inclusive fitness effect for the cannibal role. Selection on the
joint phenotype operates on the two inclusive fitnesses, weighted by their additive genetic variances. Because the variances are always positive, selective conflict occurs when the victim and cannibal inclusive fitness effects are of different signs.

Letting the phenotype z be the fitness of a gazelle \( w_i \), we get a version of Fisher’s fundamental theorem, but for change in victim fitness only:

\[
\Delta \bar{w}_i = \text{Var}(g)[1 + \hat{n} b_{w_i w_i} \beta_{g,i}] + \text{Var}(g)[\hat{n} b_{w_i w_i} + \beta_{g,i}],
\]

(5.6)

where again \( b_{w_i w_i} \) and \( \hat{n} \) both equal 1 (the latter because each individual has only one fitness). As in the multi-species case, declines in fitness can outweigh gains, a well-known result in social evolution [40]. This is for two reasons. First, victims could be selected to lose personal fitness (first term) if it gave sufficient gains to victim genes in related cannibals (second term). This is Hamilton’s altruism [1,2]. Second, in the absence of relatedness, victim fitness can still decline because of selection on cannibals (third and fourth terms), that is, because of conflict. Just as in the multi-species case, individuals acting in other roles can reduce (or sometimes increase) the fitness obtained in the focal role.

However, as pointed out to me by Piter Bijma, this does not really capture the essence of Fisher’s fundamental theorem, because when relatives are affected, it is not the change in fitness that is important, but the change in inclusive fitness [8]. So, if we let the joint phenotype z be the inclusive fitness of victims—call it \( w_i' \)—equation (5.5) becomes

\[
\Delta \bar{w}_i' = \text{Var}(g)[\hat{n} b_{w_i w_i} \beta_{g,i}] + \text{Var}(g)[\hat{n} b_{w_i w_i} + b_{w_i w_i} \beta_{g,i}]
\]

(5.7)

As in (4.11) the multiplier of \( \text{Var}(g) \) reduces to 1, though the logic is more complicated. Note that \( \hat{n} \beta_{g,i} \) is a constant, while the two \( b \) terms describe how \( w_i \) and \( w_i' \) change with a unit change in \( i \)'s inclusive fitness, \( w_i' \). Thus, the first bracket gives the change in \( w_i + \hat{n} \beta_{g,i} w_i \), which is \( i \)'s inclusive fitness, for a unit change in \( i \)'s inclusive fitness, and this is clearly 1. Similarly, the second bracket asks how a unit change in \( i \)'s inclusive fitness affects the summed inclusive fitness of its partners, which I write as \( b_{w_i w_i} \). Equations (5.6) and (5.7) can be converted to equations for change in cannibal fitness and inclusive fitness by switching the \( i \) and \( j \) subscripts throughout.

Once again, if the fitness model (here equation (5.2)) describes absolute rather than relative fitness, then the right-hand side of equations (5.3)–(5.7) should be divided by \( w_i \).

Equation (5.7) shows that if there are no cannibal genes that affect victim inclusive fitness (\( \text{Var}(g) \neq 0 \)) then the rate of change of victim inclusive fitness is equal to its additive genetic variance. However, if there are cannibal genes that affect victim inclusive fitness (\( \text{Var}(g) \neq 0 \)) and this in turn affects cannibal inclusive fitness (\( b_{w_i w_i} \neq 0 \)) there will be selection in cannibals for genes that change mean fitness in victims. If cannibals evolve to be better at catching victims, it will reduce victim fitness. Thus, this inclusive fitness formulation (5.7) captures a reason for decline in fitness—conflict—not explicitly treated in prior formulations [8]. This conflict of course remains (indeed is enhanced) in the absence of relatedness.

6. Discussion

The results derived here model how selection works on joint phenotypes, highlighting the role of conflict between two parties. Considering joint phenotypes is useful because they are the objects of conflict. Selective conflict exists when the two parties are selected to push the joint phenotype in opposite directions—when the selection terms in (4.9) or the inclusive fitness terms in (5.4) differ in sign for the two parties. The outcome of such selection depends not only on the relative magnitudes of these selection terms, but also on the genetic variances. If Fisher’s fundamental theorem of natural selection is regarded as a design principle, the versions here incorporate conflicting design criteria.

