QUADERNI

Università degli Studi di Siena
DIPARTIMENTO DI ECONOMIA POLITICA

ENNIO BILANCINI
LEONARDO BONCINELLI

Cooperation with Defection

n. 482 - Giugno 2006

Abstract - The Prisoner Dilemma is a typical structure of interaction in human societies. In spite of a long tradition dealing with the matter from different perspectives, the emergence of cooperation or defection still remains a controversial argument from both empirical and theoretical point of views. In this paper an innovative model is presented and analyzed in the attempt to provide a reasonable framing of the issue. A population of boundedly rational agents repeatedly chooses to cooperate or defect. Each agent's action affects only her interacting mates, according to a network of relationships which is endogenously modifiable since agents are given the possibility to substitute undesired mates with unknown ones. Full cooperation, full defection and coexistence of both cooperation and defection in homogeneous clusters are possible outcomes of the model. A computer program is developed with the purpose of understanding the impact of parameters values on the type of outcome. Numerous simulations are run and the resulting evidence is analyzed and interpreted.

JEL classification: C63; C88; D85
Keywords: Prisoner Dilemma; cooperation; segregation; networks; simulation

Ennio Bilancini, Department of Economics, University of Siena
Leonardo Boncinelli, Department of Economics, University of Siena

1 Introduction

The contrast between cooperation and defection is usually applied to account for the dichotomy between a behavior which is superior from a societal point of view and another one which is superior from an individual point of view. The Prisoner Dilemma game is the standard representation embedding the strategic characteristics of such a dichotomy. The emergence of cooperation in the Prisoner Dilemma is often explained by means of the Folk theorem.[1] According to it, if interaction is infinitely or indefinitely repeated and rational agents use appropriate tit-for-tat strategies, then cooperation may emerge (Fudenberg and Maskin (1986), Kreps et al. (1982)).

In some cases such argument provides a plausible explanation of the emergence of cooperation in real world interactions. The application of a Folk theorem binds explanation to be given in terms of full rationality. However, this has the drawback of making the interaction setting highly inflexible. More precisely, in its basic version the Folk theorem requires the same strategic situation to be repeated an infinite or indefinite number of times and, in particular, the same players to participate in each repetition.

Real world interactions not always fit in such a rigid picture. For instance, there are contexts where the set of players is modifiable by the act of will of some individual or group. In others, an exit option or some kind of punishment are available. There are even more complex patterns of interaction where several groups of individuals simultaneously play a Prisoner Dilemma game and any player can decide, according to some rule, whether to exit or enter one ore more groups. With respect to explaining the emergence of cooperation in these cases, the Folk theorem is of little use.

In the last two decades, there have been several attempts to extend the result of the Folk theorem by introducing more realistic assumptions about the setup of the game. For this purpose, distinct sets of conditions under which a Folk theorem holds have been identified (see Gintis (2004) for a critical survey). Although this research line is very fascinating from both technical and philosophical perspectives, in our opinion – and for what concerns the issue under consideration – it is not worth the

[1]The survival of cooperation has been studied from different perspectives. It is not the aim of this paper to provide a survey of such numerous attempts. However, it is worth mentioning, besides the cited approach in repeated games with fully rational agents, the stream of biological models where behaviors are defined as traits evolving through a selection process (Trivers (1971) and Dawkins (1976) for a broad discussion on the topic).

effort. The reason is that the more realistic are the assumptions about the setup of the game, the more demanding, and hence the less realistic, is the requisite of full rationality. The derivation of new versions of the Folk theorem based on more realistic assumptions have therefore produced the main effect of moving the scarce realism of assumptions from the sphere of the game setup to that of agents' cognitive capabilities.

Our opinion is that cooperative behaviors may be better analyzed and understood in a framework where a population of self-regarding agents interacts in a boundedly rational way – i.e. with limited cognitive capabilities – but where agents can also modify, to some extent, the structure of local relationships they are involved in.[2]

The present contribution is innovative under two respects if compared with the existing literature. The first novelty is the modeling of agents' interaction. Indeed, we give agents not only the chance to choose between cooperation and defection, but also to modify the composition of their neighborhood, making the entire pattern of interactions endogenous. For this purpose, we model the interaction structure – that is, who interacts with whom – through a network where agents are nodes and links represent bilateral interactions. We assume that each agent has a maximum number of neighbors. This is meant to take into account the physical constraint due to the finiteness of time which can be spent interacting with other people. A part from that, we impose no particular restriction on the interaction structure and, hence, our representation is more general than spatial or lattice-based ones. Locality of interaction is introduced in the usual way by assuming that the effects of agents' actions spread to all and only their neighbors. Finally, in each period of time agents are randomly given the opportunity to cease some existing relationships[3] and, if it is feasible, to form new ones.[4]

[2]The idea that locality of interaction and cooperative behaviors might be related is not novel. Among others, Eshel et al. (1998) and Bergstrom and Stark (1993) consider agents arranged in a circle each interacting with her two immediate neighbors. Jun and Sethi (2004) adopt the same spatial structure but let agents interact with a parameterized number of neighbors, varying that parameter to analyze the effects. In Eshel et al. (1998) agents are arrayed in a plane rather than along a line. Many of the models in this stream of literature take imitation as the driving force behind strategy selection.

[3]The possibility for a cooperator to disconnect a defector may be interpreted as a form of targeted punishment.

