Journal of Bioeconomics 6: 165–194, 2004

Coexistence of Strategies and Culturally-Specific Common Knowledge: An Evolutionary Analysis

ANGELO ANTOCI
DEIR, University of Sassari, Via Sardegna 58, 07100 Sassari, Italy (angelo.antoci@virgilio.it)

PIER LUIGI SACCO
DADI, IUAV, Dorsoduro 2206, Convento delle Terese, 30123 Venice, Italy (sacco@iuav.it)

LUCA ZARRI
Department of Economics, University of Verona, Viale dell'Università 4, 37129 Verona, Italy and School of Economic and Social Studies, University of East Anglia, Norwich NR4 7TJ, UK (luca.zarri@univr.it and l.zarri@uea.ac.uk)

Synopsis: We analyze social dynamics in a continuous population where randomly matched individuals have to choose between two pure strategies only ('cooperate' (C) and 'not cooperate' (NC)). Individual payoffs associated with the possible outcomes of each interaction may differ across groups, depending on the specific social and cultural context to which each agent belongs. In particular, it is assumed that three sub-populations are initially present, 'framing' the game according to the prisoner's dilemma (PD), assurance game (AG) and other regarding (OR) payoff configurations, respectively. In other words, we assume that common knowledge about the payoffs of the game is 'culturally-specific'. In this context, we examine both the adoption process of strategies C and NC within each sub-population and the diffusion process of 'types' (PD, AG and OR) within the overall community. On the basis of an evolutionary game-theoretic approach, the paper focuses on the problem of *coexistence* of PD, AG and OR groups as well as of 'nice' (C) and 'mean' (NC) strategies. We show that coexistence between C and NC is possible in the heterogeneous community under examination, even if it is ruled out in homogeneous communities where only one of the three types is present.

Key words: prisoner's dilemma, social dynamics, culturally-specific common knowledge, subjective games, evolutionary game theory

JEL classification: C72, Z1

1. Introduction

Though the prisoner's dilemma (PD) has been extensively studied under a wide variety of conditions and perspectives (see, e.g. Kandori 1982, Rubinstein 1986, Binmore & Samuelson 1992, Ellison 1994), *coexistence* of strategies has rarely been obtained as a theoretical result. Eshel et al. (1999) consider a large population with a local interaction structure, where unrelated individuals often meet with their neighbors and are allowed to occasionally change their own strategy by imitating the most successful agents belonging to the interaction neighborhood. In this framework, they define as 'unbeatable' a strategy which turns out to be robust against the invasion of a finite group of identical mutants and find that, whenever agents play either PD or chicken game (CG), cooperation is the *unique* unbeatable strategy insofar as the learning neighborhood is far larger than the interaction neighborhood. Under very different conditions, Karandikar et al. (1998) obtain a somewhat similar conclusion in a model where two agents play the PD over time and follow an aspiration-based adjustment rule: such a process leads to the eventual emergence of the mutual cooperation outcome (i.e. even in this framework, coexistence is ruled out). However, Palomino & Vega-Redondo (1999) correctly point out that such a result crucially depends on the presence of inter-agent 'feedback effects' due to the small number of players involved in the game. In their paper they set up an aspiration-based dynamic model of bounded rationality where a continuum of agents are randomly matched and play the PD: under certain conditions, their analysis brings about an interesting coexistence result, as long-run *partial* cooperation (never exceeding half of the population) emerges as the unique limit outcome of social adjustment paths. Hirshleifer & Martinez Coll (1991) analyze the dynamics related to the adoption process of four pure strategies ('cooperate', 'tit for tat', 'defect' and 'bully') within a large population of utility-maximizing agents. The subjects have to choose which strategy to adopt in a series of random pairwise matchings with other individuals belonging to the same population. The four strategies are played both with payoff configurations of PD type and of CG type. However, it is assumed that the two games are played separately: they first consider the adoption process in a population where all the agents believe they are playing a PD game (whose payoff levels are known to all) and are all rationally maximizing their own payoff; subsequently, the same process takes place within a population where agents play a CG and believe this information to be common knowledge. In such a context, the adoption process of the above behavioral options leads to the coexistence of 'nice' strategies (such as 'cooperate' and 'tit for tat') and 'mean' strategies (such as 'defect' and 'bully'). As a consequence, their predictions are consistent with the following well-known experimental[1] and empirical result: despite (normally relevant) cultural and economic differences, in many *large* social environments, a *mixture* of 'nice' and 'mean' behaviors is often observed, as almost everywhere some people are honest, for example, tend to return valuable lost items, to tip in restaurants and to queue in the markets, whereas some other people belonging to the same population do not.[2] The same holds even for more proactive and morally demanding behaviors such as volunteering, contributing to charities, donating blood without monetary reward, voting and saving unknown people at the risk of one's own life, which are normally displayed by a positive fraction of the overall community under examination. Fehr & Gächter (1999), by referring to 16 different experimental studies, show that *reciprocally* and *selfishly* motivated people turn out to systematically coexist: in particular, they argue that in all scenarios[3] both types are present in non-negligible fractions, though the former seems to prevail. In the light of these observations, Hirshleifer & Martinez Coll's (1991) coexistence result as well as Palomino & Vega-Redondo's (1999) conclusion are quite interesting, especially if we think of the

PD environment. In such a large-population framework, if we focused our attention on the adoption process of the 'classical' two pure strategies only (namely 'cooperate' and 'defect'), coexistence between 'nice' and 'mean' strategies would be ruled out.

However, against the implicit assumption of the above cited authors, it is far from obvious that individuals always interact with each other on the basis of a *clear* and *shared* perception of the overall social structure they are embedded in.[4] In other words, the notion of 'common knowledge' turns out to be highly controversial and *context-dependent*, as each individual's ability to frame the social situation she is embedded seems to be the effect of complex factors, active at both the natural and cultural level. For these reasons, individuals are likely to differ in the way they conceptualize the game they are about to play and such differences are likely to be specific to each sub-population. In this paper, we mainly focus on the cultural dimension by assuming that common knowledge about the payoff matrix has a salience (see Schelling 1960, Sahlins 1972) that is crucially dependent on the value-system characterizing the different types of players composing the overall community. In fact, it is reasonable to believe that the idea of salience not only regards focal points (sometimes described in terms of common knowledge about non-rational impulses; see, e.g. Sugden 1991)[5] but, at a deeper level, the perception of the whole payoff structure of the game, depending on the locally prevailing social norms and cultural patterns. As Wildavsky (1992, p.12) points out, 'To the extent that individuals in different cultures value the same outcomes differently, there will be a different set of payoffs and, consequently, a different model of this situation for each culture. As a very crude first approximation, one might think of culture as being implicit in a model's payoffs.'

On the basis of this approach, the relationship between rationality and salience should be somehow reversed, with respect to traditional game-theoretical frameworks where salience is a sort of 'second-best resource' agents rely upon insofar as they fail to fruitfully coordinate their (individually rational) actions. On the contrary, in this paper we claim that agents, in the first place, tend to conceptualize the game they are about to play in a strongly culturally-dependent manner; then, at a second stage, rationality comes into the picture, inducing players to choose the best strategies available on the basis of their information set. Clearly, insofar as salience regards the framing problem, a fortiori it can be claimed to concern the more specific and conventional problem of focal points emergence. Thus, it is worth investigating the possibility of coexistence between 'nice' and 'mean' strategies in a different strategic context. In particular, we will focus on the following scenario: players have to choose one out of two strategies only (either 'cooperate' or 'not cooperate'); however, the possible outcomes of random pairwise matchings are differently evaluated by single agents, i.e. individuals are heterogeneous in terms of their perception of the payoff matrix of the game they are involved in. As we pointed out above, agents are homogeneous only within specific sub-populations characterized by common socialization patterns and salient social norms[6]: as far as the framing problem of the initial multi-population community is concerned, then, inter-group heterogeneity corresponds to intra-group homogeneity. In other words, we are still assuming common knowledge about the structure of the game to be played, but such a common knowledge is 'culturally-specific': each player's expectations about his opponent's behavior are systematically biased by his own reference culture and, therefore, confirmed only insofar as he happens to be matched with players of the same 'type'.[7] Landa's (1981) interesting fieldwork and theory on Chinese merchants in Southeast Asia shows that the emergence and economic success of the so called 'ethnically homogeneous middleman group' (EHMG) in those countries is closely related to Chinese merchants' attitude to selectively interact only with members of their own dialect group. In turn, such attitude is grounded in Confucian ethics, which is a complex set of cultural norms promoting mutual trust and aid among kinsmen.

The idea of common knowledge we are referring to recalls Lewis' (1969) definition concerned with justification and not with truth: what each person has reason to believe may be dependent on 'background information', which, we claim, is likely to depend in turn on his/her reference culture and social norms. This implies that agents' 'inductive standards' will be shared within each sub-population but will differ across them. Borrowing Furnivall's (1957, pp. 304–305) terminology, we may refer to such a community as a 'plural society', a 'medley' where people 'mix but do not combine. Each group holds by its own relation, its own culture and language, its own ideas and ways. As individuals they meet, but only in the marketplace in buying and selling.' Several experimental studies focusing on the effects of cultural background on game-theoretic behavior (see, e.g. Smith & Bond 1993) confirm that a variable such as culture crucially affects the set of reference behaviors individuals have in mind when playing standard games like PD.

