Nº 333
August 2008

Documento de Trabajo

ISSN (edición impresa) **0716-7334**
ISSN (edición electrónica) **0717-7593**

On the Evolution of Reciprocal Behavior: A Game Theoretic Approach

Rodrigo Harrison
Mauricio Villena

Versión impresa ISSN: 0716-7334
Versión electrónica ISSN: 0717-7593

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA

Oficina de Publicaciones
Casilla 76, Correo 17, Santiago
www.economia.puc.cl

ON THE EVOLUTION OF RECIPROCAL BEHAVIOR: A GAME THEORETIC APPROACH

Rodrigo Harrison*
Mauricio Villena**

Documento de Trabajo Nº 333

Santiago, Agosto, 2008

*harrison@faceapuc.cl
**mauricio.villena@uai.cl

INDEX

On the Evolution of Reciprocal Behavior: A Game Theoretic Approach

Rodrigo J. Harrison* and Mauricio G. Villena†

August 2008

Abstract

Why does an altruistically inclined player behave altruistically in some contexts and egoistically or spitefully in others? This article provides an economic explanation to this question. The basic argument is centered on the idea that social norms shape our preferences through a process of cultural learning. In particular, we claim that, in contexts with a stable norm of reciprocity, an altruistic player can respond in kind to egoistic or spiteful players by behaving either egoistically or spitefully when confronting them and yet continue to be an altruistic player. This is why, instead of studying the evolution of preferences as such, in this work we analyze the evolution of social norms that indirectly determine individual preferences and behavior. Such a study requires that we distinguish between players' behavioral preferences, or those individuals show with their behavior, and players' intrinsic preferences, or those they inherently support or favor. We argue that, whereas the former can change through the evolution of social norms, in this case a reciprocity norm, the latter are not subject to evolutionary pressures and, therefore, we assume them to be given.

JEL codes: C72, A13

Keywords: Social Norms, Reciprocity, Endogenous Preferences, Asymmetric Evolutionary Game

1 Introduction

Sometimes the preferences that we show in our day to day social interactions are not those we prefer to display or those that truly reflect our intrinsic preferences. Social norms, either formal or informal, do somehow shape our preferences. That explains, for instance, why sometimes we behave differently in groups than when

*Instituto de Economia, Pontificia Universidad Católica de Chile.Casilla 76 - Correo 17, Santiago - Chile. Phone: 56/2/354 7102, Fax: 56/2/553 6472, E mail: harrison@faceapuc.cl

†Corresponding Author. Escuela de Negocios Universidad Adolfo Ibañez, Chile. Av. Diagonal las Torres 2640, Peñalolén. Santiago, Chile. Phone: 56/2/675 4636, Fax: 56/2/2784413, E-mail: mauricio.villena@uai.cl.

we are alone and why our behavior differs in diverse contexts. It can further be argued that this is a two-way relationship, with people's preferences also shaping social norms. Nevertheless, this process usually requires a longer span of time. If someone's preferences, for example, do not coincide with a social norm, it is more likely that the individual's preferences will accommodate to the social norm before the social norm adjusts to the preferences of that specific individual.

Although the relationship between social norms and preferences has been widely recognised in the economic literature (for details and further references, see Ch. 11 in Bowles (2004)), there is still much to be done in the way of formal economic modelling. This essay is intended to contribute therein, as its main aim is to formally study the effect of the evolution of social norms on individual preferences. In particular, we analyse the impact of a norm of reciprocity on preferences in a specific economic context characterised by negative externalities and strategic substitutes. Reciprocity is one of the most well-known types of social norms identified in the literature and can basically be described as the human tendency to respond to the actions of others with similar actions (see for instance Gouldner (1960); Elster (1989)).

In order to formally model the evolution of a reciprocity norm, we use the evolutionary game-theoretical concept of evolutionary stable strategy (ESS).[1] The ESS notion allows us to represent a situation in which a social norm that is pursued by the entire population cannot be invaded by an alternative "mutant" strategy, in the sense that no mutation adopted by an arbitrarily small fraction of individuals can enter and survive by attaining at least a comparable monetary payoff (Maynard Smith and Price (1973); Maynard Smith (1974 and 1982)).

Unlike most previous theoretical research on the analysis of the evolutionary stability of preferences, we suggest that what really evolve are not the agent's intrinsic preferences, but rather the norm of reciprocity.[2] In other words, we argue that an altruistic player, for example, will not become a permanently egoistic or spiteful player just because, in a competitive environment, she loses in terms of material payoffs if she does not evolve towards egoistic or spiteful preferences. Instead, we claim that, in contexts with a stable norm of reciprocity, an altruistic player can respond in kind to egoistic or spiteful players by behaving either egoistically or spitefully when confronting them and yet continue to be an altruistic player. This is why, instead of studying the evolution of preferences as such, we are interested in analyzing the evolution of social norms that indirectly determine individual preferences and behavior. Such a study requires that we distinguish between players' behavioral preferences, or those individuals show with their behavior, and players' intrinsic preferences, or those they inherently

[1] For further details about evolutionary game theory see inter alia: Van Damme (1987; 1994), Kandori (1997), Mailath (1992), Matsui (1996), Samuelson (1997), Vega-Redondo (1996), Villena and Villena (2004) and Weibull (1996; 1998).

[2] For previous work on the evolution of preferences in general, see inter alia: Bergstrom (1995), Bowles and Gintis (1999), Bester and Güth (1998), Güth (1995), Huck and Oechssler (1998), Koçkesen et al., (2000a; 2000b), OK and Vega-Redondo, F. (2000), Sethi and Somanathan (2000) and Sobel (2005).

support or favour. We argue that, whereas the former can change through the evolution of social norms, in this case a reciprocity norm, the latter are not subject to evolutionary pressures and, therefore, we assume them to be given. In this framework, intrinsic preferences can be thought of as acquired through genetic inheritance and, therefore, considered to be invariable, i.e., exogenous to the model, whereas behavioral preferences are considered to be determined by a combination of components acquired through genetic inheritance and cultural learning. In our model, it is precisely the cultural learning component in the form of a social norm that provides the dynamic side of individual preferences.[3]

The specific norm of reciprocity put forth in this work is based on the idea that the concern an individual expresses for her opponent's success depends not only on her own intrinsic preference parameter but also on the intrinsic preferences of her opponent. This definition allows for asymmetric responses from players and the key question is what degree of reciprocity will be evolutionarily stable in an economic context characterised by negative externalities and strategic substitutes and what effect it will have on the individual's behavioral preferences. Particularly, we are interested in studying the evolutionary stability of what we call symmetric reciprocal behavior, in which case an individual's regard for her opponent's payoffs will depend on her own intrinsic preference parameter and on that of her opponent in exactly the same fashion, that is, equally weighted; and in pure symmetric reciprocal behavior, in which case individuals respond in kind to their opponents regardless of the former's intrinsic preferences.