[4]Zimmermann et al. (2004) proposed a model somehow close to ours in representing the interaction structure through an evolving network. A part from other differences, it is worth underlying that in their model, unlike ours, behavior is adapted simply by imitation of the neighbor with a highest pay-off and, above all, only defecting agents are given the possibility to break a link.

The second novelty concerns the aim of the paper. Instead of being only concerned with sustainability of cooperation, we consider more valuable to investigate the joint emergence of cooperation and behavioral segregation. As a matter of fact, the literature provides evidence in favor of cooperative outcomes (Boyd and Richerson (2004)) as well as in favor of non-cooperative ones. Obviously, economists consider the latter less remarkable since defection is not a very surprising outcome once the one-shot version of a Prisoner Dilemma game is considered. Therefore, economists' focus is mainly on the emergence of cooperation and not much attention is given to whether cooperation and defection coevolve and how this may happen. In our opinion, a better insight can be obtained by looking at cooperation and defection as jointly emerging from the evolution of the interaction structure. More precisely, we expect cooperation to prevail among some agents and defection among some others, people being eventually clustered on the basis of their behavior as a consequence of many uncoordinated individual decisions. Notice that this outcome is particularly likely if relationships between cooperators tend to last more than those involving at least one defector. Interestingly, this may provide a further explanation of the sustainability of cooperation on a payoff basis. The coexistence of clusters of cooperators and defectors may prevent cooperators from deviating because if they defect there is a chance of being excluded from a cooperating cluster.

The following is a summary of the rest of the paper. Section 2 introduces the model. A population of agents is arranged in a network describing the interaction structure. In every period each agent is either a cooperator or a defector, takes a benefit for each cooperator in her neighborhood, and sustains a cost if she cooperates. Agents are randomly selected to update their behavior. When selected an agent decides about whether to cooperate or defect according to a simple boundedly rational optimization. In addition, agents are randomly given the possibility to cease their existing interactions and start new ones with unknown individuals, up to their maximum number. Section 3 provides preliminaries and definitions. In particular, an intuitive measure of behavioral clustering is introduced and applied in order to define an appropriate concept of equilibrium which takes into account the specific features of our framework.

The complexity of the situation we model prevents us from providing a complete analytical characterization of solutions. Nevertheless, some results are derived. First, the system always converges in probability towards an equilibrium. Second, in any equilibrium with at least one cooperator there is a significant lower bound for behavioral clustering. This implies that a certain degree of behavioral segregation emerges in all cases where both cooperation and defection survive. Moreover, since there exist equilibria where groups of cooperators and groups of defectors coexist and equilibria where only one behavior survives, we are interested in understanding the main determinants of either type of equilibria. For this purpose, we use computational means to investigate their frequency of emergence. In Section 4 we report the results of thousands of simulations and provide evidence for the emergence, persistence and coexistence of clusters of cooperators and clusters of defectors in a significant number of cases. Finally, in Section 5 we conclude with a summary of results. Proofs are given in Appendix A.

2 The Model

Description. Let N denote the finite agents' set with $n \equiv ||N||$ its cardinality.

Let the network of connections among agents be represented by a $n \times n$ adjacency matrix G, with its generic element g_{ij} such that $g_{ij} = 1$ if agent i is linked to agent j, $g_{ij} = 0$ otherwise. With the purpose of making G correctly represent reciprocal relationships, we impose irreflexivity, $g_{ii} = 0$ for any $i \in N$, and symmetry, $g_{ij} = 1$ implies $g_{ji} = 1$ for any $i, j \in N$. Moreover, we assume a physical constraint on the number of interactions an individual can have; we denote such a bound with m.

Two individual behaviors can be adopted, namely cooperation and defection. The vector $V \in \{0, 1\}^n$ represents the collection of behaviors over the entire population, with 1 and 0 standing for cooperation and defection respectively.

We refer to a triple (N, G, V) as an interaction state.

Let $n_i \equiv \sum_j g_{ij}$ be the number of people interacting with i – her neighbors – and $n_i^1 \equiv \sum_j g_{ij} V_j$ be the number of cooperating ones. In any interaction state (N, G, V) every agent $i \in N$ gets a benefit b from each of her cooperating neighbors. Moreover any cooperator suffers a loss of l, due to the effort of cooperating. Hence agent i's payoff may be written as

$$\pi_i(N, G, V) = bn_i^1 - V_i l$$

Dynamics. We assume a fixed population where individuals have the possibility to revise both their behavior and the composition of their neighborhood.

Time is discrete. The dynamic process undergone by the system at each time can be obtained by the sequential application of the steps illustrated in figure 1.

1. An agent is randomly selected to update her behavior.

2. Every agent can sever each of her existing links with probability $p \in (0, 1)$. All the severing decisions are taken simultaneously.

3. Once disconnections have been carried out, agents having less than m connections decide whether to enter the market for new connections. Requests are then randomly matched and satisfied if possible, until no more connections are feasible.[5]

4. Payoffs are calculated and distributed to agents.

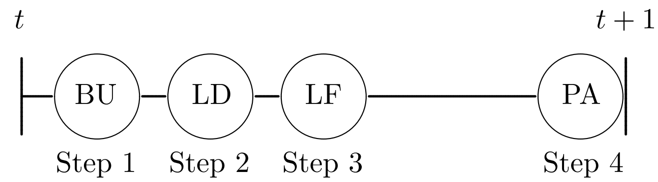

Figure 1. Behavior update (BU), link disconnection (LD), link formation (LF) and payoff assignment (PA) sequentially occur.

