An Appendix to “**The Evolution of Cooperation**”

Arnon Lotem, Michael A. Fishman and Lewi Stone.

Nature 1999, **400**:226-227

How phenotypic defection stabilises indirect reciprocity:

To explore our simulation results from a theoretical perspective we proceed as follows:

In Appendix A we present a basic game theory model which assumes that opponents are able to classify one another correctly. We show that in the absence of phenotypic defectors, the model has no ESS solution, but that in the presence of phenotypic defectors a cooperative ESS (SA) is achieved.

In Appendix B we now extend the original model to account for the possibility of uncertainty in the classification of opponents. We focus on the relevant case in which SA players misclassify other SA players as defectors and *thus fail to provide help.* We suppose that correct classification occurs with probability p. These “type 1” mistakes occur (with probability (1-p)) because SA players are capable of accumulating negative image as a result of past interactions with defectors.

In Appendix C we consider a more complex scheme for characterising uncertainty in the classification of opponents. Now we are also concerned with “type 2” mistakes which occur in the absence of information of the opponents’ image (e.g. in the first round of each generation). We assume that in such situations (eg., when the opponent’s image is “0” in the N&S model), SA players provide help, and thus may *erroneously help defectors* (as frequently happens in N&S simulation).

In Appendix D we take an alternative route to illustrate the stabilising effect of phenotypic defection. We modify N&S analytical model[2] and show that in the presence of phenotypic defectors there will be a stable cooperation by selective altruists.

General methods for the game theory model (Appendix A-C) are detailed in Appendix F.

APPENDIX A – The basic model

Similar to Nowak & Sigmund[1,2] consider three strategy types: **U**nconditional **A**ltruists (**UA**) who always cooperate; **U**nconditional **D**efector (**UD**) who never cooperate; **S**elective **A**ltruists/Discriminators (**SA**) who help other cooperators, but not defectors . We will show that **SA** is the single Evolutionary Stable Strategy (ESS) and occurs only if there are phenotypic defectors in the game. To simplify our calculations it is convenient to assume that interactions are based on a correct classification of opponents, and that the number of interactions is sufficient to provide information for such classification. (Indeed if *indirect reciprocity* plays a role in the lives of social animals, classification mechanisms should be subject to selection, and thus improves the probability of correct classification.) Nevertheless, we emphasise that very similar results emerge even when these assumptions are considerably relaxed and there are uncertainties in classification (see appendices B&C).

Denoting the (*per capita*) benefits over a lifetime by *B*, and the lifelong costs by *C* , we have the following payoff matrix:

$$\begin{array}{c c} & \begin{array}{ccc} \textbf{UA} & \textbf{UD} & \textbf{SA} \end{array} \\ \begin{array}{c} \textbf{UA} \\ \textbf{UD} \\ \textbf{SA} \end{array} & \begin{pmatrix} B-C & -C & B-C \\ B & 0 & 0 \\ B-C & 0 & B-C \end{pmatrix} \end{array} \tag{A.1}$$

For the case $B > C$ (as in N&S), system (A.1) has no ESS solutions (see analysis below).

Assume now that phenotypic defectors with the frequency of $0 < q < 1$ enter the game. The payoff matrix is then:

$$\begin{array}{c c} & \begin{array}{ccc} \textbf{UA} & \textbf{UD} & \textbf{SA} \end{array} \\ \begin{array}{c} \textbf{UA} \\ \textbf{UD} \\ \textbf{SA} \end{array} & \begin{pmatrix} (1-q)B-C & -C & (1-q)B-C \\ (1-q)B & 0 & 0 \\ (1-q)(B-C) & 0 & (1-q)(B-C) \end{pmatrix} \end{array} \tag{A.2}$$

UA is *strictly dominated* by **SA** (every term in row-3 is greater than its corresponding term in row-1), implying that **UA** will disappear from the population (Weibull 1996)[3]. **UA** can thus be neglected, leaving the following payoff matrix:

$$\begin{array}{c c} & \begin{array}{cc} \textbf{UD} & \textbf{SA} \end{array} \\ \begin{array}{c} \textbf{UD} \\ \textbf{SA} \end{array} & \begin{pmatrix} 0 & 0 \\ 0 & (1-q)(B-C) \end{pmatrix} \end{array} \tag{A.3}$$

