Electronic Supplementary Material for "The evolution of payoff matrices: providing incentives to cooperate"

Erol Akçay[1,*] and Joan Roughgarden[2]

September 30, 2010

Author affiliations

[1]: National Institute for Mathematical and Biological Synthesis (NIMBioS), 1534 White Avenue, Knoxville, TN 37996

[2]: Department of Biology, Stanford University, 371 Serra Mall, Stanford, CA 94305

[*]: Corresponding author, email: erol@nimbios.org

1 The recursion equation

With the definitions given in the main text, we can write down the recursion equations for the allele frequencies in the population. Denote the frequency of allele A with q, and assume that mating is random in the population. Thus, we have, for the frequencies f_i of the three genotypes: $f_{\mathcal{A}} = q^2$, $f_{\mathcal{H}} = 2q(1-q)$, $f_{\mathcal{B}} = (1-q)^2$. Assuming a well-mixed population where individuals are matched at random, the expected contribution of A alleles to the next generation by an AA homozygote (i.e.

genotype $\mathcal{A}$) can be written as:

$$\frac{1}{2}f_{\mathcal{A}}\Big[\Omega_{\mathcal{A}\mathcal{A},1}+\Omega_{\mathcal{A}\mathcal{A},2}\Big]+f_{\mathcal{H}}\Big[\rho_{\mathcal{A}\mathcal{H}}\Omega_{\mathcal{A}\mathcal{H},1}+(1-\rho_{\mathcal{A}\mathcal{H}})\Omega_{\mathcal{H}\mathcal{A},2}\Big]+f_{\mathcal{B}}\Big[\rho_{\mathcal{A}\mathcal{B}}\Omega_{\mathcal{A}\mathcal{B},1}+(1-\rho_{\mathcal{A}\mathcal{B}})\Omega_{\mathcal{B}\mathcal{A},2}\Big] \tag{1}$$

Similarly, we have for the contribution of A alleles from an AB heterozygote (genotype $\mathcal{H}$):

$$\begin{aligned}\frac{1}{2}\Bigg[&f_{\mathcal{A}}\Big[\rho_{\mathcal{A}\mathcal{H}}\Omega_{\mathcal{A}\mathcal{H},2}+(1-\rho_{\mathcal{A}\mathcal{H}})\Omega_{\mathcal{H}\mathcal{A},1}\Big]+\frac{1}{2}f_{\mathcal{H}}\Big[\Omega_{\mathcal{H}\mathcal{H},1}+\Omega_{\mathcal{H}\mathcal{H},2}\Big]\\&+f_{\mathcal{B}}\Big[\rho_{\mathcal{H}\mathcal{B}}\Omega_{\mathcal{H}\mathcal{B},1}+(1-\rho_{\mathcal{A}\mathcal{B}})\Omega_{\mathcal{B}\mathcal{H},2}\Big]\Bigg]\end{aligned}$$

Thus, we can write for the frequency of A alleles in the next generation, q':

$$\begin{aligned}\bar{w}p' =&f_{\mathcal{A}}\Bigg[\frac{1}{2}f_{\mathcal{A}}\Big[\Omega_{\mathcal{A}\mathcal{A},1}+\Omega_{\mathcal{A}\mathcal{A},2}\Big]+f_{\mathcal{H}}\Big[\rho_{\mathcal{A}\mathcal{H}}\Omega_{\mathcal{A}\mathcal{H},1}+(1-\rho_{\mathcal{A}\mathcal{H}})\Omega_{\mathcal{H}\mathcal{A},2}\Big]\\&+f_{\mathcal{B}}\Big[\rho_{\mathcal{A}\mathcal{B}}\Omega_{\mathcal{A}\mathcal{B},1}+(1-\rho_{\mathcal{A}\mathcal{B}})\Omega_{\mathcal{B}\mathcal{A},2}\Big]\Bigg]\\&+\frac{1}{2}f_{\mathcal{H}}\Bigg[f_{\mathcal{A}}\Big[\rho_{\mathcal{A}\mathcal{H}}\Omega_{\mathcal{A}\mathcal{H},2}+(1-\rho_{\mathcal{A}\mathcal{H}})\Omega_{\mathcal{H}\mathcal{A},1}\Big]+\frac{1}{2}f_{\mathcal{H}}\Big[\Omega_{\mathcal{H}\mathcal{H},1}+\Omega_{\mathcal{H}\mathcal{H},2}\Big]\\&+f_{\mathcal{B}}\Big[\rho_{\mathcal{H}\mathcal{B}}\Omega_{\mathcal{H}\mathcal{B},1}+(1-\rho_{\mathcal{H}\mathcal{B}})\Omega_{\mathcal{B}\mathcal{H},2}\Big]\Bigg],\end{aligned} \tag{2}$$