Agents are assumed to be boundedly rational in the following sense.

1. *Local knowledge*: besides the rules of the game, an agent knows only her behavior and the behaviors of her neighbors at instant t.

2. *Laplacian reasoning*: agents adopt Laplace decision rule which assigns equal probability to every interaction state when no information about the relative likelihood of interaction states is available.

3. *1-period-looking*: each agent takes into consideration only the payoff paid at instant $t + 1$.

Agents make their choices in the attempt to maximize expected payoff, where the latter is calculated according to the limited cognitive capabilities described above.

[5]Notice that some requests may remain unsatisfied when, among those willing to connect, there are only agents that are already linked together.

In addition, weakly dominated actions are always discarded and this is assumed to be known by all agents. Furthermore, population size is assumed to be large enough in order to obtain the negligibility of i) the fraction of agents with less than m connections after step 3, and ii) the impact of private information in the process of forming expectations about the current interaction state.

We can now solve the decision problem an agent faces in each period of time. By backward reasoning, let us begin with step 3. Suppose agent i is linked to less than m neighbors. By Laplacian reasoning she assigns a positive probability to the existence of cooperators. Hence, if i requests for a new connection then there is the possibility that a cooperator disconnects one of her neighbors, applies for a new connection and is paired off with i. Therefore, for each of i's vacant slots, requesting for a new connection weakly dominates the alternative choice because it gives a null payoff if the agent is paired off with a defector, or a positive payoff if paired off with a cooperator, contra the null payoff of being alone. As a consequence, agents with less than m connections always apply for new ones.

Next, consider step 2. Suppose agent i is selected to possibly sever a certain link ij. If j is a defector then severing weakly dominates not severing because the former allows the request for a new connection – with best and worst cases as described in step 3 – contra the null payoff of being connected to a defector. For similar reasons, if j is a cooperator then not severing weakly dominates severing.

Let us now examine step 1. Suppose agent i is given the possibility to modify her behavior. By results of step 2 the pair cooperator-cooperator does not disconnect, the pair cooperator-defector disconnects, the pair defector-cooperator does not disconnect and the pair defector-defector disconnects. Notice that by virtue of Laplacian reasoning these pairs are reputed equiprobable. Therefore, the frequency of cooperators in the market for new connections expected by agent i is $1/4$ and the benefit of severing link ij is $b/4$. Hence, the payoff of agent i may be seen as composed of two parts. The first does not depend on the chosen behavior and is equal to the sum of $n_i^1 b$, representing the total benefits accruing from cooperating neighbors, and $p(m - n_i^1)b$, representing the expected benefits due to new connections coming from i's severing decisions. If i cooperates, the second part is equal to $-l$ which represents the individual loss of cooperating; instead, if i defects, the second part is equal to the sum of $p(m - n_i^1)b/4$, representing the expected value of new connections coming from the severing decisions of i's defecting neighbors, and $-pn_i^1 b/4$, representing the expected cost of losing cooperating neighbors. Summing up, $\pi_i(1)$ and $\pi_i(0)$ are what i expects, respectively, from cooperating and defecting

$$
\begin{aligned}
\pi_i(1) &= n_i^1 b + p(m - n_i^1)b/4 - l \\
\pi_i(0) &= n_i^1 b + p(m - n_i^1)b/4 + p(m - n_i^1)b/4 - p n_i^1 b
\end{aligned}
$$

Agent i chooses to cooperate whenever $\pi_i(1) > \pi_i(0)$, that is if $n_i^1 > l/pb + m/4$. On the contrary, agent i chooses to defect whenever $n_i^1 < l/pb + m/4$. When equality holds, agent i is indifferent and, as a tie-break rule, defection is assumed. We define n^* as the threshold number of cooperators in a neighborhood which induces cooperation, namely the smallest integer greater than $l/pb + m/4$.

3 Definitions and Analytical Results

Measures. In order to investigate the evolution of cooperation we need a measure of its spreading. A simple and natural one is the fraction of cooperators in the population. Let us indicate with $n^1 \equiv \sum_i V_i$ the number of cooperators in an interaction state (N, G, V). We refer to the ratio

$$C = \frac{n^1}{n}$$

as the measure of cooperation relative to a certain interaction state.

We are also interested in the degree of behavioral clustering. Intuitively, we qualify an interaction state as highly behaviorally clustered if interactions between agents who both cooperate or defect are sensibly more frequent than interactions between cooperators and defectors. More precisely, we focus on the number of interactions between individuals behaving in the same way with respect to the number of interactions between individuals behaving differently. The following measure of behavioral clustering is adopted. Let $n^{1,1} \equiv V'GV$ be the number of cooperator-to-cooperator links and let $n^{1,0} \equiv V'G(e - V)$ be the number of cooperator-to-defector links, where e is a vector with all elements equal to 1. Let $n^{0,1}$ and $n^{0,0}$ be defined similarly. Finally, let $n^{1,01} \equiv V'Ge$ and $n^{0,01} \equiv (e - V)'Ge$ be the number of cooperator-to-anyone links and the number of defector-to-anyone links respectively. The matrix

$$B = \begin{bmatrix} n^{1,1}/n^{1,01} & n^{1,0}/n^{1,01} \\ n^{0,1}/n^{0,01} & n^{0,0}/n^{0,01} \end{bmatrix}$$

denotes the behavioral clustering of an interaction state. The first row of B represents the fraction of existing links of cooperators connecting to other cooperators – entry b_{11} – and to defectors – entry b_{12}. Similarly, the second row of B represents the fraction of existing links of defectors connecting to cooperators – entry b_{21} – and to other defectors – entry b_{22}. Clearly, the first row is undetermined when there are no cooperators while the second row is undetermined when there are only cooperators. Finally, notice that each row sums up to one.

