Working Paper Series

Department of Economics

University of Verona

Free Riders and Cooperators in Public Goods Experiments: Can Evolutionary Dynamics Explain their Coexistence?

Angelo Antoci, Paolo Russu, Luca Zarri

WP Number: 54 May 2009

ISSN: 2036-2919 (paper), 2036-4679 (online)

Free Riders and Cooperators in Public Goods Experiments:

Can Evolutionary Dynamics Explain their Coexistence?

Angelo Antoci (University of Sassari)

Paolo Russu (University of Sassari)

Luca Zarri (University of Verona)

Abstract

An oft-cited and robust result from Public Goods Game experiments is that, when subjects start playing, the aggregate level of contributions is significantly different from zero. At the same time, a sizeable proportion of players free ride from the outset. Behavioural economics has persuasively shown that these laboratory findings are compatible with the presence of motivationally *heterogeneous* agents, displaying both standard, self-centred preferences and non-standard, interdependent preferences. However, at the theoretical level, economists would prefer to account for motivational heterogeneity *endogenously*, instead of simply *assuming* it from the outset. Our work provides such endogenisation, by assuming that social evolution is driven by material payoffs only. By separately focusing on different types of 'experimentally salient' pro-social players (such as Reciprocators, Strong Reciprocators and Altruists), we are able to shed light – to our knowledge, for the first time, within the public good framework – on the evolutionary stability of two-type populations consisting of positive proportions of both 'nice' and 'mean' guys.

Keywords: Free Riding; Strong Reciprocity; Altruism; Nonstrategic Punishment; Public Goods Game; Evolutionary Game Theory.

JEL Classification: C7; D6; H8; Z1.

1 Introduction

Collective action dilemmas among humans remain central today within a large number of relevant domains where either positive or negative externalities are produced, from the use of shared resources to automobile traffic control. As we know from the economic literature specifically dealing with such issues for more than two decades, the problem of voluntary provision of public goods has not a straightforward way-out, as, at individual level, the well-known temptation to *ride free* on the generosity of others would arise and jeopardise the provision of the good itself[1]. However, despite the threat of socially inefficient equilibria emergence due to extensive free riding on the part of independent, 'rational' individuals, successful real-life attempts to privately provide various types of public goods abound in several societies, including projects such as street cleaning and lighting, child care, contributions to local security and playground maintenance. In other words, several collective action dilemmas turn out to be successfully addressed, even though the presence of free riders cannot be ruled out. Similar conclusions have been reached, in the last decades, through experimental research on private provision of public goods, mainly conducted within the well-known *Public Goods Game* (*PGG*) or *Voluntary Contribution Mechanism* (*VCM*) framework[2]. Such highly controlled experiments involve small groups of individuals who are endowed with a given sum of money and allowed to either invest (possibly part of) it in a public account ('cooperate') or keep it in a private account ('defect'). As Masclet et al. (2003) observe, this game is appealing as "it starkly isolates the conflict between self-interest and group interest and allows a simple measure of the extent of group-interested behavior" (p. 366). A key message from experimental research in this area is that economic theory overestimates the prevalence of free riding in *PGGs*: laboratory evidence indicates that *in the aggregate* individuals contribute significantly *more* than would be implied by pure self-interest alone (Rabin, 1993). This is particularly true in one-shot, nonrepeated game protocols (as well as in the early stages of repeated *PGGs*), where, in contrast with predictions based on a strong version of the free riding hypothesis, subjects tend to make nonzero contributions, of 40-60% of the Pareto efficient level (Dawes and Thaler, 1988; see also Anderson and Putterman, 2006). However, in paying attention to the aggregate level of positive contributions to the public account, we should carefully consider the fact that, in the same settings, a sizeable proportion of players *does* conform to the *Homo Oeconomicus* model, by riding free on the generosity of others (see e.g. Gintis, 2000; Fischbacher et al., 2001 and Kocher et al., 2008). In other words, the whole picture reveals that a significant degree of behavioural heterogeneity exists, within the group of players involved in the *PGG*: while some of them are 'irrationally' willing to contribute to the public account,

[1] The economic literature on voluntarily provided public goods, largely related to Olson's (1965) classical analysis of collective action, has been pioneered by the fundamental work by Bergstrom et al. (1986).

[2] The pioneering experiments have been run in the '70s and '80s, thanks to Marwell, Dawes, Orbell and Isaac and Walker (for an extensive survey, see Ledyard, 1995).

others act as free riders. The major goal of our model is to provide a satisfactory evolutionary account for such *heterogeneity* in the *PGG*, by predicting *neither universal free riding nor full cooperation*. We aim at providing an answer to the following question: does evolutionary dynamics lead to a stable coexistence of 'nice' and 'mean' guys in the *PGG*? Can we find polymorphic equilibria where free riders and (different types of) cooperators coexist, in the medium-long run? Our paper shows, to our knowledge for the first time, that within a *PGG* framework with three players simultaneously involved, *stable coexistence* between free riders and various types of (conditional and unconditional) cooperators is possible, in line with experimental evidence. Further, with regard to populations where a positive proportion of so called 'strong reciprocators' is initially present, coexistence interestingly occurs in correspondence with cost to impact ratios (of punishment) which are neither too high nor too low, as it is the case in the laboratory. By contrast, we find that when either the costs of punishing or the costs of being punished (or both) are very large, only equilibria with monomorphic populations are possible. These findings allow us to evolutionarily 'map' recent experimental results on the PGG with price-varying designs (Anderson and Putterman, 2006; Carpenter, 2007), as the diffusion of strong reciprocity and the aggregate level of cooperation turn out to crucially depend on both the 'price' and the impact of sanctioning. The structure of the remainder of the paper is as follows. Section 2 contains the structure of the model and Section 3 illustrates the social dynamics. Section 4 contains our major results and Section 5 concludes.

2 The Model

We shall analyse the following context. Let us consider a (very large) community of individuals enjoying the benefits of a given collective (i.e. non-rival and non-excludable) good. The interdependence among individuals manifests itself on the occasion of random encounters: in other words, we are introducing a game with a random matching structure. However, a key assumption is that randomly occurring encounters involve three players at a time: when three individuals are matched, their behaviour crucially affects the possibility to actually provide the collective good, as well as each other's enjoyment of the good. This reflects the structure of *PGGs*, where a group of N players (with $N > 2$) is involved. A key feature of our model is that while the *objective*, material game takes the form of the *PGG*, players do *subjectively* behave as their 'type' prescribes to do. The structure of the material *PGG* will be introduced in this section, whereas the different player types composing the heterogeneous populations under study will be illustrated in Section 3. In each three-player matching, the 'material game' to be played is the well-known *PGG*. More specifically, we assume that a single player has to make a *binary*, 'all-or-nothing' choice: he may either contribute to the public good (by giving a certain amount of money) or free ride[3].