In particular, let us assume that the whole community consists of three sub-populations (types) of payoff-maximizing individuals: (a) everybody perceives the game matrix as the classical PD payoff configuration; (b) agents believe that an assurance game (AG) is about to be played; and (c) the payoff matrix is given by the other regarding (OR) game structure. The purpose of our analysis is to consider the social dynamics taking place within such a complex environment. However, before introducing the formal model, it is preliminarily important to specify how the differences in terms of framing among the three sub-populations can be formalized and interpreted at the game-theoretic level. In order to do this, let us consider the following PD payoff matrix:

	Cooperate	Not cooperate
Cooperate	α, α	γ, δ
Not cooperate	δ, γ	β, β

where $\delta > \alpha > \beta > \gamma > 0$ and $(\delta - \alpha) < (\beta - \gamma)$. Two rational players (A and B) are involved. We define agent A as 'altruist' when her utility is given by a weighted average of her own and agent B's payoff: $U_A = (1 - w)\Pi_A + w\Pi_B$, where Π_i $(i = A, B)$ indicates i's payoff and w $(0 < w < 1)$ represents A's (as well as B's) 'degree of altruism' toward her opponent. If both players are characterized by such a utility function, the utility matrix of the (symmetric) game becomes:

	Cooperate	Not cooperate
Cooperate	α, α	$(1 - w)\, \gamma + w\, \delta, (1 - w)\, \delta + w\, \gamma$
Not cooperate	$(1 - w)\, \delta + w\, \gamma, (1 - w)\, \gamma + w\, \delta$	β, β

when $0 < w < w_1 = (\delta - \alpha)/(\delta - \gamma)$, we fall into the classic PD game, whereas when $w_1 = (\delta - \alpha)/(\delta - \gamma) < w < w_2 = (\beta - \gamma)/(\delta - \gamma)$, the AG structure emerges; finally, when $w > w_2 = (\beta - \gamma)/(\delta - \gamma)$, we obtain the OR game. In other words, the presence of three types of agents can be justified in terms of the perceived degree of altruism within one's reference sub-population: agents believe they are actually playing a PD, an AG or an OR game according to the level of w being low (equal to zero in the limit), intermediate or high (equal to one in the limit), respectively. The idea is that in a group where pro-social values are traditionally rooted and widespread, it is reasonable to assume that each agent will both *act* on the basis of a personal pro-social attitude and *expect* her 'neighbors' to be driven by the same other-regarding motivational force: formally, this implies a symmetric game with $w > w_2$ will be played by such altruistically-driven agents. The same kind of consideration holds for less socially concerned individuals[8]: as Goldschmidt (1993) remarks, selective interaction tends to bring about common evaluations, as repeated social contact induces people to internalize others' positions and goals. This view recalls Buchanan's (2000, p.1) reference to a characterization of altruism in terms of utility interdependence which sounds rather familiar to economists: 'Behavior may be motivated by a positive internal evaluation of the well being of other individuals who share close association as co-members of a group. Concentration on the evolutionary process at the level of genes clearly offers a basis for this sort of altruism on the part of biological units of a species.' Alternatively, the salience of utility matrices with low, intermediate or high values of w can be justified not in terms of individual motivational systems but as a consequence of (properly enforced) social norms: according to this explanation, players are still assumed to act on the basis of classic selfish preferences, but, at the same time, to be constrained by a culturally-specific set of pro-social norms prescribing how to behave in every feasible situation.[9] In particular, with reference to the above matrix, when $w < w_1$, it is 'as if' no pro-social norms were present or properly enforced in the group (D is the dominant strategy); when $w_1 < w < w_2$, then it is as if a norm of reciprocity or conditional cooperation were enforced and, finally, $w > w_2$ would imply a norm of unconditional cooperation (C is now the dominant strategy).

The reason for focusing on this specific set of alternative payoff configurations (PD, AG and OR) is three-fold. First, they encompass well-known and socially relevant scenarios (see Sen 1974). Second, they lend themselves to an analysis of social interaction taking place between individuals endowed with different degrees of altruism or, equivalently, between individuals conforming to different social norms (as we showed above). Third, *neither of them favors coexistence* – if taken separately from the others – between the two strategies under study. At the methodological level, this means that if coexistence were to emerge in our scenario, such a result would provide a strong argument in favor of the main thesis defended here: by allowing for heterogeneity not simply in terms of individual strategies or motivational structures but in terms of group-specific 'game framing', we are able to provide a plausible explanation about why in many real social environments 'mean' and 'nice' strategies turn out to coexist in the medium-long run. The plan of the remainder of the paper is as follows: Section 2 introduces the basic model; Section 3 develops the social dynamics; Section 4 provides some concluding comments.

2. The model

The general framework is as follows. Let us suppose that a *continuum* of agents belonging to a given community have to choose one out of H pure strategies $\{1,\ldots,H\}$ every time they interact with other individuals of the same community. Time is continuous. Individuals are distributed within M sub-populations $\{1,\ldots,M\}$, on the basis of their personal evaluation of the possible outcomes (in terms of pure strategies) of the random pairwise interactions. The M payoff configurations are assumed as exogenously given; more precisely, types that are initially present in the community may become extinct, but new types cannot be created. In this context, the outcome of an encounter between two individuals, let us call them I and II, is described by the pair (j,k), where the first and the second entry represent the pure strategies chosen by I and II respectively.

The adoption process of choices within the overall community is modeled by means of the so-called 'replicator equations' (see Taylor & Jonker 1978). Replicator dynamics are a widely adopted model of social (as well as natural) selection dynamics characterized by payoff monotonicity, i.e. the most rewarding strategies survive and spread over within the community at the expense of the others. This idea is well clarified in Heckathorn (1996, p.261): '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. Biologists employ these approaches to model evolutionary selection. However, the selection process has also been interpreted as reflecting a proceess of *observational learning* in which actors compare their own outcomes to those of their peers, imitating peers who do best (Brown et al. 1982, Boyd & Richerson 1985). In essence, actors look around to see who is doing well and take as role models those who appear most successful. When interpreted in this manner, these models can be termed *sideways-looking models* of behavior'. In the recent literature, it is possible to find some rigorous micro-foundations justification of replicator dynamics, see, e.g. Schlag (1998) and Björnerstedt & Weibull (1994). Such a sideways-looking selection mechanism affects both the size of each sub-population and the distribution of pure strategies within each sub-population. More precisely, we assume that social evolution not only operates at the strategic level, but also at a deeper, *meta-behavioral* level, by selecting the most rewarding way to conceptualize social interaction, i.e. to evaluate the outcomes associated with PD, AG and OR 'game framing'. In other words, as far as each agent is concerned, 'game framing' is not to be interpreted here as an exogenous psychological or cultural feature or as an irreversible, one-shot decision (as if, for some reasons, agents had to stick forever to a given value-system and/or set of social norms), but as an ongoing, endogenous process, affected by both the sub-population type he belongs to and the reward he gets by his choice. In this regard, we proceed along the lines indicated by Boyd & Richerson (1980). According to the authors, extending conventional sociobiology to include cultural transmission of behavior is important. In their 'dual inheritance theory' of genes and culture, they see culture as 'the transmission of the determinants of behavior from individual to individual, and thus from generation to generation, by social learning, imitation or some other similar process' (Boyd & Richerson 1980, pp. 101–102). The idea is then to test how the three different sub-populations (types) initially present within the community are evolutionarily robust in the sense of being able to attract an increasing number of adherents at the expense of the alternative ones.[10] We further assume that the payoffs corresponding to each pair (j, k) depend on the population to which individuals belong. In particular, we will focus on two very different cases:

(a) The payoff of player I (II) related to the event (j, k) depends on the population he belongs to and not on the population of the opponent player II (I). In this case, the payoff of player I, belonging to population i and related to the event (j, k), is expressed by the symbol a_{ijk}, where $i = 1, \dots, M$ and $j, k = 1, \dots, H$. Notice that if, given two populations i^* and i^{**}, $a_{i^*jk} > a_{i^{**}jk}$ holds whatever (j, k) is, then 'to belong to type i^*' is always more rewarding than 'to belong to type i^{**}'. In such a case, the social dynamics turns out to be very simple: type i^{**} becomes extinct. However, we shall mainly deal with the more general (and interesting) case in which such a strict payoff dominance does not hold.

(b) In this case, which includes the first as a particular case, we assume that the payoff of I (II) related to the result (j, k) also depends on the population to which the opponent player II (I) belongs[11]. I's payoffs are expressed by the symbol a_{ijkl}, where the index $l = 1, \dots, M$ represents the population of the opponent player. The specific meaning of this assumption will be subsequently clarified.

The dynamics under study can be interpreted as follows: the structure of the community outlines a preference ordering which is not based on outcomes (j, k), but on more complex outcomes (i, j, k, l): 'to be an individual of type i, playing the pure strategy j on the occasion of an encounter with an individual of type l, playing the pure strategy k'. In the following sections we will analyze these two cases separately. A rapidly growing literature considers payoffs as not univocally determined by the material outcomes of the strategic interaction taking place between players. Payoffs are more and more considered as the result of the complex interaction between material components, *psychological* (see e.g. Geanakoplos et al. 1989, Rabin 1993, 2002) and *normative* (see Fehr & Fischbacher 2002) considerations,[12] closely related to the social and cultural environment in which individuals act; see also Antoci et al. (2000), which contains a extensive review of such literature. In particular, this study builds on the work of Sacco & Zamagni (1996) and has various connections with it. Both contributions analyze heterogeneous communities and social dynamics based on the selection of the most rewarding strategies.

Nevertheless, there are some substantial differences between the two papers. While Sacco & Zamagni (1996) study the setting that corresponds to the above described case (a), they do not consider case (b). Moreover, here we assume that each individual can *only* recognize *ex post* the sub-population type her opponent belongs to and the pure strategy she is going to play. In contrast, Sacco & Zamagni postulate that individuals are able to recognize *ex ante* the opponent's player type and thus Nash equilibria are played at each matching. Consequently, social dynamics runs over the proportions of types only. The rationale behind the *ex post* type-recognition assumption is that actual interaction between players (fully) transmits information about the opponent's value-system, e.g. by close inspection of her 'mode of play' (see, e.g. Eibl-Eibesfeldt 1989, Landa 1999a). Whereas the *ex ante* recognition assumption can be thought of as a rational expectations assumption, the *ex post* recognition is then a perfect information disclosure assumption. Finally, the Sacco & Zamagni paper does not highlight phenomena of coexistence among the different types of players with which they deal.

Let us then consider a community made up of a set of M populations (types) of individuals, $\{1,2,\ldots,M\}$, where each individual has to choose her pure strategy from a (common) set of H pure strategies, $\{1,2,\ldots,H\}$. We shall indicate with the term 'action' the pair (i,j) where $i = 1,\ldots,M$ and $j = 1,\ldots,H$ respectively indicate the population and the pure strategy chosen by an individual.