In this article, we adopt an indirect evolutionary approach in the spirit of Bester and Güth's (1998) work on the evolution of altruistic preferences under pairwise random matching.[4] Nevertheless, instead of studying the evolutionary stability of preferences, we investigate the evolutionary stability of a reciprocity norm which, in turn, affects preferences and behavior, thus indirectly studying the evolution of preferences and behavior. In this sense, we assume that norms that induce more successful types of behavior in evolutionary terms will survive in the long run and norms yielding less successful behavior will go extinct.

Our theoretical proposal also deviates from previous works because our evolutionary model is asymmetrical in nature. The asymmetric character of our approach comes from the more realistic assumption that the intrinsic preferences of the players cannot be exactly the same. We argue that, if this were the case, there would be no room for reciprocal behavior. For instance, in a population dominated by altruistic players, that is, agents with altruistic intrinsic preferences, it would not be necessary to respond in kind. The players would just have to play according to their basic, intrinsic preferences. The same holds for populations dominated by egoistic or spiteful players. Hence, we argue that, in order to study the evolution of a norm of reciprocity and its impact on preferences, it is necessary to formulate a model that allows us to analyze what would

[3] See for instance: Bowles (1998), Boyd and Richerson (1985), Feldmand, Aoki and Kumm (1996), Heinrich and Boyd (1998), and Ross and Nisbett (1991).

[4] Other related work can be found in Güth and Yaari (1992), Güth (1995), Dufwenberg and Güth (1999) and Guttman (2000).

happen when the players are characterised by different intrinsic preferences. In this context, in order to approach the resulting asymmetric evolutionary game, we apply the work of Selten (1980) on the evolutionary stable strategy (ESS) concept under asymmetric contests.

2 A Model of Reciprocal Behavior

In the context of a pairwise random matching game, we consider two players, i and j, with subjective utility functions U_i and U_j, and monetary payoffs π_i and π_j. We assume that the subjective utility function for both players is linear on both monetary payoffs. Hence, we have:

$$\begin{aligned} U_i &= \pi_i + \beta_{ij}\pi_j \\ U_j &= \pi_j + \beta_{ji}\pi_i \\ \text{with } \beta_{ij} &= \lambda_{ij}\alpha_i + (1-\lambda_{ij})\,\alpha_j \\ \beta_{ji} &= \lambda_{ji}\alpha_j + (1-\lambda_{ji})\,\alpha_i \end{aligned} \tag{1}$$

Here we distinguish between player i's behavioral preference coefficient, denoted by β_{ij}, and her intrinsic preference parameter, denoted by α_i. These two preference parameters can coincide whenever player i's *norm reciprocity coefficient*, denoted by λ_{ij}, equals 1. Otherwise, the weight placed by individual i on the material payoffs of individual j is sensitive to the intrinsic preference parameter of the latter, namely α_j. From this formulation, it becomes clear that in our model the preferences that individuals show with their behavior, i.e. their behavioral preferences, are determined by a combination of an invariable component acquired through genetic inheritance, i.e. intrinsic preferences, and a dynamic cultural learning component given in this case by a norm of reciprocity, i.e. the norm reciprocity coefficient.

Following Levine (1998), we assume that player i's behavioral preference coefficient satisfies $\beta_{ij} \in (-1, 1)$, so each person places more weight on their own material payoff than on that of another. In order to ensure this, we restrict the reciprocity norm coefficient λ_{ij} to lie in the interval $[\frac{1}{2}, 1]$ and consider values of the intrinsic preference parameter such that $\alpha_i \in (-1, 1)$.

Let us briefly analyze the case where $\lambda_{ij} = 1$, which produces $\beta_{ij} = \alpha_i$. Here, we get the subjective utility function $U_i = \pi_i + \alpha_i\pi_j$, with $\alpha_i \in (-1, 1)$, which corresponds to the model of pure altruism of the type proposed by Levine (1998) and Ledyard (1995). In this case, if $\alpha_i > 0$ the player is said to be altruistic, since such a player has positive regard for her opponents' payoffs. In contrast, if $\alpha_i = 0$, the player is said to be selfish, that is she behaves egoistically, seeking to maximize her private success and showing no concern for the success of her partner. However if $\alpha_i < 0$, the player is said to be spiteful, showing a negative regard for her opponents. Clearly this specification includes that of Bester and Güth (1998), in which the parameter α_i is limited to the interval $[0, 1]$, being a model of altruistic and egoistic *behavior*. It can also be noted that this model allows for the analysis of spitefulness as parameter α_i can take a negative value, which means that player i can be malevolent as suggested by Bolle (2000) and Possajennikov (2000). We further extend the analysis by considering the specific cases of $\lambda_{ij} \neq 1$ and when the intrinsic preference parameters of the two players do not coincide, i.e. $\alpha_i \neq \alpha_j$. These two assumptions allow us to study what we define here as reciprocal *behavior*.

Definition 1 *Reciprocal behavior occurs whenever agent í's (j́'s) reciprocity norm coefficient,* $\lambda_{ij}, (\lambda_{ji})$ *lies in the interval* $[\frac{1}{2}, 1)$.

In this definition we present a basic concept of reciprocity. Essentially, player i shows some degree of reciprocity whenever the concern that she expresses for her opponent's success depends not only on her own intrinsic preference parameter but also on the intrinsic preferences of her opponent. It should be noted that, since we assume here that $\alpha_i \neq \alpha_j$ and that $\lambda_{ij}, \lambda_{ji} \in [\frac{1}{2}, 1)$ this definition of reciprocity allows for asymmetric responses from player i whenever $\lambda_{ij} \neq \lambda_{ji} \neq \frac{1}{2}$ which in turn implies that $\beta_{ij} \neq \beta_{ji}$. Hence, based on this basic notion of reciprocity we can put forth two additional more restrictive concepts of reciprocity related to the idea of symmetric reciprocal behavior.

- *symmetric reciprocal behavior*, between agents i and j, occurs whenever the agents´ reciprocity norm coefficients, λ_{ij} and λ_{ji}, lie in the interval $[\frac{1}{2}, 1)$ and are the same, i.e. $\lambda_{ij} = \lambda_{ji}$.
- *Pure symmetric reciprocal behavior*, between agents i and j, occurs whenever the agents´ reciprocity norm coefficients, λ_{ij} and λ_{ji}, lie in the interval $[\frac{1}{2}, 1)$ and are the same and equal to $\frac{1}{2}$, i.e. $\lambda_{ij} = \lambda_{ji} = \frac{1}{2}$.