[3]Hence, the present structure differs from a continuous contribution setting, where players are free to contribute by any desired amount of their endowments. Further, it it is worth

Hence, material consequences for the players depend on their choosing between contribute (or 'cooperate') and free ride (or 'defect') only. Since each agent has to decide whether to cooperate (C) or defect (D), each matching among three individuals will produce one of the following six outcomes: *(DDD), (DDC), (DCC), (CDD), (CDC), (CCC)*. As far as the material consequences of each matching are concerned, universal cooperation (CCC) Pareto-dominates mutual defection (DDD). Further, we assume that the public good to be provided is a *threshold* public good. The key feature of threshold public good games (Cadsby and Maynes, 1999) is the following: if a sufficiently large proportion of agents do contribute to the public good, so that the stated level of contributions is reached, the good is provided; otherwise, agents lose their contributions and the good is not provided[4]. In our model, we suppose that if N_{coop} is the number of players cooperating in each matching, $N_{coop} = 2$ is the 'critical threshold' of cooperators to be achieved for the public good to be privately provided. *Unilateral* free riding (that is, DCC), by exploiting the non-excludability of the good and the two opponents' cooperation, is then the most individually rewarding outcome, in material terms. By contrast, unilateral cooperation (that is, CDD) is the worst material outcome, for a single player (as she is unilaterally incurring costs for a public good that will not be provided). Hence, from the point of view of each individual player, the above six possible combinations of choices lead to the material consequences captured by the following payoff matrix:

$$\begin{array}{cccc} & DD & DC & CC \\ D & a & a & b \\ C & c & d & e \end{array} \quad (1)$$

Table 1 *3-player PGG matrix*

where $c < a < d \leq e < b^5$, C = *Cooperate* and D = *Defect*.

The above matrix has to be read as follows: the first row contains the material consequences for the row player when he plays D and the opponents both play D (first column); when he plays D, one opponent plays D and the third

observing that also most threshold public goods experiments assume that players can either decide not to contribute or to contribute by a given amount.

[4]As Myatt and Wallace (2008) correctly observe, it is the case that, beyond economics, threshold games are central to a large and significant sociological literature including works on crowd behaviour, the onset of riots and strikes and other social movements. For an evolutionary analysis of collective action focusing on the shape of public-good production functions, see Myatt and Wallace (2009).

[5]By assuming that e may be greater than d, we allow for the possibility that the threshold public good under study also possesses the following feature. Once a specific *provision-point* (the 'threshold') is met (Isaac et al., 1989), the amount of the public good may further increase, provided that the aggregate level of contributions increases. This is equivalent to assuming that contributions beyond the threshold levels, far from being wasted, result in further benefits to the group. In our model, this would imply that the individual payoff from playing C when the other two players also cooperate (that is, e) is greater than the individual payoff from playing C when one of the two opponents only cooperates, while the other defects (that is, d).

player plays C (second column); when he plays D and both opponents play C (third column), respectively. The second row has to be read analogously[6].

3 Social Dynamics

We analyse social dynamics by means of a (direct) evolutionary game-theoretical approach[7]. In line with the well-known 'indirect evolutionary approach' (IEA; see on this Güth and Yaari, 1992), we also suppose that what drives social evolution are *material* payoffs *only*[8]. More specifically, as we better clarify below in this section, we assume that players imitate the individuals who achieve the best performances, in purely material terms. As Ok and Vega-Redondo (2001) observe: "the evolutionary approach has been often used in economics to justify the assumption of material payoff maximization (...). However, it turns out that this line of argument is not true without some non-trivial qualifications, for it is possible that non-individualistic preferences are materially more rewarding than individualistic preferences in certain strategic environments" (pp. 232-233). However, a key difference between the IEA and our model is the following. In the standard structure of the IEA players act as utility maximisers. By contrast, we decided not to introduce players' subjective utilities, as, within an evolutionary framework, individuals can plausibly be characterised as boundedly rational players, crucially affected by social forces (rather by a 'calculating rationality') rewarding those who turn out to be the fittest within the population under study. In this light, we prefer to refer to 'non-utility-maximising types', rather than to rational agents lucidly attempting to maximise a given structure of preferences. In other words, we stick to the traditional evolutionary methodology, in which individual players are supposed to play 'hardwired' programmed strategies over their lifetimes. Therefore, player types will be directly defined in behavioural terms, rather than by referring to an underlying structure of (more or less sophisticated) preferences. In our model, player types prescribe the behavioural patterns which bring about specific material consequences; in turn, such material consequences drive social evolution, in the sense that the behavioural patterns that turn out to be more rewarding – in material terms – are imitated at the expense of less rewarding ones. However, in the light of real-life and experimental evidence on the coexistence of selfish and non-selfish agents within several strategically significant social contexts, we also assume that a certain degree of heterogeneity initially exists, as to the content of such fixed, 'wired-in' strategies. As Camerer and Fehr (2006) maintain: "heterogeneity in other-regarding preferences and bounded rationality, along with the structure of

[6]Such payoffs may be seen as the natural extension of a material *PD*, when three players (rather than two) are simultaneously involved.

[7]For a recent, interesting evolutionary analysis of a collective action problem, with a particular focus on threshold games in which players face a 'teamwork dilemma' (that is a strategic situation where the critical threshold is greater than one, like in our model), see Myatt and Wallace (2008).

[8]The IEA is 'indirect', as a view of social evolution, as what compete one with the other are 'cultural traits' (rather than genetic ones).

