Cooperation and diversity. An evolutionary approach

Bruni, Luigino and Smerilli, Alessandra

Università di Milano-Bicocca - Department of Political Economics, University of East Anglia

January 2010

Cooperation and diversity. An evolutionary approach

Luigino Bruni and Alessandra Smerilli[1]

In this paper we propose a pluralistic and multi-dimensional approach to cooperation. Specifically, we seek to show that, in certain settings, less unconditional forms of cooperation may be combined with more gratuitous ones.
Starting with the prisoner's dilemma game, the evolution of cooperation is analyzed in the presence of different strategies, which represent the heterogeneity of the forms of cooperation in civil life.
There are many behaviour patterns, though not all of them are based on self-interest and conditionality. The dynamics of cooperation are studied through the use of evolutionary games applied in contexts that are either one-shot or repetitive. One of the most important results of the paper is the conclusion that cooperation is favoured by heterogeneity.

JEL Classification: C72, C 73, D64

Key-words: cooperation, Prisoner's Dilemma, reciprocity, heterogeneity, evolutionary game theory.

1. Introduction

Civil life is essentially cooperation. Neoclassical economics offers a highly parsimonious view of cooperation based on individual self-interests and instrumental rationality. In such a vision of cooperation an agent, for instance, would never cooperate in a non-iterated prisoner's dilemma game. If instead the

[1] *We wish to thank Angelo Antoci, Benedetto Gui, Vittorio Pelligra, Pierluigi Sacco, Robert Sugden, Stefano Zamagni and Luca Zarri, for their valuable comments. A previous version of this paper was published in the Aiccon working paper series (no. 31) of the University of Forlì.*
Luigino Bruni(luigino.bruni@unimib.it) is associate professor of political economy in the Department of Economics of the University of Milan-Bicocca. Alessandra Smerilli(alessandra.smerilli@uea.ac.k) is associate professor of economics at the Auxilium University.

game is repeated, the traditional theory justifies the cooperation by evoking self-interest (this being the so-called 'folk theorem') or enforcement.

In reaction to this excessively parsimonious view of cooperation, recent years have seen development of a body of literature, the so-called'social preferences' theories, which instead seeks to explain why even in a one-shot non-cooperative game (i.e. the 'ultimatum' or 'trust game') it may be rational to play 'cooperatively'. The explanation, of which there are several variants, is a redefinition of the utility function of the agents, by introducing non material payoffs associated to norms such as inequality aversion or reciprocity, . In this way it is possible to explain the emergence of cooperative behaviour in contexts where the standard theory would exclude it.

This is the explanation of cooperation advanced by behavioural economists (see Gintis (2004) and Bowles and Gintis (2004)), who base their analyses of cooperation on the theory of *strong reciprocity* (Fehr and Gächter (2000)). By 'strong reciprocity' they mean a social norm which, in a manner costly to the individual, rewards those who behave well and punishes those who behave badly. This theory of cooperation stands in methodological and cultural opposition to the mainstream economic theory: whereas standard economics (i.e. that of Binmore, 2005) envisages nothing but self-interest and monetary incentives, strong reciprocity theory explains the emergence of cooperation on the basis of a form of altruism which does not even require the game's repetition.

In this paper we adopt a different perspective for explaining the emergence of cooperation. We propose a theory of cooperation which is less parsimonious than that of standard economics, but without embracing the strong reciprocity thesis. We put forward a pluralistic and multidimensional view of cooperation and consequently examine aspects hitherto insufficiently explored by economic and social theory. In particular, the intuition inspiring this essay is the multidimensional nature of cooperation, that is*, at the same time, one and many*; civil society flourishes if and when the different forms of cooperation are seen as complementary instead of competitive or substitute one another. In this sense,diversity fosters cooperation, as it is well know in biology. Specifically, we seek to show on the one hand that, in certain settings, less 'altruistic' or unconditional forms of cooperation may combine with more gratuitous ones and so foster a culture of generalized cooperation. On the other, too many unconditional actions will end to promote the non-cooperation.

We accordingly construct dynamic models which will enable us to analyse diverse patterns of cooperation. There are many such patterns, not all of them based on self-interest, but all of them important for understanding the dynamics of civil life.

We shall base our analysis on the Prisoner's Dilemma (PD) game, because it lends itself well to the modelling of 'difficult' cooperation: the kind that occurs in situations where there is no enforcement and where there is always an incentive for non-cooperation. We believe that these situations are frequent and relevant – although in civil society individuals play many games, not only the PD – and that they are important in the real dynamics of cooperation in civil life.

In section 2 we analyse the evolution of cooperation in a 'one-shot' context, while in section 3 we apply the evolutionary analysis to repeated games. In section 4 we concentrate on analysis of situations in which four strategies interact, also furnishing simulations. The paper concludes with a brief discussion on the results of our analysis.

2. Evolution in a one-shot game

2.1. The basic model: two strategies

The pay-off matrix of the game is the following.[2]

	C	**D**
C	$\beta - \gamma$	$- \gamma$
D	B	0

[2] The table represents a particular case which simplifies the analysis without compromising the results. As well known, for a game to be a Prisoner's Dilemma, the payoff order must be $\beta > \gamma > 0$.

It can be easily shown that both players will choose not to cooperate or defect (D) in a one-shot game, and that the outcome (0,0) will be a Nash equilibrium. In this kind of non-iterated game cooperation cannot arise unless errors are committed or the players behave irrationally.

The structure of our model is as follows. Time is continuous. We suppose that there is a continuum of agents belonging to a particular population, and that they must choose one of the J pure strategies $\{1,...,J\}$ whenever they interact with other subjects in the same population. The subjects are distributed among I sub-populations $\{1,...,I\}$, which are assigned exogenously in the sense that existing sub-populations may disappear but new ones cannot be created.

The model's dynamic is described by standard 'replication' equations. The replication dynamic is widely used in evolutionary models, which assume that the most profitable strategies proliferate in the population at the expense of others. Heckathorn (1996) describes this dynamic well:

"Based on the resulting payoffs, the actors with the most successful strategies proliferate at the expense of the less successful. This process is then repeated, generation after generation, until the system either approaches stable equilibrium or cyclical variation." (p. 261)

This dynamics is usually employed in biology to study the evolution of species on the basis of the relative *fitness*. However, in social sciences there is a different interpretation of such a selection process: it involves learning by observing and imitating the behaviour of others. In what follows, we adopt neither the biological analogy nor the memetic one (i.e. the extension of gene-based biological evolution to meme-based social evolution). Instead, we use the concept of 'expected utility' as an indicator of the success (not necessarily material) of a strategy: a success which, over time, is imitated by less successful strategies (those with less expected utility). The dynamic of the model can be represented by the replication equations:

$$\dot{p}_i = p_i(Y_i - Y) \qquad i = 1,...,N \qquad [1]$$

where p denotes the proportion of subjects for each subpopulation, Y the average payoff, and Y_i the average payoff for a subject belonging to the subpopulation i.

The dynamic is defined on the invariant simplex:

$$\Delta = \left\{ p \notin \Re^N, \sum_{i=1}^{N} p_i = 1, p_i \geq 0 \right\}$$

We shall use this analytical structure to analyse the evolutionary process that arises in a situation where there are two pure strategies, C and D, and first two, then three, and finally four subpopulations.

To begin, we assume that in a one-shot game there are only two types of agents: those who always defects (types N), and those who instead always cooperate (types G, where G means *gratuitous*). In each round, a player is randomly paired with another player ('random matching'), with whom s/he plays once, after which there is another random encounter.[3] In a situation such as this, it can be easily shown that G subjects will be invaded by the N, that is an evolutionary stable strategy.