We understand *symmetric reciprocal behavior* to be a case in which individual i´s regard for player j's payoff will depend on her own intrinsic preference parameter and that of her opponent in exactly the same fashion as player j considers them, that is giving them the same weight. This type of reciprocal *behavior* is symmetric, in the sense that the reciprocity norm parameters of players i and j, are equal, i.e. $\lambda_{ij} = \lambda_{ji}$, although asymmetries can still exist in the results, as discussed above. In fact, assuming that $\lambda_{ij} = \lambda_{ji}$, player i will not respond in kind to player j whenever $\lambda_{ij} = \lambda_{ji} \neq \frac{1}{2}$ which implies that $\beta_{ij} \neq \beta_{ji}$.

Finally, we understand *pure ymmetrical reciprocal behavior* to be the type of *behavior* in which an individual responds in kind to her opponents, regardless of her intrinsic preferences. That is, she ends up considering opponents in exactly the same way as she is regarded, i.e. $\beta_{ij} = \beta_{ji}$.

In the rest of the paper, using an indirect evolutionary framework, we will formally study the evolution of a reciprocity norm in the context of a game characterised by negative externalities and strategic substitutes. Specifically, following the formulation proposed in this section we will analise the *evolutionary stability of the reciprocity norm coefficient* λ_{ij}.

Next, we present the monetary payoffs that define success in our evolutionary game.

3 Evolutionary Success

As typically modelled in the above literature (see for instance: Bester and Güth (1998), Bolle (2000) and Possajennikov (2000)), here we assume that agents do pursue individual material payoffs and that these payoffs represent evolutionary success, i.e. fitness, in our evolutionary game theoretical framework. In this context, we consider a two player game in which x denotes individual 1's choice of action and individual 2 chooses some action y. Hence, the *monetary payoffs or evolutionary success* for players 1 and 2 are given by:

$$\pi_1 = x(1 - a(x + y)) \tag{2}$$

$$\pi_2 = y(1 - a(x + y)) \tag{3}$$

with $x \geq 0$, $y \geq 0$ and $a > 0$. Since from equations 2 and 3 we obtain that $\frac{\partial \pi_1}{\partial y} = -ax < 0$, $\frac{\partial \pi_2}{\partial x} = -ay < 0$ and $\frac{\partial^2 \pi_1}{\partial x \partial y} = \frac{\partial^2 \pi_2}{\partial x \partial y} = -a < 0$, it transpires that in this formulation the *strategic environment shows negative externalities and strategic substitutes*. Given the payoff's functional forms, this implies that if all players are payoff maximisers then the equilibrium vector of individual choices of action are unique, interior, symmetric, and inefficient.

This general monetary payoff specification can represent many *social dilemma* contexts in which the individual choice that maximises individual payoffs differs from the one that maximises the group payoffs. The simplest example is a *production game with negative externalities*. This could be the case of common property resource exploitation in which players 1 and 2 can exploit the resource at effort levels x and y with associated (quadratic) costs given by $a(x + y)x$, and $a(x + y)y$ respectively. Other examples can be given by *standard games of oligopolistic competition*. In particular a Cournot symmetric duopoly with zero production costs can be specified with this formulation by considering x and y as the firms' quantities choices, with a reservation price and parameter a equal to 1, from which the price becomes $1 - x - y$.

From equations 1, 2, and 3 the subjective utility functions for players 1 and 2 can be written as:

$$U_1 = x(1 - a(x + y)) + \beta_{12} y(1 - a(x + y)) \tag{4}$$

$$U_2 = y(1 - a(x + y)) + \beta_{21} x(1 - a(x + y)) \tag{5}$$

Now, we can define the following non-cooperative game:

$$\Gamma = (\{1, 2\}, \{x, y\}, \{U_1, U_2\})$$

Maximizing the subjective utility functions of players 1 and 2 given by equations 4 and 5 we obtain the Nash Equilibrium profile (x^*, y^*) for Γ:

$$x^* = \frac{1-\beta_{12}}{a\,(4-(1+\beta_{12})(1+\beta_{21}))} \tag{6}$$

$$y^* = \frac{1-\beta_{21}}{a\,(4-(1+\beta_{21})(1+\beta_{12}))} \tag{7}$$

From equations 6 and 7, we can explore the direction of the strategic effect of the behavioral preference coefficients, β_{12} and β_{21}, on the players equilibrium strategies, x^* and y^*. Since $\frac{\partial x^*}{\partial \beta_{21}} = \frac{(\beta_{12}+1)(1-\beta_{12})}{a(\beta_{12}+\beta_{21}+\beta_{12}\beta_{21}-3)^2} > 0$ (resp. $\frac{\partial y^*}{\partial \beta_{12}}$), the more altruistically inclined a player is, the higher her equilibrium action. In this case, the strategic effect has a negative impact on the altruistic player's success. Indeed, player j's altruism induces player i to choose a more altruistic action, and this reduces player j's success. This is so, since in this case a higher action from player i implies a lower return to player j given our assumption of decreasing returns.

Introducing equations 6 and 7 into 2 and 3, we obtain the following indirect material payoff value functions:

$$\pi_1^* = \frac{(1-\beta_{12})\,(1-\beta_{12}\beta_{21})}{a\,(4-(1+\beta_{12})(1+\beta_{21}))^2} \tag{8}$$

$$\pi_2^* = \frac{(1-\beta_{21})\,(1-\beta_{12}\beta_{21})}{a\,(4-(1+\beta_{12})(1+\beta_{21}))^2} \tag{9}$$

Where, from equation 1, we know that

$$\begin{aligned} \beta_{12} &= \lambda_{12}\alpha_1 + (1-\lambda_{12})\,\alpha_2, \text{ and} \\ \beta_{21} &= \lambda_{21}\alpha_2 + (1-\lambda_{21})\,\alpha_1 \end{aligned}$$

In order to study the evolutionary stability of the reciprocity norm parameter λ_{ij}, for a vector $\alpha = (\alpha_1, \alpha_2)$ we define the following evolutionary game: $\Gamma^*(\alpha) = (\{1,2\}, \{\lambda_{12}, \lambda_{21}\}, \{\pi_1^*, \pi_2^*\})$. It can be noted that $\Gamma^*(\alpha)$ is an asymmetric game, in which the asymmetry comes from the *intrinsic preference* parameter, since we have assumed that $\alpha_1 \neq \alpha_2$. In the next section, we study the evolutionary stability of the reciprocity norm parameter in an asymmetric context. We base our analysis on the work of Selten (1980) who defines and establishes conditions for the existence of evolutionary stable (*behavior*) strategies in asymmetric contexts in general.

4 Indirect Evolution of Reciprocal Behavior

Formally, $\Gamma^*(\alpha)$ defines a pairwise random matching game between two large populations of individuals. Each population is characterised by its own intrinsic preference parameter. Hence, the pair $\alpha = (\alpha_1, \alpha_2)$ fully describes two different populations, each one of them with players showing homogenous intrinsic preferences.

