Selection in structured populations

Erol Akçay
Biol 417, Spring 2019

April 1, 2019

So far, we concerned ourselves with social interactions that happen between individuals who meet each other at random: the probability that a focal Hawk-player played another Hawk for example, was simply equal to the overall frequency of Hawks in the population. Frequently, however, interactions are not between randomly meeting individuals, but rather occur within some *population structure*. This means that a focal individual, depending its type, may have a lower or higher probability of interacting with different types of individuals.

Population structure is a very general term, signifying slightly different (but related) things in ecology and evolution. For instance, in ecology, a population structure (or structured population) might refer to age- or stage-structure we have seen before. In population genetics, population structure usually refers to spatial structure (when age- or stage-structure is meant, the phrase "class structured" is frequently used).

Biologically, a frequent cause for this phenomenon is that populations are distributed in space, and in many species, offspring disperse a finite (and sometimes not great) distance from their parents. Therefore, intuitively, when an individual interacts with those around them, it is likely that they are to some degree related to each other, and therefore might have the same behavioral tendencies, etc. Such populations are called *viscous populations*. Another way non-random interactions might come about is if individuals recognize different types (e.g., their relatives), and preferentially interact with some. Such kin-recognition is also clearly present in many species (mostly in the relatively weak form of recognizing familiar individuals, who are more likely to be kin than strangers). Whatever the cause, non-random, or assortative interactions can change the course of evolution.

Hamilton's rule

Let us consider the simple, single shot donation (or Prisoner's Dilemma) game, given by the matrix below, with only pure, unconditional strategies.

$$\begin{array}{ccc} & \text{Donate} & \text{Don't Donate} \\ \text{Donate} & b-c & -c \\ \text{Don't donate} & b & 0 \end{array} \tag{1}$$

Now, instead of individuals just bumping into each other randomly, assume that each individual has probability r with interacting with someone of their own strategy, and $(1-r)$ of interacting with someone at random. Now the expected payoffs to Donate and Don't Donate are:

$$w_C = r(b-c) + (1-r)(p(b-c) - (1-p)c) \tag{2}$$

$$w_D = r \times 0 + (1-r)(pb + (1-p) \times 0) \tag{3}$$

A sufficient condition.

For Donate to be an ESS, we need $w_C - w_D > 0$. Calculating that difference, we find the condition for the cooperators to increase to be:

$$\begin{aligned} r(b-c) + (1-r)(p(b-c) - (1-p)c) - (1-r)pb &> 0 \\ r(b-c) - (1-r)c &> 0 \\ rb - c &> 0\,. \end{aligned} \tag{4}$$

This is the celebrated Hamilton's rule for a two-person interaction (Hamilton, 1964).

The derivation above is in many ways the simplest possible one for Hamilton's rule. It makes an important point very apparent: that Hamilton's Rule is simply the expression of the mean *personal* fitness of the carriers of an altruistic variant being higher than that of the non-altruistic variant. The reason that altruist fare better (if they do) is simple that when everyone tends to interact with their own type (with probability r), altruist tend to interact with other altruists and receive the benefits of their donations. Thus, Hamilton's rule simply becomes an expression of whether altruists have enough additional altruists in their environment (relative to the environment of the non-altruists) to compensate them for the cost of their altruistic behavior.

Another thing to note about condition (4) is that it is independent of the frequency of Donate (or C). This is surprising, since we set out writing it simply by requiring $w_C - w_D > 0$ at any frequency. It turns out, this is a general property of a class of games (or interactions) called "additive" interactions. In additive interactions, the payoff (or fitness) of an individual can be written as the sum of two (or more) terms; one only depending on the genotype of the focal individual, and the other(s) only on that of the social partner(s). If a game is additive, then the condition for one type to do better than the other is independent of the frequency of types.

The above derivation of Hamilton's rule is simple, but its simplicity is somewhat deceptive: we pulled that assortment probability r out of a hat, and have no idea what produces it. That is frequently is a very important biological question. In the next section, we will look at another way of arriving at Hamilton's rule that will generalize it and also clarify what the general meaning of relatedness coefficient should be.