social interactions, determine when collective outcomes are close to predictions based on rationality and self-regarding preferences, or are far from those predictions (...) The new models of heterogeneous social preferences and bounded rationality explain these puzzling results in a unifying way because they explicitly take heterogeneity and incentive interactions between different types of individuals into account. Therefore, they can explain when Economic Man dominates aggregate outcomes and when he fails to do so" (p. 52). With specific reference to a motivational force such as reciprocity, Fehr and Gächter (1999), by focusing on sixteen different experimental studies, interestingly show that *reciprocally* and *selfishly* motivated people turn out to systematically coexist: in all scenarios both types are present in non-negligible proportions, though the former seems to prevail. In a similar vein, several recent papers indicate that the existence of distinct types of players emerges experimentally (see e.g. Fehr and Fischbacher, 2005 and Kocher et al., 2008). Fischbacher and Gächter (2006), by means of a new methodological strategy, both account for the existence of types in the lab and, through a direct test of the role of social preferences in voluntary cooperation, show that the interaction of heterogeneous types plays a key role in accounting for the dynamics of cooperation[9]. In the light of this, we claim that a sensible modelling choice is to restrict the strategy space to 'focal' types only and, more specifically, to the few behavioural patterns whose quantitative significance in relevant strategic interaction settings is 'revealed' and increasingly confirmed by carefully controlled economic experiments. In particular, as far as experimental games on sociality are concerned, available lab evidence indicates that the issue of heterogeneity ought to be taken seriously and that (a) a significant proportion of subjects are *self-interested* and (b) a sizeable proportion act *cooperatively* (e.g. play C in the PD or make positive contributions in $PGGs$ or in Dictator Games). However, as far as the latter, broad category is concerned (i.e. 'cooperative' players), experiments indicate that different forms of pro-sociality seems to be at work, within the strategic contexts investigated so far. The major goal of our model is to shed light on the evolutionary fate of heterogeneous populations where selfish and non-selfish players are initially present, by focusing on both the possibility to actually provide the public good and the (associated) equilibrium population mixture of player types.

3.1 Four Behavioural Types: Egoists, Strong Reciprocators, Reciprocators and Altruists

Although previous research clearly indicates that many people act as if they were driven by social preferences, not everybody does. In fact, most experimental studies show that a substantial number of subjects behave in a standard, purely selfish manner. A key question, therefore, is how the "heterogeneity of motives at the individual level can be captured by parsimonious models and how the different individual motivations interact" (Fehr and Fischbacher, 2005, p.155). The two authors also point out that available empirical evidence reveals

[9]For a similar methodological strategy, see Kurzban and Houser (2005).

that the four quantitatively most important types of social preferences that have been established so far are strong reciprocity, inequity aversion, unconditional altruism and spitefulness (or envy). In the present model, we rule out distributional social preferences such as inequity aversion[10] and spitefulness[11] and focus instead on *strong reciprocity* and *unconditional altruism.* More specifically, in line with Antoci and Zarri (2008), we explicitly draw a distinction between (simple) reciprocators and strong reciprocators and separately investigate all three pro-social types of players (namely, altruists, reciprocators and strong reciprocators) within 2-type populations where public good provision privately occurs and opportunistic players who ride free on others' generosity are also (initially) present.

3.1.1 Egoists

In all the three 2-type populations we explore in this paper, we assume that a positive proportion of the initial population is composed by classic selfish players. With reference to our 3-player *PGG*, we define a *Selfish* player (SEL; or 'Egoist'), as an individual who always plays D (i.e. he never contributes to the public good to be privately provided).

3.1.2 Altruists

In the 3-player *PGG* under study, an *Altruistic* player (ALT) is a player who always plays C (i.e. she always contributes to the public good to be provided). In other words, ALTs are unconditional cooperators who systematically contribute to the public account, regardless of what their opponents do, in each matching.

3.1.3 Reciprocators

A reciprocator (REC) here closely resembles Gintis' (2000) *Homo reciprocans*, who "comes to strategic interaction with a propensity to cooperate, responds to cooperative behavior by maintaining or increasing his level of cooperation, and responds to noncooperative behavior by retaliating against the 'offenders', even at a cost to himself, and even when he could not reasonably expect future personal gains to flow from such retaliation" (pp. 251-252). Hence a REC, unlike an ALT or a SEL, has no dominant strategies but is willing to act as a conditional cooperator, by displaying, according to his opponent's type, both positive and negative reciprocity (see on this Rabin, 1993). Reciprocators can be seen as players driven by a hardwired norm of conditional cooperation leading them to prefer the joint cooperation outcome to outcomes in which the partners get

[10]According to Fehr and Fischbacher (2005), available experimental evidence indicates that strong reciprocity is quantitatively more important than inequity aversion.

[11]Fehr and Fischbacher (2005) note that spiteful choices seem to be quantitatively less important than reciprocal choices, at experimental level. Further, spitefulness cannot account for the stylised fact that the same people are often ready to *help* others by bearing personal costs in some contexts, but are also willing to *harm* others in other contexts (p. 155).

exploited – while SELs act *as if* they were driven by the opposite preference ordering over these two types of outcomes[12]. As far as lab evidence on conditional cooperation is concerned, it is interesting to notice that in Brosig's (2002) experimental one-shot *PD* most subjects did not exploit their partners when they expected them to play cooperatively, even under conditions of anonymity. As he notes: "when subjects believed that their partner would cooperate, they hesitated to maximize their own payoff through defection" (p. 284). Analogously, Gintis (2000) observes that experimental subjects often report retrospectively that they were conditional cooperators, in the game protocol. Guttman (2003) develops a theory in which 'reciprocator types' survive in a competitive, evolutionary environment where they have to compete with 'opportunistic types' who act by maximising material payoffs and, as a consequence, prefer to exploit their opponents over jointly cooperating. However, in his model, where players are randomly matched and play a *PD* game (rather than a *PGG*), reciprocators and opportunists are defined not by their strategies, but by their *preferences* and, as a consequence, he analyses social dynamics by adopting the IEA, while (as we clarified above) we define our player types directly in behavioural terms and adopt a classic direct evolutionary approach[13].

3.1.4 Strong Reciprocators

As we clarified above, a key feature of reciprocity is the propensity to respond to defection by defecting. However, as Gintis (2000) interestingly points out, when other forms of punishment of 'unfair' opponents are available, *Homo reciprocans* is often willing to use them. In other words, it is very likely that if institutional features allow it, a conditional cooperator will behave in such a context as a Strong Reciprocator (SR), that is as a conditional cooperator who both responds to defection (cooperation) with defection (cooperation) *and* incurs costs in order to *explicitly* punish defectors. For example, this occurs in Ostrom et al. (1992). A significant level of punishment has been found also by Fehr and Gächter (2000): in that well-known study, unlike in Ostrom et al. (1992), the design included the possibility of explicit retaliation but ruled out *strategic* behaviour, since the two authors ensured that group composition changed in every period, so that subjects were aware that punishers could not benefit from costly retaliation. The main reason why we decided to also explore a 2-type population where SELs interact with SRs is that experiments suggest

[12]In this regard, Fehr and Fischbacher (2005) observe: "Economists and biologists defined the term 'reciprocity' in the past in different ways. Biologists think of reciprocity, or 'reciprocal altruism', as tit-for-tat strategies in repeated interactions (Trivers 1971; Axelrod and Hamilton 1981). Some economists (Binmore 1998) use the term in a similar way. During the last ten years, however, an increasing number of contributions have shown that reciprocal behavior also exists in sequentially structured one-shot interactions. Reciprocity in one-shot interactions cannot be explained on the basis of selfish motives" (p. 184).