Following Selten (1980) we can define an ex-ante extensive form symmetric game $G(\alpha)$ associated with $\Gamma^*(\alpha)$. In this game, symmetric players are randomly selected to play the underlying game $\Gamma^*(\alpha)$, their strategies being conditioned to which type of player they have been assigned, either type α_1 or type α_2. In other words, they are assigned to play the "role" of one of the two possible types of agents in the original asymmetric game. Since the player type in this case is fully described by the intrinsic preference parameter, this allocation is equivalent to assigning the player to a given population α_1 or α_2. Like Selten (1980), we assume that each player is assigned to one of these two different populations with probability $\frac{1}{2}$. However, we also suppose here that in this ex ante extensive form game the two symmetric players are assigned to the same population with probability 0. That is, the two agents can not play the same "role" or face a symmetric ex post game $\Gamma^*(\alpha)$. Consequently, the game $G(\alpha)$ shows identical *a priori* probabilities that each agent be assigned to each type of player or role and each position in the game is occupied by one of the two roles defined here.

In this setting let us define a behavioral strategy. A behavioral strategy for a specific player is a pair $\lambda = (\lambda^{\alpha_1}, \lambda^{\alpha_2}) \in [\frac{1}{2}, 1]^2$, where λ^{α_1} is used in "position" α_1 and λ^{α_2} is used in "position" α_2. In $G(\alpha)$, nature plays first and selects agents to play the underlying game $\Gamma^*(\alpha)$; each agent plays one of the two types we have defined here. Figure 1 below describes this basic setting.

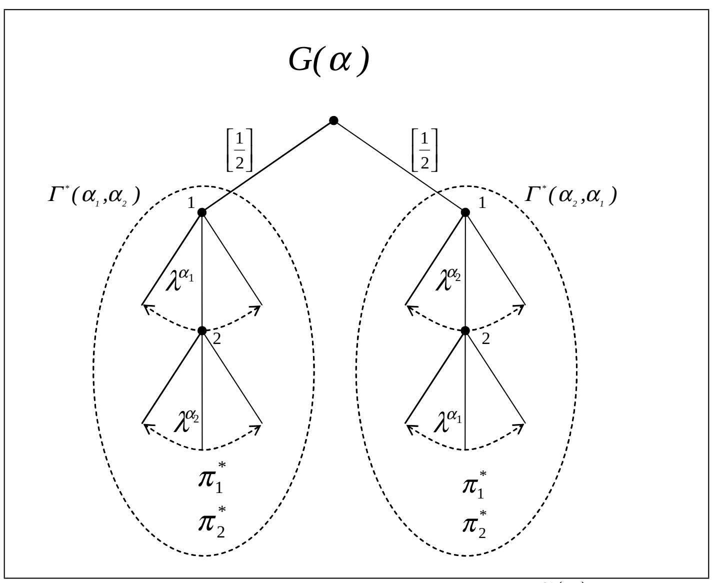

Figure 1: Ex ante symmetric game $G(\alpha)$.

From Selten (1980) (see also Weibull (1996) for further details) we know that a behavioral strategy λ in $G(\alpha)$ is evolutionarily stable if and only if it is a strict Nash equilibrium profile of $\Gamma^*(\alpha)$. Hence any ESS for $G(\alpha)$ will be entirely described by a strict equilibrium profile of $\Gamma^*(\alpha)$. More formally, an asymmetric evolutionary stable strategy for $G(\alpha)$ on the reciprocity norm parameters is a vector $(\lambda^{\alpha_1}, \lambda^{\alpha_2})$ satisfying the condition of being a strict Nash equilibrium profile of $\Gamma^*(\alpha)$.

5 Results

In this section we prove the equilibrium properties of the game described above. We begin presenting a lemma showing the existence of a strict Nash equilibrium in the underlying game $\Gamma^*(\alpha)$. Following Selten (1980) we will then be able to characterize the equilibium profile of the ex ante extensive form symmetric game $G(\alpha)$.

Let us first define the set: $A \equiv \{\alpha = (\alpha_1, \alpha_2) \subseteq (-1,1)^2 : \frac{\alpha_1+1}{3\alpha_1-1} \leq \alpha_2 \leq -\frac{2}{3} - \alpha_1\}$, and the function: $\phi(\alpha_1, \alpha_2) \equiv \frac{1}{2}\sqrt{\frac{(\alpha_1+\alpha_2-2)(3\alpha_1+3\alpha_2+2)}{3(\alpha_2-\alpha_1)^2}}$, then:

Lemma For every $\alpha \in A$, the game $\Gamma^*(\alpha)$ has a symmetric strict Nash equilibrium described by the profile:

$$\lambda = (\frac{1}{2} + \phi(\alpha_1, \alpha_2), \frac{1}{2} + \phi(\alpha_1, \alpha_2))$$

Proof of Lemma See Appendix.

Now, the following proposition characterises the equilibrium profile of $G(\alpha)$. This profile will in turn allow us to describe the agents'behavior in the underlying game $\Gamma^*(\alpha)$.

Proposition For every $\alpha \in A$, the game $G(\alpha)$ has a unique evolutionary stable behavioral strategy, described by $((\frac{1}{2} + \phi(\alpha_1, \alpha_2), \frac{1}{2} + \phi(\alpha_1, \alpha_2))$. This profile induces *symmetric reciprocal behavior* in $\Gamma^*(\alpha)$.

Proof of Proposition From the lemma we know that, for every $\alpha \in A$ the profile $\lambda = (\frac{1}{2} + \phi(\alpha_1, \alpha_2), \frac{1}{2} + \phi(\alpha_1, \alpha_2))$, is the unique symmetric strict Nash equilibrium of $\Gamma^*(\alpha)$ and from Selten (1980) (see Weibull (1996) proposition 2.18) we know then, that λ is the unique symmetric evolutionary stable behavioral strategy of $G(\alpha)$.■

From the proposition, the only evolutionary stable behavioral strategy in $G(\alpha)$ is one that induces *symmetric reciprocal behavior*. This implies that individual 1's regard for player 2's payoff will depend on her own intrinsic preference parameter and that of her opponent in exactly the same fashion as player 2 considers them, that is, giving them the same weight. Specifically, in this case the resulting strategy profile in $\Gamma^*(\alpha)$ is symmetric and produces the following behavioral preference parameters:

$$\beta_{12}^{*} = \frac{(\alpha_1+\alpha_2)}{2} + (\alpha_1-\alpha_2)\phi(\alpha_1,\alpha_2)$$
$$\beta_{21}^{*} = \frac{(\alpha_1+\alpha_2)}{2} - (\alpha_1-\alpha_2)\phi(\alpha_1,\alpha_2)$$