2.2. Three strategies with recognizability

We now introduce a third type, which adds to the previous ones: the T type (from tit-for-tat[4]). 'Cooperate' is not the T type's dominant strategy; but on encountering and recognizing a G type, s/he *does not exploit G but prefers to cooperate*. The T type's decision to cooperate is conditioned by the 'assurance' (to use Sugden's [2003] term) that the adversary too will cooperate.

What does the dynamic analysis tell us in this situation? Two cases should be distinguished: if the T type recognizes the other players, s/he has a perfect signal

---

[3] The game will always be non-repeated, because even if the players meet again in the future they will not recognize each other.

[4] See Axelrod (1984), Sugden (2004).

with which to discriminate between N and G types, so that the cooperative solution is sustainable in the long period. Instead, if T does not have a perfect signal, and may therefore commit errors, thereby being at risk of cooperating with N types and of not cooperating with other T types and with G types, the results are different: according to the value of the payoffs and to the probability of error, there may arise a non-cooperation equilibrium or the survival of only T types, with G types destined for 'extinction'.

We begin with the simpler case in which the T type is assumed to receive a perfect signal.

The hypotheses are therefore that:

a. there exist three types: Ns (who always play D), Gs (who always play C), and Ts (who play C with those that they identify as Gs or Ts, and D with Ns); the payoffs are those set out in table 1;

b. the probability of encountering a type i is p_i, i = N,G,T, and $p_t = 1 - p_n - p_g$;

c. the expected utilities are therefore:

$$U_n = p_n(0) + p_g(\beta) + p_t(0)$$
$$U_g = p_n(-\gamma) + p_g(\beta - \gamma) + p_t(\beta - \gamma)$$
$$U_t = p_n(0) + p_g(\beta - \gamma) + p_t(\beta - \gamma)$$

The payoffs can be arranged in a matrix, which we call A:

$$A \equiv \begin{bmatrix} 0 & \beta & 0 \\ -\gamma & \beta-\gamma & \beta-\gamma \\ 0 & \beta-\gamma & \beta-\gamma \end{bmatrix}$$

The system's dynamic can be written as follows:

$$\begin{aligned} \dot{p}_n &= p_n\left[(Ap)_1 - {}^t\mathrm{p}\cdot\mathrm{Ap}\right] \\ \dot{p}_g &= p_g\left[(Ap)_2 - {}^t\mathrm{p}\cdot\mathrm{Ap}\right] \qquad [2] \\ \dot{p}_t &= p_t\left[(Ap)_3 - {}^t\mathrm{p}\cdot\mathrm{Ap}\right] \end{aligned}$$

where ${}^t\mathrm{p} \equiv (p_n, p_g, p_t)$ and $(Ap)_l$ with $l = 1,2,3$ is the l-th component of the vector **Ap** and therefore corresponds to the expected payoff (as in hypothesis e.), while **^{t}pAp** is the average payoff.

The space of the dynamic is the simplex:

$$\Delta = \{p \in \Re^3 : p \geq 0 \text{ e } p_n + p_g + p_t = 1\} \qquad [3]$$

For the dynamic analysis we draw on Bomze (1983) on the replication dynamic: the analytical procedure is set out in the Appendix.

It is easy to show that only G types and T types will survive in time, while N types will become extinct: this is a well known result. Consider the simplex in Figure 1: the side on which $p_n=0$ (i.e. only Gs and Ts exist) consists entirely of fixed points. The presence of T types therefore means that the equilibrium may settle at cooperation. The most interesting result is that the final proportion of G types will be greater, the smaller their proportion at the beginning of the game. On following the trajectories within the simplex, in fact, we find that those starting from a point where the proportion is high finish at a point where the final proportion is low. The explanation for this is straightforward: G types are ‘preyed upon’ by the Ns, so that the larger their number at the beginning, the more nutritious the Ns’ ‘diet’ will be. Instead, if there are only a few G types at the beginning, the Ns will have scant prey and will succumb (because they cannot prevail over the T types).

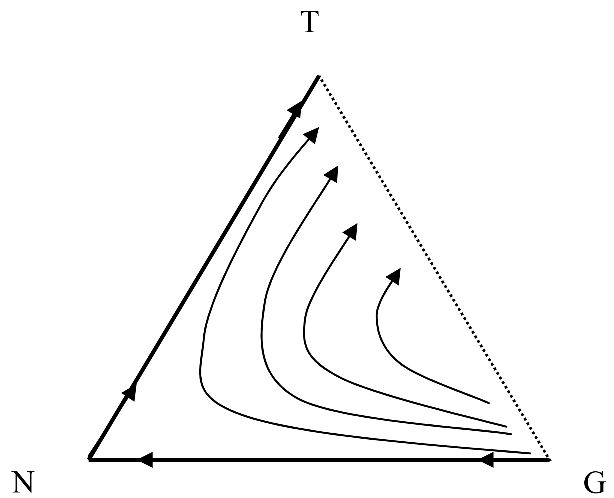

Figure 1

G types – the unconditional cooperators – can survive and assert themselves over time is there exists a mechanism (in this case the perfect signal utilized by T types) with which N types can be recognized.

We shall now see what happens if the hypothesis of perfect recognizability is abandoned.

2.3. Imperfect recognizability

If the T type does not have a perfect signal, the situation becomes more complicated. Let us assume a probability *r* that the T type is mistaken in identifying the type with which s/he is dealing.

Hence, if a T type is matched with an N type, s/he will play N with probability *r* and C with probability *(l – r)*; if instead s/he is matched with a G type s/he will play C with probability *r* and D with probability *(1 – r)*, and so on.

We therefore introduce a new hypothesis (d) in place of hypothesis c.

Hypothesis d. The expected utilities of the three types are

$$U_n = p_n(0) + p_g\beta + p_t(1-r)\beta$$

$$U_g = p_n(-\gamma) + p_g(\beta-\gamma) + p_t[r(\beta-\gamma) + (1-r)(-\gamma)]$$

$$U_t = p_n(1-r)(-\gamma) + p_g[r(\beta-\gamma) + (1-r)\beta] + p_t\{r[r(\beta-\gamma) + (1-r)(-\gamma)] + (1-r)[r\beta]\}$$

We alter matrix A for the dynamic analysis, while everything else remains the same as before.

As said, in a situation of this kind, N types or T types will survive over time if the probability that T types will recognize the other players is relatively high; whereas if this probability is low the T types will become extinct as well, so that nothing changes with respect to a situation in which only N types and G types are present.

As shown by figures 2a and 2b, G types are anyway destined for extinction, though matters are different for N types and T types. If the probability is low, all the trajectories simply converge on a situation in which only N types survive. If the probability is high, the final outcome depends on the initial situation. In other words, in the lower part of the figure (which, according to how the simplex is constructed, signifies relatively many Ns, many Gs and few Ts), Ns will prevail over both the Gs and the Ts. In the upper part, where there are initially more Ts and fewer Ns, Ts will prevail. The G types will always become extinct, even in a world where N types do not exist, because of the possibility that the T types will fail to recognize them. For if T types do not recognize G types, they will not cooperate with them.