Denoting the frequency of **UA**, **UD**, and **SA** as x_1, x_2, and x_3 respectively, the replicator equation for system (1.3) is:

$$x_3' = (1-q)(B-C)(1-x_3)x_3$$

implying

$$x_3(t) \xrightarrow[t\to\infty]{} 1 \text{ whenever } x_3(0) > 0.$$

Since $B > C$, system (2.2) has a single globally attractive ESS solution **SA**.

No ESS in the absence of phenotypic defectors (system A.1):

Here we show that in the absence of phenotypic defectors (D=0), the model described in system A.1. has no ESS solutions (see Appendix D for a description of general methods).

Let us denote the frequency of **UA**, **UD**, and **SA** by x_1, x_2, and x_3 respectively.
The replicator equations for the system with payoff matrix P given by 1.1 are:

$$x_1^{'} = f_1(\mathbf{x}) \equiv (\mathrm{e}_1 - \mathbf{x}^{\mathrm{t}}) \circ \mathrm{P} \circ \mathbf{x} = [(B-C)x_3 - C](1 - x_1 - x_3)x_1$$

$$x_2^{'} = -x_1^{'} - x_3^{'}$$

$$x_3^{'} = f_3(\mathbf{x}) \equiv (\mathrm{e}_3 - \mathbf{x}^{\mathrm{t}}) \circ \mathrm{P} \circ \mathbf{x} = (B-C)(1 - x_1 - x_3)x_3^2$$

where $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is the standard basis of $\Re^3$.
This set of replicator equations has two stationary potential ESS solutions (obtained as described in Appendix D):
a point solution $\mathbf{x}_1 = \mathbf{e}_2$, and a set solution $\mathbf{x}_\xi \in \Omega = \{\xi\, \mathbf{e}_1 + (1-\xi)\, \mathbf{e}_3 \mid 0 \le \xi \le 1\}$.
We test these stationary solutions using the ESS criterion (D1, of Appendix D). One sees that:

$$\mathbf{x}_1 \neq \mathrm{e}_3 : \Gamma_0(\mathbf{x}_1, \mathrm{e}_3) = 0, \text{and } \Gamma_1(\mathbf{x}_1, \mathrm{e}_3) = -(B-C) \quad \text{(A.4)}$$

Hence $\mathbf{x} = \mathbf{x}_1$ is not an ESS.

To test the set solution Ω, choose:

$$\mathbf{x}^{\#} = \frac{2B-C}{2B}\mathbf{e}_1 + \frac{C}{2B}\mathbf{e}_3 \in \Omega \;, \mathbf{e}_2 \notin \Omega$$
$$\text{and } \Gamma_0(\mathbf{x}^{\#}, \mathbf{e}_2) = -\frac{C}{2} < 0 \quad \text{(A.5)}$$

Hence Ω is not an Evolutionary Stable set.

We conclude that in the absence of phenotypic defectors the game has no ESS solution.

APPENDIX B

We now consider our original model (Appendix A) when the assumption of correct classification of opponents is relaxed. Recall that in the N&S model, the level of uncertainty in the classification of opponents is an outcome of the specific "image scoring” mechanism assumed. However, rather than postulating a specific mechanism that produces some level of mistakes in classification, here we take an alternative approach which directly introduces the probability of such mistakes as a parameter in the model.

Among the three strategies, only selective altrusits (SA) base their responses on classification of opponents and therefore they are the only players that are likely to make mistakes. Hence we first consider the possibility that SA players will not help those SA players who they have erroneously misclassified as defectors. This “type 1” mistake will happen when SA players accumulate negative image as a result of past interactions with defectors (with real genetic defectors, phenotypic defectors, or other SA players perceived as defectors).