where $\bar{w}$ denotes the mean fitness of the population:

$$\begin{aligned}\bar{w} =&f_{\mathcal{A}}\left[\frac{1}{2}f_{\mathcal{A}}\Big[\Omega_{\mathcal{A}\mathcal{A},1}+\Omega_{\mathcal{A}\mathcal{A},2}\Big]+f_{\mathcal{H}}\Big[\rho_{\mathcal{A}\mathcal{H}}\Omega_{\mathcal{A}\mathcal{H},1}+(1-\rho_{\mathcal{A}\mathcal{H}})\Omega_{\mathcal{H}\mathcal{A},2}\Big]+f_{\mathcal{B}}\Big[\rho_{\mathcal{A}\mathcal{B}}\Omega_{\mathcal{A}\mathcal{B},1}+(1-\rho_{\mathcal{A}\mathcal{B}})\Omega_{\mathcal{B}\mathcal{A},2}\Big]\right]\\&+f_{\mathcal{H}}\left[f_{\mathcal{A}}\Big[\rho_{\mathcal{A}\mathcal{H}}\Omega_{\mathcal{A}\mathcal{H},2}+(1-\rho_{\mathcal{A}\mathcal{H}})\Omega_{\mathcal{H}\mathcal{A},1}\Big]+\frac{1}{2}f_{\mathcal{H}}\Big[\Omega_{\mathcal{H}\mathcal{H},1}+\Omega_{\mathcal{H}\mathcal{H},2}\Big]+f_{\mathcal{B}}\Big[\rho_{\mathcal{H}\mathcal{B}}\Omega_{\mathcal{H}\mathcal{B},1}+(1-\rho_{\mathcal{H}\mathcal{B}})\Omega_{\mathcal{B}\mathcal{H},2}\Big]\right]\\&+f_{\mathcal{B}}\left[f_{\mathcal{A}}\Big[\rho_{\mathcal{A}\mathcal{B}}\Omega_{\mathcal{A}\mathcal{B},2}+(1-\rho_{\mathcal{A}\mathcal{B}})\Omega_{\mathcal{B}\mathcal{A},1}\Big]+f_{\mathcal{H}}\Big[\rho_{\mathcal{H}\mathcal{B}}\Omega_{\mathcal{H}\mathcal{B},2}+(1-\rho_{\mathcal{H}\mathcal{B}})\Omega_{\mathcal{B}\mathcal{H},1}\Big]+\frac{1}{2}f_{\mathcal{B}}\Big[\Omega_{\mathcal{B}\mathcal{B},1}+\Omega_{\mathcal{B}\mathcal{B},2}\Big]\right]\end{aligned} \tag{3}$$