[13]A further important difference between the two papers is that while, as we clarify below in this section, we suppose that the time horizon is infinite, Guttman (2003) combines the IEA with a *finitely* repeated game: within such evolutionary environment, he finds that both the reciprocator and the opportunist type will optimally cooperate over most of the stages of their careers.

that the presence of explicit, targeted punishment opportunities crucially affects the final aggregate outcomes (see Fehr and Gächter, 2000 and Ones and Putterman, 2007). In other words, we do this in the light of the fact that experimental research clearly indicates that the final outcome of a given game is extremely sensitive to the specific features of the punishment options available to the players (Masclet et al., 2003; Anderson and Putterman, 2006). Specifically, as far as punishment is concerned, the two different treatments of the *PGG* analysed in Fehr and Gächter (2000) make clear that the overall levels of cooperation in the population seem to critically depend on the possibility or not, for the subjects, to have access to direct targeted punishment, rather than to implicit punishment (in the form of low contributions to the public good to be provided) only[14]. In this light, in the 3-player *PGG* under study, a SR is both willing to (conditionally) cooperate and incurs costs in order to punish defectors: hence, we assume that the propensity to incur costs in order to explicitly punish defectors is what differentiates SRs from RECs[15]. Real-life cases of nonstrategic punishment occur when participants in bloody family feuds seek revenge even when it is extremely costly to do so and when people decide to walk away from profitable transactions whose terms they believe to be unfair (see on this Frank, 1987).

3.2 Information Assumptions and Replicator Dynamics

However, a key point is that information assumptions play a crucial role, for both reciprocators and strong reciprocators. In this light, we investigate strong reciprocators' behaviour by separately assuming ex ante recognition and ex post recognition of the opponent's type. As we will see in Section 4, this will allow us to consider three types of SRs, that is *brave*, *consequentialist* and *vengeful* SRs. As far as RECs are concerned, we suppose that the opponent's type can be recognised ex ante, so that we separately focus on consequentialist and vengeful RECs only. We further suppose that time is continuous and the population is modelled as a continuum of players. As anticipated above, a random matching structure exists, so that 3-player interactions continuously occur over time. As far as such 3-player matchings are concerned, the material game that individuals play is the *PGG* illustrated in Section 2[16]. On the whole, we (separately) focus on three 2-type populations of boundedly rational players. Hence, x and $(1-x)$ indicate the proportions of individuals of the types SEL and SR; SEL and REC and SEL and ALT, respectively. Following Taylor and Jonker (1978), we suppose that the growth rates of the proportions are given by the well-known replicator

[14]Fehr and Gächter (2002) interestingly find that a clear relationship exists between contributing and punishing: in their *PGG* experiment, 75% of the punishment acts carried out by the 240 subjects have been executed by above-average contributors.

[15]The analysis of this case is important, since, as Fehr and Fischbacher (2005) maintain: "empirical evidence clearly suggests that in the domain of payoff-decreasing or punishing behavior, strong negative reciprocity is the dominant motive" (p. 155).

[16]Miller and Andreoni's (1991) work is the first theoretical paper aimed at providing experimental evidence on the *PGG* with evolutionary foundations.

equations[17]:

$$\dot{x} = x(1-x)\left[E\Pi_{SEL}(x) - E\Pi_i(x)\right]$$

where $E\Pi_{SEL}$ are the expected payoffs of type SEL and $E\Pi_i$ are the expected payoffs of type i = SR, REC, ALT. Expected payoffs are calculated by using conditional probabilities, as we will see below.

Replicator dynamics are a widely adopted model of social (as well as natural) selection dynamics characterised by payoff monotonicity, where the most rewarding strategies survive and spread over at the expense of less rewarding ones (see on this also Weibull, 1995).

4 Results

In this section, we present our major results, with regard to the three 2-type populations we focus on, namely SEL-SR (Case 1), SEL-REC (Case 2) and SEL-ALT (Case 3). However, before separately considering these cases, a preliminary clarification is in order. In both Case 1 and Case 2, a positive proportion of conditional cooperators exists, in the initial population, and whenever three (either strong or simple) reciprocators meet, both CCC and DDD may arise, as an equilibrium. With regard to such homogeneous matchings, we suppose, in line with other works (see e.g. Guttman, 2003 and Antoci and Zarri, 2008), that when three (either strong or simple) reciprocators meet, a social norm prescribes that the Pareto-superior (that is, CCC) outcome will prevail.

4.1 Case 1. SEL-SR Population

Let us first investigate a mixed population initially composed of Egoists (SELs), who systematically defect, and Strong Reciprocators (SRs), who are both willing to conditionally cooperate and ready to costly punish 'mean' guys, that is unconditional defectors[18]. Here we consider three subcases, where different information and behavioural assumptions regarding SRs are separately introduced. As Ok and Vega-Redondo (2001) correctly point out, the answer to the question concerning how the (expected) material payoffs of the individualistic and non-individualistic agents compare in equilibrium at various population compositions crucially depends, inter alia, on the extent of information agents have on their opponent's type.

[17]However, the results of the paper remain the same under any other payoff-monotonic adoption dynamics (see Weibull 1995).

[18]Also Sethi and Somanathan (2005) focus on this 2-type population, in their evolutionary analysis of common property resource use.

4.1.1 Subcase 1.1. Ex Post Recognition and Brave Strong Reciprocators

In this first subcase, we suppose that SRs do *not* recognise their opponents' type *ex ante*, so that they bravely play C in each matching (as ALTs do)[19]. However, we also assume that, after cooperating, they can recognise their opponents' type (ex post recognition assumption) and that if SRs see that their opponent is a SEL, they are willing to incur material costs in order to punish her. More specifically, we believe it is plausible to assume also that the extent of costs for both punishers and punishees critically depends on the number of SELs and SRs involved in the 3-player matching: when a single SR meets two SELs, the costs are given by $\lambda > 0$ (for the punisher) and $\varepsilon > 0$ (for the punishees), respectively, whereas when two SRs meet a SEL, the costs are given by $\pi > 0$ and $\eta > 0$, respectively. As a consequence, depending on the number of SELs and SRs being present in a single 3-player matching, at the individual level the PGG will lead to one of the (material) outcomes captured by the matrix below:

$$\begin{array}{cccc} & SEL, SEL & SEL, SR & SR, SR \\ SEL & a & a-\varepsilon & b-\eta \\ SR & c-\lambda & d-\pi & e \end{array} \tag{2}$$