Let us denote the probability of correct classification of a **SA** by another **SA** by $0 < p \leq 1$. The (*per capita*) benefits of receiving help over a lifetime by *B*, and the lifelong costs of donating favours by $C < B$. In these terms we have the following payoff matrix.

$$\begin{array}{c c} & \begin{array}{ccc} \textbf{UA} & \textbf{UD} & \textbf{SA} \end{array} \\ \begin{array}{c} \textbf{UA} \\ \textbf{UD} \\ \textbf{SA} \end{array} & \begin{pmatrix} B-C & -C & B-C \\ B & 0 & 0 \\ B-C & 0 & p(B-C) \end{pmatrix} \end{array} \tag{B.1}$$

Proceeding as in Appendix A, we can show that *system (B.1) does not have any ESS solutions*.

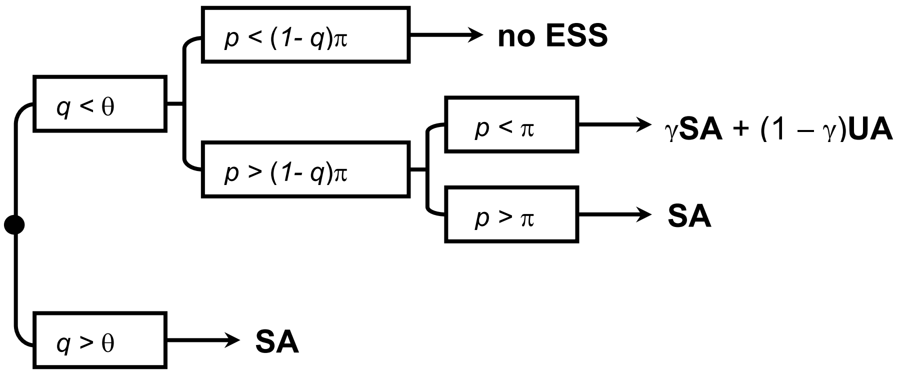
Now let us introduce phenotypic defectors at frequency $0 < q < 1$. Then we have a 3×3 interaction matrix, P—as given by (B.1), for the three strategies plus a 3×1 matrix, P’, for their interactions with phenotypic defectors. And therefore we have the payoff function:

$$(1-q)\left[\mathbf{x} \circ \mathrm{P} \circ \mathbf{x}^{\mathrm{t}}\right] + q\left[\mathbf{x} \circ \mathrm{P}'\right] = \mathbf{x} \circ \left[(1-\mathrm{q})\mathrm{P}_{*1} + q\mathrm{P}', \ldots, (1-q)\mathrm{P}_{*1} + q\mathrm{P}'\right] \circ \mathbf{x}^{\mathrm{t}} =$$

$$\mathbf{x} \circ [(1-\mathrm{q})\begin{pmatrix} \mathrm{B-C} \\ \mathrm{B} \\ \mathrm{B-C} \end{pmatrix} + q\begin{pmatrix} \mathrm{-C} \\ 0 \\ 0 \end{pmatrix}, \ldots, (1-\mathrm{q})\begin{pmatrix} \mathrm{B-C} \\ 0 \\ \mathrm{p(B-C)} \end{pmatrix} + q\begin{pmatrix} \mathrm{-C} \\ 0 \\ 0 \end{pmatrix}] \circ \mathbf{x}^{\mathrm{t}} = \mathbf{x} \circ \overline{\mathrm{P}} \circ \mathbf{x}^{\mathrm{t}} \tag{B.2}$$

$$\text{where } \overline{\mathrm{P}} = \begin{array}{c c} & \begin{array}{ccc} \textbf{UA} & \textbf{UD} & \textbf{SA} \end{array} \\ \begin{array}{c} \textbf{UA} \\ \textbf{UD} \\ \textbf{SA} \end{array} & \begin{pmatrix} (1-q)B-C & -C & (1-q)B-C \\ (1-q)B & 0 & 0 \\ (1-q)(B-C) & 0 & p(1-q)(B-C) \end{pmatrix} \end{array}$$

Using the methods described in Appendix D we can show that the properties of system (B.2) depend on the magnitudes of *p* and *q*. (Full details are available upon request. Also see Appendix C, in which we provide a full analysis of a more general model.) The following scheme summarises the results for all possible values of the parameters. Here endpoints represent the ESS associated with the region of the parameter space described by the inequalities on the path leading from the origin (•) to that endpoint.