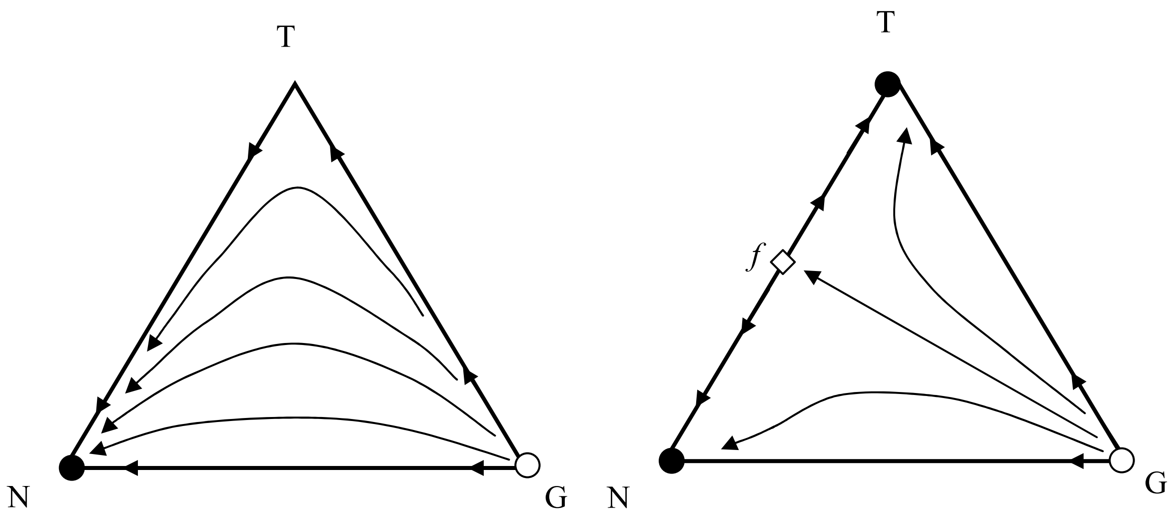

Figure 2. Case 2a: $r < \dfrac{\beta}{2\beta - \gamma}$, case 2b: $r \geq \dfrac{\beta}{2\beta - \gamma}$

Proposition 1. If hypotheses a, b and d hold, and if the T types recognize with probability r the types with which they play, then only N types will survive over time if $r < \dfrac{\beta}{2\beta - \gamma}$.

N types or T types are able to survive if $r \geq \dfrac{\beta}{2\beta - \gamma}$.

Probability *r* therefore depends on the ratio between the utility obtained by not cooperating at the expense of those who cooperate, and the utility of those who cooperate with those who do likewise. The threshold value of *r* (the minimum value for T types to be able to survive over time) is lower, the smaller the value of γ: that is, the smaller the 'exploitation advantage' or the 'cost of coherence'. In the second example, 2b, the possible equilibrium depends, other conditions remaining equal, on the position of the fixed point *f*. It can be shown, in fact, that if r tends to its minimum value, i.e. $r \to \frac{\beta}{2\beta - \gamma}$, the point tends to shift towards the vertex T.[5] This means that as *r* diminishes, the equilibrium of only Ns becomes more likely (all the trajectories in the simplex starting from a point situated lower than *f* lead to N).

A similar result has been obtained by Bowles and Gintis (2004). In their model, however, three different types interact: the selfish type, the unconditional cooperator, and the reciprocator, who punishes the selfish type at his/her own expense if recognizing him/her. With the base version of their model Bowles and Gintis obtain the survival, at asymptotic level, of only selfish types. Then by varying the parameters and using simulations, they obtain high levels of cooperation.

3. The dynamic game

We saw in the previous section that G types had little hope of surviving in the evolutionary context described.

We now see what happens if the game is repeated. We assume, that is to say, that associated with every random encounter is a repeated interaction *with the same person*. This interaction may be of greater or lesser duration according to a parameter, π, which denotes the probability that the game will continue for another round[6]. After a series of interactions with the same person, another random encounter occurs, and the (repeated) game resumes with another (randomly matched[7]) partner.

---

[5] The coordinates of point *f* are $f \equiv \left(\frac{2r\beta - r\gamma - \beta}{2r\beta - 2r\gamma - \beta + \gamma}, \quad 0, \quad \frac{\gamma(1-r)}{2r\beta - 2r\gamma - \beta + r} \right)$, and as r tends to its minimum value, the former coordinate (p_n) tends to 0.

[6] We are hence in a context of indefinitely repeated game.

[7] We adopt the simplifying hypotheses that the repeated game finishes for all players at the same moment, but the results do not change if the game finishes earlier for one of them, because s/he is re-assigned to another game.

We are well aware that if the game is repeated, the possible strategies are infinite. We consequently restrict our analysis to four strategies, differentiating the T strategy (of the previous section) into two strategies, which we shall call (following Sugden 2004) B (= Brave) and C (= Cautious). B and C strategies are a kind of trigger strategies: they stop to cooperate if they see the other player defecting.

The strategies considered are therefore the following:

1. N: never cooperate. N is a highly important strategy because analysis of cooperation dynamics becomes non-banal precisely when non-cooperation scenarios are possible.
2. G: always cooperate.
3. C: cooperate with a player who cooperated in the previous round; do not cooperate with a player who did not cooperate in the previous round, and *begin by not cooperating*. If these cautious types are to cooperate, they must have obtained cooperation in the previous round. When Cs encounter other Cs or Ns, they never cooperate. An immediate consequence ensues: in a world with only Cs and Ns, cooperation will never be possible, and it will not be possible to distinguish Cs from Ns because they behave in exactly the same way.
4. B: this strategy has the same structure as C, the only difference being that B begins by cooperating. B stands for 'Brave', in fact. Bs are players who begin by cooperating (and therefore risk being 'exploited' by Ns or Cs in the first round). But if in the second round they do not receive cooperation, nor will they cooperate.

If we use p_n, p_b, p_g, p_c to denote the probabilities of encountering, respectively, an N, B, G or C type, the expected utilities in a world with these four possible strategies are:

$$U_n = p_n(0) + p_b\beta + p_g\frac{\beta}{1-\pi} + p_c(0) \qquad [3]$$

An N type will never cooperate with other N types and with C types who begin by not cooperating and do not cooperate if the other player did not cooperate in the first round, whence $p_n(0)$, $p_c(0)$. If the N type encounters a B type, s/he will obtain β in the first round because B began with an act of cooperation, but the subsequent payoffs will be equal to 0 because B will stop cooperating from the second round onwards. Finally, if N encounters a G, s/he will obtain β in every round[8] because G will always cooperate.

$$U_b = p_n(-\gamma) + p_b \frac{(\beta - \gamma)}{1-\pi} + p_g \frac{(\beta - \gamma)}{1-\pi} + p_c(-\gamma + \beta\pi) \qquad [4]$$

The B type begins with an act of cooperation and continues to cooperate if the adversary in the first round has responded by cooperating. Cooperation is assured with other B types and with G types, but not with N types, or even with C types.[9]

$$U_g = p_n \frac{-\gamma}{1-\pi} + p_b \frac{(\beta-\gamma)}{1-\pi} + p_g \frac{(\beta-\gamma)}{1-\pi} + p_c\left(\frac{\beta-\gamma}{1-\pi} - \beta\right) \qquad [5]$$

A G type will therefore always cooperate with Bs and with Gs, and with Cs from the second round onwards, while Gs will let themselves be 'exploited' by Ns.

$$U_c = p_n(0) + p_b(\beta - \gamma\pi) + p_g\left(\frac{\beta-\gamma}{1-\pi} + \gamma\right) + p_c(0) \qquad [6]$$

Finally, a C type will not cooperate with Ns and Cs, and s/he will cooperate with Gs from the second round onwards. With Bs, C types will receive β in the first round, given that Bs begins with an act of cooperation, and (- γ) in the second round. From the third round onwards Cs will obtain 0.