Table 2. *Material Payoffs in a SEL-SR population (Subcase 1.1)*

In this subcase, expected payoffs are:

$$E\Pi_{SEL}(x) = ax^2 + (a-\varepsilon)x(1-x) + (b-\eta)(1-x)^2$$

$$E\Pi_{SR}(x) = (c-\lambda)x^2 + (d-\pi)x(1-x) + e(1-x)^2$$

and replicator dynamics can be written as follows:

$$\dot{x} = x(1-x)\left[\alpha(\varepsilon)x^2 + \beta(\varepsilon,\lambda)x + \gamma(\varepsilon)\right] \tag{3}$$

where:

$$\begin{aligned} \alpha(\varepsilon) &:= b-e+d-c-\varepsilon = \overline{\alpha}-\varepsilon, \qquad \overline{\alpha} > 0 \\ \beta(\varepsilon,\lambda) &:= a-d+2(e-b)+\lambda+3\varepsilon = \overline{\beta}+\lambda+3\varepsilon, \qquad \overline{\beta} < 0 \\ \gamma(\varepsilon) &:= b-e-2\varepsilon = \overline{\gamma}-2\varepsilon, \qquad \overline{\gamma} > 0 \end{aligned}$$

Moreover, let us define $\Delta(\varepsilon,\lambda) := (\beta(\varepsilon,\lambda))^2 - 4\alpha(\varepsilon)\gamma(\varepsilon)$ and $\lambda_0 := 2\overline{\gamma} + d - a - 2\sqrt{(\overline{\gamma}+d-c)\overline{\gamma}}$.

On the basis of simple algebraic calculations, we obtain the following proposition:

Proposition 1 *The fixed point $x^* = 1$ always exists and is always attractive. Further,*

[19] This attitude resembles Sugden's (1986) notion of 'brave' reciprocity.

i.) if $\varepsilon < \dfrac{b-e}{2}$*,* $\Delta(\epsilon, \lambda) < 0$*, (Region* ***1*** *in Figure 1(a)), the dynamic regime is represented by case* ***1,*** *in Figure 1(b);*

ii.) if $\varepsilon < \dfrac{b-e}{2}$*,* $\Delta(\epsilon, \lambda) > 0$*, (Region* ***2*** *in Figure 1(a)), the dynamic regime is represented by case* ***2*** *in Figure 1(b);*

iii.) if $\varepsilon > \dfrac{b-e}{2}$*, (Region* ***3*** *in Figure 1(a)), the dynamic regime is represented by case* ***3*** *in Figure 1(b)*

The value of λ_0 is strictly positive (as it is in Figure 1(a)) if $b - e < \frac{(d-a)^2}{4(a-c)}$. By contrast, if λ_0 is negative, the curve $\Delta(\varepsilon, \lambda) = 0$ crosses the ε-axis in $\varepsilon_0 > 0$ and the tangency point (that always exists) between the line $\varepsilon = \frac{b-e}{2}$ and this curve shifts upwards.

4.1.2 Subcase 1.2. Ex Ante Recognition and Consequentialist Strong Reciprocators

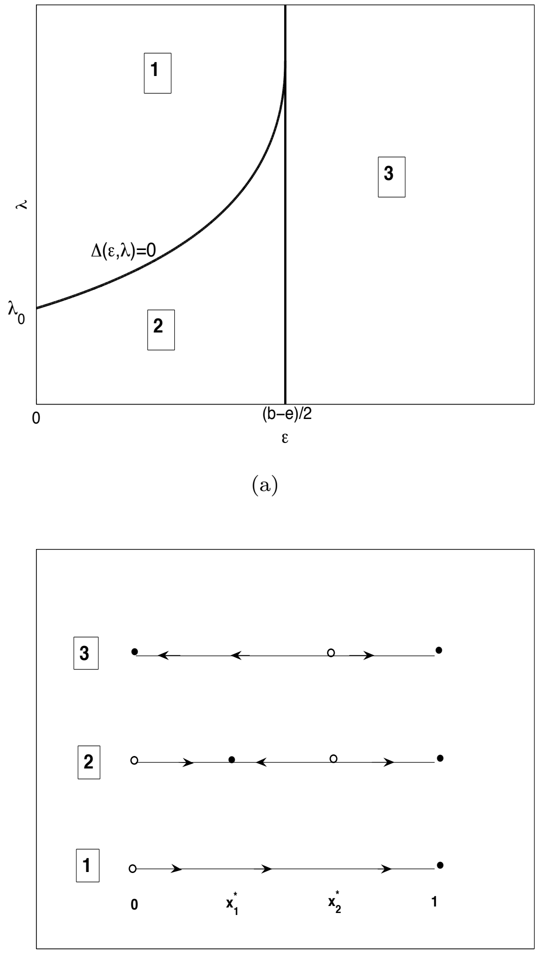
Here we suppose that, unlike in Subcase 1.1., SRs can recognise the opponents' type *ex ante*. In this regard, it has been argued that players tend to subconsciously signal their type via facial expressions and other emotional factors (Frank, 1988). Hence, we assume that, when meeting a SEL and another SR, SRs decide to play C, as in that case (with two cooperating agents) they would reach the 'critical threshold' for the public good to be provided. In other words, what we are considering here is a variant of Strong Reciprocity in which a *consequentialist* component plays a key role. Further, like in Subcase 1.1., we suppose that SELs are explicitly (and costly) punished by SRs (the levels of costs are the same as in the previous subcase):

$$\begin{array}{cccc} & SEL, SEL & SEL, SR & SR, SR \\ SEL & a & a-\varepsilon & b-\eta \\ SR & a-\lambda & d-\pi & e \end{array} \tag{4}$$

Table 3. *Material Payoffs in a SEL-SR population (Subcase 1.2)*

Let us observe the key differences, compared to the previous subcase. The new curve $\Delta(\varepsilon, \lambda) = 0$ now is always above the curve depicted in Figure 1: it crosses the λ-axis in correspondence with the point $\lambda_{02} = 2\overline{\gamma} + d - a - 2\sqrt{(\overline{\gamma} + d - a)\overline{\gamma}}$ which is greater than λ_0, whereas the tangency point with the vertical line is the same as before. Further, the content of the previous footnote does not hold, in this subcase, as λ_{02} is always positive. In conclusion, the region with an attractive internal fixed point (Region 2) in Subcase 1.2 is always larger than the one we found in Subcase 1.1. The subsequent Figures show this. Clearly, all the parameters are the same.