Equations (2) and (3) indicate that the evolutionary dynamics of the two alleles are determined by the sums $(\rho_{ij}\Omega_{ij,1} + (1-\rho_{ij})\Omega_{ji,2})$. We can therefore define $\alpha_{ij} = (\rho_{ij}\Omega_{ij,1} + (1-\rho_{ij})\Omega_{ji,2})$ as the interaction coefficient between genotypes i and j. Thus, our model of payoff matrix evolution becomes equivalent to the "pairwise interaction model" of frequency dependent selection (**?**). The interaction coefficients α_{ij} represent the expected payoff of genotype i from a pairing with genotype j, weighted by the probabilities of assuming either role. We can then write the recursion equation in terms of the interaction coefficients α_{ij}:

$$\bar{w}q' = f_{\mathcal{A}}\left[f_{\mathcal{A}}\alpha_{\mathcal{A}\mathcal{A}} + f_{\mathcal{H}}\alpha_{\mathcal{A}\mathcal{H}} + f_{\mathcal{B}}\alpha_{\mathcal{A}\mathcal{B}}\right] + \frac{1}{2}f_{\mathcal{H}}\left[f_{\mathcal{A}}\alpha_{\mathcal{H}\mathcal{A}} + f_{\mathcal{H}}\alpha_{\mathcal{H}\mathcal{H}} + f_{\mathcal{B}}\alpha_{\mathcal{H}\mathcal{B}}\right]. \tag{4}$$

2 Multi-locus model

To see how robust our polymorphism result is to the assumption of a single locus, we constructed a diploid model with multiple loci with additive effects of each locus on the side-payment and recombination between loci. It is notoriously difficult to analyze such models analytically, and hence we explored the behavior of this model numerically (the Mathematica code we used is provided as supplementary material).

Briefly, the model considers a n loci that have two alleles each, labeled A and B, and assume that each A allele contributes an fixed sidepayment σ_A to the payoff of the opponent at (D, C) and causes its carrier to incur a cost $\chi_A = \sigma_A$. Individuals are matched randomly, and after playing the game, each individual produces gametes in proportion to their payoff. At this stage, recombination takes place with probability r, and a cross-over point is chosen assuming that each locus is equidistant from its neighbors. Gametes than fuse randomly to produce the individuals of the next generation.

Our results can be summarized as follows: We can distinguish between two cases, depending on whether the minimum side-payment to switch the NE from DD to DC requires an even or odd number of A alleles. Label this minimum number of alleles as n_A In the first case, a homozygote individual can achieve the minimum side-payment required and pay the least cost possible. Under

such conditions, our simulations show that one of the haplotypes with the allele A at $n_A/2$ positions will fix in the population, and hence, no polymorphism will be maintained. On the other hand, if $n_A/2$ is odd, there is no such genotype. Consequently, our simulations show that the population will reach a stable polymorphism mainly between two haplotypes, one having A at $(n_A + 1)/2$ positions and the other at $(n_A - 1)/2$ positions. In these cases, the pattern of variation seen in the one-locus case will reappear: since the haplotypes differ in a single locus, the polymorphism is effectively a single-locus one.

Interestingly, high recombination between loci promotes convergence to the stable outcome. When recombination is low, convergence to the equilibrium happens quite slowly, and a number of haplotypes are retained in the population for a long time. In the case with no recombination, many haplotypes are retained in the population. This is unsurprising, because all loci have equal effects on the game, and hence two haplotypes that have the same number of A alleles have the same phenotype. Therefore, they are selectively neutral against each other. Adding recombination, however, means that the haplotypes get broken down, and in the end, only one of them survives the combined effect of selection and recombination.

Overall, these results suggest that the polymorphism does not disappear entirely with multiple loci, but its generality needs to be qualified. As the number of loci grows large, the available genetic variation in the side-payment will be better approximated by a continuous variable, and hence it will be increasingly likely that the homozygote of a haplotype will make the minimum side-payment required to shift the NE, and will consequently go to fixation. However, when the number of loci is not too big, there will be potential for genetic polymorphism in the payoff matrix, and all the associated effects on behavioral diversity and variation in actual payoffs is expected to be observed.