[8] The expected utility associated to this interaction is hence $\beta + \beta\pi + \beta\pi^2 + \ldots$, and then $\frac{\beta}{1-\pi}$.

[9] The payoff $p_c(-\gamma + \beta\pi)$ depends on the fact that B cooperates the first time and C responds by not cooperating; B will therefore have (- γ), but C will cooperate in the second round, because B has cooperated in the first. From the third round onwards the payoff will be 0.

3.1. Evolutionary analysis

In order to analyse the evolution in dynamic terms, we consider three strategies at a time (so that we can use simplexes).

After the first game, it is likely that the proportion of players adopting the winning strategy will increase in future pairings: that is, the winning strategy will be imitated by others. This will be the basis for our *both repeated and evolutionary* analysis.

It will be assumed in the analysis that $\pi > \frac{\gamma}{\beta}$.[10]

3.1.1. First case: N, C, G

We begin the analysis with B types omitted.

The replication dynamic can be represented with the following simplex:

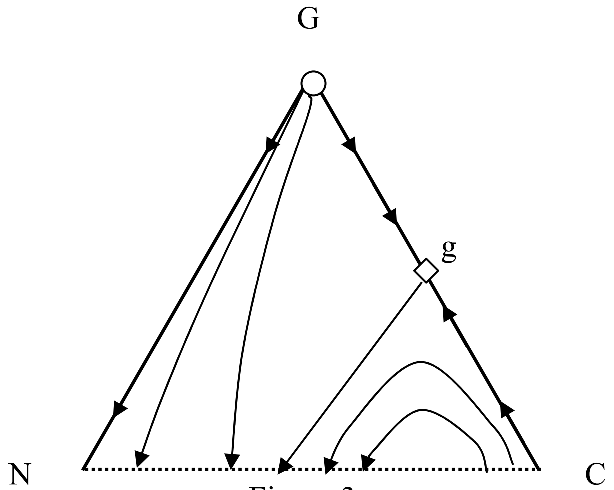

Figure 3

[10] We imagine, in fact, that two players agree to cooperate in each round. If they abide by the agreement, the expected utility of each player is $\frac{\beta-\gamma}{1-\pi}$. However, if one of the players breaks the agreement, the other will no longer cooperate. Thus a player who breaks the agreement in the first round will receive β, but from the second round onwards s(/he will always receive 0. The condition for cooperation agreements (without enforcement) to come about is: $\frac{\beta-\gamma}{1-\pi} > \beta$, and hence $\pi > \frac{\gamma}{\beta}$.
On this see also Sugden (2004)

When strategies N, C and G are present, the outcome may be one of the multiple fixed points along the line NC, which signifies *non-cooperation*. If only strategies G and C are present, the outcome may be a unique combination of C and G that depends on the position of the point (in this case a saddle point), *g*, i.e.:

$$g \equiv \left(0, \quad \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi} \quad \frac{\beta\pi-\gamma}{(\beta-\gamma)\pi} \right).$$

Other conditions remaining equal, if $\pi \to 1$ – so that the likelihood of continuing with the same person initially encountered is very high – the point shifts towards vertex G.

This result strikes us as important: only G types are able somehow to activate Cs, who without Gs would always be confined to a world of non-cooperation.

The following proposition therefore holds:

Proposition 2. In a world in which the types or strategies N, G, C are present, the replication dynamic has two different outcomes: a combination of C and G (fixed point g) only if p_n *is equal to 0, or a combination along the line of fixed points N, C (and consequently non-cooperation).*

Without the presence of B types – who always begin with an act of cooperation – it is unlikely that virtuous cooperation mechanisms will be triggered.

3.1.2. Second case: N, B, C

Another interesting case is that in which G types are absent. Here too, non-cooperation is a probable equilibrium. The other equilibrium is the one where only B strategies survive. In a three-strategy world in which only Ns, Cs and Bs are present, in fact, Ns and Cs will never cooperate, and moreover the Ns will have no Gs to exploit. Instead, the Bs will cooperate only and exclusively with each other, obtaining a greater payoff – if the game lasts for a long time – than that received by the Ns and the Cs.

Here too, as shown by figure 4, the possible long-period equilibrium depends on the coordinates of the fixed point *f*.

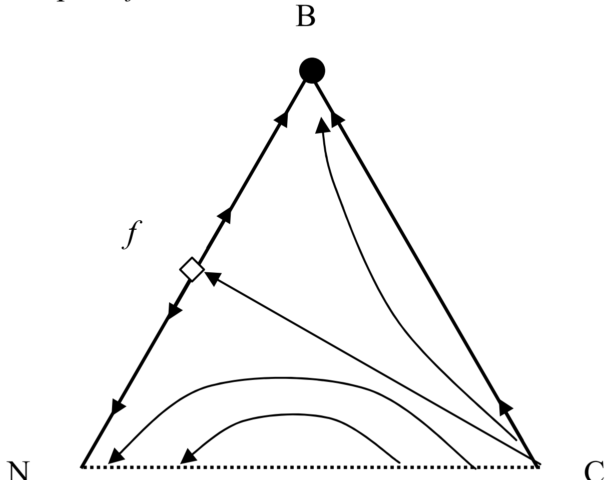

Figure 4

All the points of departure in the simplex lying below the trajectory from C to *f* will evolve towards a non-cooperative equilibrium if N and C are present.

Proposition 3. *In a world in which the types or strategies N, C, B are present, the replication dynamic has two different outcomes: the survival of B strategies alone, or a combination along the line of fixed points N, C (and consequently non-cooperation).*

The coordinates of point *f* are now:

$$f \equiv \left(\frac{\beta\pi - \gamma}{(\beta - \gamma)\pi} \quad 0 \quad \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi} \right)$$

It is evident that if $\pi \to 1$, the point tends to shift towards the N vertex, so that that greater the probability of the game continuing, the more likely it becomes that Bs will prevail and that the cooperative outcome will occur. In a world without G types, Cs do not begin to cooperate. We may say that the sacrifice of the Gs somehow restores cooperation potential to Cs, for without their presence the only possible form of cooperation is that between B types. To be noted is that B types begin with an act of cooperation. In their absence, a non-cooperative equilibrium would arise.

3.1.3. Third case: N, B, G

The simplex relative to this third and final case shows that, depending on the point of departure and the position of fixed point f on the side NB, there will be a different final equilibrium, which may be a combination of G and B, or a world consisting only of Ns. Matters are different when the three types instead coexist in the population at time 1 (when the dynamic begins). In this case, non-reciprocity, i.e. an equilibrium consisting of only N types, may prevail.

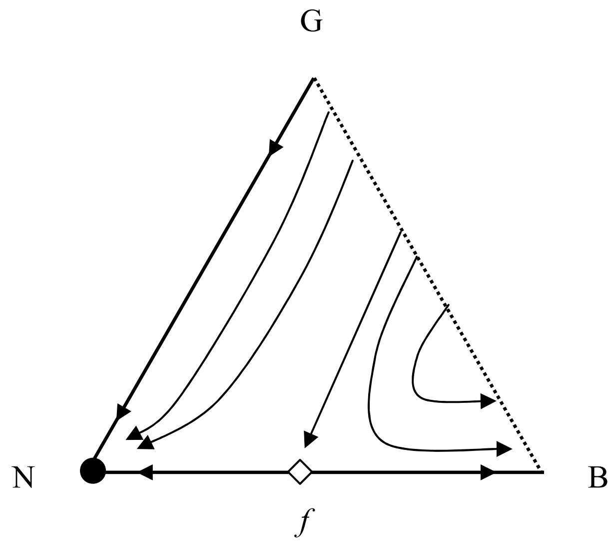

Figure 5

Proposition 4. In a world in which the strategies N, B, G are present, two equilibria are possible: the survival of only types N and a coexistence of B types and G types along the line of fixed points on the B-G side. Which of the two equilibria will come about depends on the position of the fixed point f along the N-B side.

