

Appendix from J. A. Fletcher and M. Zwick, “Unifying the Theories of Inclusive Fitness and Reciprocal Altruism”

(Am. Nat., vol. 168, no. 2, p. 000)

Supporting Details

Here we provide additional information about the generalization of Hamilton’s rule (HR) and some of the mathematical details used in our models and analysis. Main headings here correspond to those in the text.

The Progressive Generalization of Hamilton’s Rule

Queller developed his version of HR after failures of Hamilton’s version (eq. [2]) were reported (Cavalli-Sforza and Feldman 1978; Boorman and Levitt 1980; Karlin and Matessi 1983). Queller (1985) concluded that these failures were due to either genotype/phenotype differences or nonadditive fitness functions. To accommodate differences in genotype and phenotype frequencies, he proposed equation (3). An earlier version of this equation proposed by Orlove and Wood (1978) and highlighted by Michod and Hamilton (1980) included the phenotype of actors and others in order to accommodate varying degrees of dominance in heterozygote individuals. In this earlier version, there was still a fixed (i.e., environment-independent) relationship between each diploid genotype and its corresponding phenotype. In contrast, Queller’s interpretation allows phenotypes to vary with local circumstances and thus accommodates conditional behavior as well as interactions among nonrelatives.

Hamilton’s Rule Applied to a Classic Reciprocal Altruism Model

First, we show how the change in fraction of TFT players (ΔQ) in figure 1 is calculated, given the starting parameter values. Then, we show how we apply Queller’s version of Hamilton’s rule (HR) to make accurate predictions about the same populations.

Calculating ΔQ for Figure 1

In a binomial distribution of a population of TFT and ALLD players, we can calculate the frequency of each type of pairing on the basis of the initial fraction of TFT in the population (Q) as follows:

$$\begin{aligned}f_{\text{TT}} &= Q^2, \\f_{\text{TA}} &= 2Q(1 - Q), \\f_{\text{AA}} &= (1 - Q)^2,\end{aligned}$$

where f_{TT} , f_{TA} , and f_{AA} represent the frequency of TFT-TFT, TFT-ALLD, and ALLD-ALLD pairings, respectively.

Let w_{CC} , w_{CD} , w_{DC} , and w_{DD} represent the four fitness payoff values of a two-player game matrix such that the payoff is for the player whose behavior is the first subscript and whose partner exhibits the behavior given by the second subscript. In terms of our fitness parameters (see table 2),

$$\begin{aligned}w_{\text{CC}} &= w_0 + b - c + d, \\w_{\text{CD}} &= w_0 - c, \\w_{\text{DC}} &= w_0 + b, \\w_{\text{DD}} &= w_0.\end{aligned}$$

Now, given our knowledge of the behaviors of these two strategies in each pairing type and the number of iterations (i), we can calculate the new fraction of TFT players (Q') after one generation. Because fitness payoffs are proportional to offspring number, Q' is the total expected payoff for TFT players divided by the total expected payoff for all players. (For convenience, nonoverlapping generations are assumed.) We scale these benefits by the fraction of each pairing type in both the numerator and denominator:

$$Q' = \frac{f_{\text{TFT}}(2iw_{\text{CC}}) + f_{\text{TAT}}[w_{\text{CD}} + (i-1)w_{\text{DD}}]}{f_{\text{TFT}}(2iw_{\text{CC}}) + f_{\text{TAT}}[w_{\text{CD}} + w_{\text{DC}} + 2(i-1)w_{\text{DD}}] + f_{\text{ATA}}(2iw_{\text{DD}})}. \quad (\text{A1})$$

Now, given Q , i , b , c , and w_0 (and d in cases where fitness is nonadditive), we can calculate $\Delta Q = Q' - Q$. For figure 1, fitness is additive ($d = 0$). Note also that the base fitness value (w_0) will factor out of all right-hand-side terms and cancel in equation (A1); thus, the change in Q is independent of w_0 .

Applying Hamilton's Rule

To check that Queller's version of HR works, we want to know whether this inequality is satisfied when $\Delta Q > 0$ and not satisfied when $\Delta Q < 0$. For convenience, we rearrange equation (3) and solve it as an equality in order to determine the b value where Queller's version of HR predicts $\Delta Q = 0.0$. We call this b^* :

$$b^* = \frac{\text{Cov}(G_A, P_A)}{\text{Cov}(G_A, P_O)} c. \quad (\text{A2})$$

To calculate the covariance terms, we need G_A , P_O , and P_A values for different types of pairings. Table A1 gives these values.

Table A1
 G_A and predicted P_A and P_O values for different pairings lasting i iterations

| Actor (A) | Opponent (O) | G_A | P_A | P_O |
|-----------|--------------|-------|-------|-------|
| TFT | TFT | 1 | 1 | 1 |
| TFT | ALLD | 1 | $1/i$ | 0 |
| ALLD | TFT | 0 | 0 | $1/i$ |
| ALLD | ALLD | 0 | 0 | 0 |