Hamilton's rule derived from the Price equation

Recall that the Price equation (neglecting the transmission bias term) is given by:

$$\bar{w}\Delta\bar{g} = \mathrm{cov}(w, g)\ , \tag{5}$$

where g denotes the genotype of an individual, i.e., the trait that offspring inherit from their parents without change. In a haploid, two allele model, for example, $g = 0$ for one allele and 1 for the other. Now, suppose fitness of individuals is the result of their interactions with one partner (i.e., a two-person game). In that case, the fitness $w_\bullet$ of a focal individual will be a function of its own genotype $g_\bullet$ and that of its partner, $g_\circ$. In general, $w_\bullet$ can depend on the two gs in very complex ways. For the moment let us assume that the fitness of a focal individual, $w_\bullet$ is given by:

$$w_\bullet = w_0 + \beta_{wg_\bullet} g_\bullet + \beta_{wg_\circ} g_\circ \ , \tag{6}$$

where $\beta_{wg_\bullet}$ and $\beta_{wg_\circ}$ are just constants, and w_0 is a baseline fitness. The β coefficients describe the contribution to the focal individuals' fitness of the focal individuals' own genotype and that of its partner. For example, in the donation game above, if we adopt the convention that Donate means $g = 1$ and Not Donate $g = 0$, then we have $\beta_{wg_\bullet} = -c$ and $\beta_{wg_\circ} = b$.

Substituting this fitness function into equation (5), and expanding the covariance, we have:

$$\begin{aligned} \bar{w}\Delta\bar{g} &= \text{cov}(w_0 + \beta_{wg_\bullet} g_\bullet + \beta_{wg_\circ} g_\circ \ , g_\bullet) \\ &= \beta_{wg_\bullet} \text{cov}(g_\bullet, g_\bullet) + \beta_{wg_\circ} \text{cov}(g_\circ, g_\bullet) \\ &= (\beta_{wg_\bullet} + \beta_{wg_\circ} \beta_{g_\circ g_\bullet}) \text{var}(g) \ , \end{aligned} \tag{7}$$

where in the last line we used $\text{cov}(x, x) = \text{var}(x)$, and $\beta_{g_\circ g_\bullet} = \text{cov}(g_\circ, g_\bullet)/\text{var}(g)$ is the regression coefficient of $g_\circ$ on $g_\bullet$. Since a variance and $\bar{w}$ are always positive, (7) tells us that the mean genotypic value of a population will increase (i.e., in a haploid population, the frequency of the allele for which $g = 1$) if the term in the parenthesis is positive, and decrease otherwise. In the additive donation game above, the condition for the phenotype to increase then becomes:

$$\beta_{g_\circ g_\bullet} b - c > 0 \ . \tag{8}$$

Comparing (8) to (4), we see $\beta_{g_\circ g_\bullet} = r$, but note that in deriving condition (8) we made no assumption about how individuals are matched to their partners. The message of (8) is that however the matching process happens, it is the linear regression of the partner genotype on the focal individuals' genotype that matters for the action of selection. This gives us a definition of relatedness as a regression coefficient, which we'll call R:

$$R = \beta_{g_\circ g_\bullet} \equiv \frac{\text{cov}(g_\circ, g_\bullet)}{\text{var}(g)} \tag{9}$$

The regression definition of relatedness is very general, and is the relevant quantity independent of what is creating structured or assortative interactions. Whenever you express fitness as a function of genotypes, you end up calculating the regression between focal individual and partner genotypes in some way.

Inclusive fitness

In the above derivation with the Price equation, we again used an individuals' own personal fitness $w_\bullet$ (which is a function of both the focal individuals' and its partner's genotypes) to derive Hamilton's Rule. This way of arriving at Hamilton's Rule is sometimes called the "direct," or "neighbor-modulated" fitness approach in social evolution theory (Wenseleers et al., 2010; Marshall, 2015). It contrasts with another approach that gets more press, called "inclusive fitness," where one constructs a fitness function that is a function of only the focal individuals' genotype, and treats this fitness function as being maximized. The way one does that is, in Hamilton's own words (Hamilton, 1964, p. 8):

It should really be "partner-modulated," but that's the name that stuck.