As the simplex is constructed here, considering that the position of N in terms of fraction of the population is (1.0.0) and the position of B is (0,1,0), the fixed point *f* has the following coordinates:

$$f \equiv \left(\frac{\beta\pi - \gamma}{(\beta - \gamma)\pi}, \quad \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi}, \quad 0 \right)$$

The position of point *f* therefore depends on β and γ, and on the value of π. In particular, for $\pi \to \frac{\gamma}{\beta}$, point *f* will approach B. If instead $\pi \to 1$, point *f* will shift towards N. With a small value of π, *ceteris paribus*, the likelihood that only N types will prevail is very high; instead, with a very high π, it very likely that the final equilibrium will be the one in which B types and G types coexist.

For every intermediate value between the two extremes, the final equilibrium will depend on the point of departure: if this is a point to the left of the trajectory leading from side B-G to point *f*, then the tendency is an equilibrium of only Ns; vice versa, if the point of departure is to the right of the trajectory, the outcome will be a coexistence of Bs and Gs. Note that points to the left are characterized, amongst other things, by a lower percentage of Bs than of Gs. It is therefore important that B types be relatively more than Gs and Ns for the B-G equilibrium to come about. In short, evident here is the delicate role of G strategies: if there are too many of them, they foster the emergence of N types over Bs. Metaphors aside, in a population where non-cooperation is possible, if there are too many unconditional acts, not only are they likely to become extinct, but they will also extinguish the possibility of cooperation, for an equilibrium consisting of non-generalized cooperation.

At the same time, the coordinates of point *f* also depend on *β* and *γ*. The value of *γ* is the one which most clearly tells us what the social rewards structure is. A high *γ* denotes a culture which penalizes reciprocity, while a high (*β*-*γ*) denotes a culture which rewards it. In fact, if the first coordinate is high, point *f* tends to N (the same happens if the second coordinate is low), while if it is low *f* tends to B.

This is because the coordinate of N is directly proportional to *β*: while both coordinates depend on (*β*-*γ*), the sign of *γ* is negative in the coordinate of N and positive in the coordinate of B. This tells us that the more a society, *ceteris paribus*, makes reciprocity of G and B type costly, the more likely the prevalence of non-cooperation becomes.

4. In a four-dimensional world

Thus far we have compared three strategies at a time, and we have analysed their dynamic evolution. The question now is what changes if the four strategies N, B, G, C interact simultaneously.

In the four-strategies case, the replication dynamic can be depicted by a three-dimensional simplex.:

$$\Delta = \{p \in \Re^4 : p \geq 0 \ \text{ e } \ p_n + p_b + p_g + p_c = 1\}$$

In this case matrix A becomes:

$$A = \begin{pmatrix} 0 & \beta & \dfrac{\beta}{1-\pi} & 0 \\ -\gamma & \dfrac{\beta-\gamma}{1-\pi} & \dfrac{\beta-\gamma}{1-\pi} & \beta\pi-\gamma \\ \dfrac{-\gamma}{1-\pi} & \dfrac{\beta-\gamma}{1-\pi} & \dfrac{\beta-\gamma}{1-\pi} & \dfrac{\beta\pi-\gamma}{1-\pi} \\ 0 & \beta-\pi\gamma & \dfrac{\beta-\pi\gamma}{1-\pi} & 0 \end{pmatrix} \qquad [7]$$

The vector ${}^{t}\mathbf{p} \equiv (p_m, p_b, p_w, p_c)$, so that the system of equations becomes:

$$\begin{aligned} \dot{p}_n &= p_n\left[(Ap)_1 - {}^{t}\mathrm{p}\cdot \mathrm{Ap}\right] \\ \dot{p}_b &= p_b\left[(Ap)_2 - {}^{t}\mathrm{p}\cdot \mathrm{Ap}\right] \\ \dot{p}_g &= p_g\left[(Ap)_3 - {}^{t}\mathrm{p}\cdot \mathrm{Ap}\right] \\ \dot{p}_c &= p_c\left[(Ap)_4 - {}^{t}\mathrm{p}\cdot \mathrm{Ap}\right] \end{aligned} \qquad [8]$$

Given that analysis of the system of differential equations [8] would be highly complex, here we only report the frontier conditions (those in which at least one strategy is extinct). Following the examples of Hirshleifer and Martinez Coll (1991), and of Antoci, Sacco and Zarri (2004), we may represent the surface (or frontier) of Δ on the plane. The simplex Δ can be imagined as having a triangular base N,C,B, and G as its upper vertex (if the simplex in figure 6 were drawn three-dimensionally, the three vertices G would become a single upper vertex).

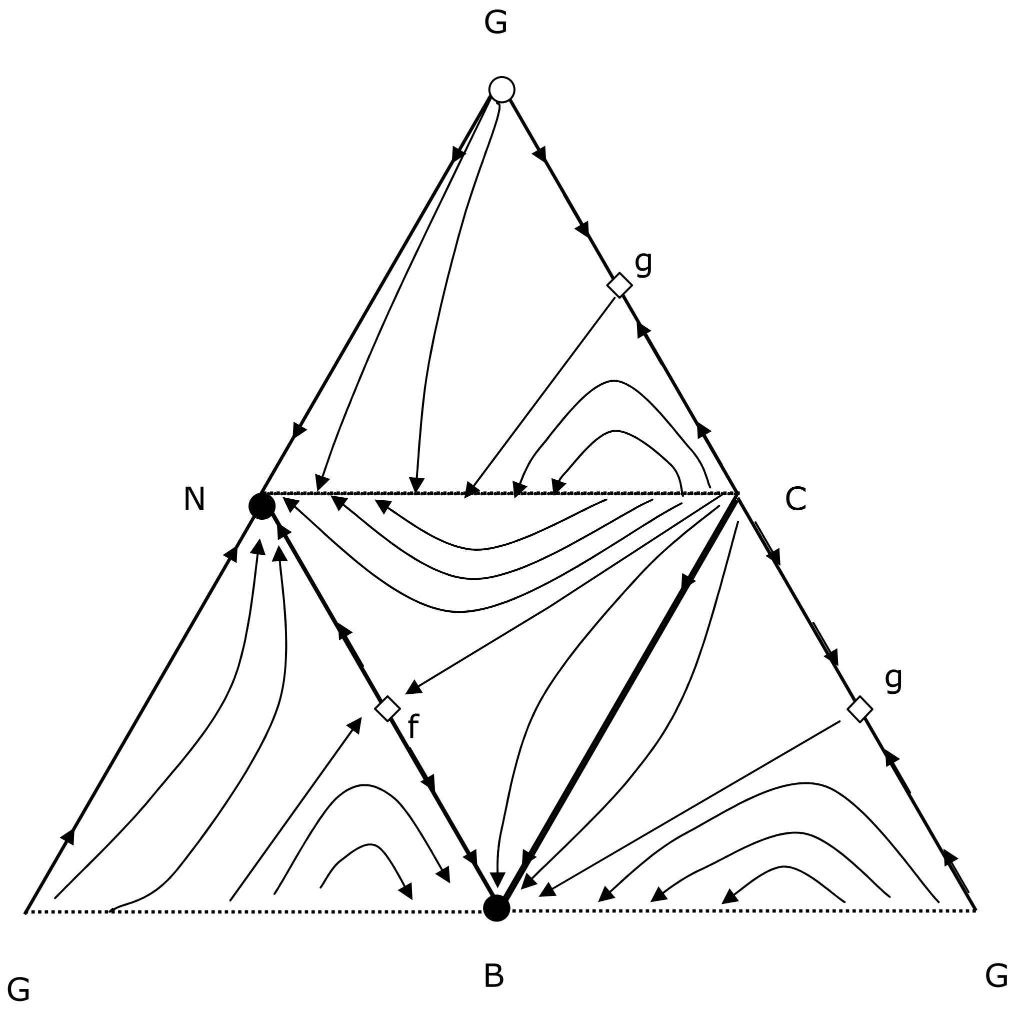

Figure 6

Figure 6 shows that there are four possible equilibrium combinations:

- a combination of G and B, i.e. cooperation

- a combination of N and C, i.e. non-cooperation
- the extinction of all the strategies except N
- the extinction of all the strategies except B.