Here $\theta = \dfrac{B-C}{B}$, $\pi = \dfrac{(1-q)B-C}{(1-q)(B-C)} > 0 \Leftrightarrow q < \theta$ and $\gamma = \dfrac{qC}{(1-p)(1-q)(B-C)}$

In conclusion, while in the absence of phenotypic defectors there is no stable solution to the model, the *introduction of phenotypic defectors can, under some conditions, stabilise a population of **S**elective **A**ltruists*.

We briefly comment on the two conditions which lead to an ESS of SA players in this model, as specified in the above diagram.

The first condition ($q > \theta$) requires that either the frequency of phenotypic defectors (q) or the cost of cooperation (C) will be relatively large. In simple terms, when q is large, SA players can do better than UA players by saving the accumulated small costs when refusing to help many phenotypic defectors. When C is large SA players have the advantage over UA since they save high costs by refusing to help phenotypic defectors.

When q is small ($q < \theta$), the second condition requires ($p > \pi$) i.e., the probability (p) of correct classification of a SA by another SA will be sufficiently high (e.g. when B =1, C = 0.35, q = 0.2, then p > 0.87).

Note that the first condition above is inconsistent with our simulation results. It predicts a stable cooperation by selective altruists even when the cost of cooperation is high. Our simulation, on the other hand, failed to produce cooperation when the cost of cooperation was high, no matter how many phenotypic defectors we added (see Figure 1e). The reason for this discrepancy is that the model above ignores a second type of mistake in opponents' classification, which was in fact common in the simulation: in the absence of information on opponents' image (e.g. in the first round of each generation), discriminate altruists (k=0) occasionally provided help to defectors. In Appendix C below we shell extend the above model to account also for this type of mistakes.

Appendix C

We now extend the model described in Appendix B to consider "type 2 mistakes" which may occur when SA players have no information about their opponents' image (e.g. first round in a model of non-overlapping generations, where all players have image of "0" at birth). We assume that in situations where information about image is lacking, SA players provide help, and thus may *erroneously help defectors* (as frequently occurs in the N&S simulation model).

As in Appendix B, we denote $0 < p < 1$ as the probability of **SA** classifying correctly another **SA** and thus helping him. In addition, we denote $0 < r < 1$ as the probability of **SA** helping a defector. We now proceed with the analysis.

I) ***No phenotypic defectors***.

In these terms the payoff matrix is given by:

$$\begin{array}{cc} & \begin{array}{ccc} \textbf{UA} & \textbf{UD} & \textbf{SA} \end{array} \\ \begin{array}{c} \textbf{UA} \\ \textbf{UD} \\ \textbf{SA} \end{array} & \begin{pmatrix} B-C & -C & B-C \\ B & 0 & rB \\ B-C & -rC & p(B-C) \end{pmatrix} \end{array} \qquad \text{(C.I.1)}$$

System (C.1) has five potential ESS solutions.

$$\mathbf{x}_1 = \mathbf{e}_1 \quad \mathbf{x}_2 = \mathbf{e}_2 \quad \mathbf{x}_3 = \mathbf{e}_3$$

$$\mathbf{x}_4 = (1-\alpha_0)\mathbf{e}_2 + \alpha_0 \mathbf{e}_3 \quad : \alpha_0 = \frac{rC}{(p-r)(B-C)}$$

$$\mathbf{x}_5 = (1-\beta_0-\chi_0)\mathbf{e}_1 + \beta_0\mathbf{e}_2 + \chi_0\mathbf{e}_3 \qquad \text{(C.I.2)}$$

$$\beta_0 = \frac{(1-p)(B-C)}{(1-r)^2 B} \text{ and } \chi_0 = \frac{C}{(1-r)B}$$

Using the methods described in Appendix D, we now examine which of these are ESS solutions:

$$\mathbf{x}_1 \neq \mathbf{e}_3 \text{ but } \Gamma_0(\mathbf{x}_1, \mathbf{e}_2) = -C \qquad \text{(C.I.3)}$$