2.1. Independence of the opponent's type

Let us examine, in the first place, the 'conventional' case in which each player's payoff does not depend on the population of the opponent player but only on the pure strategy followed by the latter. Let x_{ji} be the proportion (w.r.t. the whole community) of individuals belonging to population i and following pure strategy j; thus

$$x_j \equiv \sum_{i=1}^{M} x_{ji}$$

represents the proportion of the community playing pure strategy j. The proportions x_{ji} and x_j can be interpreted as the probabilities that the opponent player respectively follows action (i,j) and pure strategy j. Each individual knows the opponent player's type *ex post* only, i.e. after both players have played their pure strategies. Therefore, individuals are not able to play best responses, but each individual, in each instant of time, is ‘programmed’ to play one action only. The expected payoff Y_{ji} of the action (i,j) is

$$Y_{ji} \equiv \sum_{k=1}^{H} a_{ijk} x_k \tag{1}$$

where $i = 1, \ldots, M$ and $j = 1, \ldots, H$. The mean payoff Y of the community is

$$Y \equiv \sum_{j=1}^{H} \sum_{i=1}^{M} x_{ji} Y_{ji} \tag{2}$$

2.2. Dependence on the opponent’s type

In this context, the payoffs of individuals of population *i* depend not only on the pure strategy played by the opponent player, but also on the population to which the opponent player belongs, i.e. on the action he chooses.[13] The rationale of this methodological choice is as follows: we assume that a given outcome (i.e. pair of strategies) can bring about different values in terms of overall individual payoff (i.e. ‘utility’) according to how each agent evaluates his opponent’s ‘game framing’, which in turn, as we previously clarified, crucially depends on the specific social norms and cultural patterns characterizing each sub-population. In particular, it seems reasonable to assume that for an AG player cooperating when the opponent defects determines a *lower payoff* if the opponent is a PD agent rather than an AG agent or an OR agent, as the AG player knows that PD players are selfish (or, equivalently, act as if they were driven by anti-social norms) and, unlike OR agents, tend to exploit their opponents. In other words, whereas defection of an AG or OR opponent may be interpreted as a *mistake*, defection of a PD player is to be interpreted as *intentional*. Consequently, the cooperative player is more harmed by the latter occurence than by the former. More precisely, we now consider the payoff of a player of type *i* playing pure strategy *j* when matched with a player of type *l* playing pure strategy *k*: a_{ijkl}. As above, x_{ji} represents the proportion of individuals belonging to population *i* playing pure strategy *j*, and the expected payoff of playing *j* by an individual of the type *i* is

$$Y_{ji} \equiv \sum_{k=1}^{H} \sum_{l=1}^{M} x_{lk} a_{ijkl} \tag{3}$$

where $i = 1, \ldots, M$ and $j = 1, \ldots, H$; and the mean payoff of the community is

$$Y \equiv \sum_{j=1}^{H} \sum_{i=1}^{M} x_{ji} Y_{ji} \tag{4}$$

The analysis of this simple strategic context will allow us to show that:

(1) Even though these specific payoff configurations (if taken separately) do not generate coexistence-favoring social dynamics, coexistence may take place in the heterogeneous community under examination;
(2) Even in the simplest strategic context (two pure strategies), social dynamics may turn out to be quite complex, unlike the case of a homogeneous community where individuals have to choose between two pure strategies only;
(3) By considering only initial aggregate proportions of agents choosing between 'cooperate' (C) and 'not cooperate' (NC) (i.e. failing to take into account the types of individuals initially adopting each strategy), rather misleading predictions may be obtained; that is, the limit outcomes of the social dynamics may be radically different for very similar initial aggregate proportions.

3. Social dynamics

As we anticipated above, the selection mechanism of actions is modelled by means of the so-called 'replicator equations' (Taylor & Jonker 1978):

$$\dot{\mathbf{x}}_{ji} = x_{ji}\left(Y_{ji} - Y\right) \tag{5}$$

where $i = 1,\ldots,M$ and $j = 1,\ldots,H$. The recent literature contains several independent approaches to a rigorous micro-foundations derivation of the replicator process (See Antoci & Sacco 1995, Sacco 1994 for a concise survey). Following Weibull (1995), we can obtain (5) as follows. Assume that the number of individuals in the community is very large; let $p_{ji}(t) \geq 0$ be the number of individuals choosing action (j,i) and let $p(t) \equiv \sum_{j=1}^{H}\sum_{i=1}^{M} p_{ji}(t)$ be the community size; thus $x_{ji}(t) = p_{ji}(t)/p(t)$. Let us suppose that all individuals have a background fitness, measured as the number of offsprings per time unit $\beta \geq 0$ and a death rate $\delta \geq 0$ that are independent of their performance in the game under study. Augmenting this 'biological' replicator process by the corrective factor Y_{ji}, population dynamics can be represented as follows:

$$\dot{p}_{ji} = p_{ji}(\beta + Y_{ji} - \delta)$$

It is easy to show (see Weibull 1995, pp. 72, 73) that such dynamics imply dynamics (5) for population shares. Dynamics (5) are defined on the invariant simplex:

$$\Delta = \left\{ \mathbf{x} \in \Re^{MH},\ \sum_{j=1}^{H}\sum_{i=1}^{M} x_{ji} = 1,\ x_{ji} \geq 0 \right\}$$

Notice that the states of the community in which all individuals choose the same action are fixed points under dynamics (5). The other fixed points are the states of the community in which the actions representing a positive proportion of the community yield the same expected payoffs, i.e. there exists a constant Y^* such that

$$Y_{ji} = Y^*$$

for every action (j, i) such that $x_{ji} > 0$. If payoffs depend on the opponent's type, such condition can be explicitly written as follows:

$$Y_{ji} = \sum_{k=1}^{H} \sum_{l=1}^{M} a_{ijkl} x_{lk} = Y^* \tag{6}$$

By contrast, if payoffs do not depend on the opponent player's type, the condition above can be written as

$$Y_{ji} = \sum_{k=1}^{H} a_{ijk} x_k = Y^* \tag{7}$$

where x_k is the proportion in the community of players playing pure strategy k.

System (7) is a linear system where unknowns are represented by x_k, with $k = 1, 2, \ldots, H$, where x_k is the proportion of players playing pure strategy k. Therefore, we cannot have more than H unknowns.[14] Further, such a system does not generically admit solution if the number of actions is greater than the number of pure strategies that are played in the community. This implies that the fixed points we may generically observe are those with H actions at most, where H is the number of pure strategies that are initially present in the community. It also follows that the maximum number of sub-populations that may coexist at a fixed point is equal to the number of available pure strategies. This means that the degree of complexity of the social structure is closely related to the number of pure strategies available. Such a result does not hold with respect to the linear system defined by condition (6), where unknowns are represented by x_{ji}, where $j = 1, 2, \ldots, H$ and $i = 1, 2, \ldots, M$. As a consequence, in the latter scenario, we cannot rule out that a fixed point (generically) exists where all $H \times M$ actions coexist.

By means of this analytical framework, we will analyze the process of cultural evolution taking place within a large community in which there are two pure strategies only (C and NC), and three sub-populations in total. In population 1 individuals have PD payoffs. Let us recall that if we have two players, I and II, the four possible outcomes of the PD game (from the point of view of player I) are ordered as follows:

$$(\mathrm{NC}, \mathrm{C}) \succ (\mathrm{C}, \mathrm{C}) \succ (\mathrm{NC}, \mathrm{NC}) \succ (\mathrm{C}, \mathrm{NC})$$

where the first entry of each pair represents the strategy chosen by I and $\succ$ indicates strict preference. If we assign indices 1 and 2 to strategies NC, and C respectively, PD payoffs satisfy the following inequalities:

$$a_{112} > a_{122} > a_{111} > a_{121}.$$

In population 2, individuals have AG payoffs (see Sen 1967), i.e.:

$$(\mathrm{C},\mathrm{C}) \succ (\mathrm{NC},\mathrm{C}) \succ (\mathrm{NC},\mathrm{NC}) \succ (\mathrm{C},\mathrm{NC})$$

and consequently

$$a_{222} > a_{212} > a_{211} > a_{221}.$$

In this game, players show both *positive reciprocity* (being kind to those who have been kind to them) and *negative reciprocity* (by retaliating if they have been hurt), that is their propensity to cooperate is conditional on their opponent's behavior. Fehr & Gächter (1999) and Fehr & Fischbacher (2002) show that there is strong experimental and empirical evidence that agents exhibit both types of reciprocation and that this behavior occurs even in one-shot encounters between strangers and when retaliation is costly and yields neither present nor future material rewards.[15] For surveys of experimental results documenting the frequency of reciprocity in Ultimatum Bargaining Games, Gift-Exchange Games and Trust Games, see, e.g. Güth et al. (1982), Camerer & Thaler (1995), Fehr et al. (1993) and Roth (1995).

Finally, in population 3 individuals have OR payoffs, i.e.:

$$(\mathrm{C},\mathrm{C}) \succ (\mathrm{C},\mathrm{NC}) \succ (\mathrm{NC},\mathrm{C}) \succ (\mathrm{NC},\mathrm{NC})$$

and consequently

$$a_{322} > a_{321} > a_{312} > a_{311}.$$

In PD and OR populations, the strategies NC and C respectively (strictly) dominate the alternative strategy. Therefore, without loss of generality (see, e.g. Weibull 1995), we can analyze dynamics (5) by assuming that no player in these populations chooses the dominated strategies. We shall indicate by $\mathrm{NC}_{\mathrm{PD}}$ and $\mathrm{NC}_{\mathrm{AG}}$ the actions 'to be a PD individual playing strategy NC' and 'to be an AG individual playing strategy NC', respectively; and by C_{OR} and C_{AG} the actions 'to be an OR individual playing strategy C' and 'to be an AG individual playing strategy C', respectively.

3.1. Bistable dynamics

Let us first analyze a numerical example in which payoffs do not depend on the opponent's type and coexistence is ruled out. Let us consider the payoff structure given by the matrix shown in Table 1 (from the point of view of the row player). Dynamics (5) run over four variables, x_{11}, x_{21}, x_{22} and x_{32}, representing the proportions of individuals following actions $\mathrm{NC}_{\mathrm{PD}}$, $\mathrm{NC}_{\mathrm{AG}}$, C_{OR} and C_{AG}, respectively, and can be written as follows:

$$\begin{aligned}
\dot{x}_{11} &= x_{11}\left[(Ax)_1 - {}^{\mathrm{t}}\mathbf{x}\cdot\mathbf{Ax}\right] \\
\dot{x}_{21} &= x_{21}\left[(Ax)_2 - {}^{\mathrm{t}}\mathbf{x}\cdot\mathbf{Ax}\right] \\
\dot{x}_{22} &= x_{22}\left[(Ax)_3 - {}^{\mathrm{t}}\mathbf{x}\cdot\mathbf{Ax}\right] \\
\dot{x}_{32} &= x_{32}\left[(Ax)_4 - {}^{\mathrm{t}}\mathbf{x}\cdot\mathbf{Ax}\right]
\end{aligned} \qquad (8)$$

Table 1. Independence of the opponent's type: bistable dynamics.

	NC_{PD}	NC_{AG}	C_{AG}	C_{OR}
NC_{PD}	4	4	8	8
NC_{AG}	5	5	7	7
C_{AG}	3	3	9	9
C_{OR}	2	2	10	10

where ${}^{t}\mathbf{x} \equiv (x_{11}, x_{21}, x_{22}, x_{32})$, $\mathbf{A}$ is the payoff matrix

$$\begin{bmatrix} 4 & 4 & 8 & 8 \\ 5 & 5 & 7 & 7 \\ 3 & 3 & 9 & 9 \\ 2 & 2 & 10 & 10 \end{bmatrix}$$

and $(Ax)_r$ is the rth component of the vector $\mathbf{Ax}$ and corresponds to the expected payoff of the action. Finally, $\mathbf{xAx}$ is the mean payoff. In this specific case, the state space of dynamics (8) is the (three-dimensional) simplex

$$\Delta = \{\mathbf{x} \in \Re^4\colon\ \mathbf{x} \geq 0 \text{ and } x_{11} + x_{21} + x_{22} + x_{32} = 1\}$$

By the illustrative device adopted in Hirshleifer & Martinez Coll (1991), we can represent the edges of Δ (i.e. the boundary of the simplex in which at least one action is extinct) in the plane (see Figure 1). Thus the simplex Δ can be imagined as based on the triangle NC_{PD}–NC_{AG}–C_{AG}, while C_{OR} is the upper vertex, that in which all actions are extinct except for C_{OR} (by drawing the edges in the three-dimensional euclidean space, all the C_{OR} vertices in Figure 1 will come together).