4.1. Some simulations

Which of these equilibria are more likely depends on the initial conditions. To furnish a clearer idea of the dynamic, we now report some simulations. They have been obtained by setting various initial conditions for the system. We assigned the following values to the parameters:

$\beta = 2$, $\gamma = 1$, $\pi = 4/5$

The first graph shows the evolution over time of the strategies when the initial conditions state: $p_n = p_b = p_g = p_c = 0.25$.

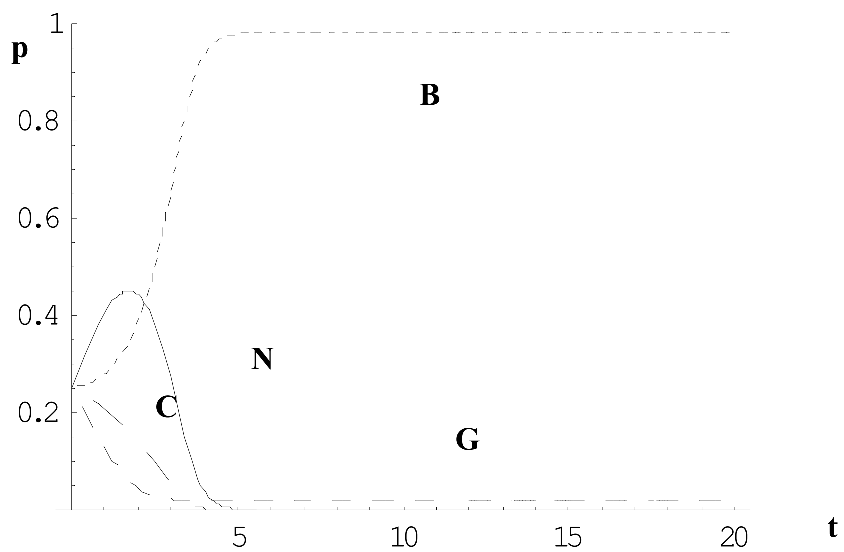

Figure 7

In this case the final equilibrium is of the B-G type where the proportion of G is very small. What happens if we change the initial conditions? The next graph illustrates a situation where the initial proportions are $p_n = 0.25$, $p_b = 0.25$, $p_g = 0.1$

$p_c = 0.4$. We have left the proportions of B and G unaltered, but we have increased Cs with respect to Gs.

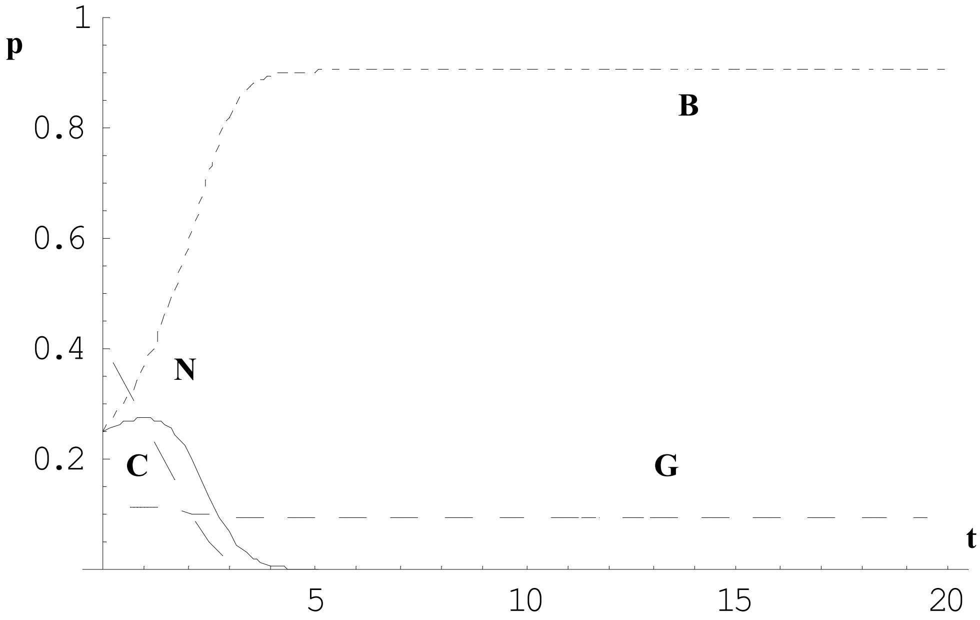

Figure 8

Interestingly, a greater proportion of Cs, although it does not improve their chances of 'survival', helps the development of Gs, which in this case remain constant over time. We saw in section 3.1.1 that only G types are able to activate Cs; we may now state that Cs are essential for the survival of Gs. The importance of the role performed by Cs (which in the three-strategy world seemed almost irrelevant) also emerges from the following graph, which has been constructed with the following initial proportions: $p_n = 0.4$, $p_b = 0.3$, $p_g = 0.1$ $p_c = 0.2$. In this case the Ns are initially in a greater proportion than Bs, and there are more Cs than Gs.

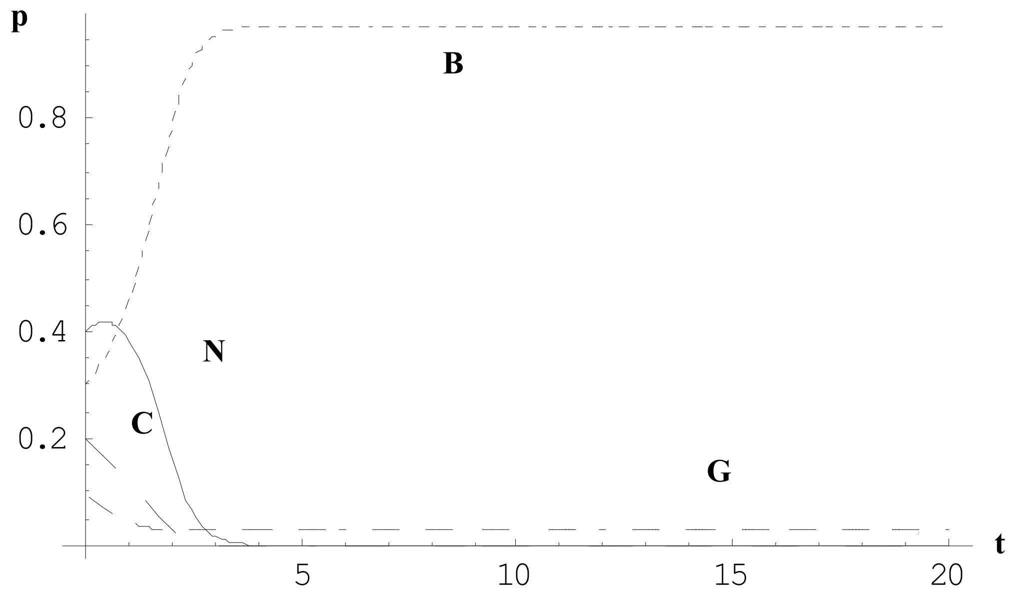

Figure 9

Hence, cooperation may prevail even if there are initially more Ns than Bs, provided that there is a sufficient number of Cs.