Hence $\mathbf{x}_1$ is not an ESS.

$$\mathbf{x}_2 = \mathbf{e}_2$$

$$\Gamma_0(\mathbf{x}_2, \mathbf{x}) = C(x_1 + r x_3) \qquad \text{(C.I.4)}$$

and

$$\Gamma_1(\mathbf{x}_2, \mathbf{x}) = -(B-C)[(1-r)x_1 + (p-r)x_3]x_3$$

Hence $\mathbf{x}_2 \equiv$ **UD is an ESS whenever *r > 0.***

$$\mathbf{x}_3 \neq \mathbf{e}_1 \text{ but } \Gamma_0(\mathbf{x}_3, \mathbf{e}_1) = -(1-p)(B-C) \qquad \text{(C.I.5)}$$

Hence $\mathbf{x}_3$ is not an ESS.

$$\begin{gathered} \mathbf{x}_4 \neq \mathbf{e}_2 \\ \text{but} \\ \Gamma_0(\mathbf{x}_4, \mathbf{e}_2) = 0 \text{ and } \Gamma_1(\mathbf{x}_4, \mathbf{e}_2) = -rC\alpha_0 \end{gathered} \qquad \text{(C.I.6)}$$

Hence $\mathbf{x}_4$ is not an ESS.

$$\begin{gathered} \mathbf{x}_5 \neq (1-\chi_0)\mathbf{e}_1 + \chi_0 \mathbf{e}_3 \equiv \mathbf{x}_\chi \\ \\ \Gamma_0(\mathbf{x}_5, \mathbf{x}_\chi) = \Gamma_1(\mathbf{x}_5, \mathbf{x}_\chi) = 0 \end{gathered} \qquad \text{(C.I.6)}$$

Hence $\mathbf{x}_5$ is not an ESS.

In conclusion: In the *absence of phenotypic defectors*, and when both type-1 and type-2 mistakes are allowed for (r > 0; p < 1), *there is a single ESS solution* **UD** i.e., the frequency of defectors is unity. Since an ESS is an asymptotically stable equilibrium solution on the set **X**, and **X** is closed; by the corollary to the *Poincaré–Bendixon Theorem*, all solutions will converge to $\mathbf{e}_2$.

II With *phenotypic defectors*.

$$(1-q)\begin{pmatrix} B-C & -C & B-C \\ B & 0 & rB \\ B-C & -rC & p(B-C) \end{pmatrix} + q\begin{pmatrix} -C \\ 0 \\ -rC \end{pmatrix} =$$

$$\begin{pmatrix} (1-q)B-C & -C & (1-q)B-C \\ (1-q)B & 0 & (1-q)rB \\ (1-q)(B-C)-qrC & -rC & (1-q)p(B-C)-qrC \end{pmatrix} \qquad \text{(C.II.1)}$$

There are six potential ESS solutions to be checked using the methods described in Appendix D:

$$\begin{gathered} \mathbf{x}_1 = \mathbf{e}_1 \\ \\ \mathbf{x}_1 \neq \mathbf{e}_2 \text{ but } \Gamma_0(\mathbf{x}_1, \mathbf{e}_2) = -C \end{gathered} \qquad \text{(C.II.2)}$$

Hence $\mathbf{x}_1$ is not an ESS.

$$\mathbf{x}_2 = \mathbf{e}_2$$

$$\Gamma_0(\mathbf{x}_2, \mathbf{x}) = C(x_1 + rx_3) \qquad \text{(C.II.3)}$$

and

$$\Gamma_1(\mathbf{x}_2, \mathbf{x}) = -(1-q)(B-C)[(1-r)x_1 + (p-r)x_3]x_3$$

Hence $\mathbf{x}_2 \equiv$ **UD** is an ESS whenever $r > 0$.