Proposition 2 *In this subcase the dynamic regimes of the system are the same as the ones of the dynamic regime (3). Further, Region 2 in Figure 1(a), other things equal, is larger than in Subcase 1.1.*

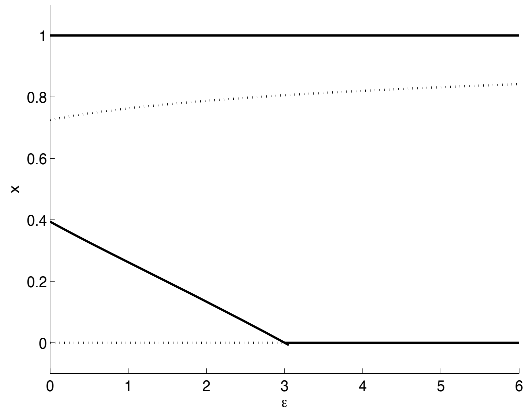


(a)

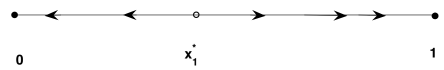
(b)

Figure 1: Dynamic regimes in subcase 1.1.

Proof. Notice simply that the curve $\Delta(\varepsilon, \lambda) = 0$ shifts parallelly upwards. ■



(a) Subcase 1.1



(b) Subcase 1.2

Figure 2: Fixed points values, varying the parameter ε. Repulsive (respectively, attractive) fixed points are represented by dotted (respectively, continuous) curves. The parameters' values are: $b = 26$, $e = 20$, $d = 20$, $a = 5$, $c = 1$, $\lambda = 3.5$.

4.1.3 Subcase 1.3. Ex Ante Recognition and Vengeful Strong Reciprocators

Like in Subcase 1.2, also in this third subcase we assume that SRs can recognise the opponents' type ex ante. However, we suppose here that such players, far from reasoning consequentialistically, can be characterised as 'vengeful' SRs, in the sense that they are willing to systematically sanction SELs not only by means of explicit punishment (like in the previous subcases), but also *implicitly*, that is by *always defecting* in each 3-player matching where at least *one* selfish agent is present, regardless of the material consequences that this choice will bring about. Hence, the material payoff matrix now takes the following form:

	SEL, SEL	SEL, SR	SR, SR
SEL	a	$a-\varepsilon$	$a-\eta$
SR	$a-\lambda$	$a-\pi$	e

(5)

Table 4. *Material Payoffs in a SEL-SR population (Subcase 1.3)*

Here the dynamics are represented by the equation:

$$\dot{x} = x(1-x)\left[\alpha(\varepsilon)_1 x^2 + \beta(\varepsilon,\lambda)_1 x + \gamma(\varepsilon)_1\right] \tag{6}$$

where:

$$\alpha(\varepsilon)_1 := -\overline{\gamma}_1 - \varepsilon < 0$$
$$\beta(\varepsilon,\lambda)_1 := 2\overline{\gamma}_1 + \lambda + 3\varepsilon > 0$$
$$\gamma(\varepsilon)_1 := -\overline{\gamma}_1 - 2\varepsilon < 0$$

with $\overline{\gamma}_1 := e - a > 0$.

Proposition 3 *For any parameter choice, the dynamic regime here is the one illustrated in Figure 3*

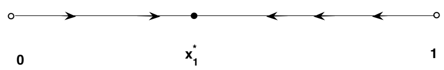


Figure 3: Dynamic regime in Subcase 1.3

4.2 Egoists and Strong Reciprocators can Coexist

On the whole, in a SEL-SR population where a 3-player *PGG* is continuously played, we found that the attractive fixed points critically depend on both information and behavioural assumptions concerning Strongly Reciprocal players. In particular, coexistence between SRs and SELs cannot occur when ex ante recognition holds and SRs are supposed to act as 'vengeful' players who defect whenever their 3-player interactions include at least one selfish player (Subcase 1.3). By contrast, coexistence of Egoists and Strong Reciprocators may be observed if we suppose that Strong Reciprocators can recognise the opponents' type *ex post only* and act *bravely* in each matching. Further, a similar conclusion can be reached even when SRs are supposed to be able to recognise their opponents' type *ex ante* (*à la* Frank), provided that they are driven by a *consequentialist* attitude: in this scenario, coexistence arises as an equilibrium outcome, insofar as, in 3-player interactions involving one SR and two SELs, the costs for both the punisher and the punishees are not too large (see Figure 1(a))[20].

4.3 Case 2. SEL-REC Population

Let us now turn to a mixed population initially composed of Egoists (SELs) and (Simple) Reciprocators (RECs) who, unlike SRs, do *not* incur costs in order to *explicitly punish* defectors. In other words, defecting is the only means of (implicit) punishment of their opponents that RECs have access to. Here we assume that players can recognise their opponents ex ante. However, analogously to the case of Strong Reciprocators, we separately analyse two subcases in which these non-selfish players are supposed to act as Consequentialist and Vengeful RECs, respectively.

4.3.1 Subcase 2.1. Ex Ante Recognition and Consequentialist Reciprocators

	SEL, SEL	SEL, REC	REC, REC	
SEL	a	a	b	(7)
REC	a	d	e	

Table 5. *Material Payoffs in a SEL-REC population (Subcase 2.1)*

In this subcase, $\varepsilon = \lambda = 0$. Hence, we are on the origin of the axis of Region 2, where the internal fixed point $x_1^* = \dfrac{b-e}{b-e+d-a} < 1$ always exists and the other internal fixed point coincides with $x_2^* = 1$. Therefore, the following proposition holds:

Proposition 4 *For any parameters choice, the dynamic regime is the one illustrated in Figure 4.*

4.3.2 Subcase 2.2. Ex Ante Recognition and Vengeful Reciprocators

	SEL, SEL	SEL, REC	REC, REC	
SEL	a	a	a	(8)
REC	a	a	e	

Table 6. *Material Payoffs in a SEL-REC population (Subcase 2.2)*

[20]This is in line with a key feature of most experimental games, where (unlike *PGGs* with price-varying designs such as Anderson and Putterman, 2006 or Carpenter, 2007) punishment costs are held fixed and supposed not to be too large.

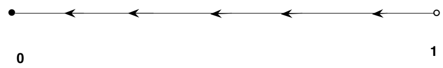


Figure 4: Dynamic regime in Subcase 2.1.

This subcase can be seen as Subcase 1.3. 'degenerating' without internal fixed points.

Proposition 5 *For any parameters choice, the dynamic regime is the one illustrated in Figure 5.*

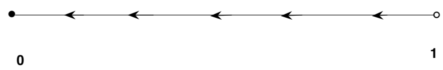


Figure 5: Dynamic regime in Subcase 2.2.

Within a SEL-REC population where a 3-player PGG is continuously played, a mixture of SELs and RECs arises when SELs initially coexist with Consequentialist RECs. By contrast, insofar as Vengeful RECs interact with SELs, evolutionary dynamics leads to a monomorphic population composed exclusively of reciprocators.