In Figure 1, the dynamics on the edges are obtained by means of Bomze's (1983) classification technology for two-dimensional replicator equations. Following Bomze's symbols, a dotted line represents a line of fixed points (pointwise fixed), a full dot • represents a fixed point which is locally attractive, whereas saddle points are indicated by their insets and outsets (stable and unstable manifolds, respectively). Only some representative trajectories are sketched. From Figure 1, we can see that social dynamics bring about a 'bistable dynamics', i.e. the only attractive fixed points are the vertices C_{OR} and NC_{AG} and their attraction basins are separated by a two-dimensional (repulsive) pointwise fixed set in the interior of Δ, whose intersection with the edges is given by the pointwise fixed lines shown in Figure 1. If large enough proportions of individuals choosing action C_{OR} (NC_{AG}) are initially present, then all actions except for C_{OR} (respectively, NC_{AG}) become extinct. The social structure that eventually emerges is very simple: a single population playing only one pure strategy is present. In the following example, we consider dynamics starting from a payoff structure which strongly favors coexistence.

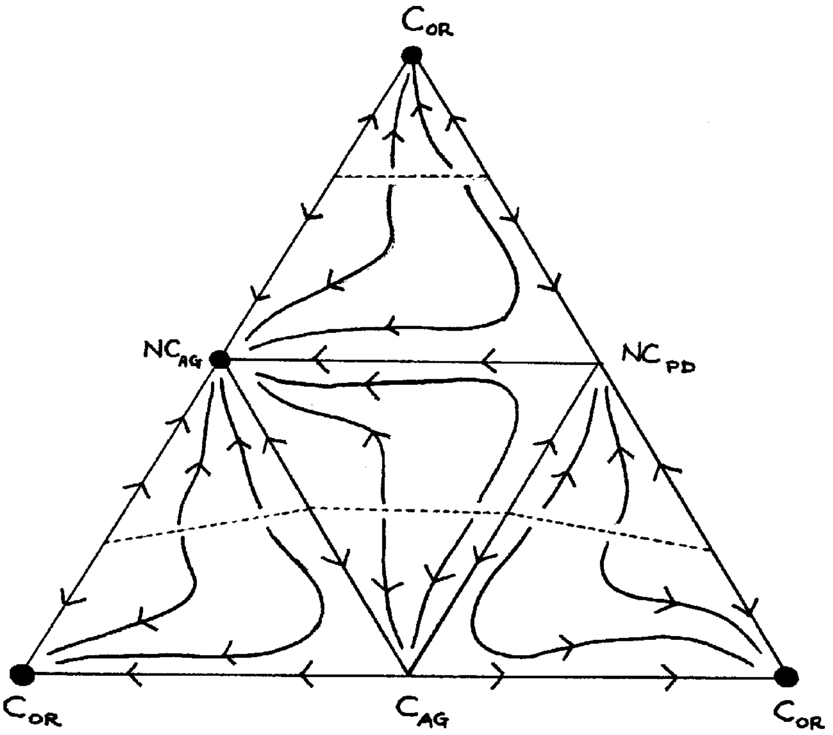

Figure 1. Bistable dynamics. Coexistence is ruled out here as the only attractive fixed points are the vertices C_{OR} and NC_{AG}.

Table 2. Independence of the opponent's type: coexistence.

	NC_{PD}	NC_{AG}	C_{AG}	C_{OR}
NC_{PD}	1	1	12	12
NC_{AG}	5	5	9	9
C_{AG}	3	3	10	10
C_{OR}	6	6	7	7

3.2. Coexistence-favoring payoffs

Let us consider the payoff matrix shown in Table 2, where we still assume that players' payoffs do not depend on the opponent's type. In this example, the highest payoff level is reached by a PD individual when matched with an individual playing the pure strategy C. On the other hand, the payoff of a PD individual is very low when he is matched with an individual playing NC. This rules out the emergence of a

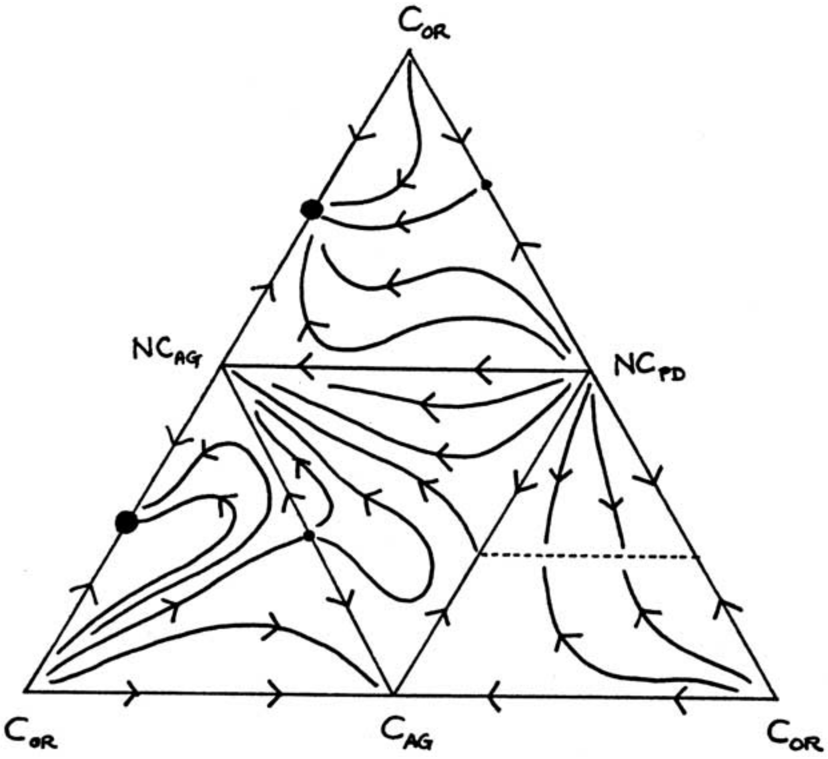

Figure 2. C_{OR}–NC_{AG} coexistence. Coexistence between C_{OR} and NC_{AG} players emerges.

homogeneous community where only the PD sub-population exists. On the contrary, OR individuals assign a relatively high payoff to the outcome (C, NC), while their payoff associated to the outcome (C, C) is relatively low. In this case, as in the previous one, we cannot expect a homogenous OR-type community to emerge, as such a community would turn out to be extremely vulnerable with respect to a population of PD players. The other entries of this coexistence-favoring payoff matrix can be interpreted in a totally analogous way. The phase portrait at the edges of the simplex Δ is represented in Figure 2.

It is easy to verify that there are no fixed points in which more than three actions coexist. Thus, by a well-known result (see Weibull 1995), under dynamics (8) trajectories always approach the edges represented in Figure 2. In this figure, we can notice that, starting from 'almost all' the initial distributions of actions in the community, the social dynamics reach a fixed point in which AG and OR-type sub-populations coexist playing NC_{AG} and C_{OR}, respectively. It may be interesting to note that such a social scenario may be compatible with a group selection perspective, although such possibility is not explicitly tackled in the present paper. Think, e.g. of Rubin's (2000, p. 21) claim that 'Altruism in contemporary group

Table 3. Independence of the opponent's type: bitable dynamics and coexistence.

	NC_{PD}	NC_{AG}	C_{AG}	C_{OR}
NC_{PD}	4	4	13	13
NC_{AG}	6	6	7	7
C_{AG}	5	5	12	12
C_{OR}	3	3	13.5	13.5

selection models is associated with faster growth of altruistic groups. This can occur if such altruism leads to cooperation in some PD. I call altruism that would lead to faster growth "efficient altruism". (...) But such altruism must also be associated with monitoring the recipient to avoid free riding, for societies that allowed excessive shirking and free riding would not have grown as fast as others.' Rubin's considerations provide an interesting interpretation of coexistence occurring between altruistic cooperators (i.e. C_{OR} players) and players choosing not to cooperate as a form of punishment of defecting behaviors (i.e. NC_{AG} players). In the last example of this section, we focus on a mixed case where a configuration of bistable dynamics is merged with a different one where coexistence arises. Let us examine the payoff matrix shown in Table 3. This payoff structure is characterized by the fact that both C_{OR} and NC_{AG} individuals perform well with opponent players of the same type. Therefore, the vertices C_{OR} and NC_{AG} are both locally attractive (see Figure 3). However, if the proportions of NC_{PD} and C_{AG} individuals are large enough, a two-population community in which only these two types coexist may emerge. As above, it is easy to verify that no fixed points exist with more than two actions. Thus, Figure 3 represents the 'limit' dynamics of (8). We shall discuss this case further in the last section, when some features of aggregate dynamics will be analyzed.

Let us now consider the payoff matrix shown in Table 4, in which payoffs depend not only on the strategic choices of the two players but also on the opponent player's type. Notice that we have now introduced two parameters, α and β, where $0 < \alpha < 1$ and $\beta > 0$, which index the qualitative modifications of social dynamics. Such a matrix shows that whereas PD player's payoffs are completely independent of the population his opponent belongs to, both AG and OR players' payoffs crucially depend on their opponent's type. In particular, an AG player gets a lower payoff when cooperating with a defecting PD rather than with a defecting AG, as a consequence of the negative 'psychological externality' due to PD agents' anti-cooperative attitude. Likewise, AG players can be expected to be 'happier' when defecting with a defecting PD rather than with a defecting AG. Further, AG-type agents get a higher payoff when cooperating with a cooperating OR rather than with a cooperating AG. The rationale behind these payoff differences can be explained as follows: OR players are perceived as *more trustworthy* agents as they tend to cooperate unconditionally, that is to never defect regardless of their opponent's behavior. In other words, we can plausibly imagine a sort of 'moral ranking' among the three types, according to which – as far as the opponent's choices are concerned – OR behavior is preferable to

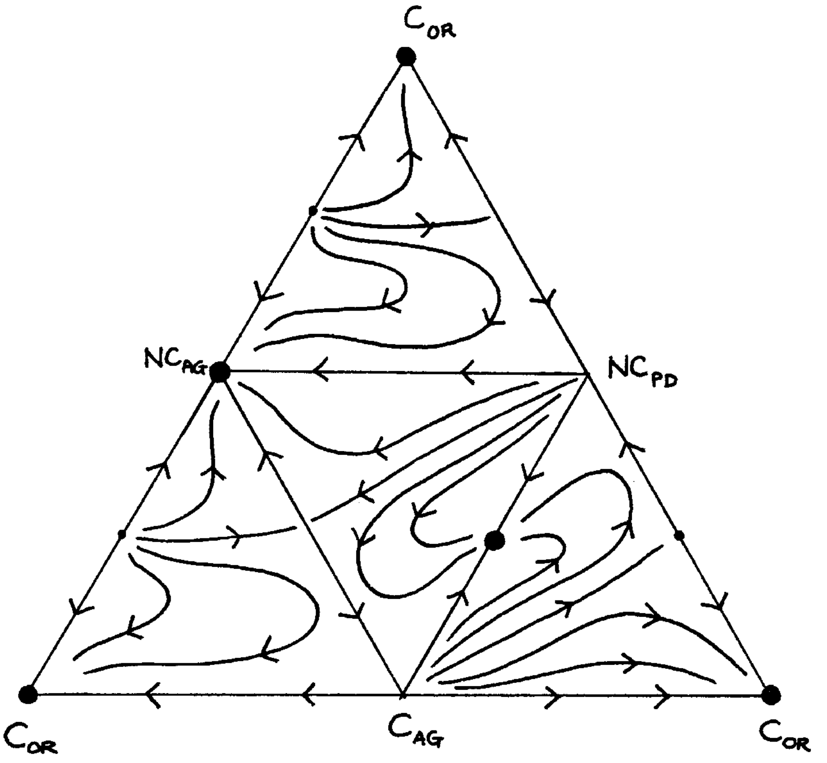

Figure 3. Bistable dynamics and C_{AG}–NC_{PD} coexistence. This scenario represents a mixed case where a configuration of bistable dynamics (as it is the case in Figure 1) is merged with a configuration where coexistence arises (as it is the case in Figure 2).