> Inclusive fitness may be imagined as the personal fitness which an individual actually expresses in its production of adult offspring as it becomes after it has been first stripped and then augmented in a certain way. It is stripped of all components which can be considered as due to the individual's social environment, leaving the fitness which he would express if not exposed to any of the harms or benefits of that environment. This quantity is then augmented by certain fractions of the quantities of harm and benefit which the individual himself causes to the fitnesses of his neighbours. The fractions in question are simply the coefficients of relationship appropriate to the neighbours whom he affects...

So, instead of tracking the total fitness of individuals, we track the fitness effects of a focal individual on itself only, and then add to that the effect of the focal individual on its partner (rather than the partner on the focal individual, as before), weighted by the relatedness. One can show mathematically that the inclusive fitness approach, done right, is equivalent to the direct fitness approach. In the particular case above, since the interaction is purely symmetric, the correspondence is immediate. The effect of having $g = 1$ instead of 0 is $-c$, the effect of the focal individual on its partner is b, and the relatedness coefficient is symmetric, so again equal to R. Hence, we have:

Indeed, this was a major contribution of Hamilton's paper; the other one was to come up with the neighbor-modulated approach itself!

Again, in this case; not true in general

$$w_{IF} = -c + Rb\,, \tag{10}$$

Technically, w_{IF} should be called "inclusive fitness effect," because it is the change in inclusive fitness if one "flipped" the focal individual's genotype from 0 to 1. But to many a theorist's dismay, the word "effect" gets dropped quite often.

which is exactly the same as (8).

Essentially, direct and inclusive fitness methods are different ways of accounting for fitness effects and attributing them to the correct genotypes. However, the direct fitness approach has a conceptual simplicity. While you have to make sure you correctly account for its social environment's effects, there is less scope for "mis-placing" a fitness effects, or miscalculating a relatedness coefficient, because you are always dealing with one focal individual's fitness (or rather, the expectation of it (Akçay and Van Cleve, 2016)). Inclusive fitness formulations tend to provide more scope for making errors in attributing fitness effects, or sometimes even not calculating relatedness coefficients correctly. In the simple cases (like the one above), both approaches are perfectly fine, but in more complicated ones (starting with non-additive interactions which we will talk about

This is probably the reason why people started using "direct" instead of "neighbor-modulated" over the years.

below, but even more so with class-structured populations), inclusive fitness accounting can get pretty hairy.

On the other hand, inclusive fitness has one virtue (at least in the eyes of some): by accounting for the fitness effects of the genotype of a single individual, it gives us a quantity that is at least on the surface similar to the classical models of population genetics that is a function of only one genotype. I happen to think that similarity is only superficial, since both direct and inclusive fitness methods are functions of both the focal individual's genotype and its covariance with that of social partners. Inclusive fitness sweeps this dependence into the relatedness term to make it look *as if* it only depends on one genotype, whereas direct fitness explicitly acknowledges that fitness is affected by multiple genotypes. There are some questions where the *as if*-thinking has merit, but in general, I'm in favor for the explicit accounting.

Non-additive interactions

Now we move beyond the additive game case. In particular, consider the following modification of the donation game:

$$\begin{array}{ccc} & \text{Donate} & \text{Don't Donate} \\ \text{Donate} & b+d-c & -c \\ \text{Don't donate} & b & 0 \end{array} \quad , \tag{11}$$

where we added a new parameter, d, that denotes the additional payoff increment a player gets when both players donate. This parameter is sometimes called the "synergy" parameter: if $d > 0$, then two donations are better than twice one donation. However, d can also be negative, in which case, the value of the second donation is less than that of the first. Now, if we want to express the fitness of individuals as a result of playing this game in a manner similar to (6), we need more than two terms, since whether an individual also gets the d-increments depends on the combination of the two genotypes. We can write fitness here as:

$$w_\bullet = w_0 + bg_\bullet - c_\circ + dg_\bullet g_\circ \; , \tag{12}$$

Note gs are either 0 or 1, so the third term is non-zero only when both $g_\bullet$ and $g_\circ$ are 1, i.e., when they both donate. Putting this fitness expression into the Price equation (5), we obtain:

$$\begin{aligned} \bar{w}\Delta g &= \mathrm{cov}(w_0 + bg_\circ - cg_\bullet + dg_\bullet g_\circ, g_\bullet) \\ &= b\mathrm{cov}(g_\circ, g_\bullet) - c\mathrm{cov}(g_\bullet, g_\bullet) + d\mathrm{cov}(g_\bullet g_\circ, g_\bullet) \\ &= (Rb - c + \beta_{g_\bullet g_\circ, g_\bullet} d)\mathrm{var}(g) \; , \end{aligned} \tag{13}$$

where $\beta_{g_\bullet g_\circ, g_\bullet} = \mathrm{cov}(g_{\bullet g_\circ}, g_\bullet)/\mathrm{var}(g)$ is the regression of the *product* of partner and focal individual genotypes on the focal genotype. As before, since the vari-

$\beta_{g_\bullet g_\circ, g_\bullet}$ is sometimes called the "triplet relatedness," since it is equivalent to drawing three individuals from the population with the same genotypes.

ance of g has to be zero, whether the mean genetic value of the population will increase or decrease will depend on the sign of the parenthesis. To evaluate it, we now need to calculate $\beta_{g_\bullet g_\circ, g_\bullet}$. To do that, let's return to our simple model of assortment, where there was a constant probability r that a focal individual interacted with its own type. The covariance $\mathrm{cov}(g_\bullet g_\circ, g_\bullet)$ is then given by (p is the frequency of the "Donate" allele):

$$\begin{aligned} \mathrm{cov}(g_\bullet g_\circ, g_\bullet) &= \mathrm{E}[g_\bullet^2 g_\circ] - \mathrm{E}[g_\bullet g_\circ]\mathrm{E}[g_\bullet] \\ &= p(r + (1-r)p) - p^2(r + (1-r)p) \\ &= p(1-p)(r + (1-r)p) \ , \end{aligned} \tag{14}$$

which divided by $\mathrm{var}(g) = p(1-p)$ yields:

$$\beta_{g_\bullet g_\circ, g_\bullet} = r + (1-r)p \ . \tag{15}$$

Thus, Hamilton's Rule for the increase of the Donate genotype with non-additive fitness effects is given by:

$$rb - c + d(r + (1-r)p) > 0 \tag{16}$$

Note that, in contrast to the case with additive effects, we could not get rid of the p from the final condition: non-additive social selection is inherently frequency dependent. In particular, since $0 \leq r \leq 1$, the left-hand side of (16) is decreasing in d when d is negative, and increasing in d is positive. That means that for $d < 0$, it is possible the condition is satisfied for low p (when cooperation is rare) but not for high p (when cooperation is common). That would be negative frequency dependence. This would yield a stable polymorphism where Donate and Not Donate coexist. Conversely, when $d > 0$, it is possible that cooperation cannot increase when rare, but can go to fixation when common; in other words both Donate and Not Donate might be evolutionarily stable (positive frequency dependence).