Equilibrium. We proceed to define and comment an equilibrium notion which is appropriate to the dynamics into analysis. Let us indicate with (N, G^t, V^t) the interaction state at time t.

DEFINITION 1 *[Equilibrium]*

An interaction state $(N, G^{\bar{t}}, V^{\bar{t}})$ *is an equilibrium if and only if*

$$\begin{array}{ll} 1) & \forall t > \bar{t},\ \ V^t = V^{\bar{t}} \\ 2) & \forall t > \bar{t},\ \ n^{1,1}(N, G^t, V^t) = n^{1,1}(N, G^{\bar{t}}, V^{\bar{t}}) \end{array}$$

The first condition requires the constancy over time of agents' behaviors. This implies that in equilibrium the measure of cooperation C must be constant. The second condition is meant to capture the notion of stability for the relevant aspects of the interaction network G. Connections involving al least a defector will never be stable while connections between cooperators only will never be broken. Therefore, it seems reasonable to define an equilibrium notion only with respect to the latter, more precisely by requiring the infeasibility of new connections between cooperators. In conclusion, if what concerns is an aggregate and impersonal description of an interaction state, then the second condition of the above definition seems to capture the gist of network stability.

Equilibria: existence and convergence. At this stage we deal in greater detail with equilibria, investigating the issues of both existence and convergence and providing a further characterization of their properties.

The existence of at least one equilibrium is easily established by considering a limit case. Consider any interaction state where $n^1 = 0$. Any agent i will never change her behavior because $n_i^1 = 0 < n^*$. Since cooperators do not exist and will never exist, the number of links between cooperators is trivially constant and equal to zero in any period from now on.

Absolute convergence of the system to some equilibrium state is not ensured. However, we can prove convergence in probability. In order to get such a result we crucially exploit the finiteness of the state space and the positiveness of probability associated to any finite sequence of states. The detailed proof is given in the appendix.

Proposition 1 (Convergence in Probability)
As time goes to infinity, any interaction state converges almost surely to an equilibrium.

Proposition 1 tells us that sooner or later an equilibrium interaction state emerges. By focusing on equilibria we can assess the long run behavior of the system. Therefore, we turn to the investigation of equilibrium characteristics.

For $n^* > m$ and for $n^* = 0$ any equilibrium interaction state (N, G, V) must satisfy, respectively, the condition $C = 0$ and $C = 1$.[6] For $0 < n^* \leq m$ we cannot exclude any value of C for equilibria, ranging in principle from 0 to 1. In addition, notice that the system we are dealing with is non-ergodic, meaning that initial conditions matter for equilibrium selection and, in particular, for the value that C will assume.

As regards behavioral clustering we have already noticed that the rows of B sum up to one, which allows to restrict attention to b_{11} and b_{22}. However, the same reasons behind the restriction of network stability to connections among cooperators only, suggest to consider b_{11} the opportune index of behavioral clustering. Therefore, we turn our attention to the range of values b_{11} can take. Trivially, if $C = 0$ then b_{11} is indeterminate. Moreover, in equilibrium any cooperator is satisfied with her current choice implying that all cooperators must have at least n^* cooperating neighbors. Hence, if $C > 0$ then $b_{11} \geq n^*/m$. This bound can be refined exploiting the fact that i) the number of cooperators with less than m cooperating neighbors is at most $(n^* - 1)$, because otherwise some defector could become a cooperator, and ii) there must be at least $(n^* + 1)$ cooperators, since there exists a cooperator and she must have at least n^* cooperating neighbors. In the proof of Proposition 2 we carry out such refinement obtaining a function of n^1, n^* and m whose infimum is easily computed.

[6]The case $n^* = 0$ is considered for completeness, but it is impossible in our model of individual decision-making since $n^* > l/pb + m/ > 0$.

Proposition 2
Any equilibrium interaction state with $C > 0$ must satisfy

$$b_{11} \geq \frac{3}{4}$$

Notice that if there are only cooperators b_{11} is trivially equal to 1.

The analytical results we have exposed do not provide a full explanation of the subject we are addressing. In our model there are equilibria where clusters of cooperators and clusters of defectors coexist and equilibria where only one behavior survives. We refer to the former as *mixed* equilibria and to the latter as *pure* ones. Our aim is to understand which conditions – that is, which parameters values – favor the emergence of segregated clusters of cooperators and defectors instead of the achievement of complete cooperation or defection. By proposition 2 we know that in mixed equilibria there is a high degree of behavioral clustering.

However, because of the complexity of the dynamic system under consideration, we found extremely difficult to obtain an analytical characterization of the effects that parameters have on the frequency of appearance of mixed and pure equilibria. For this reason we run thousands of simulations in order to collect data suitable for analysis by induction.

4 Simulation results

A first examination of the dynamic rules of the system brings us to the following observations:

- the threshold n^* matters for individual choice of behavior and hence presumably affects where the system tends to,
- parameters l, b, p and m determine the value of n^*,
- the probability p is also relevant for the rate of renewal of connections, possibly having further non-trivial effects on aggregate outcomes,
- population size n and neighborhood dimension m may affect the emerging configuration of the system for combinatorial reasons.