Table 4. Dependence on the opponent's type.

	NC_{PD}	NC_{AG}	C_{AG}	C_{OR}
NC_{PD}	1	1	12	12
NC_{AG}	5	4	9	9
C_{AG}	3	$3+\alpha$	10	11
C_{OR}	6	7	8	$8+\beta$

AG behavior which in turn can be considered as 'morally superior' to PD behavior. Therefore, from the above matrix, OR players prefer to cooperate with a defecting AG agent rather than with a defecting PD-type agent and with a cooperating OR agent rather than with a cooperating AG-type. As anticipated above, when payoffs depend on the opponent player's type, far more complex social structures are likely to emerge; in particular, fixed points with more than two actions are not ruled out in generic cases. In our example, it is easy to verify that a fixed point P, in which all the actions are present, exists if and only if $\beta < 1/2$; further, it is always locally attractive (see the mathematical appendix) and its coordinates are:

$$(x_{11}^*, x_{21}^*, x_{22}^*, x_{32}^*) = \frac{1}{16 - 7\alpha\beta - 4\beta}(6 - 3\alpha\beta, 1 - 2\beta, 5 + \alpha - 4\alpha\beta - 2\beta, 4 - \alpha)$$

Thus, a social configuration with three sub-populations playing four actions can be locally attractive under dynamics (8). Notice that, in such a configuration, C_{AG} and NC_{AG} individuals coexist, whereas such a coexistence pattern is ruled out in a community where only an AG population is initially present. The dynamics driven by the above payoff matrix is interesting also because, as parameter values vary, a relatively rich classification of cases can emerge. In the following figures, we only sketch the 'representative' ones. Since fixed points with more than two actions may exist under such a payoff matrix, their stability cannot be checked by reference to Bomze's classification only; it is also necessary to use the standard procedure of local analysis. We consider the following four cases.

Case (*a*): For $\alpha = 1/4$ and $\beta \geq 4$, the fixed point P does not exist; thus trajectories always approach the edges of Δ. The dynamics on the edges is given in Figure 4. We

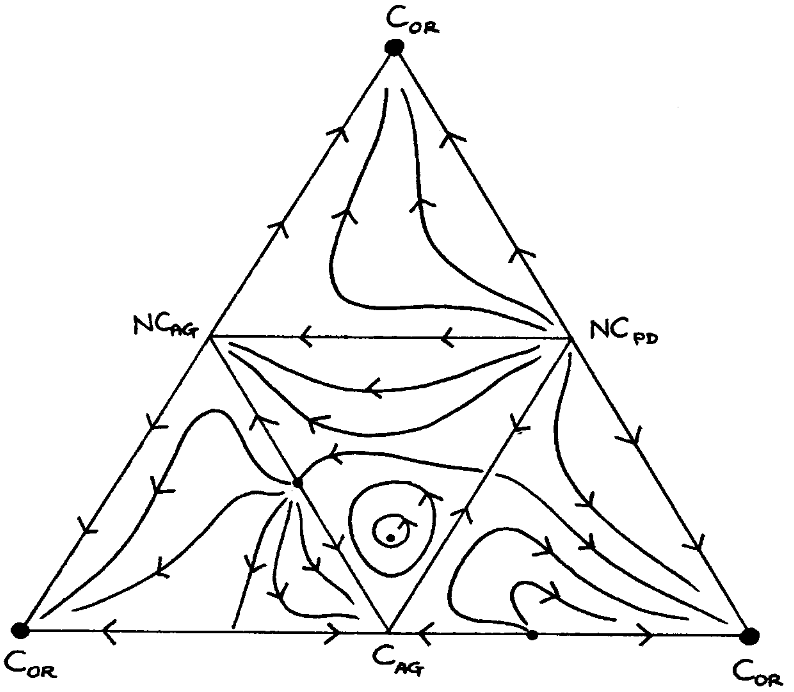

Figure 4. Case (a): $\alpha = 1/4$ and $\beta \geq 4$. 'Almost all' the trajectories approach the vertex C_{OR}. If the OR population is extinct (see triangle NC_{PD}–NC_{AG}–C_{AG}), we have a fixed point surrounded by closed trajectories.

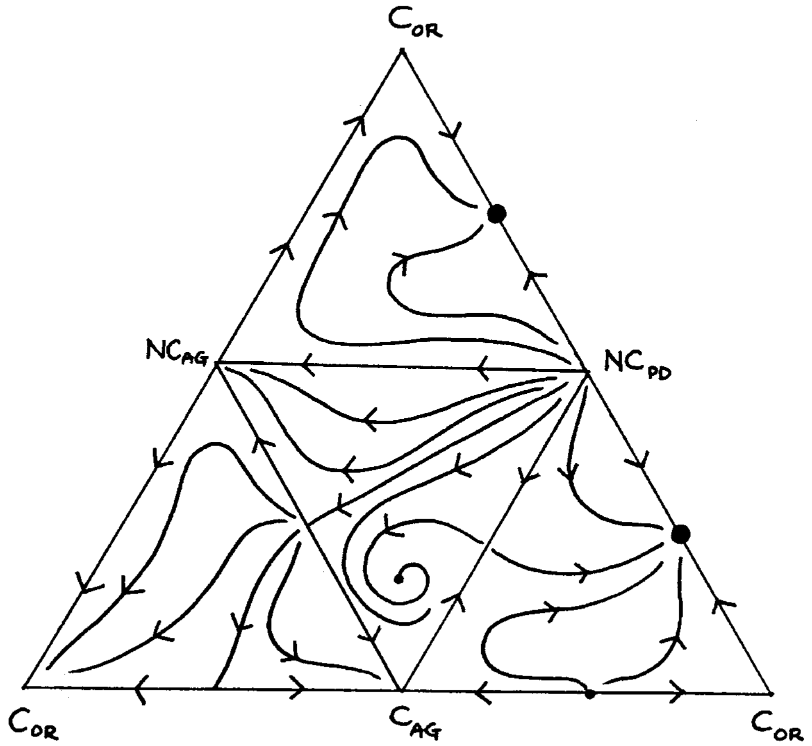

Figure 5. Case (b): $\alpha > 1/4$ and $3 < \beta < 4$. 'Almost all' trajectories approach a fixed point in which C_{OR} and NC_{PD} players coexist. The fixed point in the interior of the triangle $NC_{PD}-NC_{AG}-C_{AG}$ is a saddle point.

can observe that 'almost all' the trajectories approach the vertex C_{OR}. Notice that, in this case, the action C_{OR} performs better against itself than the action NC_{PD} against C_{OR}. Notice also that, if the OR population is extinct (see triangle $NC_{PD}-NC_{AG}-C_{AG}$) we have a fixed point surrounded by closed trajectories. However, it is easy to see that such trajectories become repulsive when the OR population is introduced into the community (see the mathematical appendix).

Case (*b*): For $\alpha > 1/4$ and $3 < \beta < 4$, the fixed point P does not exist and the dynamics on the edges is shown in Figure 5. In this case, 'almost all' trajectories approach a fixed point in which both C_{OR} and NC_{PD} coexist. In the mathematical appendix we show that the fixed point in the interior of the triangle $NC_{PD}-NC_{AG}-C_{AG}$ (which is attractive on the edges) is a saddle point, i.e. it is unstable.

Case (*c*): For $\alpha = 1/4$ and $1 \leq \beta < 3/2$, the fixed point P does not exist and the dynamics on the edges are shown in Figure 6. In this case, the fixed point in which both C_{OR} and NC_{PD} coexist becomes unstable; the fixed point in the interior of the triangle $NC_{PD}-NC_{AG}-C_{AG}$ remains repulsive, while almost all the trajectories are attracted by the fixed point in the interior of the triangle $NC_{PD}-C_{AG}-C_{OR}$.

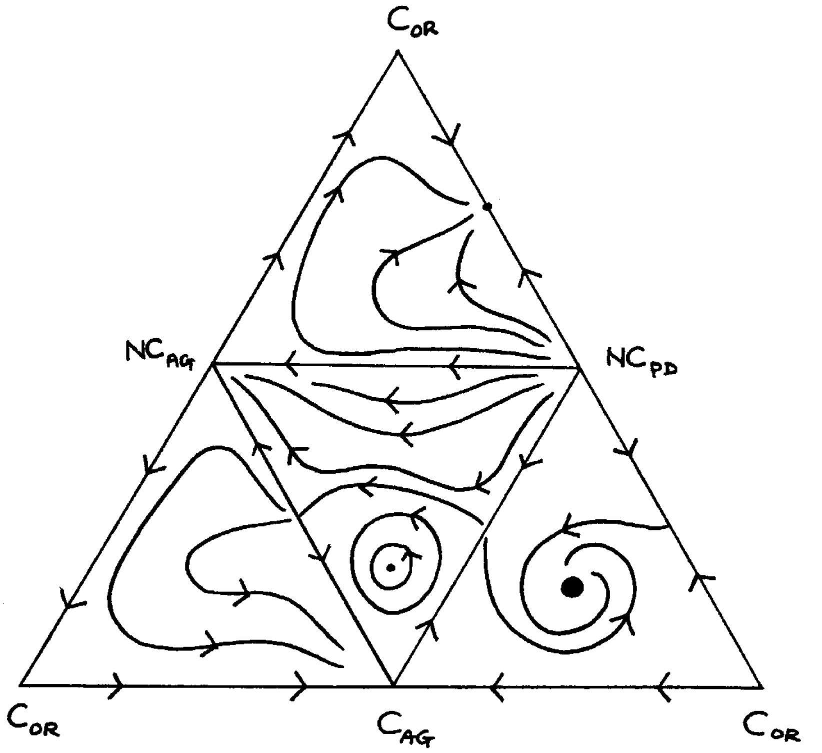

Figure 6. Case (c): $\alpha = 1/4$ and $1 \leq \beta < 3/2$. Almost all the trajectories are attracted by the fixed point in the interior of the triangle NC_{PD}–C_{AG}–C_{OR}.