$$\mathbf{x}_3 = \mathbf{e}_3$$

$$\Gamma_0(\mathbf{x}_3, \mathbf{x}) = qC(\rho_1 - r)x_1 + [B - q(B-C)](\rho_2 - r)x_2 \qquad \text{(C.II.4)}$$

$$\rho_1 = 1 - \frac{(1-p)(1-q)(B-C)}{qC} \text{ and } \rho_2 = \frac{(1-q)(B-C)}{C + (1-q)(B-C)} p$$

Hence $\mathbf{x}_3 \equiv$ **<u>SA is an ESS whenever r < min{ρ1, ρ2},</u>** or equivalently, whenever any of the following conditions holds:

$$\text{(i) } q < \theta, \pi < p < \pi\sigma, \text{ and } r < \rho_1$$

$$\text{(ii) } q < \theta, \pi\sigma < p, \text{ and } r < \rho_2$$

$$\text{(iii) } q > \theta \left(\pi\sigma < 0 < p\right) \text{ and } r < \rho_2$$

$$\theta = \frac{B-C}{B} \equiv \theta, \pi = \frac{(B-C) - qB}{(1-q)(B-C)}, \text{ and } \sigma = \frac{B - q(B-C)}{(1-q)B}$$

,,,,,,

$$\mathbf{x}_4 = (1-\alpha_q)\mathbf{e}_2 + \alpha_q \mathbf{e}_3 \quad : \alpha_q = \frac{rC}{(p-r)(1-q)(B-C)}$$

$$\mathbf{x}_4 \neq \mathbf{e}_2 \qquad \text{(C.II.5)}$$

but

$$\Gamma_0(\mathbf{x}_4, \mathbf{e}_2) = 0 \text{ and } \Gamma_1(\mathbf{x}_4, \mathbf{e}_2) = -rC\alpha_q$$

Hence $\mathbf{x}_4$ is not an ESS.

$$\mathbf{x}_5 = (1 - \beta_q - \chi_q)\mathbf{e}_1 + \beta_q \mathbf{e}_2 + \chi_q \mathbf{e}_3$$

$$\beta_q = \frac{(1-p)(B-C) - q(1-r)^2 B}{(1-r)^2(1-q)B} : \chi_q = \frac{C}{(1-r)(1-q)B} \qquad \text{(C.II.6)}$$

$$\mathbf{x}_5 \neq (1-\chi_q)\mathbf{e}_1 + \chi_q \mathbf{e}_3 \equiv \mathbf{x}_\chi$$

but

$$\Gamma_0(\mathbf{x}_5, \mathbf{x}_\chi) = \Gamma_1(\mathbf{x}_5, \mathbf{x}_\chi) = 0$$

Hence $\mathbf{x}_5$ is not an ESS.

,,,,,,,

$$\mathbf{x}_6 = (1-\eta)\mathbf{e}_1 + \eta\mathbf{e}_3 \quad : \eta = \frac{q(1-r)C}{(1-p)(1-q)(B-C)}$$

$$\Gamma_0(\mathbf{x}_6, \mathbf{x}) = \left[q(1-r)^2 B - (1-p)(B-C)\right]\frac{Cx_2}{(1-p)(B-C)} \qquad \text{(C.II.7a,b,c)}$$

$$\Gamma_0(\mathbf{x}_6, \mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = (1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_3 \equiv \mathbf{x}_\lambda$$

$$\Gamma_1(\mathbf{x}_6, \mathbf{x}_\lambda) = (1-p)(1-q)(B-C)(\eta-\lambda)^2$$

Hence $\mathbf{x}_6$ ≡ **UA** ⊕ **SA** <u>**is an ESS**</u> whenever any of the following conditions holds:

(⊕ denote a mixed ESS solution)

$$\text{(i)}\ q < \theta,\ (1-q)\pi < p < \pi,\ \text{and}\ \left(\rho_1 < 0 <\right) r < \rho_3 = 1 - \sqrt{\frac{(1-p)(B-C)}{qB}}$$

$$\text{(C.II.7d,e,f)}$$

$$\text{(ii)}\ q < \theta,\ \pi < p < \pi\sigma,\ \text{and}\ \rho_1 < r < \rho_3$$

In conclusion: Without phenotypic defectors, the only ESS solution is of a population of unconditional defectors (UD). However, the introduction of phenotypic defectors creates two additional ESS solutions that allow stable cooperation, depending on parameter values and initial conditions. In fact, a shift between two solutions (**UD** and **SA**) may be illustrated by our computer simulation in Figure 1b.