Nevertheless, these equations imply that, although this type of reciprocal *behavior* is symmetric in the sense that the reciprocity norm parameters of players 1 and 2 are equal there can still be asymmetries in the responses of each player. Particularly, player 1 will not respond in kind to player 2 whenever $\lambda^{\alpha_1} = \lambda^{\alpha_2} \neq \frac{1}{2}$ which implies that $\beta_{12}^{*} \neq \beta_{21}^{*}$. Hence, in order to ensure *pure symmetric reciprocal behavior* in $\Gamma^{*}(\alpha)$, the expression $\phi(\alpha_1,\alpha_2)$ must be equal to 0. It can be easily shown that $\phi(\alpha_1,\alpha_2)=0$ whenever $\alpha_2+\alpha_1+\frac{2}{3}=0$, and so, we can formally state that for all $\alpha=(\alpha_1,\alpha_2)$ such that $\alpha_2+\alpha_1+\frac{2}{3}=0$ there exists a unique evolutionary stable behavioral strategy, described by $((\frac{1}{2},\frac{1}{2})$ in $G(\alpha)$. This profile induces *pure symmetric reciprocal behavior* in $\Gamma^{*}(\alpha)$, implying that an individual will respond in kind to her opponents, regardless of their intrinsic preferences. In other words, she will end up considering her opponents in exactly the same way as she is regarded, i.e. $\beta_{12}^{*}=\beta_{21}^{*}=\frac{(\alpha_1+\alpha_2)}{2}$.

The results implied by the proposition can be more easily followed through a graphical analysis, see Figure 2 below. The dashed zone in Figure 2 shows the range of α for which there exists a unique evolutionary stable behavioral strategy in $G(\alpha)$. Clearly, the whole dashed area shows a set of pairs of intrinsic preferences that induce *symmetric reciprocal behavior* in $\Gamma^{*}(\alpha)$. However, only the line depicted by $\alpha_2+\alpha_1+\frac{2}{3}=0$, implies a family of vectors that induce *pure symmetric reciprocal behavior* in $\Gamma^{*}(\alpha)$.

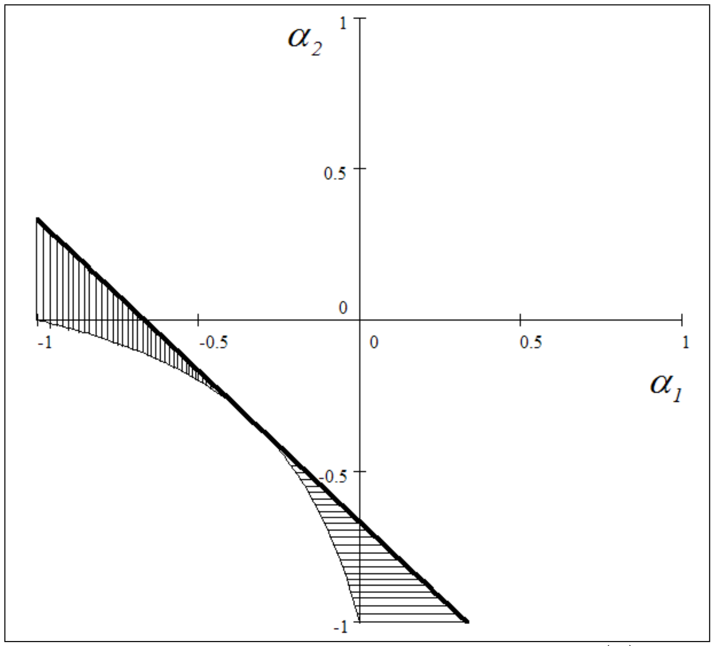

Figure 2: Evolutionary stable behaviour in $G(\alpha)$

The following corollary can also be inferred from the previous results.

Corollary For all $\alpha = (\alpha_1, \alpha_2)$ such that $\frac{\alpha_1+1}{3\alpha_1-1} \leq \alpha_2 \leq -\frac{2}{3} - \alpha_1$, $\beta^*_{12}, \beta^*_{21} < 0$■

Proof of Corollary See Appendix.

The Corollary implies that in the context of our basic game characterised by negative externalities and strategic substitutes the resulting behavioral preference coefficients will be negative, denoting a spiteful *behavior* from the players. This result relates our findings to those of Possajennikov (2000). In terms of our framework, Possajennikov's results imply that $\phi(\alpha_1, \alpha_2) = 0$ which in turn entail $\lambda = \frac{1}{2}$ whenever $\alpha_2 + \alpha_1 + \frac{2}{3} = 0$. Hence, it can be inferred that $\beta^*_{12} = \beta^*_{21} = -\frac{1}{3}$, which means that Possajennikov's results imply *pure symmetric reciprocal behavior*, being therefore a particular case of the results implied by the proposition. Here, it becomes clear that our model allows us to extend the previous analysis considering not only *pure symmetric reciprocal behavior* with $\beta^*_{12} = \beta^*_{21} < 0$, but also *symmetric reciprocal behavior* with $\beta^*_{12} \neq \beta^*_{21} < 0$. In terms of Figure 2, it can be argued that Possajennikov's results are limited to α defined by the line $\alpha_2 + \alpha_1 + \frac{2}{3} = 0$, which implies a set of α that only induce *pure symmetric reciprocal behavior* in $\Gamma^*(\alpha)$. Our results instead also include the whole dashed area which implies a family of profiles that induce *symmetric reciprocal behavior* in $\Gamma^*(\alpha)$. This type of reciprocal *behavior* allows for a wider range of evolutionarily stable results in social dilemma settings in which not only the behavioral preference coefficients of the players are different and negative, i.e. $\beta^*_{12} \neq \beta^*_{21} < 0$, but in which the intrinsic preference parameters of the players are also different, i.e. $\alpha_1 \neq \alpha_2$.This implies that an altruistic player when confronting a spiteful player will also end up behaving as a spiteful player (and viceversa), but with a different intensity, depending on the reciprocity norm coefficient λ. This will in turn also entail different choices of actions x and y. Table 1 below shows some evolutionarily stable results that respond to *symmetric reciprocal behavior*.