All the sub-populations coexist in this fixed point and, as above, social dynamics reach a fixed point in which both strategies C and NC coexist.

Case (*d*): For $\alpha = 1/4$ and $1/4 \leq \beta < 1/2$, the locally attractive fixed point P exists; at such a point, no population becomes extinct and AG individuals play both C and NC. The dynamics on the edges (shown in Figure 7) is analogous to that of Figure 6; however, in this case, the fixed point in the interior of the triangle NC_{PD}–C_{AG}–C_{OR} becomes a saddle point (see the mathematical appendix).

The local attractivity of the fixed point *P* does not imply its global attractivity and, in the interior of the state space Δ, other attractors may exist. However, even if in this case *P* may be a global attractor, it is certainly possible to construct *ad hoc* payoff matrices according to which dynamics (8) have a strange attractor in the interior of the state space Δ. In such a case, OR, AG and PD sub-populations in this community coexist, all playing pure strategies C and NC, although the dynamics never reach a fixed point. Furthermore, the outcome of the social dynamics can be critically dependent on initial distributions of actions in the community; in such a case, social dynamics is unpredictable, at least from a deterministic point of view. To build these

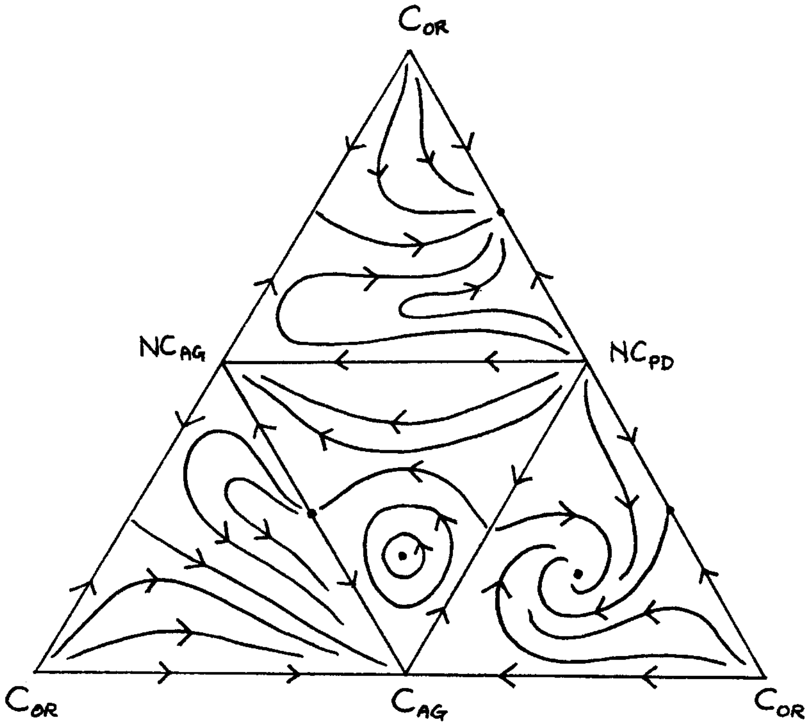

Figure 7. Case (d): $\alpha = 1/4$ and $1/4 \leq \beta < 1/2$. The dynamics on the edges is analogous to that of Figure 6; however, in this case, the fixed point in the interior of the triangle $\mathrm{NC_{PD}}$–$\mathrm{C_{AG}}$–$\mathrm{C_{OR}}$ becomes a saddle point.

ad hoc matrices, see Schnabl et al. (1991); in particular, see matrices (7)–(9) of their paper.

4. Dynamics of aggregate variables and concluding remarks

In order to stress the importance of our results, let us recall that for symmetric two-player games with two pure strategies (e.g. NC and C) played in a homogeneous community:

$$\begin{array}{ccc} & \mathrm{NC} & \mathrm{C} \\ \mathrm{NC} & a_{11} & a_{12} \\ \mathrm{C} & a_{21} & a_{22} \end{array}$$

we can only have four (generic) dynamic regimes under replicator dynamics (see Weibull 1995, pp. 74–76):

(i) For $a_{11}>a_{21}$, $a_{12}>a_{22}$ the pure strategy NC (strictly) dominates C; in this case, the share of individuals choosing NC approaches the value 1 when time goes to infinity;

(ii) For $a_{11}<a_{21}$, $a_{12}<a_{22}$ the opposite case holds.

(iii) For $a_{11}>a_{21}$, $a_{12}<a_{22}$ both the pure population states, in which all the individuals respectively play NC or C, are locally attractive fixed points; their attraction basins are separated by a repulsive fixed point in which both strategies are played;

(iv) For $a_{11}<a_{21}$, $a_{12}>a_{22}$ there is a globally attractive fixed point where both strategies coexist.

In such a context (two pure stategies and a homogeneous community), only payoff configuration (iv) admits coexistence between NC and C. According to the others, we expect to see individuals playing only one strategy after transient dynamics. Therefore, the coexistence of strategies can be explained only through very restrictive assumptions over individuals' (homogeneous) payoffs. This prediction is rather unrealistic in a world where we generally observe coexistence between 'nice' and 'mean' strategies. Hirshleifer & Martinez Coll (1991) assume homogeneous payoffs but add to the set of pure strategies related to the games PD and CG two 'reactive' strategies, such as 'tit for tat' (a nice strategy) and 'bully' (a mean strategy). The games PD and CG are played separately, i.e. they first consider replicator dynamics in a PD environment and then study dynamics for the CG. They show that, in this context, dynamics exhibit very interesting features. More specifically, they show that dynamics can be substantially more complex than the dynamics of regimes (i)–(iv) and that we can expect, within both payoff environments, coexistence between nice and mean strategies. The complexity of dynamics studied by Hirshleifer & Martinez Coll is a direct consequence of the assumption that players are able to play reactive strategies.

In our paper we have taken a different route by postulating that all individuals in the community play two (non-reactive) pure strategies only but, at the same time, they are heterogeneous with respect to their way of framing the game, which is culturally-specific (i.e. specific to each sub-population). Furthermore, we obtain coexistence results even when each individual has payoffs that do not favor coexistence, i.e. a type (i), a type (ii) or a type (iii) individual payoffs configuration. A further relevant result is related to dynamics in Figure 3. Let us consider Figure 8 in which aggregate dynamics associated to the phase portrait of Figure 3 are sketched. More precisely, in Figure 8 we have on the horizontal axis the aggregate proportion of individuals playing NC, $x_1 = x_{11}+x_{21}$, while on the vertical one we represent the aggregate proportion of individuals playing C, $x_2 = x_{22}+x_{32}$. The aggregate dynamics are represented in the segment in which x_1 and x_2 are positive and $\sum_{j=1,2} x_j = 1$. Clearly, through each point of the aggregate phase space in Figure 8 we do not have uniqueness of trajectories, i.e. for every initial pair of (aggregate) proportions, we may have different dynamic regimes according to the underlying social structure of the community. In Figure 8, the full dots (•) indicate aggregate coordinates corresponding to the locally attractive fixed points in Figure 3. Notice that, in this case, we have three attracting fixed points, one of which is characterized by coexistence. From Figure 3, we can also observe that these fixed points are not in

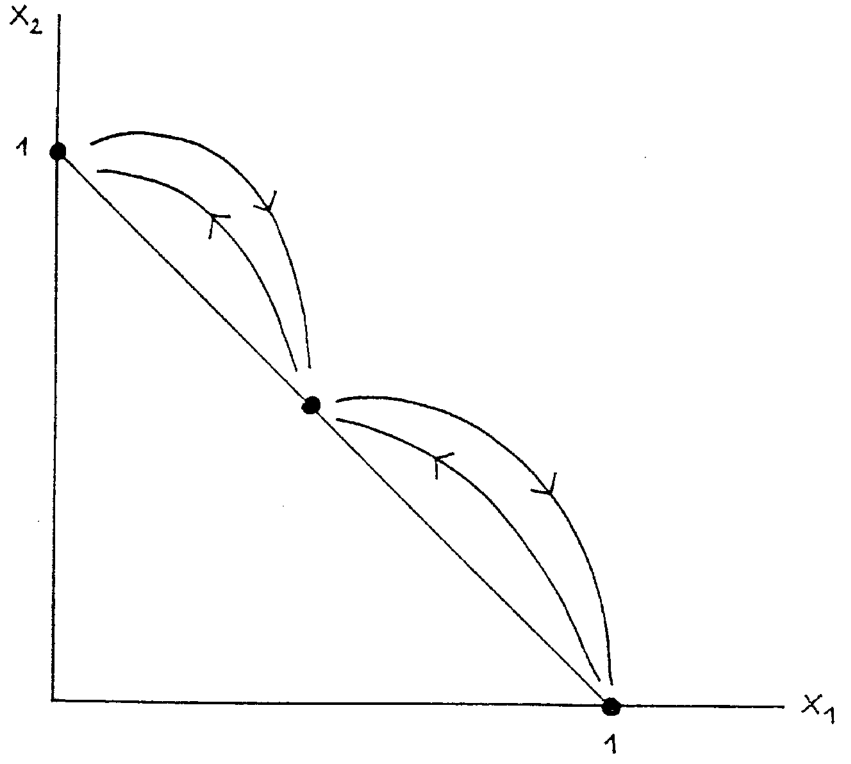

Figure 8. Aggregate dynamics associated to the phase portrait of Figure 3.

general locally attracting for aggregate dynamics (as the arrows in Figure 8 indicate). In fact, even if the aggregate distribution of pure strategies is near the fixed point with coexistence in Figure 8, we can choose a social structure playing such an aggregate distribution which, under dynamics (8), reaches one of the two fixed points without coexistence (and *vice versa*). Therefore, if we make predictions taking into account the initial values of the aggregate distribution of pure strategies only, these are likely not to be correct.