The analysis above presents some features that are qualitatively similar to our simulation results. First, under some conditions, the introduction of phenotypic defectors can stabilise a population of Selective Altruists. As in the earlier model of Appendix B, for SA to be an ESS, the probability of correct classification of SA by another SA has to be sufficiently high (under all three ESS conditions specified in system (C.II.4) above, p has to be greater than π, or greater than πσ). In addition, here, the probability of type-2 classification mistakes (that SA help UD) has to be lower than a critical value ($r<\rho_1$ or $r<\rho_2$).

These ESS conditions may account for the stability of discriminate altruism in our simulation. For example, under parameter values analogous to those we used in the simulation (B=1, C=0.35, q=0.2) SA will be an ESS if $p>0.94$ and $r<0.56$. In fact these are not unreasonable estimates of the type-1 and type-2 mistakes occurring in the simulation model where all players have zero image at the beginning of each generation. Discriminate altruists help all players in the first round, making the probability of an SA player with k=0 to develop a negative image (and consequently to be classified as a defector) very low. Hence the probability of an SA misclassifying an SA, $p > 0.94$ does not seem unreasonable for the simulation. Consider now type-2 errors: when the SA population is close to ESS i.e., the population mostly consist of discriminators some of which are phenotypic defectors. Hence the frequency r of helping a defector (type-2 mistake) will largely depend on and correspond to the frequency of phenotypic defectors present, which is $q=0.2$ in our simulation. This is well within the ESS condition $r<0.56$.

Finally, the fact that our simulations failed to produce cooperation when the cost of cooperation was high (which was inconsistent with the results of our earlier model in Appendix B), is predicted by the current model. This is simply because ρ_2 decreases when the cost of cooperation increases (see the definition of ρ_2 in the ESS conditions above). In simple terms, the increase in the cost of cooperation causes SA players to lose more each time they help a defector by mistake, and as a result, a smaller frequency of type-2 mistakes r can be tolerated.

APPENDIX D:

We follow Nowak & Sigmund analytical model[2], equations 26,27 (page 567):

$$\begin{aligned} P_1 &= bx_3 - 2c \\ P_2 &= 0 \\ P_3 &= bx_3 + cx_2 - 2c \end{aligned} \qquad \text{(NS:26,27)}$$

To introduce *phenotypic defectors* at frequency $0 < q < 1$, we carry out the following transformation $x_1 \to (1-q)y_1$, $x_2 \to (1-q)y_2 + q$, $x_3 \to (1-q)y_3$.

$$\begin{aligned} \overline{P}_1 &= (1-q)by_3 - 2c \\ \overline{P}_2 &= 0 \\ \overline{P}_3 &= (1-q)by_3 + (1-q)cy_2 - (2-q)c \end{aligned} \qquad (1.1)$$

Now $\overline{P}_3 - \overline{P}_1 = qc + (1-q)cy_2$ i.e., $\overline{P}_1 < \overline{P}_3$, or altruism is *strictly dominated* by selective altruism. Exclusion of the strictly dominated strategy (Weibull, 1996) yields a reduced system.

$$\begin{aligned} &\pi_2 = 0\,,\ \pi_3 = (1-q)(b-c)z - c \\ &y_3 = z \ \text{ and } \ y_2 = 1-z \end{aligned} \qquad (1.3)$$

System (1.3) yields the following replicator equations

$$\begin{aligned} &z' = (\pi_3 - \pi_2(1-z) - \pi_3 z)z = \pi_3(1-z)z = (1-q)(b-c)(z-\zeta)(1-z)z \\ &\zeta = \frac{c}{(1-q)(b-c)} \in (0,1) \Leftrightarrow 2c < b \ \text{ and } \ q < \frac{b-2c}{b-c} = \theta \end{aligned} \qquad (1.4)$$

Thus, since $c < b$, we have three possibilities.