Social evolution theorists have found it useful to distinguish potential and actual conflicts [41,42]. Potential conflict exists over the range of possible selection regimes that would lead to different signs of selection on the two parties. For the two-species equation (4.9), potential conflict exists for any values of the joint phenotype \( z \) that would affect the fitness of gazelles and cheetahs in opposite directions (\( b_{w_i z} > 0 \) or \( b_{w_i z} < 0 < b_{z w_i} \)). Within a species with two roles affecting the same joint phenotype (equation (5.4)) it is the opposite signs effect of the joint phenotype on inclusive fitness (\( \hat{n} b_{w_i z} + \hat{n} b_{z w_i} \beta_{g,i} \), and \( \hat{n} b_{w_i z} + \hat{n} b_{z w_i} \)) that determine potential conflict.

Actual conflict depends on the effects of real genes that actually create selection. Potential conflict may not result in actual conflict if one party has no power to affect the joint phenotype. The equations derived here describe the process that leads to actual conflict but they do not fully describe actual conflict. The process of selective conflict described here depends on segregating genetic variation but much of the actual conflict observed in nature presumably results from variation previously fixed by selection. Most of the genes underlying gazelle and cheetah conflicts—for example, genes underlying speed, agility and perception—are presumably fixed. Actual phenotypic conflict can be defined as occurring when two parties push a joint phenotype in opposite directions, as the result of either current or past selective conflict. Such phenotypic conflict can occur even when one party is currently depleted of genetic variation for the trait. Cheetahs may have little genetic variation left for increasing their speed, but they nevertheless use their accumulated speed genes, through past selective conflict, to capture escaping gazelles.

The equations are agnostic with respect to the size and direction of fitness effects, so they can represent the evolution of cooperation as well as conflict. When the fitness effects of the joint phenotype in multi-species interactions have the same sign or when the inclusive fitness effects in same-species interactions have the same sign, then both parties are being selected in the same direction. I have emphasized conflict because it has been relatively neglected in indirect-genetic-effect models and because conflict is likely the biggest driver of Fisher’s deterioration of the environment. Moreover, the division into effects owing to different species or to different roles within species emphasizes that potential for conflict generally remains even over potentially cooperative or mutualistic traits. When the fitness effects have the same sign, selection operating on both parties will push the joint trait in a common direction until it reaches a point where it is no longer beneficial for one party, at which point selective conflict may commence. For example, a pea plant and its rhizobial
symbiont may both benefit from the nitrogen provided by the latter, but the legume may try to extract more [43].

Ultimately, conflicts are about fitness, and if we use the fitness of either party as the joint phenotype, the equations become versions of Fisher’s fundamental theorem of natural selection, extended to multiple parties. Fisher showed that fitness changes at the rate of the additive genetic variance for fitness [7] and the result has been generalized to inclusive fitness [30], but the new versions derived here emphasize that it can also change as a function of genetic variances of all the parties that have an effect on that organism and that fitness can decline because of conflict. Previous versions of the fundamental theorem emphasize how an individual’s own genes are selected to influence its (inclusive) fitness; the versions derived here add in the effects of genes residing in others.

Additional generalizations of these results are desirable, for example combining both within- and between-species effects, unequal generation times, age structure, correlated traits, gene–environment correlations, overlapping fitnesses and nonlinear effects on phenotypes and fitness. These results are also still partial in the sense of ignoring change owing to other factors. They also assume that the effects of the predictors remain constant in the next generation. The models do not include the entire effect of the biotic environment, only the part that arises from change in gene frequency of other species. Fitness might also decline (or increase) owing to changes in the population sizes of the various parties [44]. But the main point of Fisher’s fundamental theorem is to capture the effects of the adaptive engine of evolution [18] and here it is done much more completely by including the sometimes potentially conflicting adaptive engines of multiple parties.

The message can be illustrated through another model for fitness increase that, like Fisher’s theorem, has been both useful and controversial: Wright’s adaptive landscape. Imagine that each of the parties has its own fitness landscape. Instead of a lonely mountaineer steadily climbing his peak we have multiple mountaineers, each climbing his own peak, but roped to the others. As one climbs, he often drags another down from his peak. While this metaphor should not be pushed too far, it does suggest that Dobzhansky [45] may have been wrong in proposing that life is concentrated near fitness peaks. Instead, there are forces keeping the valleys and lower slopes populated and, because their inhabitants are constantly pulling and being pulled, this is where much of evolution occurs.

Acknowledgements. I thank Alan Templeton, Joan Strassmann, Devin Dobias, Anya Plutynski, Alan Grafen and especially Piter Bijma for comments on the manuscript.

Funding statement. This work was supported by the National Science Foundation under grants no. IOS-1256416 and DEB-1146375 and the John Templeton Foundation.

References


effects on community structure. Evolution 60, 991 – 1003.