4.4 Case 3. SEL-ALT Population

Finally, let us focus on a population where only Egoists (SELs), who systematically defect, and Altruists (ALTs), who systematically cooperate, are present.

	SEL, SEL	SEL, ALT	ALT, ALT
SEL	a	a	b
ALT	c	d	e

(9)

Table 7. *Material Payoffs in a SEL-ALT population*

In such context, the dynamics are:

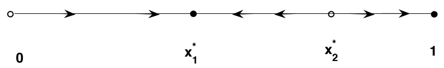
$$\dot{x} = x(1-x)(\alpha x^2 + \beta x + \gamma)$$

where:

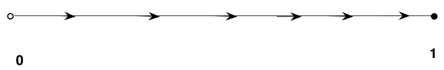
$$\begin{aligned}\alpha &:= d + b - c - e > 0 \\ \beta &:= -2b + 2e + a - d < 0 \\ \gamma &:= b - e > 0\end{aligned}$$

In suct context, the following proposition holds:

Proposition 6 *If* $|\beta|^2 - 4\alpha\gamma > 0$*, the dynamic regime is represented in Figure 6(a); if* $|\beta|^2 - 4\alpha\gamma < 0$*, the dynamic regime is represented in Figure 6(b).*



(a)



(b)

Figure 6: Dynamic regimes in Case 3.

5 Concluding Remarks

The growing success obtained in the last decades by behavioral economics crucially depends on its ability to convincingly show, both theoretically and experimentally, that within several economically interesting contexts, subjects are often willing to 'irrationally' cooperate, reward and punish others even at a cost and regardless of any expectation of economic benefits. These forms of costly cooperation, rewarding and sanctioning, which in the laboratory turn out to be often carried out also in one-shot interactions, in the last period of finitely repeated interactions and in perfect stranger treatments (in which no subject encounters another more than once) can then reasonably be qualified as genuinely *nonstrategic*[21] – at least to a significant extent –, as they differ

[21] Neuroeconomic evidence has been increasingly showing that the desire to punish is often strong and nonstrategic, and that it is a largely automatic response triggered by emotional

in a fundamental respect from the (usually complicated) punishment strategies explored in repeated games where self-interested players only are involved (like the ones studied within the classic Folk Theorem literature). The model presented in the previous sections provides well-known stylised experimental facts with evolutionary foundations, with reference to 2-type populations in which selfish free riders and (different types of) unselfish agents are initially present. Our major result is that, like in Guttman (2003), the equilibrium population, far from being monomorphic, may turn out to be a *stable mixture* of selfish and non-selfish types[22]. To our knowledge, this is the first paper where such mixed composition of the equilibrium population emerges with regard to a *PGG* setting. Further, while other theoretical works find that reciprocators coexist with egoists in equilibrium, in our contribution Selfish players may coexist in the long-run not only with Reciprocators, but also with Strong Reciprocators and with Altruists (see Case 1 and Case 3 in Section 4, respectively), that is with unconditional – rather than conditional – cooperators. Moreover, our results indicate that coexistence of 'nice' and 'mean' guys in equilibrium may occur under both ex ante and ex post type recognition: under both information assumptions, we found that a mix of Egoists and Strong Reciprocators can prevail, in the medium-long run. Such coexistence results are in line not only with a real-life widespread phenomenon such as the simultaneous presence of 'nice' and 'mean' guys in the same social context, but also with a robust finding emerging from experiments dealing with social preferences: the systematic coexistence, within a given experimental framework, of selfish and non-selfish behaviours (see on this Fehr and Gaechter, 1999 and the considerations developed in the Introduction, in this paper). More specifically, these results appear to be consistent with laboratory evidence from *PGG* experiments, where partial cooperation emerges as a robust finding in *PGGs* (both in one-shot versions and in the earlier rounds of repeated *PGGs)* and behaviourally heterogeneous players seem to be systematically involved.

Finally, as far as SEL-SR populations are concerned, we found that the above results critically depend on the costs and impact of punishment. In this regard, a new but significant experimental research area has been developing, in the last years, within the so called 'economics of punishment' literature, with the aim of investigating the sensitivity of nonstrategic punishment to standard economic incentives (see Egas and Riedl, 2005; Anderson and Putterman, 2006 and Carpenter, 2007). In particular, by means of price-varying designs of the *PGG*, these authors ask whether punishers are unresponsive to the costs of punishing or whether also nonstrategic sanctions, like ordinary goods, obey the law

forces. At the same time, experiments indicate that such nonstrategic responses are far more frequent when the punisher perceives the opponent's type as unfair (see Singer and Fehr, 2005).

[22] It is important to remark that we are able to reach this conclusion with reference to a *PGG* framework where 3-player matchings continuously occur within 2-type populations, whereas Antoci and Zarri (2008) show that coexistence never occurs with *pairwise* random matchings and different 3-type populations (such as Altruists, Egoists and various forms of Strong Reciprocators, that have been separately considered and supposed to initially interact with the former two player types).

of demand. The three independently undertaken studies interestingly provide the same (somewhat paradoxical) answer to such key question: *the demand for nonstrategic punishment displays the usual downward slope with respect to price*[23]. Hence, both cooperation rates and quantity of punishment in the PGG critically depend inter alia on the cost parameters. We evolutionarily map this conclusion as also in our model we may end up with dramatically different equilibrium outcomes (both in terms of cooperation rates and SEL-SR proportions) depending on the costs to punisher and punished. Further, Anderson and Putterman (2006) interestingly find that punishment occurs even at higher cost to the punisher than to the target. This is in line with our evolutionary results: in our analysis, coexistence of SELs and SRs occurs in correspondence with relatively low levels of both costs of punishing and costs of being punished. Whenever either the costs to punisher or the costs of being punished (or both) are extremely large, coexistence is ruled out, as an equilibrium outcome. However, as it is evident from Region 2 in Figure 1(a), coexistence of Egoists and Strongly Reciprocal players in equilibrium is compatible with ratios of costs to punisher and punished greater than one.

References

[1] Anderson C.M., Putterman L., 2006, Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism, Games and Economic Behavior, 54, 1-24.

[2] Andreoni J., 1988, Why free ride? Strategies and learning in public goods experiments, Journal of Public Economics, 37, 3, 291-304.

[3] Andreoni J., Miller J.H., 1993, Rational co-operation in the finitely repeated prisoner's dilemma: experimental evidence, Economic Journal, 103, 570-585.

[4] Antoci A., Zarri L., 2008, On punishing non-punishers. When (not so) nice guys deserve the stick, not the carrot, mimeo, University of Verona.