Appendix

The dynamics under system (8) for payoffs which do not depend on the opponent player, can be completely analyzed by Bomze's (1983) results. When payoffs depend on the opponent player, we obtain fixed points in which more than two actions coexist. In these cases, to analyze local stability, it is necessary to linearize system (8) around these fixed points. To this end, we use the well-known correspondence between replicator equations and Lotka–Volterra systems (see Hofbauer & Sigmund 1988). In particular, in this case, we have that the transformation

$$T:(y,z,w)\to(x_{11},x_{21},x_{22},x_{32}) = \left(\frac{1}{1+y+z+w},\frac{y}{1+y+z+w},\frac{z}{1+y+z+w},\frac{w}{1+y+z+w}\right)$$

maps the trajectories under Lotka–Volterra equations:

$$\begin{aligned}\overset{\bullet}{y} &= y[4+3y-3z-3w]\\ \overset{\bullet}{z} &= z[2+(2+\alpha)y-2z-w]\\ \overset{\bullet}{w} &= w[5+6y-4z+(\beta-4)w]\end{aligned}\tag{9}$$

onto those of replicator equations (8). The inverse transformation of T is:

$$T^{-1}:(x_{11},x_{21},x_{22},x_{32})\to(y,z,w)=\left(\frac{x_{21}}{x_{11}},\frac{x_{22}}{x_{11}},\frac{x_{32}}{x_{11}}\right)$$

System (9) has a unique fixed point at which y, z and w are strictly positive if and only if $\beta < 1/2$. In this case, the fixed point has coordinates

$$(y^*,z^*,w^*)=\frac{1}{3(\alpha\beta-2)}(2\beta-1,4\alpha\beta+2\beta-\alpha-5,\alpha-4)$$

Otherwise, it has no fixed points where all the actions are simultaneously present in the community. In the original coordinates, the above fixed point becomes

$$(x^*_{11},x^*_{21},x^*_{22},x^*_{32})=\frac{1}{16-7\alpha\beta-4\beta}(6-3\alpha\beta,1-2\beta,5+\alpha-4\alpha\beta-2\beta,4-\alpha)$$

The Jacobian matrix J of (9), evaluated at (y^*, z^*, w^*), has entries J_{ij}:

$$J_{11}=3y^*,\quad J_{12}=-3y^*,\quad J_{13}=-3y^*$$

$$J_{21}=(2+\alpha)z^*,\quad J_{22}=-2z^*,\quad J_{23}=-z^*$$

$$J_{31}=6w^*,\quad J_{32}=-4w^*,\quad J_{33}=(\beta-4)w^*$$

By Routh–Hurwitz criterion (see, e.g., Beavis & Dobbs 1989, p. 134), a necessary and sufficient condition for local asymptotic stability of (y^*, z^*, w^*) is that Tr$J < 0$, Det$J < 0$ and Det$\tilde{J} < 0$ where

$$\tilde{J}\equiv\begin{bmatrix}J_{11}+J_{22} & J_{23} & -J_{13}\\ J_{32} & J_{11}+J_{33} & J_{12}\\ -J_{31} & J_{21} & J_{22}+J_{33}\end{bmatrix}$$

It is easy to see that this system meets such conditions.

A fixed point with $y = 0$ and $z, w > 0$, i.e. a fixed point in which only $x_{21} = 0$, exists if and only if $\beta < 3/2$. In this case, it has the following coordinates:

$$(\bar{y}, \bar{z}, \bar{w}) = \left(0, \frac{1}{2-\beta}, \frac{3-2\beta}{2(2-\beta)}\right)$$

The Jacobian matrix evaluated at $(\bar{y}, \bar{z}, \bar{w})$ is:

$$\begin{bmatrix} 4 - 3\bar{z} - 3\bar{w} & 0 & 0 \\ (2+\alpha)\bar{z} & -2\bar{z} & -\bar{z} \\ 6\bar{w} & -4\bar{w} & (\beta - 4)\bar{w} \end{bmatrix}$$

Notice that the eigenvalue in the direction of the interior of the simplex Δ:

$$4 - 3\bar{z} - 3\bar{w} = \frac{1-2\beta}{2(2-\beta)}$$

is strictly positive if and only if $\beta > 1/2$, i.e. when the interior fixed point exists in the simplex. We have a fixed point with $z = 0$ and $y, w > 0$, corresponding to the fixed point in which only $x_{22} = 0$, if and only if $\beta < 1/4$. In this case, it has coordinates

$$(\hat{y}, \hat{z}, \hat{w}) = \left(\frac{1-4\beta}{3(2+\beta)}, 0, \frac{3}{2+\beta}\right)$$

and the relative Jacobian matrix

$$\begin{bmatrix} 3\hat{y} & -3\hat{y} & -3\hat{y} \\ 0 & 2 + (2+\alpha)\hat{y} - \hat{w} & 0 \\ 6\hat{w} & -4\hat{w} & (\beta - 4)\hat{w} \end{bmatrix}$$

has the following strictly positive eigenvalue in the direction of the interior of the simplex:

$$2 + (2+\alpha)\hat{y} - \hat{w} = \frac{5 + \alpha - 2\beta - 4\alpha\beta}{3(2+\beta)}.$$

We always have a fixed point with $w = 0$ and $y, z > 0$ (i.e. with only $x_{32} = 0$) and it has the following coordinates:

$$(\hat{y}, \hat{z}, \hat{w}) = \left(\frac{2}{3\alpha}, \frac{4\alpha + 2}{3\alpha}, 0\right)$$

with the Jacobian matrix

$$\begin{bmatrix} 3\hat{y} & -3\hat{y} & -3\hat{y} \\ (2+\alpha)\hat{z} & -2\hat{z} & -\hat{z} \\ 0 & 0 & 5 + 6\hat{y} - 4\hat{z} \end{bmatrix}$$

We can see that it has the following strictly positive eigenvalue in the direction of the interior of the simplex:

$$5 + 6\hat{y} - 4\hat{z} = \frac{4 - \alpha}{3\alpha}.$$

Acknowledgements

We are grateful to the editor, an anonymous referee, Robert Sugden, participants in the Fourth International METU Conference in Economics, Ankara, September 2002; participants in the First Conference of Italian Association of Cognitive Sciences on ‘Cognitive Sciences and Economics’, Rovereto, University of Trento, September 2002; participants in the Second Workshop on ‘Dynamic Methods for Economics and Finance’, University of Urbino, September 2002; participants in the International Business & Economics Conference, Las Vegas, October 2002 and seminar audiences at University of East Anglia, University of Bologna and University of Verona for comments and suggestions on previous versions of the paper. The usual disclaimer applies.

Notes

1. Andreoni & Miller (1993) and Cooper et al. (1996) set up experiments where people play the PD game sequentially with randomly changing opponents and find that while a minority of players act selfishly, the majority adopt nonselfish behaviors.
2. Such an observation seems to be valid across countries and social contexts; on the contrary, in the light of empirical and experimental evidence, what appears to be strongly culturally-specific is the *relative frequency* with which the two types of behavior are observed.
3. The experimental settings quoted by Fehr & Gächter (1999) include PD, Investment Game, Public Goods Game and Trust Game. Fehr & Fischbacher (2002) argue that a large body of experimental evidence systematically refutes the self-interest hypothesis suggesting that several subjects exhibit preferences for reciprocal fairness.
4. The notion of ‘subjective game’ has been recently elaborated in order to account for agents’ limited knowledge about the structure of the ‘objective’ game to be played (see, e.g. Kalai & Lehrer 1995, Matsushima 1998b, Oechssler & Schipper 2003). In this light, with reference to the two game-theoretical strands of literature mentioned so far (the one dealing with the coexistence of strategies issue and the one focused on subjective games) we believe that our paper lies somewhere in-between, as it aims at accounting for coexistence of conflicting strategies occurring within an evolutionary environment where, as it will be subsequently clarified, individual perceptions of the game turn out to be biased due to cultural factors.
5. Binmore (1994, p. 140) points out that ‘A society’s pool of common knowledge — its culture, provides the informational input that individual citizens need to coordinate on *equilibria* in the games that people play. (...) An analyst ignorant of this data would not necessarily be able to predict the equilibrium on which members of the society would coordinate in a specific game. He might therefore categorize the equilibrium selection criteria that the society uses as arbitrary. However, the criteria will not seem arbitrary to those within the society under study.’
6. It seems reasonable to assume that agents belonging to the same group have passed through similar socialization processes and therefore share common values and tend to conform to the same

(population-specific) social norms: the majority of human customs and behaviors appears to be the consequence of complex processes of cultural evolution. Binmore (1994) remarks that 'A society's culture consists of more than the shared knowledge that we all belong to the same species. Vast amounts of historical data are enshrined in its customs and traditions'.

7. 'A community of rational individuals is held together by the pool of common knowledge that I shall call its *culture*. The gossamer threads of shared knowledge and experience may seem flimsy bonds with which to hold a society together when compared with the iron shackles of duty and obligation postulated by traditional ethical theories. However, one must remember that the iron shackles of the traditionalists exist only in their imaginations, and even the most gossamer of real threads is more substantial than an iron shackle that is only imagined. Moreover, like Gulliver in Lilliput, we are bound by so many threads that even real shackles could fulfill their function with no greater efficiency' (Binmore 1994). With reference to the idea of 'cultural bias', see Douglas (1982).
8. For empirical evidence, see Ayres & Siegelman (1995) and Rapaport (1995) showing that market outcomes appear to systematically depend on the ethnicity of the parties involved; at the experimental level, Weimann (1994) observes that in a repeated public good game framework American students turn out to be less cooperative than Germans, while Ockenfels & Weimann (1999) find that eastern Germans are far more selfish than western subjects.
9. Regarding the interpretation in terms of culturally-specific motivational systems, it is important to clarify that we do not need to assume that OR players are actually driven by *genuinely altruistic* concerns: we can equivalently interpret their conceptualization of the game as the effect of a sophisticated 'as if' calculating morality, letting them to implement the cooperative outcome and so to efficiently pursue their original selfish goals (see Sen 1974 for this intuition and Mueller 1986). On this view, people are assumed to choose the most efficient among alternative motivational structures, perceived as competing 'happiness technologies' (see Menicucci & Sacco 1997). An analogous explanation may be provided for AG players.
10. In other words, we are assuming that a form of learning takes place over time. As Oechssler & Schipper (2003, p. 137) remark, 'The theoretical and experimental literature on learning in games has substantially increased in recent years .... By and large, however, this literature is concerned with *learning how to play a game* rather than with *learning about a game*. That is, the question of how players perceive a game has rarely been addressed so far. A normal form game consists of the set of players, the set of possible strategies, and a payoff function for each player. Learning about a game therefore means that players, who have incomplete knowledge about some of these elements, learn about those elements while playing the game' (emphasis added). As far as our paper is concerned, we may assert that here learning is both 'about how to play a given game' and 'about the game itself': insofar as their opponents perform better, players imitate both their 'game framing' and the strategies they play.
11. Referring to Granovetter's (1985) work, Sacco & Zamagni (1996) remind us that the network of social relations which individual behaviors are embedded represents one of the key factors affecting agents' goals and motivations.
12. 'In psychological games the payoff to each player depends not only on what every player does but also on what he thinks every player believes, and on what he thinks every player believes, and on what he thinks they believe others believe, and so on. (...) the traditional theory of games is not well suited to the analysis of such belief-dependent psychological considerations as surprise, confidence, gratitude, disappointment, embarassment, and so on' (Geanakoplos et al. 1989, pp. 60–61). The above recalled interaction could, even radically, modify the 'purely material' payoff structure and, consequently, the choices determined by them, as we showed in section 1 by illustrating how different levels of the 'degree of altruism' w can lead to alternative payoff configurations, such as PD, AG and OR (see Taylor 1987 for a rigorous analysis).
13. Similarly, Banerjee & Weibull (1995) set up a 'discriminating players' model where agents are able to identify their opponents' type and to consequently act on the basis of an 'opponent-sensitive' logic of play.