$\boldsymbol{\alpha}_1$	$\boldsymbol{\alpha}_2$	$\boldsymbol{\lambda}_{12}=\boldsymbol{\lambda}_{21}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$
0.02	−1.00	0.97	−0.007	−0.973
0.10	−0.96	0.85	−0.060	−0.797
−0.02	−0.92	0.99	−0.022	−0.915
0.10	−0.80	0.65	−0.211	−0.483
−0.06	−0.76	0.97	−0.084	−0.732
0.06	−0.76	0.67	−0.211	−0.483
−0.10	−0.67	0.98	−0.113	−0.662
−0.02	−0.67	0.70	−0.213	−0.480
−0.10	−0.63	0.91	−0.151	−0.584
−0.14	−0.59	0.98	−0.151	−0.584
−0.10	−0.59	0.78	−0.211	−0.483
−0.14	−0.55	0.83	−0.211	−0.4823
−0.18	−0.51	0.92	−0.211	−0.482 3
−0.84	−0.02	0.95	−0.797	−0.060

Table 1: Symmetric reciprocal behavior (Results for a=1)

The examples presented in Table 1 show that, when a reciprocity norm is stable, an altruistic (spiteful) individual confronting a spiteful (altruistic) agent can show spiteful (altruistic) preferences with her behavior, despite being intrinsically altruistic (spiteful). As also shown in Table 1, the stable reciprocity norm in this context does not necessarily imply that the individual's choices of actions are the same. In fact, if one player is more spiteful than the other, she receives what could be interpreted as a punishment, being the recipient of behavior that is more spitefully inclined than her opponent's intrinsic preferences indicated. This kind of behavior also works the other way around, since a more altruistic player receives a reward in the form of more altruistic behavior than suggested by her opponent's intrinsic preferences.

6 Concluding Remarks

Why does an altruistically inclined player behave altruistically in some contexts and spitefully in others? This article provides an economic explanation to this question. The basic argument is centred on the idea that social norms shape our preferences through a process of cultural learning. If a player is altruistically inclined in a context in which a social norm of reciprocity is in place, she can behave spitefully towards a spiteful player and altruistically towards an altruistic player. The advantage of this approach in comparison to previous theoretical research on the analysis of the evolutionary stability of preferences is that it allows us to explain contextual behavior without assuming that the intrinsic preferences of the player change constantly; what changes (evolves, in this case) are the norms of behavior that indirectly determine preferences and behavior.

In particular, we have used an indirect evolutionary framework in this paper to formally examine the evolutionary stability of a reciprocity norm and its impact on individual preferences and behavior. The main conclusion is that, in a specific economic context characterised by negative externalities and strategic substitutes, evolutionary stable behavioral strategies are consistent with our definition of symmetric reciprocal behavior for certain values of the players' intrinsic preference parameters. This implies that the individual's regard for her opponent's payoff will depend on her own intrinsic preference parameter and that of her opponent in exactly the same fashion, that is, giving both the same weight. Although this type of reciprocal behavior is symmetric in the sense that the reciprocity norm parameters of both players are equal, there can still be asymmetries in the responses of each player. Specifically, if the reciprocity norm parameters are equal but different from $\frac{1}{2}$, the behavioral preference coefficient of both players will differ, which implies asymmetrical responses in terms of different individual choices of action (see Table 1). The scenario in which individuals respond in kind to their opponents, regardless of their intrinsic preferences, what we called here pure symmetric reciprocal behavior, is, therefore, only a special case that depends on the players' intrinsic preference parameters satisfying a specific and more restricitve condition (see Figure 2).

Another conclusion that can be inferred from this study is that, in the context of our basic game characterised by negative externalities and strategic substitutes, the resulting behavioral preference coefficients will be negative, denoting spiteful behavior from the players. This relates our findings to those of Possajennikov (2000), which can be considered to be a particular case of the results implied in this work. Our model, nevertheless, extends previous analyses, considering not only pure symmetric reciprocal behavior, but also symmetric reciprocal behavior. It can be argued that this type of reciprocal behavior allows for a wider range of evolutionarily stable results in social dilemma settings in which not only the behavioral preference coefficients of the players are different and negative, but in which the intrinsic preference parameters of the players are also different. This possibility is precisely what allows us to provide an explanation of why, in some contexts, an altruistic player confronting a spiteful player will also end up behaving as a spiteful player (and vice versa) but with a different intensity, depending on the player's reciprocity norm coefficient, which, in turn, will also entail different choices of actions.

Finally, future extensions of this research include replicating the analytical framework presented in this work under an economic context characterised by positive externalities and strategic complements and experimental testing of the hypotheses inferred from this theoretical model, providing empirical evidence about the evolution of social norms and individual preferences.

7 Appendix

Proof of Lemma Given a game $\Gamma^*(\alpha)$, from equations (8) and (9) we have that:

$$\frac{\partial \pi_i^*}{\partial \lambda_{ij}} = \frac{\partial \pi_i^*}{\partial \beta_{ij}}\frac{\partial \beta_{ij}}{\partial \lambda_{ij}} + \frac{\partial \pi_i^*}{\partial \beta_{ji}}\frac{\partial \beta_{ji}}{\partial \lambda_{ij}} = 0 \tag{10}$$

We know that $\frac{\partial \beta_{ji}}{\partial \lambda_{ij}} = 0$ and $\frac{\partial \beta_{ij}}{\partial \lambda_{ij}} = \alpha_i - \alpha_j$, since $\alpha_i \neq \alpha_j$ then $\frac{\partial \pi_i^*}{\partial \lambda_{ij}} = \frac{\partial \pi_i^*}{\partial \beta_{ij}(\lambda_{ij})} = 0$

Rearranging trems in the first order conditions we have:

$$\begin{aligned}
\frac{\partial \pi_1^*}{\partial \lambda_{12}} &= \frac{\partial \pi_1^*}{\partial \beta_{12}} = \frac{(3\beta_{12}\beta_{21} - \beta_{21} - \beta_{12} - 1)(\beta_{21} - 1)}{a(\beta_{12} + \beta_{21} + \beta_{12}\beta_{21} - 3)^3} = 0 \\
\frac{\partial \pi_2^*}{\partial \lambda_{21}} &= \frac{\partial \pi_2^*}{\partial \beta_{21}} = \frac{(3\beta_{12}\beta_{21} - \beta_{21} - \beta_{12} - 1)(\beta_{12} - 1)}{a(\beta_{12} + \beta_{21} + \beta_{12}\beta_{21} - 3)^3} = 0
\end{aligned}$$

where $\beta_{12} = \beta_{12}(\lambda_{12})$ and $\beta_{21} = \beta_{21}(\lambda_{21})$.

By assumption $a > 0$ and since $\beta_{ij} \in (-1, 1)$ we know that

$$a(\beta_{12} + \beta_{21} + \beta_{12}\beta_{21} - 3) \neq 0$$

and

$$\begin{aligned}
(\beta_{21} - 1) &\neq 0 \\
(\beta_{12} - 1) &\neq 0
\end{aligned}$$

therefore, both first order conditions are reduced to the following condition:

$$(3\beta_{12}\beta_{21} - \beta_{21} - \beta_{12} - 1) = 0 \tag{11}$$

with $\beta_{12} = \lambda_{12}\alpha_1 + (1 - \lambda_{12})\alpha_2$ and $\beta_{21} = \lambda_{21}\alpha_2 + (1 - \lambda_{21})\alpha_1$.