5. Conclusions

We may now draw some conclusions. We started from the hypothesis that there are only two strategies – N and G – and we saw, in line with the literature, that in this case cooperation has no chance of surviving. We also saw that in random encounters where the game lasts for only one round, cooperation is possible in the presence of three strategies (N, G, T) but only if T types have a high probability of recognizing the types with which they interact. But cooperation often involves repeated encounters between the same people, and who recognize each other. Our analysis in this case showed that strategy T is entirely ineffectual unless recognizability is presumed. Recognizability is important in civil life because it underpins the reputation games in which we cooperate because we recognize the others and they recognize us. But recognition is not always possible, especially in the great societies of the globalized world. For this reason we extended our analysis by abandoning the recognizability hypothesis and advancing other reasons for the emergence of spontaneous (or without enforcement) cooperation.

We summarize our results as follows:

(a) *The 'crucial' role of G types*. We have seen at various points in our analysis that G types should not be too numerous, because if they are they compromise themselves and also the survival, for example, of Bs. In populations where non-cooperation is possible (which is the case of all real ones), unconditional acts are essential, but when too numerous, they become counter-productive.

(b) G types perform a vital role, for only they can activate the cooperation of Cs. Without the presence of G types, Cs would never experience cooperation and therefore would never respond with an act of cooperation. G types are consequently valuable, but they should be protected. The success of numerous forms of cooperation – from firms to families – depends also, and sometimes above all, on the presence of a small number of unconditional reciprocators able to activate people who would never be so activated if they only interacted with conditional cooperators.

(c) *Alliances: C types*. These are 'activated' by Gs, but at the same time their presence is highly beneficial to Gs because it increases their expected utility. Gs, in fact, cooperate with Bs and with Cs, but they are exploited by Ns. In a four-strategy world, Cs protect the Gs against extinction.

Cooperation is therefore favoured by heterogeneity or diversity.

From a mathematical point of view, it might be objected that G types are not necessary. The onset of cooperation would only require slightly more sophisticated Bs. But this was not the purpose (i.e. to study which strategies favour cooperation) for which the model was conceived. Our analysis started from the assumption that behaviours like G exist in civil society. (And who could deny the presence in the real world of unconditional actions? Even Binmore (2006) with his orthodoxy and anthropological parsimony admits their existence). Our model has sought to analysis the conditions under which unconditional actions can not only survive but also perform a virtuous civil role.

Appendix

Proof of proposition 1:
The expected utilities are:

$$U_n = p_n(0) + p_g\beta + p_t(1-r)\beta$$

$$U_g = p_n(-\gamma) + p_g(\beta-\gamma) + p_t[r(\beta-\gamma)+(1-r)(-\gamma)]$$

$$U_t = p_n(1-r)(-\gamma) + p_g[r(\beta-\gamma)+(1-r)\beta] + p_t\{r[r(\beta-\gamma)+(1-r)(-\gamma)]+(1-r)[r\beta]\}$$

The matrix of payoffs is:

$$A = \begin{pmatrix} 0 & \beta & \beta - r\beta \\ -\gamma & \beta-\gamma & r\beta-\gamma \\ r\gamma-\gamma & \beta - r\gamma & r\beta - r\gamma \end{pmatrix}$$

Adding a constant to each column of A does not change the dynamics, so we subtract the first row:

$$\begin{pmatrix} 0 & 0 & 0 \\ -\gamma & -\gamma & 2r\beta-\beta-\gamma \\ r\gamma-\gamma & -r\gamma & 2r\beta - r\gamma - \beta \end{pmatrix}$$

We know that::
$\beta > \gamma > 0$

Following Bomze (1983), proposition 1 (p. 210) :
1. the eigenvalue of the corner N in direction N-G is proportional to ($-\gamma$) , then is <0
2. the eigenvalue of the corner N in direction N-T is proportional to ($r\gamma - \gamma$), then is <0
3. the eigenvalue of the corner G in direction G-N is proportional to (γ), then is >0
4. the eigenvalue of the corner G in direction G-T is proportional to ($-r\gamma + \gamma$), then is >0
5. the eigenvalue of the corner T in direction T-G is proportional to $(r\gamma - \gamma)$, then is <0
6. the eigenvalue of the corner T in direction T-N is proportional to ($\beta + r\gamma - 2r\beta$). This value could be positive or negative, depending on the value of *r*, and then we must distinguish between two cases:
   First case: $r < \dfrac{\beta}{2\beta-\gamma}$, eigenvalue >0
   Second case: $r \geq \dfrac{\beta}{2\beta-\gamma}$, eigenvalue <0.

FIRST CASE
Proposition 2 (Bomze, p. 210) shows that there aren't any fixed point on the N-G side and on the N-T side.
Proposition 5 (pag. 211) shows that there aren't any fixed point on the side G-T, and proposition 6 shows that internal fixed points do not exist.

SECOND CASE

Proposition 2 (p. 210) tells us that there exists a fixed point *f* on the side N-T, in fact the quantity $(r\gamma-\gamma)(2r\beta-r\gamma-\beta)$ is negative. The eigenvalues associated to the fixed point are proportional to γ (positive) in direction NT, and to the quantity $\frac{af-cd}{f}$, that is $\frac{-\gamma(2r\beta-r\gamma-\beta)-(2r\beta-\gamma-\beta)(r\gamma-\gamma)}{2r\beta-r\gamma-\beta}$ in the other direction, then is negative.

Proof of proposition 2:
The expected utilities are:

$$U_n = p_n(0)+p_c(0)+p_g\frac{\beta}{1-\pi}$$
$$U_c = p_n(0)+p_c(0)+p_g\frac{\beta-\gamma\pi}{1-\pi}$$
$$U_g = p_n\left(\frac{-\gamma}{1-\pi}\right)+p_c\frac{\beta\pi-\gamma}{1-\pi}+p_g\frac{\beta-\gamma}{1-\pi}$$

The matrices are:

$$\begin{pmatrix} 0 & 0 & \frac{\beta}{1-\pi} \\ 0 & 0 & \frac{\beta-\pi\gamma}{1-\pi} \\ \frac{-\gamma}{1-\pi} & \frac{\beta\pi-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} \end{pmatrix} \text{ and: } \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{-\pi\gamma}{1-\pi} \\ \frac{-\gamma}{1-\pi} & \frac{\beta\pi-\gamma}{1-\pi} & \frac{-\gamma}{1-\pi} \end{pmatrix}$$

Following proposition 1 (p. 210) we may say:
1. the eigenvalue of the corner N in direction N-C is equal to 0
2. the eigenvalue of the corner N in direction N-G is proportional to $\frac{-\gamma}{1-\pi}$, and then is negative
3. the eigenvalue of the corner C in direction C-N is equal to 0
4. the eigenvalue of the corner C in direction C-G is proportional to $\frac{\beta\pi-\gamma}{1-\pi}$, then is positive (we have supposed that $\pi>\frac{\gamma}{\beta}$)
5. the eigenvalue of the corner G in direction G-C is proportional to $\frac{\gamma-\pi\gamma}{1-\pi}$ and then is positive
6. the eigenvalue of the corner G in direction G-N is proportional to $\frac{\gamma}{1-\pi}$, and then is positive.