In addition, we need the genotype average, which is Q , and the phenotype average, which is

$$\bar{P} = \frac{f_{\text{TFT}}2i + f_{\text{TAT}}}{2i}.$$

This is the fraction of C behaviors over the total number of behaviors and is compared with both P_O and P_A to calculate covariance. Then, expanding equation (A2) by using the definition of covariance and grouping the covariance summation by the frequency of pairings gives

$$b^* = \frac{f_{\text{TFT}}2(1-Q)(1-\bar{P}) + f_{\text{TAT}}[(1-Q)(1/i-\bar{P}) + (0-Q)(0-\bar{P})] + f_{\text{ATA}}2(0-Q)(0-\bar{P})}{f_{\text{TFT}}2(1-Q)(1-\bar{P}) + f_{\text{TAT}}[(1-Q)(0-\bar{P}) + (0-Q)(1/i-\bar{P})] + f_{\text{ATA}}2(0-Q)(0-\bar{P})} c, \quad (\text{A3})$$

where each pair of parentheses in the numerator represents $(G_A - Q)(P_A - \bar{P})$ and each pair of parentheses in the denominator represents $(G_A - Q)(P_O - \bar{P})$, given the different circumstances in different pairings.

Table A2 gives the b^* values calculated using equation (A3) and the same parameters used in figure 1. Notice that Queller's version of HR exactly predicts this “tipping point” in the conditions under which TFT can increase or not in each population. The corresponding values in table A2 are indicated in figure 1 with arrows.

Table A2

Equilibrium benefit level b^*
(where $\Delta Q = 0.0$) as predicted
by Queller's version of HR for
the examples shown in figure 1

| Q | i | b^* |
|-----|-----|-----------|
| .15 | 15 | 1.476190 |
| .15 | 4 | 3.222222 |
| .20 | 2 | 6.000000 |
| .15 | 2 | 7.666666 |
| .10 | 2 | 11.000000 |

Iterations and Synergy

Here we show in general how the original fitness parameters (unprimed) are related to the resulting fitness parameters (primed) in the resulting cumulative payoff matrix (table 3) after i iterations of pairings in a population of TFT and ALLD players. Using similar notation as above for the cumulative payoff matrix and starting with the ALLD-ALLD pairing, we find the following:

$$w_{AA} = w'_0 = iw_{DD} = iw_0.$$

Now, given that TFT cooperates in the first iteration,

$$w_{TA} = w'_0 - c' = w_{CD} + (i - 1)w_{DD} = w_0 - c + iw_0 - w_0,$$

and similarly,

$$w_{AT} = w'_0 + b' = w_{DC} + (i - 1)w_{DD} = w_0 + b + iw_0 - w_0.$$

Therefore, b and c remain unchanged ($b' = b$ and $c' = c$), whereas $w'_0 = iw_0$. Finally,

$$w_{TT} = w'_0 + b' - c' + d' = iw_{CC} = i(w_0 + b - c + d).$$

Substituting for w'_0 , b' , and c' and subtracting iw_{CC} from both sides give

$$b - c + d' = i(b - c + d), \text{ or}$$

$$d' = i(b - c + d) - (b - c).$$

Similar to the calculation of b^* above, given a binomial distribution and known values of all the fitness parameters, we can find the equilibrium TFT proportion (Q^*) where $\Delta Q = 0$. Given $i = 10$, $b = 4$, $c = 1$, and $d = 0$, $Q^* = 0.037037$ or $1/27$. Now, if we analyze the same situation but assume $d' = 27$ and $i = 1$ (therefore $G_A = P_A$, $G_O = P_O$), the predicted equilibrium point is the same: $Q^* = 1/27$. This value holds for all benefit levels. (Of course, given that there is no mutation in this model, $Q = 1.0$ and $Q = 0.0$ are also equilibrium points for all b .)

Hamilton's Rule Applied to Cooperation across Species

We calculate Q'_1 and Q'_2 in a similar way as above, where pairings are strictly heterospecific. For instance, in the case where the cooperative type is TFT in both species, the pairings are TFT_1 - TFT_2 , TFT_1 - $ALLD_2$, $ALLD_1$ - TFT_2 , and $ALLD_1$ - $ALLD_2$, where subscripts indicate species. Again, assuming random binomial pairings, the frequency of these pairings can be calculated from Q_1 and Q_2 and the frequency of behaviors predicted for each type in each pairing. To generate figure 3, we simply iterate by feeding the new Q' values (as Q) into calculations for

the next generation. We also have to adjust this technique for different types of cooperative strategies, such as TF2T, Pavlov, and ALLC.

Calculating whether Queller's version of HR is satisfied, each generation is also done similarly as above, where when focal actors are from one species, others are always from the other species, and different instances of HR (HR_1 and HR_2) are calculated for each species. Again, the different types of cooperating strategies must be accommodated when calculating the phenotype of actors and others and the overall average phenotype for each type of pairing.

Additional Literature Cited in Appendix

- Boorman, S. A., and P. R. Levitt. 1980. The genetics of altruism. Academic Press, New York.
- Cavalli-Sforza, L. L., and M. W. Feldman. 1978. Darwinian selection and "altruism." *Journal of Theoretical Biology* 14:263–281.
- Karlin, S., and C. Matessi. 1983. The eleventh R. A. Fisher memorial lecture: kin selection and altruism. *Proceedings of the Royal Society of London B* 219:327–353.
- Michod, R. E., and W. D. Hamilton. 1980. Coefficients of relatedness in sociobiology. *Nature* 288:694–697.
- Orlove, M. J., and C. L. Wood. 1978. Coefficients of relationship and coefficients of relatedness in kin selection: a covariance form for the RHO formula. *Journal of Theoretical Biology* 73:679–686.