Looking for a symmetric equilibrium, we set $\lambda_{12} = \lambda_{21} = \lambda$. Therefore the condition (11) becomes:

$$3\alpha_1\alpha_2 - \alpha_2 - \alpha_1 - 1 + \lambda 3(\alpha_2 - \alpha_1)^2 - \lambda^2 (3)(\alpha_2 - \alpha_1)^2 = 0$$

whose solutions are $\lambda^{\pm} = \frac{1}{2}\left(1 \pm \sqrt{\frac{(\alpha_1+\alpha_2-2)(3\alpha_1+3\alpha_2+2)}{3(\alpha_2-\alpha_1)^2}}\right)$.

We know that $(\alpha_1 + \alpha_2 - 2) < 0$ then it should be the case that $3\alpha_1 + 3\alpha_2 + 2 \leq 0$, then we have two cases to analyze:

Case 1. If $3\alpha_1 + 3\alpha_2 + 2 = 0$ implies that $\lambda^{\pm} = \frac{1}{2}$ and $\beta_{12} = \beta_{21}$, i.e. both solutions turn out to be the same and equal to $\frac{1}{2}$. Hence if $\alpha = (\alpha_1, \alpha_2) \subseteq (-1, 1)^2$ such that $\alpha_1 + \alpha_2 = -\frac{2}{3}$, $\Gamma^*(\alpha)$ has a unique symmetric strict Nash equilibrium described by the profile $(\lambda^{\alpha_1} = \frac{1}{2}, \lambda^{\alpha_2} = \frac{1}{2})$.

Case 2. If $3\alpha_1 + 3\alpha_2 + 2 < 0$ implies that $\alpha_2 < -(\alpha_1 + \frac{2}{3}))$. Therefore $\sqrt{\frac{(\alpha_1+\alpha_2-2)(3\alpha_1+3\alpha_2+2)}{3(\alpha_2-\alpha_1)^2}} > 0$ which implies $\lambda^+ > \frac{1}{2}$ and $\lambda^- < \frac{1}{2}$. By assumption $\lambda \in [\frac{1}{2}, 1]$, then the only feasible result to analize is $\lambda^+ = \frac{1}{2}\left(1 + \sqrt{\frac{(\alpha_1+\alpha_2-2)(3\alpha_1+3\alpha_2+2)}{3(\alpha_2-\alpha_1)^2}}\right)$.

Finally we have to impose the restriction $\frac{1}{2}\left(1 + \sqrt{\frac{(\alpha_1+\alpha_2-2)(3\alpha_1+3\alpha_2+2)}{3(\alpha_2-\alpha_1)^2}}\right) \leq 1$, which implies that $\alpha_2(3\alpha_1-1) \leq \alpha_1+1)$. Recalling that $3\alpha_1+3\alpha_2+2 < 0$, then we have three cases:

a. if $\alpha_1 = \frac{1}{3}$ imkplies that $\alpha_2 < -(\alpha_1 + \frac{2}{3}) = -1$ then $(3\alpha_1 - 1) \neq 0$.

b. if $(3\alpha_1 - 1) > 0$ implies that $\alpha_1 > \frac{1}{3}$ therefore $0 > 3\alpha_1 + 3\alpha_2 + 2 > 1 + 3\alpha_2 + 2 \rightarrow -1 > \alpha_2$ which implies $(3\alpha_1 - 1) \not> 0$.

c. if $(3\alpha_1 - 1) < 0$ then $\alpha_2 \geq \frac{\alpha_1+1}{(3\alpha_1-1)}$ which is the only feasible case.

Hence for all $\alpha = (\alpha_1, \alpha_2) \subseteq (-1, 1)^2$ such that $\frac{\alpha_1+1}{3\alpha_1-1} \leq \alpha_2 \leq -\frac{2}{3} - \alpha_1$, $\Gamma^*(\alpha)$ has a unique symmetric strict Nash equilibrium described by the profile

$\lambda = (\frac{1}{2}+\phi(\alpha_1, \alpha_2), \frac{1}{2}+\phi(\alpha_1, \alpha_2))$ with $\phi(\alpha_1, \alpha_2) = \frac{1}{2}\sqrt{\frac{(\alpha_1+\alpha_2-2)(3\alpha_1+3\alpha_2+2)}{3(\alpha_2-\alpha_1)^2}}$
■

proof of the corollary

Case 1: $\beta^*_{12} = \frac{(\alpha_1+\alpha_2)}{2} + (\alpha_1-\alpha_2)\phi(\alpha_1, \alpha_2)$. By contradiction suppose $\beta^*_{12} \geq 0$ then $\phi(\alpha_1, \alpha_2) \geq -\frac{(\alpha_1+\alpha_2)}{2(\alpha_1-\alpha_2)}$.

Suppose first that $(\alpha_1 - \alpha_2) < 0$, since $\alpha_1 \neq \alpha_2$ and $-\frac{(\alpha_1+\alpha_2)}{2} > 0$ then $(\alpha_1-\alpha_2)\phi(\alpha_1, \alpha_2) > 0$. Given that $\phi(\alpha_1, \alpha_2) \in [0, \frac{1}{2}]$ then $(\alpha_1-\alpha_2)\phi(\alpha_1, \alpha_2) < 0$ which is a contradiction. Suppose now that. $(\alpha_1 - \alpha_2) > 0$, since $\phi(\alpha_1, \alpha_2) = \frac{1}{2}\sqrt{\frac{(\alpha_1+\alpha_2-2)(3\alpha_1+3\alpha_2+2)}{3(\alpha_2-\alpha_1)^2}}$ and $\alpha_2 \neq \alpha_1$we ge~~t~~ $\alpha_1+\alpha_2+1 \leq 0$. But $\alpha_2+\alpha_1 \leq -\frac{2}{3}$ then $\alpha_1 + \alpha_2 + 1 > 0$ which is a contradiction.

Case 2: $\beta^*_{21} = \frac{(\alpha_1+\alpha_2)}{2} - (\alpha_1-\alpha_2)\phi(\alpha_1, \alpha_2)$. By contradiction suppose $\beta^*_{21} \geq 0$ then $\phi(\alpha_1, \alpha_2) \geq \frac{(\alpha_1+\alpha_2)}{2(\alpha_1-\alpha_2)}$.