Following proposition 2 (pag. 210) we know that N-C is pointwise fixed.

Proposition 5 (p. 211) tells us that there exists a fixed point g (saddle point) on the side G-C, in fact the quantity *(e – b)(f – c)* is negative, and the eingenvalues associated to the fixed point are proportional to:

1. $-\frac{(e-b)(c-f)}{e-b+c-f}$, that means $\frac{-\left(\frac{\beta\pi-\gamma}{1-\pi}\right)\left(\frac{\gamma-\pi\gamma}{1-\pi}\right)}{\frac{\beta\pi-\pi\gamma}{1-\pi}}$: this quantity is negative;

2. $\frac{bf-ce}{e-b+c-f}$, that is positive.

Proof of proposition 3:

Expected utilities:

$U_n = p_n(0) + p_c(0) + p_b(\beta)$

$U_c = p_n(0) + p_c(0) + p_b(\beta - \gamma\pi)$

$U_b = p_n(-\gamma) + p_c(\beta\pi - \gamma) + p_b\left(\frac{\beta-\gamma}{1-\pi}\right)$

Matrices:

$$\begin{pmatrix} 0 & 0 & \beta \\ 0 & 0 & \beta-\gamma\pi \\ -\gamma & \beta\pi-\gamma & \frac{\beta-\gamma}{1-\pi} \end{pmatrix} \text{ and: } \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\gamma\pi \\ -\gamma & \beta\pi-\gamma & \frac{\beta\pi-\gamma}{1-\pi} \end{pmatrix}$$

1. the eigenvalue of the corner N in direction N-C is equal to 0
2. the eigenvalue of the corner N in direction N-B is proportional to $-\gamma$, and then is negative
3. the eigenvalue of the corner C in direction C-N is equal to 0
4. the eigenvalue of the corner C in direction C-G is proportional to $\beta\pi-\gamma$, and then is positive
5. the eigenvalue of the corner B in direction B-C is proportional to $\frac{-\beta\pi+\gamma-\pi\gamma+\pi^2\gamma}{1-\pi}$ and then is negative
6. the eigenvalue of the corner B in direction B-N is proportional to $\frac{\gamma-\beta\pi}{1-\pi}$, and then is negative

Following proposition 2 (p. 210) we may say:

- the side N-C is pointwise fixed
- On the side N-B there exists an unique fixed point *f*; the eigenvalues of f are positively proportional to:

  γ (positive)

  $\frac{0-(-\gamma\pi)(-\gamma)}{\frac{\beta\pi-\gamma}{1-\pi}}$ (negative).

  The fixed point has coordinates(Bomze 1983, pag. 204):

  $$p_m = \frac{1}{1+\frac{\gamma}{\frac{\beta\pi-\gamma}{1-\pi}}} = \frac{\beta\pi-\gamma}{(\beta-\gamma)\pi}$$

  $p_c = 0$

$$p_b = \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi}$$

We know that do not exist fixed points on the side C-B (prop. 5) and that do not exist internal fixed points (prop.6).

Proof of proposition 4:
Expected utilities:

$$U_n = p_n(0) + p_b\beta + p_g\frac{\beta}{1-\pi}$$

$$U_b = p_n(-\gamma) + p_b\frac{\beta-\gamma}{1-\pi} + p_g\frac{\beta-\gamma}{1-\pi}$$

$$U_g = p_n\left(\frac{-\gamma}{1-\pi}\right) + p_b\frac{\beta-\gamma}{1-\pi} + p_g\frac{\beta-\gamma}{1-\pi}$$

Matrices:

$$\begin{pmatrix} 0 & \beta & \frac{\beta}{1-\pi} \\ -\gamma & \frac{\beta-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} \\ \frac{-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} \end{pmatrix}, \text{ and: } \begin{pmatrix} 0 & 0 & 0 \\ -\gamma & \frac{\beta\pi-\gamma}{1-\pi} & \frac{-\gamma}{1-\pi} \\ \frac{-\gamma}{1-\pi} & \frac{\beta\pi-\gamma}{1-\pi} & \frac{-\gamma}{1-\pi} \end{pmatrix}$$

1. the eigenvalue of the corner N in direction N-B is proportional to $-\gamma$ and then is negative
2. the eigenvalue of the corner N in direction N-G is proportional to $\frac{-\gamma}{1-\pi}$ and then is negative
3. the eigenvalue of the corner B in direction B-N is proportional to $-\frac{\beta\pi-\gamma}{1-\pi}$ and then is negative
4. the eigenvalue of the corner B in direction B-G is equal to zero
5. the eigenvalue of the corner G in direction G-B is equal to zero
6. the eigenvalue of the corner G in direction G-N is equal to $\frac{\gamma}{1-\pi}$, and then is positive

We know that there esists a fixed point on the side N-B (prop.2) , and that the eigenvalues of the fixed point are positively proportional to :

γ , then positive

$\dfrac{\left(\frac{\beta\pi-\gamma}{1-\pi}\right)\left(\frac{-\gamma}{1-\pi}\right)-(-\gamma)\left(\frac{\beta\pi-\gamma}{1-\pi}\right)}{\frac{\beta\pi-\gamma}{1-\pi}}$, that becomes: $\frac{-\gamma\pi}{1-\pi}$ and then is negative.

The fixed point has coordinates (Bomze 1983, pag. 204):

$$p_m = \frac{1}{1+\frac{\gamma}{\frac{\beta\pi-\gamma}{1-\pi}}} = \frac{\beta\pi-\gamma}{(\beta-\gamma)\pi}$$

$$p_b = \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi}$$

$$p_w = 0$$

We also know that the side B-G is pointwise fixed.

REFERENCES

Antoci A., Sacco P. e Zarri L. (2004) "Coexistence of Strategies and Culturally-Specific Common Knowledge: An Evolutionary Analysis", *Journal of Bioeconomics*, vol. 6, pp. 165-194

Binmore K. (2006) *Natural Justice*, Oxford University Press, USA

Binmore K. (2006) "Why do people cooperate?", *Politics, Philosophy and Economics*, vol. 5 (1), pp. 81-96.

Bomze I. (1983) "Lotka-Volterra Equation and Replicator Dynamics: A Two-Dimensional Classification", *Biological Cybernetics*, vol. 48, pp. 201-211.

Bowles S. e Gintis H. (2004) "The evolution of strong reciprocity: cooperation in a heterogeneous population." *Theoretical Population Biology*, vol. 65, pp. 17-28.

Fehr E. and Gachter S. (2000) "Fairness and Retaliation: The Economics of Reciprocity", *Journal of Economic Perspectives*, 14, pp. 159-181.

Gintis H (2004) "Modeling Cooperation Among Self-Interested Agents: A Critique", *The Journal of Socio-Economics*, 33, pp. 311-322.

Heckathorn D. (1996) "The dynamics and dilemmas of collective action", *American Sociological Review*, vol. 61, pp. 250-277.

Hirshleifer J., Martinez Coll J. (1991) "The limits of reciprocity", *Rationality and Society*, vol. 3, pp. 35-64.

Sugden R. (2003) "The logic of team reasoning", *Philosophical explorations*, vol. 6, pp. 165-181.

Sugden R. (2004) *The economics of rights, cooperation and welfare*, second edition, Palgrave Macmillian, London.