3 Mutations with negative incentives

The analysis in the main text is predicated on a particular type of mutant that provides a positive incentive to its opponent for switching its behaviour. However, mutants can also arise that change the payoff matrix in other ways, for example those that provide negative incentives by depressing their opponents' payoff at the (D, D) outcome. Suppose that such a mutant depresses the payoff of playing D by μ, paying again an unconditional cost $\chi(\mu) > 0$. Then the game $G_{\mathcal{BH}}$ between the resident homozygotes and the mutant heterozygote is the following:

$$G_{\mathcal{BH}}: \qquad \begin{array}{cc} & \text{Role 2} \\ & \begin{array}{cc} C & D \end{array} \\ \text{Role 1} \begin{array}{c} C \\ D \end{array} & \begin{array}{cc} r, r-\chi(\mu) & s, t-\chi(\mu) \\ t, s-\chi(\mu) & p-\mu, p-\chi(\mu) \end{array} \end{array} .$$

As in the previous case, the only way such a mutant can invade is when $\mu > p-s$, and $\chi(\mu) < t-p$. If invasion is successful, the game between heterozygotes again becomes a Hawk-Dove game with two alternative NE. The difference between this case and the invasion of positive incentives is that the length of the Pareto boundary, i.e. the difference between the best and worst efficient outcome for a given player, is not reduced as a result of such negative incentives. In other words, the overall magnitude of conflict is not reduced with invasion of such mutants, even though they lead the population away from a prisoners' Dilemma game. However, there is still local alignment of interest: both NE are preferred by both players to any unilateral deviation by either of them. The results with respect to the polymorphisms are unchanged: the homozygote mutant again results in no additional change in the NE, while paying a higher cost. Hence, the heterozygote enjoys an advantage.

4 The asymmetric prisoners' dilemma game

We briefly consider the case of a resident game that is asymmetric and illustrate that such games can in fact be invaded and swept out of the population by certain mutants. As an example, suppose that the game $G_{\mathcal{BB}}$ is given by:

$$G_{\mathcal{BB}}: \qquad \begin{array}{cc|cc} & & \multicolumn{2}{c}{\text{Role 2}} \\ & & C & D \\ \hline \text{Role 1} & C & 3,3 & 0,4.5 \\ & D & 5,0 & 1,0.5 \end{array} .$$

Suppose that the mutant allele A is characterized by $\sigma = \chi(\sigma) = \frac{3}{4}$. Now, the NE of the game $G_{\mathcal{HB}}$ is shifted to (D, C), but the sidepayment is not enough to shift the NE of the game $G_{\mathcal{BH}}$, which stays at (D, D). Thus, the interaction coefficient $\alpha_{\mathcal{HB}} = \frac{1}{2}(4.25 + 0.25) = 2.25$, whereas $\alpha_{\mathcal{BB}} = 0.75$. Thus, the mutant allele will invade. The difference with the case above, however, is that there is still scope for improvement when the mutant is in Role 2. This scope is taken advantage of by the AA homozygote, which makes a sidepayment of 1.5, and thus can shift the NE in games where it is in Role 2.

For the general asymmetric prisoners' dilemma, denote the payoffs to Role 1 player with the same symbols as in (7) in the main text, and the payoffs to Role 2 player with primed versions (i.e. the payoffs for the outcome (C, D) are (s, t'); for (D, D), (p, p'), and so on). Also, assume (without loss of generality) that $p' - s' < p - s$, so that it takes a smaller sidepayment to shift the NE when in Role 1. One can show that the conditions the allele A has to satisfy for being able to both invade the population fixed for B and subsequently go to fixation are:

$$\begin{aligned} p - s > \sigma \quad &> \max\Big[p' - s', \tfrac{p-s}{2}\Big] \\ \min\Big[\frac{t-p}{2}, \frac{t'-s'-2\sigma}{4}\Big] \quad &> \chi(\sigma)\,. \end{aligned} \tag{5}$$

The first condition imposes both a lower and an upper limit to the sidepayment σ: if σ is too low, it will fail to shift the NE even in Role 1 and thus the mutant will pay a cost without receiving a benefit, and cannot invade. If σ is too high on the other hand, the mutant heterozygote shifts the NE in both roles and therefore, the homozygote only adds to the cost that the individual incurs without additional benefits. The second condition signifies that the cost χ needs to be low enough such that the mutant ends up benefiting from single shifts of NE.