Suppose first that $(\alpha_1 - \alpha_2) > 0$ since $\alpha_1 \neq \alpha_2$ and $\frac{(\alpha_1+\alpha_2)}{2} < 0$ then $(\alpha_1-\alpha_2)\phi(\alpha_1, \alpha_2) < 0$. Given that $\phi(\alpha_1, \alpha_2) \in [0, \frac{1}{2}]$ then $(\alpha_1-\alpha_2)\phi(\alpha_1, \alpha_2) > 0$ which is a contradiction. Suppose now that. $(\alpha_1 - \alpha_2) < 0$, since $\phi(\alpha_1, \alpha_2) = \frac{1}{2}\sqrt{\frac{(\alpha_1+\alpha_2-2)(3\alpha_1+3\alpha_2+2)}{3(\alpha_2-\alpha_1)^2}}$and $\alpha_2 \neq \alpha_1$we get $\alpha_1 + \alpha_2 + 1 \leq 0$. But $\alpha_2 + \alpha_1 \leq -\frac{2}{3}$ then $\alpha_1 + \alpha_2 + 1 > 0$ which is a contradiction.■

8 References

Bergstrom (1995) On the Evolution of Altruistic Ethical Rules for Siblings, *American Economic Review*, 85 (1): 58-81.

Bester, H. and Güth, W. (1998) Is Altruism Evolutionarily Stable? *Journal of Economic Behavior and Organization*, 34 (2): 193: 209.

Bolle, F. (2000) Is Altruism Evolutionarily Stable? And Envy and Malevolence? -Remarks on Bester and Güth, *Journal of Economic Behavior and Organization* 42(1): 131-133.

Bowles, S. (1998) Endogenous preferences: The cultural consequences of markets. and other economic institutions, *Journal of Economic Literature* 36: 75–111.

Bowles, S. and Gintis, H. (1999) *The Evolution of Strong Reciprocity*, mimeo, University of Massachusetts at Amherst.

Bowles, S. (2004) *Microeconomics: Behavior, Institutions and Evolution*, New Jersey: Princeton University Press

Boyd, R. and Richerson, PJ (1985) *Culture and the evolutionary process.* Chicago: University of Chicago Press.

Dufwenberg M. and Güth, W. (1999). Indirect Evolution vs. Strategic Delegation: a Comparison of Two Approaches to Explaining Economic Institutions. *European Journal of Political Econom*y 15: 281-295.

Elster, J. (1989) Social norms and economic theory, *Journal of Economic Perspectives* 3: 99-118.

Feldmand, M.W., Aoki, K. and Kumm, J. (1996) *Individual Versus Social Learning: Evolutionary Analysis in a Fluctuacting Environment*, Santa Fe Institute Working Paper 96-05-031.

Gouldner, A. (1960). The norm of reciprocity: A preliminary statement, *American Sociological Review*, 25: 176-177

Güth, W. (1995) An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives, *International Journal of Game Theo*ry, 24: 323-344.

Güth W. and M. Yaari (1992) *Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach.* In Witt, U. (ed.), Explaining Forces and Changes: Approaches to Evolutionary Economics, University of Michigan Press.

Guttman, J. (2000) On the Evolutionary Stability of Preferences for Reciprocity. *European Journal of Political Economy* 16: 31-50.

Heinrich, J. and Boyd, R. (1998) The Evolution of Conformist Transmission and Between-Group Differences. *Evolution and Human Behavior*, 19: 215-242.

Huck, S. and Oechssler, J. (1998) The Indirect Evolutionary Approach to Explaining Fair Allocations, *Games and Economic Behavior* 28: 13-24.

Kandori, M. (1997)Evolutionary Game Theory in Economics, in Kreps, D. M. and Wallis, K. F. (eds.) *Advances in Economics and Econometrics: Theory and Applications*, Seventh World Congress of the Econometric Society, Vol. 1. Cambridge: Cambridge University Press.

Koçkesen, L. Ok, E. A. and Sethi, R. (2000a) The Strategic Advantage of Negatively Interdependent Preferences, *Journal of Economic Theory* 92 (2): 274-299.

Koçkesen, L. Ok, E. A. and Sethi, R. (2000b) Evolution of Interdependent Preferences in Aggregative Games, *Games and Economic Behavior*, 31: 303-310.

Ledyard, J. O. (1995) *Public Goods: A Survey of Experimental Research*, in Handbook of Experimental Economics, ed. John Kagel and Alvin E. Roth, Princeton University Press.

Levine, D. (1998) Modeling Altruism and Spitefulness in Experiments, *Review of Economic Dynamics*, 1: 593-622.

Mailath, G. J. (1992) Introduction: Symposium on Evolutionary Game Theory, *Journal of Economic Theory*, 57: 259-277.

Matsui, A. (1996) On Cultural Evolution: Social Norms, Rational Behavior, and Evolutionary Game Theory, *Journal of the Japanese and International Economies*, 10: 262-294.

Maynard Smith, J. (1974) The Theory of Games and the Evolution of Animal Conflicts, *Journal of Theoretical Biology*, 47: 209-221.

Maynard Smith, J. (1982) *Evolution and The Theory of Games*, Cambridge: Cambridge University Press.

Maynard Smith, J. and Price, G. (1973) The Logic of Animal Conflicts. *Nature*, 246: 15-18.

Ok E.A. and Vega-Redondo, F. (2000) *On the Evolution of Individualistic Preferences: Complete versus Incomplete Information Scenarios*, mimeo, New York University.

Possajennikov, A. (2000) On the Evolutionary Stability of Altruistic and Spiteful Preferences, *Journal of Economic Behavior and Organization*, 42, 125-12.

Ross, L. and Nisbett, R. E. (1991) *The Person and the Situation: Perspectives of Social Psychology*, New York: McGraw Hill.

Samuelson, L. (1997) *Evolutionary Games and Equilibrium Selection*, Cambridge MA: MIT Press.

Selten, R. (1980) A Note on Evolutionarily Stable Strategies in Asymmetric Animal Conflicts, *Journal of Theoretical Biology* 84: 93-101.

Sethi, R. and Somanathan, E. (2000) Preference Evolution and Reciprocity, *Journal of Economic Theory* 97(2): 273-297.

Sobel, J. (2005) Interdependent Preferences and Reciprocity, *Journal of Economic Literature*, Vol. XLIII: 392.436.

Van Damme, E. (1987) *Stability and Perfection of Nash Equilibria*, Berlin: Springer Verlag.

Van Damme, E. (1994) Evolutionary Game Theory, *European Economic Review*, 38: 847-858.

Vega-Redondo, F. (1996) *Evolution, Games and Economic Behavior*, Oxford: Oxford University Press.

Villena, MG and Villena, MJ (2004) Evolutionary Game Theory and Thorstein Veblen's Evolutionary Economics: Is EGT Veblenian? *Journal of Economic Issues*, 38(3): 585-610.

Weibull, J. (1996) *Evolutionary Game Theory*, London, England: The MIT Press.

Weibull, J. (1998) *What Have We Learned From Evolutionary Game Theory So Far?*, Mimeo, Stockholm School of Economics.