Let us concentrate first on n^* and p. Since p also indirectly influences the system by modifying n^*, we choose to counterbalance this effect by adequately adjusting l and b. Unlike m, in fact, l and b do not exert any other influence on the system and therefore they suit such adjusting role. For the same reason l and b are of no interest apart from their effect on n^*. Three hundreds simulations have been run for several vectors of parameters values, where n^*, p (and consequently l/b) vary while n and m are kept fixed at 30 and 7 respectively.[7]

Table 2 in appendix B shows the number of mixed, pure cooperating and pure defecting equilibria evidencing that

- p is positively correlated with the emergence of mixed equilibria,
- the frequency of mixed equilibria first increases as n^* increases and then it decreases when n^* gets over 4.

These results can be explained as follows. Notice that the possibility to sever a link is exploited only when that link connects to a defector. Hence, the higher is p the greater is the robustness of cooperator-to-cooperator links compared with other kinds of links. Therefore, a high value of p favors the formation of self sustaining clusters of cooperators. Moreover, by isolating cooperators p also hinders cooperation from spreading over the entire population. As regards n^*, notice that 4 cooperators represents the middle value in the range of variation of n^*.[8] Then, it seems plausible that the closer n^* to its middle value the higher the probability neither cooperation nor defection prevails.

With the aim of better understanding the type of influence exerted by n^* and p, we carry out some simple OLS estimates with different specifications using p and $|n^* - 4|$ as basic regressors. In particular, we were doubtful whether the influences of $|n^* - 4|$ and especially p were more than proportional and whether there was a separable joint effect. As table 1 shows, the explanatory power of the model measured by R^2 decreases when one, the other or both regressors are modified by an exponential[9] transformation. Moreover, the explanatory power (obviously) increases when $p \cdot |n^* - 4|$ is added, but very slightly and the new regressor is not significantly different from zero. For these reasons we conclude that the linear dependence seems to best fit and a separable joint effect is unlikely to exist.

[7] A higher number of simulations might have been run and/or greater values for n and m might have been used. However, we noticed that by progressively raising n and m the qualitative meaning of results was not changing while the convergence time was dramatically increasing. Moreover, after 300 runs we found that results varied very slightly.

[8] In fact, an agent can always cooperate, or cooperate if she has a number of cooperating neighbors at least equal to 1, or 2, or 3, or 4, or 5, or 6, or 7, or never cooperate. Hence, the range of variation of n^* counts 9 different possibilities (since $m = 7$) with 4 its middle value.

[9] We tried other types of tranformations getting similar results which, therefore, have not been reported.

Table 1. OLS estimates

regressors	coeff.	prob.	R^2
c	58.99	0.00	0.758
p	67.24	0.00	
$\lvert n^* - 4\rvert$	-34.57	0.00	
c	24.56	0.03	0.732
$\exp p$	38.49	0.00	
$\lvert n^* - 4\rvert$	-34.57	0.00	
c	58.01	0.00	0.735
p	67.24	0.00	
$\exp \lvert n^* - 4\rvert$	-9.55	0.00	
c	23.59	0.46	0.709
$\exp p$	38.49	0.00	
$\exp \lvert n^* - 4\rvert$	-9.55	0.00	
c	53.48	0.00	0.766
p	85.37	0.00	
$\lvert n^* - 4\rvert$	-29.98	0.00	
$p \cdot \lvert n^* - 4\rvert$	-15.11	0.27	

Let us now try to establish which contribution n and m give to the emergence of mixed equilibria. Computer simulations become extremely time-demanding when both n and m increase, posing serious constraints to the extent of our investigations. For this reason, we restrict our attention to the impact of m/n, which we imagine is the key parameter here. A further problem is constituted by the fact that, as n increases, the velocity of connections renewal increases relatively to the velocity of behavior update if just one agent per period is allowed to modify her behavior, as the model setup provides. Since we are interested in the net effect of m/n we counterbalance this by proportionally raising the number of agents who are allowed to change behavior.

Three hundreds simulations have been run for various population sizes and for three couples of p and n^*, while keeping m fixed to 7. Results are in tables 3, 4 and 5, suggesting that an increase in population size implies a slight increase in the number of mixed equilibria. Intuition provides ambiguous arguments for this result. If n rises then more clusters can form but each possible cluster is less likely. In any case, the outlined influence of m/n requires further investigation to be better understood.

5 Conclusions

In this paper we show how clusters of cooperators and clusters of defectors can emerge from a single population, as the outcome of many uncoordinated individual decisions. The key element we introduce is the individuals' ability to affect the composition of their neighborhood. In particular, agents have the chance to substitute undesired neighbors. Such enrichment of the strategic framework have important consequences. First, individuals cease relationships with defectors and preserve those with cooperators. Therefore, the only source of instability for interactions between cooperators is a change of behavior in favor of defection. Second, individuals benefit if disconnected by defecting mates and lose if disconnected by cooperating mates. Hence, the value of cooperation (defection) is positively (negatively) affected by the number of cooperators in the neighborhood, generating a sort of conformity effect and increasing the likelihood that cooperators do not change behavior. These two facts imply that cooperators are likely to aggregate in clusters, segregating themselves from the rest of population. More precisely, we show that the system converges almost surely to an equilibrium where, if there is at least a cooperator, then not less than 3/4 of all relationships cooperators have are with other cooperators.