[5] Bergstrom T.C., L. Blume, H. Varian., 1986, On the private provision of public goods, Journal of Public Economics, 29, 25-49.

[6] Brosig J., 2002, Identifying cooperative behavior: some experimental results in a prisoner's dilemma game, Journal of Economic Behavior and Organization, 47, 275-290.

[23] Anderson and Putterman (2006) show that this result is robust to controlling for the significant effect of the level of free riding and to latent heterogeneity in the propensity to punish free riders. As they observe: "despite the fact that emotions may be the proximate cause of punishment, as argued by Fehr and Gächter and suggested by punishments at price:cost rations of 1:1 or more, the responsiveness of punishment to price suggests that the phenomenon still behaves like a taste for a good being traded off against its money cost, which represents other foregone satisfactions" (p. 20).

[7] Burlando R., Hey J.D., 1997, Do Anglo-Saxons free ride more?, Journal of Public Economics, 64, 41-60.

[8] Cadsby C.B., Maynes E., 1999, Voluntary provision of threshold public goods with continuous contributions: experimental evidence, Journal of Public Economics, 71, 53-73.

[9] Camerer C., Fehr E., 2006, When does 'economic man' dominate social behavior?, Science, 311, 47-52.

[10] Carpenter J.P., 2007, The demand for punishment, Journal of Economic Behavior and Organization, 62, 522-542.

[11] Dawes R., Thaler R., 1988, Anomalies: cooperation, Journal of Economic Perspectives, 2, 187-197.

[12] Egas M., Riedl A., 2005, The economics of altruistic punishment and the demise of cooperation, Tinbergen Institute Discussion Paper n. 065/1, Amsterdam.

[13] Fehr E., Fischbacher U., 2005, The economics of strong reciprocity, in Gintis H., Bowles S., Boyd R., Fehr E. (eds), Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life, Cambridge (Mass.) and London, MIT Press, 151-191.

[14] Fehr E., Gächter S., 1999, Reciprocal fairness, heterogeneity, and institutions, Paper presented at the AEA Meeting, New York.

[15] Fehr E., Gächter S., 2000, Cooperation and punishment, American Economic Review, 90, 4, 980-994.

[16] Fehr E., Gächter S., 2002, Altruistic punishment in humans, Nature, 415, 137-140.

[17] Fischbacher U., Gächter S., 2006, Heterogeneous social preferences and the dynamics of free riding in public goods, IZA Discussion Paper N. 2011.

[18] Fischbacher U., Gächter S., Fehr E., 2001, Are people conditionally cooperative? Evidence from a public goods experiment, Economics Letters, 71, 397-404.

[19] Frank R., 1987, If homo economicus could choose his own utility function, would he want one with a conscience?, American Economic Review, 77, 4, 593-604.

[20] Frank R., 1988, Passions within Reason. The Strategic Role of the Emotions, New York, Norton.

[21] Gintis H., 2000, Evolving Game Theory, Princeton, Princeton University Press.

[22] Güth W., Yaari, 1992, Explaining reciprocal behavior in simple strategic games: an evolutionary approach, in Witt U. (ed.), Explaining Process and Change: Approaches to Evolutionary Economics, Ann Arbor, University of Michigan Press, 23-34.

[23] Guttman J.M., 2003, Repeated interaction and the evolution of preferences for reciprocity, Economic Journal, 113, 631-656.

[24] Isaac R.M., Schmidtz D., Walker J.M., 1989, The assurance problem in a laboratory market, Public Choice, 62, 217-236.

[25] Isaac R.M., Walker J.M., Williams A.W, 1994, Group size and the voluntary provision of public goods, Journal of Public Economics, 54, 1-36.

[26] Kocher M., Cherry T., Kroll S., Netzer R.J., Sutter M., 2008, Conditional cooperation on three continents, Economics Letters, 101, 175-178.

[27] Kurzban R., Houser D., 2005, Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations, Proceedings of the National Academy of Sciences, 102, 5, 1803-1807.

[28] Ledyard J., 1995, Public goods: A survey of experimental research, in Kagel J., Roth A. (eds.), Handbook of Experimental Economics, Princeton, Princeton University Press.

[29] Masclet D., Noussair C., Tucker S., Villeval M.C., 2003, Monetary and nonmonetary punishment in the voluntary contributions mechanism, American Economic Review, 93, 1, 366-380.

[30] Miller J.H., Andreoni J., 1991, Can evolutionary dynamics explain free riding in experiments?, Economics Letters, 36, 9-15.

[31] Myatt D.P., Wallace C., 2008, When does one bad apple spoil the barrel? An evolutionary analysis of collective action, Review of Economic Studies, 75, 499-527.

[32] Myatt D.P., Wallace C., 2009, Evolution, teamwork and collective action: production targets in the private provision of public goods, Economic Journal, 119, 61-90.

[33] Ok E.A., Vega-Redondo F., 2001, On the evolution of individualistic preferences: An incomplete information scenario, Journal of Economic Theory, 97, 231-254.

[34] Ockenfels A., Weimann J., 1999, Types and patterns: an experimental East-West-German comparison of cooperation and solidarity, Journal of Public Economics, 71, 275-287.

[35] Olson M., 1965, The Logic of Collective Action: Public Goods and the Theory of Groups, Cambridge (Ma.), Harvard University Press.

[36] Ones U., Putterman L., 2007, The ecology of collective action: A public goods and sanctions experiment with controlled group formation, Journal of Economic Behavior and Organization, 62, 495-521.

[37] Ostrom E., Walker J., Gardner R., 1992, Covenants with and without a sword: self-governance is possible, American Political Science Review, 86, 404-417.

[38] Rabin M., 1993, Incorporating fairness into game theory and economics, American Economic Review, 83, 1281-1302.

[39] Sethi R., Somanathan E., 2005, Norm compliance and strong reciprocity, in Gintis H., Bowles S., Boyd R., Fehr E. (eds), Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life, Cambridge (Mass.) and London, MIT Press., 229-250.

[40] Singer T., Fehr E., 2005, The neuroeconomics of mind reading and empathy, American Economic Review, 95, 2, 340-345.

[41] Sugden R., 1986, The Economics of Rights, Cooperation and Welfare, New York, Palgrave Macmillan.

[42] Taylor P., Yonker L., 1978, Evolutionary stable strategies and game dynamics, Mathematical Biosciences 40, 145-156.

[43] Weibull J., 1995, Evolutionary Game Theory, MIT Press, Cambridge.

[44] Weimann J., 1994, Individual behavior in a free riding experiment, Journal of Public Economics, 54, 185-200.