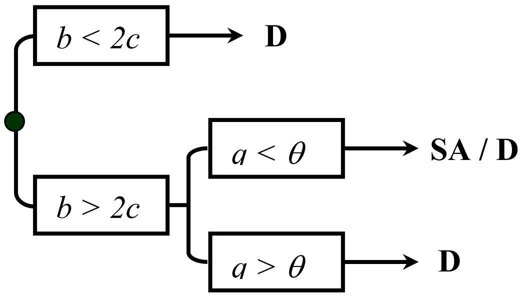
(i) $b < 2c \Rightarrow \theta < 0 < q \Rightarrow \zeta > 1 \Rightarrow z' < 0$, $\quad \forall z \in (0,1)$ i.e., system (1.4) converges to z = 0. Hence, system (1.4); and hence system (1.1); has a unique ESS solution **D** for defection.

(ii) If $b > 2c \Rightarrow 0 < \theta < 1$ and $q < \theta$, then

- (iia) $z(0) < \zeta$ implies convergence to z = 0 i.e., defection.
- (iia) $z(0) > \zeta$ implies convergence to z = 1 i.e., to an ESS solution **SA** for selective altruism.

(iii) If $b > 2c$ and $q > \theta$, then again defection is the only ESS.

These results can be summarized in the following scheme:



Appendix F: General methods

Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be the standard basis of $\Re^3$, and let us define the strategy set by

$$\mathbf{X} = \left\{ x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 \;\middle|\; 0 \le x_1, x_2, x_3 \le 1 \text{ and } x_1 + x_2 + x_3 = 1 \right\}$$

That is, $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ will be used to denote *pure strategies* **UA**, **UD**, and **SA**, and their convex combinations will denote *mixed strategies*.

The ESS Criterion (Cressman, 1992)

x* ∈ X is an ESS if for any x ∈ X : x ≠ x* implies

$$\Gamma_0(\mathbf{x}^*, \mathbf{x}) \equiv (\mathbf{x}^* - \mathbf{x})^t \circ \mathrm{P} \circ \mathbf{x}^* > 0$$

or

$$\Gamma_0(\mathbf{x}^*, \mathbf{x}) = 0 \text{ and } \Gamma_1(\mathbf{x}^*, \mathbf{x}) \equiv (\mathbf{x}^* - \mathbf{x})^t \circ \mathrm{P} \circ (\mathbf{x} - \mathbf{x}^*) > 0$$

A set $\Omega \subset \mathbf{X}$ is *Evolutionary Stable* (Cressman, 1992)[4]

If for every $\mathbf{w} \in \Omega$ and every $\mathbf{x} \in \mathbf{X} : \Gamma_0(\mathbf{w}, \mathbf{x}) \geq 0$

and

$$\mathbf{x} \notin \Omega \text{ implies } \begin{array}{c} \Gamma_0(\mathbf{w}, \mathbf{x}) > 0 \\ \text{or} \\ \Gamma_0(\mathbf{w}, \mathbf{x}) = 0 \text{ and } \Gamma_1(\mathbf{w}, \mathbf{x}) > 0 \end{array}$$

where P is the payoff matrix of the game.

$$x_j^{'} = \left[(\mathbf{e}_{\mathrm{j}} - \mathbf{x})^{\mathrm{t}} \circ \mathrm{P} \circ \mathbf{x}\right] x_j \quad , \forall$$

To find the ESS solutions, we use the fact that every ESS is a steady state solution of the *replicator* equations associated with the game prescribed by the payoff matrix P. We find these solutions, then test them with the ESS criterion.

References

1. Nowak, M.A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573-577 (1998).

2. Nowak, M.A. & Sigmund, K. The dynamics of indirect reciprocity. *J. theor. Biol.* **194**, 561-574 (1998).

3 Weibull, J. W. *Evolutionary Game Theory*. MIT Press, Cambridge, Mass. (1996).

4. Cressman, R. (1992). *The Stability Concept of Evolutionary Game Theory*. Berlin: Springer-Verlag.