Furthermore, we investigate the frequency of emergence of equilibria where both cooperation and defection survive and, in particular, how it depends on parameters values. By means of simulations two main determinants are found.[10] The first is the rate of links renewal, whose increase has the effect of raising the instability of relationships involving at least one defector, hence decreasing the relative instability of relationships between cooperators and favoring their isolation. The second is the ratio between the threshold for cooperation and the neighborhood size, whose distance from the half makes more likely the survival of a single behavior. Intuitively, if either too many or too few cooperating neighbors are required to make cooperation convenient then it is likely that, respectively, either everybody defects or everybody cooperates. In addition, we found evidence of a slight impact of the relative maximum number of relationships that individuals can have. In particular, a smaller size of neighborhoods with respect to that of population seems to increase the emergence of behavioral clusters. This result is not totally satisfying. In fact, although a smaller relative size of neighborhoods allows for more clusters, each of them is less likely. We suspect there may be combinatorial issues behind

[10]Simulations are done using an *ad hoc* computer program developed by the authors. Both the executable file and C++ source codes are available on request.

this outcome and, in any case, in order to have a better understanding of the phenomenon more simulations must be run with greater sizes of both population and neighborhoods.

The next step along this line of research is to introduce idiosyncratic elements into agents' decision problems and to investigate in which equilibrium states the system is likely to spend most time. Our suggestion is to allow for random perturbations of both behavior and connections, taking into account the possibility for agents to make all kinds of mistakes and making the system ergodic. This would permit the study of the stochastically stable distribution, allowing for substantial selection among the vast set of equilibria.

A Proofs

Proof of Proposition 1. Let $Q^T(N,G,V)$ be the probability that, starting from an interaction state (N,G,V), the system will be in an equilibrium state in T periods.[11]

Notice that, if $Q^T(N,G,V) \geq q > 0$ for any (N,G,V), we have

$$\forall l \geq 1, \quad 0 \leq \prod_{m=1}^{l} \left(1 - Q^T\left(g^{T(m-1)}, V^{T(m-1)}\right)\right) \leq (1-q)^l$$

where the term in the middle of the above expression is the probability the system will not converge to a stable state in lT periods. Clearly, taking the limit for $l \to \infty$ such a probability goes to 0.

We are left to show that $Q^T(N,G,V) \geq q > 0$ for any (N,G,V) and we will do that in three steps; the first two steps allow to assert that with positive probability a state with certain properties is reached in a finite number of periods whatsoever the initial state, while the third step simply consists of recognizing that the state that has been reached is indeed an equilibrium. Let us first introduce some of definitions which will be used in the following.

The set of *always cooperating cooperators* is $C(N,G,V) \equiv \{i \in N : V_i^t = 1, \forall t \geq 0\}$, the collection of those players that are cooperating in the current state (N,G,V) and will surely be cooperating i any future state according to the dynamics described in the paper.

A *sub-state* (M,G,V) with $M \subseteq N$ is the restriction of a state to a certain subset of players where only modalities of and links between them are considered. Finally, a sub-state of always cooperating cooperators $(C(N,G,V),G,V)$ is called *unmodifiable* if and only if $(C(N,G^t,V^t),G^t,V^t)(C(N,G,V),G,V)$ for all $t \geq 0$, that is the set of always cooperating cooperators remains the same forever and no connections are created or destroyed between them.

Step I. There exist $t_1(N,G,V)$ and $\alpha_1(N,G,V) > 0$ such that starting from (N,G,V) the probability to be after t_1 periods in a state (N,G',V') such that $(C(N,G',V'),G',V')$ is unmodifiable is at least α_1.

[11] Here and in the following (N,G,V) has to be intended as $(\bar{N},G,V)$, where only G and V are left to vary while N is exogenously fixed and constant over time. Moreover, (N,G,V) without any apix is used as shorthand for (N,G^0,V^0), that is the interaction state at time 0.

Let us prove the above statement. Ad absurdum, suppose that starting from (N, G, V) for all t the probability to be after t periods in a state (N, G', V') such that $(C(N, G', V'), G', V')$ is unmodifiable is 0. Therefore, the current sub-state $(C(N, G, V), G, V)$ is not unmodifiable; this means that there exists $\tilde{t}$ such that $(C(N, G^{\tilde{t}}, V^{\tilde{t}}), G^{\tilde{t}}, V^{\tilde{t}}) \neq (C(N, G, V), G, V)$ with positive probability $\tilde{\alpha}$. The sub-state $(C(N, G^{\tilde{t}}, V^{\tilde{t}}), G^{\tilde{t}}, V^{\tilde{t}})$ does not have to be unmodifiable either, therefore applying the same reasoning as before another modifiable sub-state is obtained after a certain length of time with positive probability. This sequence of modifiable sub-states has to be infinitely long. However, this sequence does not admit cycles, because any always cooperating cooperator will always be a cooperator, and any link between always cooperating cooperators will remain forever since a link between cooperators is never destroyed and they will always remain cooperators. The infiniteness of a sequence without cycles is in contradiction with the finiteness of the state space.

There exist $\bar{t_1}$ and $\bar{\alpha_1} > 0$ such that starting from any (N, G, V) the probability to be after $\bar{t_1}$ periods in a state (N, G', V') such that $(C(N, G', V'), G', V')$ is unmodifiable is at least $\bar{\alpha_1}$.