Reciprocity with population structure

Let's return to the reciprocity model from the evolutionary game theory lecture, with the TFT replacing the unconditional cooperators:

In the notation of the notes for the evolutionary game theory lecture, we have $r = b - c$, $t = b$, $p = 0$, $s = -c$.

$$\begin{array}{ccc} & \text{TFT} & \text{Always D} \\ \text{TFT} & \frac{1}{1-\delta}(b-c) & -c \\ \text{Always D} & b & 0 \end{array} \ , \tag{17}$$

Now, the TFT genotype (which we will assign $g = 1$) gets more than $b - c$ when playing against itself (since $0 < \delta < 1$, so $1/(1-\delta) > 1$. We can therefore see that the synergy term $d = \frac{1}{1-\delta}(b-c) - (b-c) = \frac{\delta}{1-\delta}(b-c)$ is positive. Substituting it into the non-additive version of Hamilton's rule (16), we get:

$$rb - c + (r + (1-r)p)\frac{\delta}{1-\delta}(b-c) > 0 \ . \tag{18}$$

First, if we set $r = 0$ (as we implicitly did in the last lecture), we see that the left-hand side is $-c$ when $p \approx 0$ and $\frac{\delta}{1-\delta}b - \frac{1}{1-\delta}c$ when $p \approx 1$, which gives us the same stability conditions as the last lecture.

Good exercise to check this.

When $r > 0$, we have for $p \approx 0$ the left-hand side becoming $rb(\frac{1}{1-\delta}) - c(\frac{1-(1-r)\delta}{1-\delta})$. Thus, adding assortment (increasing r adds both to the benefit and cost experienced by a rare reciprocator, but adds more to the benefit. This means that for

$$r > \frac{c(1-\delta)}{b - \delta c} \, , \tag{19}$$

The TFT strategy will be able to invade a population of unconditional Defectors. It is easy to check that the condition for fixation also gets easier to satisfy. Thus, we have shown that population structure can help reciprocity to get established (Axelrod and Hamilton, 1981; Van Cleve and Akçay, 2014).

Partial regression, non-additivity, and inclusive fitness

In the last section, we saw that if fitness effects are non-additive, you generally get frequency dependent dynamics that might mean a trait like cooperation might be favored when rare but disfavored when common. But remember the Hamilton's rule we derived before then is a frequency independent statement. This suggests that Hamilton's rule (and inclusive fitness) are restricted quantities that only apply to additive games. There are indeed many people who believe that. But there are also some that counter that a more general version of Hamilton's rule can be written that does not require additivity to hold. The way to do that is to express fitness in terms of least-squares partial regression coefficients:

$$w_\bullet = w_0 + \beta_{wg_\bullet} g_\bullet + \beta_{wg_\circ} g_\circ + \epsilon \, . \tag{20}$$

Equation (20) looks almost the same as equation (6). Both express fitness as an additive function of the genotypic values of self and partner. The most obvious difference is that last term, ϵ, which is an error term, which points to a very deep change in meaning between the two equations: whereas (6) posits that fitness in fact *is* an additive function of the genotypic values, (20) only posits that we can try to predict fitness with an additive function, but admits that sometimes we will be off by an amount, which is given by ϵ. Now if we substitute this into the Price equation, we get:

$$\bar{w}\Delta g = \beta_{wg_\bullet} \mathrm{cov}(g_\bullet, g_\bullet) + \beta_{wg_\circ} \mathrm{cov}(g_\circ, g_\bullet) + \mathrm{cov}(\epsilon, g_\bullet) \, . \tag{21}$$

So far, we haven't said what the $\beta_{wg_\bullet}$ and $\beta_{wg_\circ}$ are. Obviously, we'd like to choose them to minimize the error, and also now we have a term in the Price equation that we have no idea how to deal with: the covariance of errors with the genotype. As it turns out, we can kill two birds in one stone: if we choose the two βs such that the square of all the ϵs is minimized, then the errors are uncorrelated with $g_\bullet$, and thus the last term vanishes. Then we get the Hamilton's rule saying $\Delta g > 0$ when:

$$\beta_{wg_\bullet} + \beta_{wg_\circ}\beta_{g_\circ,g_\bullet} > 0 \tag{22}$$

Note, however, that (22) is somewhat deceptive: now all three βs in it are regression coefficients, and in particular, they look like they are frequency independent. But in fact they are not: the least-squares linear regression coefficients depend on the frequency of the different pairings, which depend on the frequency of types. This can be seen by writing the expected error:

$$\mathrm{E}[\epsilon] = P_{00}w_{00}^2 + P_{01}(w_{01} - \beta_{wg_\bullet})^2 + P_{10}(w_{10} - \beta_{wg_\circ})^2 + P_{11}(w_{11} - \beta_{wg_\bullet} - \beta_{wg_\bullet})^2 \ , \tag{23}$$

where P_{ij} is the probability of the focal individual being type j and the partner of type i, and w_{ij} is the corresponding fitness of the focal individual. The first order condition to maximize $\mathrm{E}[\epsilon]$ is that the derivative with respect to the betas vanish, which yields for $\beta_{wg_\bullet}$:

$$-2P_{01}(w_{01} - \beta_{wg_\bullet}) - 2P_{11}(w_{11} - \beta_{wg_\bullet} - \beta_{wg_\circ}) = 0 \ . \tag{24}$$

This condition (and the corresponding one for $\beta_{wg_\circ}$ are satisfied independent of the Ps only when the terms in the parentheses are both zero; in other words, when the error terms are always zero, i.e., when fitness can be perfectly expressed with an additive function, i.e., when fitness is additive.

Therefore, the regression definition of Hamilton's rule, though perfectly valid and general, does not get rid of frequency dependence. We merely swept it under the rug, or perhaps, to use a related figure of speech with a more positive spin, we tidied up the room by putting away ugly stuff in elegant boxes. Now, as your parents might attest, tidying up the room can be a positive thing even if you didn't get rid of all the junk. But it is a mistake to pretend that by tidying up the room, one got rid of the junk.

Multi-level selection

Another perspective to study evolution in (group) structured populations is called multi-level selection. In this approach, one still accounts for the effects of individual phenotypes (or genotypes) on both one's own fitness and that of others, but now does it in a slightly different formulation. In particular, consider a population subdivided into groups (or demes) of size N in which social interactions take place. This can signify different biological scenarios. For example, all N individuals in the group might be actually interacting with all others, or it could be pairwise interactions with the partner drawn randomly from within group. In either case, we can write the fitness of a focal individual as follows:

Multi-level selection has a long and contentious history, and is still being debated. This is in part because multi-level selection is affiliated with a somewhat nebulous set of ideas pertaining to group selection; the notion that populations can change through groups (instead of individuals) leaving different numbers of offspring. Here, we will treat a "vanilla" version of MLS where there will be no question of its legitimacy, as we will see.

$$w_\bullet = \beta_{w\Delta g_\bullet}\Delta g_\bullet + \beta_{w\bar{g}_G}\bar{g}_G \ , \tag{25}$$

where $\Delta g_\bullet$ refers to the deviation of the focal individual's genotype from the group mean, $\bar{g}_G$. As with the neighbor-modulated fitness approach, this equation can be interpreted as a literal description of fitness: that fitness actually depends on the mean and the deviation of the focal individual from the mean, or it can be supplemented with a residual term that signifies that it is meant to be a statistical best fit to some unspecified fitness distribution. Substituting this fitness expression into the Price equation, we get:

$$\bar{w}\Delta\bar{g} = \beta_{w\Delta g_\bullet}\mathrm{cov}(\Delta g_\bullet, g) + \beta_{w\bar{g}_G}\mathrm{cov}(\bar{g}_G, g) \tag{26}$$

Now, we have by definition $g_\bullet = \bar{g}_G + \Delta g_\bullet$, which we can substitute into the covariance terms to get:

$$\bar{w}\Delta g = \beta_{w\Delta g_\bullet}(\mathrm{cov}(\Delta g_\bullet, \bar{g}_G) + \mathrm{cov}(\Delta g_\bullet, \Delta g_\bullet)) + \beta_{w\bar{g}_G}(\mathrm{cov}(\bar{g}_G, \bar{g}_G) + \mathrm{cov}(\bar{g}_G, \Delta g_\bullet)) \ . \tag{27}$$