14. $x_j = 0$ if there are no other players in the community deciding to adopt strategy j. Further, the maximum number of equations is $N \times M$, as this is determined by the number of actions (i, j) coexisting at the fixed point (i.e. such that $x_{ji} > 0$).
15. Sahlins (1972) refers to a similar selective attitude describing human societies where the same agent consistently displays a cooperative attitude toward people he feels 'close' to as well as a payoff-maximizing or even hostile attitude towards people he perceives as 'strangers'. Landa (1999b, p. 279), referring to Sober & Wilson's (1998) famous book on the evolution of unselfish behavior, remarks that 'discriminating altruists will choose other altruists and the resulting assortive interaction is a mechanism for the evolution of altruism of the group. The group of like-minded altruists in the group will punish those members who do not cooperate. This kind of conscious choice of cooperating partners on the basis of individual and group identity, and the punishment meted out to those who violate the norms of the group, is exactly the same kind of behavior exhibited by Chinese merchants in Southeast Asia.'

References cited

Andreoni, James & John Miller. 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. Economic Journal 103:570–585.

Antoci, Angelo & Pier Luigi Sacco. 1995. A public contracting evolutionary game with corruption. Journal of Economics 61:89–122.

Antoci Angelo, Pier Luigi Sacco & Stefano Zamagni. 2000. The ecology of altruistic motivations in triadic social environments. Pp. 335–351 in L.A. Gérard-Varet, S.C. Kolm & J. Mercier Ythier (ed.) The Economics of Reciprocity, Giving and Altruism, IEA Conference Volume. Macmillan, London.

Ayres, Ian. & Peter Siegelman. 1995. Race and gender discrimination in bargaining for a new car. American Economic Review 85(3):304–319.

Banerjee, Abhijit Vinayak & Jörgen Weibull. 1995. Evolutionary selection and rational behaviour. Pp. 343–363 in A. Kirman & M. Salmon (ed.) Learning and Rationality in Economics. Basil Blackwell, Oxford.

Beavis, Brian & Ian Dobbs. 1989. Optimization and stability theory for economic analysis. Cambridge University Press, Cambridge.

Binmore, Ken. 1994. Game Theory and the Social Contract. Volume I: Playing Fair. The MIT Press, Cambridge.

Binmore, Ken & Larry Samuelson. 1992. Evolutionary stability in repeated games played by finite automata. Journal of Economic Theory 57:278–305.

Björnerstedt, Jonas & Jörgen Weibull. 1994. Nash equilibrium and evolution by imitation, mimeo. Delta, Paris.

Bomze, Immanuel. 1983. Lotka-Volterra equation and replicator dynamics: a two-dimensional classification. Biological Cybernetics 48:201–211.

Boyd, Robert & Peter J. Richerson. 1980. Sociobiology, culture and economic theory. Journal of Economic Behavior and Organization 1:97–121.

Boyd, Robert & Peter J. Richerson. 1985. Culture and the evolutionary process. University of Chicago Press, Chicago.

Brown, Joel, Michael Sanderson & Richard Michod. 1982. Evolution of social behavior by reciprocation. Journal of Theoretical Biology 99:319–339.

Buchanan, James M. 2000. Group selection and team sports. Journal of Bioeconomics 2:1–7.

Camerer, Colin & Richard Thaler. 1995. Ultimatums, dictators, and manners. Journal of Economic Perspectives 9:209–219.

Cooper, Russell, Douglas DeJong, Robert Forsythe & Thomas Ross. 1996. Cooperation without reputation: experimental evidence from prisoner's dilemma games. Games and Economic Behavior 12:187–218.

Douglas, Mary. 1982. In the active voice. Routledge, London.

Eibl-Eibesfeldt, Irenaeus. 1989. Human ethology. Aldine de Gruyter, New York.

Ellison, Glenn. 1994. Cooperation in the prisoner's dilemma with anonymous random matching. Review of Economic Studies 61:567–588.

Eshel, Ilan, Emilia Sansone & Avner Shaked. 1999. The emergence of kinship behavior in structured populations of unrelated individuals. International Journal of Game Theory 28:447–463.

Fehr, Ernst & Urs Fischbacher. 2002. Why social preferences matter – The impact of non-selfish motives on competition, cooperation and incentives. The Economic Journal 112:1–33.

Fehr, Ernst & Simon Gächter. 1999. Reciprocal fairness, heterogeneity, and institutions. Paper presented at the AEA Meeting in New York, 3–5 January 1999.

Fehr, Ernst, Georg Kirchsteiger & Arno Riedl. 1993. Does fairness prevent market clearing? An experimental investigation. Quarterly Journal of Economics 108:437–460.

Furnivall, John S. 1957. Colonial policy and practice: a comparative study of Burma and Netherlands India. New York University Press.

Geanakoplos, John, David Pearce & Ennio Stacchetti. 1989. Psychological games and sequential rationality. Games and Economic Behavior 1:60–79.

Goldschmidt, Walter. 1993. On the relationship between biology and anthropology. Man 28:341–359.

Granovetter, Mark. 1985. Economic action and social structure: the problem of embeddedness. American Journal of Sociology 91:481–510.

Güth, Werner, Rolf Schmittberger & Bernd Schwarze. 1982. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization 3:367–388.

Heckathorn, Douglas. 1996. The dynamics and dilemmas of collective action. American Sociological Review 61:250–277.

Hirshleifer, Jack & Juan Carlos Martinez Coll. 1991. The limits of reciprocity. Rationality and Society 3:35–64.

Hofbauer, Josef & Karl Sygmund. 1988. The theory of evolution and dynamical systems. Cambridge University Press, Cambridge.

Kalai, Ehud & Ehud Lehrer. 1995. Subjective games and equilibria. Games and Economic Behavior 8: 123–163.

Kandori, Michihiro. 1992. Social norms and community enforcement. Review of Economic Studies 59: 63–80.

Karandikar, Rajeeva, Dilip Mookherjee, Debraj Ray & Fernando Vega-Redondo. 1998. Evolving aspirations and cooperation. Journal of Economic Theory 80:292–331.

Landa, Janet T. 1981. A theory of the ethnically homogeneous middleman group: an institutional alternative to contract law. Journal of Legal Studies 19:349–362.

Landa, Janet T. 1999a. Bioeconomics of some nonhuman and human societies: new institutional economics approach. Journal of Bioeconomics 1:95–113.

Landa, Janet T. 1999b. The law and bioeconomics of ethnic cooperation and conflict in plural societies of Southeast Asia: a theory of Chinese merchant success. Journal of Bioeconomics 1:269–284.

Lewis, David. 1969. Convention. A philosophical study. Harvard University Press, Cambridge.

Matsushima, Hitoshi. 1998b. Toward a theory of subjective games. Mimeo, University of Tokio.

Menicucci, Domenico & Pier Luigi Sacco. 1997. Evolutionary dynamics with λ-players. Mimeo, Department of Economics, University of Florence.

Mueller, Dennis. 1986. Rational egoism versus adaptive egoism as fundamental postulate for a descriptive theory of human behavior. Public Choice 51:3–23.

Ockenfels, Axel & Joachim Weimann. 1999. Types and patterns: an experimental East-West-German comparison of cooperation and solidarity. Journal of Public Economics 71:275–287.

Oechssler, Jörg & Burkhard Schipper. 2003. Can you guess the game you're playing? Games and Economic Behavior 43:137–152.

Palomino, Frédéric & Fernando Vega-Redondo. 1999. Convergence of aspirations and (partial) cooperation in the prisoner's dilemma. International Journal of Game Theory 28:465–488.

Rabin, Matthew. 1993. Incorporating fairness into game theory and economics. American Economic Review 83(5):1281–1302.

Rabin, Matthew. 2002. A perspective on psychology and economics. European Economic Review 46: 657–685.

Rapaport, Carol. 1995. Apparent wage discrimination when wages are determined by nondiscriminatory contracts. American Economic Review 85(5):1263–1277.

Roth, Alvin. 1995. Bargaining experiments. Pp. 253–348 in A. Roth & J. Kagel (ed.) Handbook of Experimental Economics. Princeton University Press, Princeton.

Rubin, Paul H. 2000. Group selection and the limits to altruism. Journal of Bioeconomics 2:9–23.

Rubinstein, Ariel. 1986. Finite automata play the repeated prisoner's dilemma. Journal of Economic Theory 39:83–96.

Sacco, Pier Luigi. 1994. Discussion of Björnerstedt and Weibull's 'Nash equilibrium and evolution by imitation'. Pp. 172–181 in K.J. Arrow, E. Colombatto, M. Perlman & C. Schmidt (ed.) Rationality in Economics. Macmillan, London.

Sacco, Pier Luigi & Stefano Zamagni. 1996. An evolutionary dynamic approach to altruism. Pp. 265–300 in F. Farina, F. Hahn & S. Vannucci (ed.) Ethics, Rationality and Economic Behavior. Clarendon Press, Oxford.

Sahlins, Marshall. 1972. Stone age economics. De Gruyter, New York.

Schelling, Thomas. 1960. The strategy of conflict. Harvard University Press, Cambridge.

Schlag, Karl Hermann. 1998. Why imitate, and if so, how? Journal of Economic Theory 78:130–156.

Schnabl, Wolfgang, Peter Stadler, Christian Forst & Peter Schuster. 1991. Full characterization of a strange attractor. Physica D 48:65–90.

Sen, Amartya. 1967. Isolation, assurance and the social rate of discount. Quarterly Journal of Economics 81:112–125.

Sen, Amartya. 1974. Choice, orderings and morality. Pp. 54–67 in S. Körner (ed.) Practical Reason. Blackwell, Oxford.

Smith, Peter & Michael Bond. 1993. Social psychology across cultures. Harvester, Hemel Hampstead.

Sober, Elliot & David Sloan Wilson. 1998. Unto others. Harvard University Press, Cambridge.

Sugden, Robert. 1991. Rational choice: a survey of contributions from economics and philosophy. The Economic Journal 101:751–785.

Taylor, Michael. 1987. The possibility of cooperation. Cambridge University Press, Cambridge.

Taylor, Peter & Leo Jonker. 1978. Evolutionarily stable strategies and game dynamics. Mathematical Biosciences 61:51–63.

Weibull, Jörgen. 1995. Evolutionary game theory. The MIT Press, Cambridge.

Weimann, Joachim. 1994. Individual behavior in a free riding experiment. Journal of Public Economics 54:185–200.

Wildavsky. Aaron. 1992. Indispensable framework or just another ideology? Rationality and Society 4:8–23.