For the proof of this statement it is sufficient that $\bar{t_1}$ is the maximum $t_1(N, G, V)$ for any (N, G, V), and $\bar{\alpha_1}$ is the minimum $\alpha_1(N, G, V)$ for any (N, G, V), with the existence of $\bar{t_1}$ and $\alpha_1(N, G, V)$ ensured by the finiteness of the state space. In order to be convinced notice that, given $t_1(N, G, V)$ and $\alpha_1(N, G, V)$, then trivially for any $t \geq t_1$ the probability to be after t periods in a state whose sub-state of always cooperating cooperators is unmodifiable is at least α_1.

Step II. If $(C(N, G, V), G, V)$ is unmodifiable then there exist $t_2(N, G, V)$ and $\alpha_2(N, G, V) > 0$ such that starting from (N, G, V) the probability to be after t_2 periods in a state (N, G', V') such that if $i \notin C(N, G', V')$ then $V_i = 0$ is at least α_2.

Suppose not and take a state which has the minimum number of cooperators among those states reachable with positive probability. Such a state exists by the finiteness of the state space. There will be cooperators who are not belonging to $C(N, G, V)$, by the absurd hypothesis, and none of them can be willing to change behavior, since otherwise another state with an inferior number of cooperators would be reachable with positive probability. Any cooperator has therefore a sufficient number of cooperators to voluntarily cooperate. Because connections between cooperators are never broken, those cooperators will always be cooperating and the set $C(N, G, V)$ would not be unmodifiable, against the initial hypothesis.

For all (N, G, V) if $(C(N, G, V), G, V)$ is unmodifiable then there exist $\bar{t_2}$ and $\bar{\alpha_2} > 0$ such that starting from (N, G, V) the probability to be after $\bar{t_2}$ periods in a state (N, G', V') such that if $i \notin C(N, G', V')$ then $V_i = 0$ is at least $\bar{\alpha_2}$.

The proof of this statement consists of a simple check. Let $\bar{t_2}$ be the maximum $t_2(N, G, V)$ for any (N, G, V) such that $C(N, G, V)$ is unmodifiable, and let $\bar{\alpha_2}$ be the minimum $\alpha_2(N, G, V)$ for any (N, G, V) such that $C(N, G, V)$ is unmodifiable, with the existence of $\bar{t_2}$ and $\bar{\alpha_2}$ ensured by the finiteness of the state space. In order to see why this is true, notice that given $t_2(N, G, V)$ and $\alpha_2(N, G, V)$, then for any $t \geq t_2$ the probability to be after t periods in a state with all defectors except always cooperating cooperators is at least α_2, since no cooperator can emerge after that all non always cooperating agents have become defectors, otherwise being linked only with always cooperating agents and hence always cooperating herself.

Step III. Starting from any state (N, G, V) in $\bar{t_1} \cdot \bar{t_2}$ periods with at least probability $\bar{\alpha_1} \cdot \bar{\alpha_2}$ the system will reach a state (N, G^*, V^*) where the sub-state $(C(N, G^*, V^*), G^*, V^*)$ is unmodifiable and every non always cooperating cooperator is a defectors. Such a state (N, G^*, V^*) is an equilibrium, according to definition 1. In fact, always cooperating cooperators will cooperate forever, no cooperator can emerge among defectors, and no new connection between cooperators can be established since $(C(N, G^*, V^*), G^*, V^*)$ is unmodifiable. Hence, by setting $T = \bar{t_1} \cdot \bar{t_2}$ and $q = \bar{\alpha_1} \cdot \bar{\alpha_2}$ we get the desired result. Q.E.D.

Proof of Proposition 2. As previously defined, n^1 denotes the number of cooperators. Moreover, let $\hat{n}^1$ indicate the number of cooperators who have a full neighborhood of cooperators, and let $\tilde{n}^1$ indicate the remaining ones, $\tilde{n}^1 \equiv n^1 - \hat{n}^1$.

Being in equilibrium, any cooperating agent is willing to cooperate, and hence she has at least n^* cooperating neighbors. Therefore $n^{1,1} \geq \tilde{n}^1 n^* + \hat{n}^1 m$.

If $C > 0$ then $n^1 \geq 1$. Moreover, since any cooperator has at least n^* cooperating neighbors $n^1 \geq n^* + 1$. In equilibrium it is also true that at most $n^* - 1$ cooperators have a non full neighborhood, $\tilde{n}^1 \leq n^* - 1$, in order for any defector not to have a chance to become a cooperator. Clearly, given n^1 the higher $\tilde{n}^1$ the lower the bound for $n^{1,1}$, so $n^{1,1} \geq (n^* - 1)n^* + (n^1 - n^* + 1)m$. Moreover, at least one cooperator must have a full cooperating neighborhood, implying that at least $m+1$ cooperating agents exist, $n^1 \geq m + 1$.

The number of cooperator-to-anyone links, denoted by $n^{1,01}$, is limited by the number of cooperators multiplied by the maximum neighborhood size, $n^{1,01} \leq n^1 m$. The following bound

$$b_{11} = \frac{n^{1,1}}{n^{1,01}} \geq \frac{(n^* - 1)n^* + (n^1 - n^* + 1)m}{n^1 m} \tag{1}$$

is increasing in n^1 and, being interested in its minimum value, we set $n^1 = m+1$. Therefore,

$$b_{11} \geq \frac{n^*(n^* - 1) + m(m - n^* + 2)}{(m+1)m} \tag{2}$$

It is easy to check that the above expression, considered as a function of n^*, gets its minimum value for $n^* = (m+1)/2$. By simple substitution into the expression (2), we get that

$$b_{11} \geq \frac{3m^2 + 6m - 1}{4(m+1)m} > \frac{3}{4} \tag{3}$$

Q.E.D.