Collecting terms, we get:

$$\bar{w}\Delta\bar{g} = \beta_{w\Delta g_\bullet}\left\langle \mathrm{var}_G(\Delta g)\right\rangle + \beta_{w\bar{g}}\mathrm{var}(\bar{g}) + (\beta_{w\Delta g_\bullet} + \beta_{w\bar{g}})\mathrm{cov}(\bar{g}, \Delta g_\bullet) \tag{28}$$

This is one possible so-called multi-level selection partition of selection (unlike the neighbor-modulated one, there isn't a single one). This partition has three components: the first is the average within-group variance in deviations from group mean. The angled brackets and the subscript G on the variance denote that it is the mean within-group variance in the deviation. That corresponds to selection within groups, sometimes called "selection at the individual level." The second term has the variance of $\bar{g}$, i.e., the mean group genotype, corresponding to selection between groups. Finally, we have a covariance term between group mean and deviation from the mean. We can expand that term further:

But the levels of selection terminology can be confusing

$$\mathrm{cov}(\bar{g}_G, \Delta g_\bullet) = \left\langle \mathrm{E}_G[(\Delta g - 0)(\bar{g}_G - \bar{g})]\right\rangle = \left\langle \mathrm{E}_G[\Delta g \bar{g}_G]\right\rangle \ , \tag{29}$$

where we used the facts that the global (as well as within-group) mean of Δ_g is by definition 0, and that of $\bar{g}_G = \bar{g}$, and the angled brackets are again expectations over groups. Thus, the last term is the expectation of the individual deviation weighted by the group mean, which is akin to a transmission bias term.

Whether this decomposition of selection or the neighbor-modulated one (or some other one) is useful depends on what biological processes one is interested in. At the end of the day, the Price equation is simply an identity. Its power lies in its flexibility to separate different kinds of effects one is interested in, but with that great power comes great responsibility to choose your partition wisely.

The easy way to make (one type of) kin selection model

The above discussion focused on the general concepts and interpretation of inclusive fitness, Hamilton's rule, and multi-level selection. Now let's turn to a more practical question: suppose we have a *phenotype* that has social effects (i.e., affects the fitness of both its carrier and someone else). How will that phenotype evolve in a structured population with non-random interactions? Notice that in the above, we mainly concerned ourselves with genotypes, usually haploid, but in all cases discrete. Here, the phenotype can be a continuous variable (such as body size, or contribution to a public good). In that case, we might want to predict the "optimal" value of such a phenotype that would be favored by selection. It turns out, answering this question can be easy, provided we are willing to make some approximations.

We will see a more rigorous derivation of this kind of optimality modeling that includes potential ecological feedbacks in the Adaptive Dynamics lecture, but here I present a somewhat hand-vavy version with population structure.

To fix the setting, consider the public goods game introduced in the game theory lecture, where N individuals invest into creating or defending a resource (e.g., a territory, an extra-cellular matrix, a common nest) that all individuals benefit from. These investments will be individually costly, but benefits accrue to everyone, there is a conflict of interest. Suppose the benefit function is given by $B(\sum_j a_j)$ where a_j are the contributions of the jth individual, and the cost to individual i is given by $C(a_i)$, with $C(\cdot)$ an increasing function of the contribution a_i. The total payoff to individual i is then:

$$w_i = B(\sum_j a_j) - C(a_i) \; . \tag{30}$$

Now, how do we find the optimum investment? Suppose a_i is a trait with true inheritance; then we can use the Price equation with only the covariance term to write down the change in the mean contribution of the population, $\bar{a}$:

$$\bar{w}\Delta\bar{a} = \text{cov}(w_i, a_i) = (\beta_{wa_i} + \beta_{wa_j}\beta_{a_i a_j})\text{var}(a) \; . \tag{31}$$