B Tables

Table 2. The Effect of p and n^* for $n = 30$, $m = 7$

p	l/b	n^*	Mixed Eq.	Coop. Eq.	Def. Eq.	% Mixed Eq.
0.03	0	2	3	297	0	1%
0.05	0	2	12	288	0	4%
0.09	0	2	29	271	0	10%
0.14	0	2	33	267	0	11%
0.22	0	2	29	271	0	10%
0.35	0	2	40	260	0	13%
0.60	0	2	75	225	0	25%
0.95	0	2	145	155	0	48%
0.03	0.01	3	3	295	2	1%
0.05	0.02	3	27	272	1	9%
0.09	0.03	3	109	189	2	36%
0.14	0.04	3	192	108	0	64%
0.22	0.08	3	231	68	1	77%
0.35	0.10	3	279	21	0	93%
0.60	0.15	3	291	9	0	97%
0.95	0.25	3	297	3	0	99%
0.03	0.05	4	102	94	104	34%
0.05	0.09	4	146	60	94	49%
0.09	0.12	4	226	11	63	75%
0.14	0.18	4	251	7	42	84%
0.22	0.30	4	274	0	26	91%
0.35	0.50	4	287	0	13	96%
0.60	0.80	4	299	0	1	99%
0.95	1.20	4	299	0	1	99%
0.03	0.07	5	3	3	294	1%
0.05	0.14	5	12	1	287	4%
0.09	0.23	5	18	0	282	6%
0.14	0.32	5	46	0	254	15%
0.22	0.55	5	90	0	210	30%
0.35	0.85	5	156	0	144	52%
0.60	1.50	5	229	0	71	76%
0.95	2.50	5	279	0	21	93%
0.03	0.11	6	0	0	300	0%
0.05	0.19	6	0	0	300	0%
0.09	0.31	6	0	0	300	0%
0.14	0.58	6	0	0	300	0%
0.22	0.85	6	1	0	299	0%
0.35	1.30	6	0	0	300	0%
0.60	2.00	6	41	0	259	14%
0.95	4.00	6	95	0	205	32%

Table 3. The Effect of Relative Neighborhood Size for $m = 7$, $p = 0.09$, $n^* = 3$

m/n	**Mixed Eq.**	**Coop. Eq.**	**Def. Eq.**	**% Mixed Eq.**
0.23	116	182	2	39%
0.12	132	168	0	44%
0.08	135	165	0	45%
0.06	147	153	0	49%
0.05	129	171	0	43%
0.04	146	154	0	49%
0.03	150	150	0	50%

Table 4. The Effect of Relative Neighborhood Size for $m = 7$, $p = 0.03$, $n^* = 4$

m/n	**Mixed Eq.**	**Coop. Eq.**	**Def. Eq.**	**% Mixed Eq.**
0.23	75	120	105	25%
0.12	140	66	94	47%
0.08	173	40	87	58%
0.06	220	23	57	73%
0.05	244	18	38	81%
0.04	255	6	39	85%
0.03	261	6	33	87%

Table 5. The Effect of Relative Neighborhood Size for $m = 7$, $p = 0.22$, $n^* = 5$

m/n	**Mixed Eq.**	**Coop. Eq.**	**Def. Eq.**	**% Mixed Eq.**
0.23	104	0	196	35%
0.12	114	0	186	38%
0.08	128	0	172	43%
0.06	137	0	163	46%
0.05	115	0	185	38%
0.04	135	0	165	45%
0.03	138	0	162	46%

References

T. Bergstrom and O. Stark. "How Altruism Can Prevail in an Evolutionary Environment". *American Economic Review Papers and Proceedings*, 83:149–155, 1993.

R. Boyd and P. J. Richerson. *The Origin and Evolution of Cultures*. Oxford University Press, New York, 2004.

R. Dawkins. *The Selfish Gene*. Oxford Univresity Press, Oxford, 1976.

I. Eshel, L. Samuelson, and A. Shaked. "Altruists, Egoists, and Hooligans in a Local Interaction Model". *American Economic Review*, 88:157–179, 1998.

D. Fudenberg and E. Maskin. "Neighborhood Structure and the Viability of Cooperation". *Econometrica*, 54:533–554, 1986.

H. Gintis. "Modeling Cooperation Among Self-Interested Agents: A Critique". *The Journal of Socio-Economics*, 33:695–714, 2004.

T. Jun and R. Sethi. "Equilibrium Concepts for Models with Social Interactions", 2004. Mimeo. Columbia University.

D. M. Kreps, P. Milgrom, J. Roberts, and R. Wilson. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma". *Journal of Economic Theory*, 27: 245–252, 1982.

R. L. Trivers. "The Evolution of Reciprocal Altruism". *Quarterly Review of Biology*, 46:35–57, 1971.

M. G. Zimmermann, V. M. Eguíluz, and M. S. Miguel. "Cooperation, Adaptation and the Emergence of Leadership", 2004. Mimeo.