The second step follows simply from using the contribution a_i in place of the $g_\bullet$ and a_j (where $j \neq i$) in place of $g_\circ$ in equation (7). Ok, but where do we get these linear regression terms β_{wa}s when we are only given fitness as a continuous function of the contributions? If the function was linear in all the a_i and a_js, it would be easy, since whatever the linear coefficients in front of a_i and a_js were, those would automatically be our βs. In other words, if we could write the fitness w_i as:

$$w_i = \sum_j b a_j - c a_i \; , \tag{32}$$

then we'd have $\beta_{wa_i} = b - c$ and $\beta_{wa_j} = (N-1)b$. Note the appearance of the $N-1$, where N is the group size, which comes due to the fact that each partner j gives the same benefit b to the focal individual, and in a group of size N there are $N-1$ partners. Replacing $\beta_{a_i a_j}$ with r, the relatedness, we could then write the condition for the contributions to increase in the population as:

$$b - c + r(N-1)b > 0 \; . \tag{33}$$

Ok, this is if the functions B and C were linear, but in general they may not

Note that for this model to be a social dilemma, we require $b < c$ but $Nb > c$, since b is the *per capita* benefit from contribution of one individual. If this were greater than the private cost, the question would be somewhat boring, since it would be individually optimal to invest regardless of what others are doing. Conversely, if $Nb < c$, then the public good actually cannot benefit anyone, even if everyone contributed.

be. But we have a handy trick to deal with non-linear functions: we pretend that they are linear, which works well as long as we confine our attention to a restricted range.

In particular, suppose the population has currently mean contribution a_r (subscript r here stands for "resident"). Further, suppose that the genetic variation in the contribution a is low, such that most individuals phenotypes can be written as $a_r \pm \delta$, where δ is small. In that case, we can Taylor-expand the functions B and C around the resident value a_r:

$$B(\sum_j a_j) \approx B(Na_r) + \sum_j \frac{\partial B}{\partial a_j}\bigg|_{a_j=a_r} \delta$$

$$C(a_i) \approx C(a_r) + C'(a)\big|_{a_j=a_r} \delta$$

Under this approximation, clearly we can pretend that the fitness functions are locally linear in a_i and a_j, and use $\beta_{wa_i} = \frac{\partial B}{\partial a_i} - C'(a_i)$ and $\beta_{wa_j} = (N-1)\frac{\partial B}{\partial a_j}$ to get the condition for increase of a as:

$$\frac{\partial B}{\partial a_i} - C'(a_i) + r(N-1)\frac{\partial B}{\partial a_j}\bigg|_{a_i=a_j=a_r} > 0\,. \tag{34}$$

If this condition holds, a mutant contribution level a slightly larger than a_r can invade the population, and because of the approximate additivity of the payoffs, it will also be able to go to fixation. If the inequality is reversed, a mutant a slightly less than a_r will be able to invade and fix. A candidate ESS a_r must give rise to neither condition, and therefore must have:

$$\frac{\partial B}{\partial a_i} - C'(a_i) + r(N-1)\frac{\partial B}{\partial a_j}\bigg|_{a_i=a_j=a_r} = 0\,. \tag{35}$$

This is the first order ESS condition. Satisfying it is a necessary but not sufficient for a contribution level a_r to be evolutionarily stable. We will learn more about the first and second order ESS conditions in the adaptive dynamics lecture.

References

Akçay, E., and J. Van Cleve. 2016. There is no fitness but fitness, and the lineage is its bearer. Phil. Trans. R. Soc. B 371:20150085.

Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. Science 211:1390–1396.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. Journal of Theoretical Biology 7:1–16.

Marshall, J. A. R. 2015. Social Evolution and Inclusive Fitness Theory: An Introduction. Princeton University Press.

Van Cleve, J., and E. Akçay. 2014. Pathways to social evolution: reciprocity, relatedness, and synergy. Evolution 68:2245–2258.

Wenseleers, T., A. Gardner, and K. R. Foster. 2010. Social evolution theory: a review of methods and approaches. In T. Székely, A. J. Moore, and J. Komdeur, eds., Social behaviour: genes, ecology and evolution, pages 132–158. Cambridge University Press, Cambridge.