Altruism, Efficiency and Endogenous Preference Formation in an Economy with Public Goods*

Francisco Marhuenda† and Felipe Pérez-Martí‡

July 31 2001
This version: June 2004

Abstract

We reconsider the problem of altruism and efficiency in a two agent economy with public goods. We show that welfare comparisons can be made, from the point of view of both selfish and altruistic agents, regarding selfish vs. altruistic allocations. We obtain uniqueness of altruistic equilibria for a given altruistic (or selfish) profile, and completely characterize the set of altruistic equilibria for all possible such profiles. For a given profile which produces an inefficient allocation due to market failure, there exists a profile with higher altruistic parameters such that the new implied altruistic equilibrium is Pareto improving with respect to the previous one for both, the original, as well as for the new economy. These results are used to propose a mechanism of preference formation by which altruism can arise. The mechanism consists in a seduction game in which the members of the group *fall in love* with each other in a subgame perfect equilibrium.

1 Introduction

> I consider that the golden rule requires that if I like a program[1] I must share it with other people who like it.
>
> GNU Manifesto, Richard Stallman

*We appreciate comments by John Conley and David Wildasin. The second author appreciates the help of FONACIT, Venezuela, project S1-99000107.

†Universidad Carlos III de Madrid, Ap. 99. E-03080. Spain. Ph. and Fax: 34-96/5903614 Email: Marhuend@eco.uc3m.es

‡Instituto de Estudios Superiores de Administración (IESA). Av. IESA, San Bernardino, Caracas, Venezuela. Email: felipe.perez@iesa.edu.ve

[1] A software program, which is a public good.

The idea that altruism can enhance efficiency has been present in economic literature since the seminal papers of J. R. Barro (1974 [3]) and G. Becker(1974 [4]). We shall not review here the extensive literature relating more or less directly to this question.

Nonetheless, there are some pending issues in the economic analysis of public goods, such as a proper analysis of welfare comparisons when altruism increases (and, therefore, preferences change) or proposing realistic models which explain how altruism may arise.

As far as we are aware of, Andreoni (1989, [1]) introduced, for the first time in the literature of public goods, altruistic preferences, building on the *impure* public goods formulation of Cornes-Sandler (1984, [10]). Andreoni's paper deals with issues of efficient government policies when people are altruists, but does not address the issue of efficiency improvements when altruism increases.

As it is well known, incentive problems with selfish individuals makes the provision of public goods often impossible or inefficient. The introduction of altruism in the model, even though considered an obvious way of "resolving" those incentives problems, poses some theoretical issues, not addressed before, regarding preference comparisons. In fact, efficiency is defined for a given set of preferences, so that the change of those preferences, which occurs when people become altruistic, might render the issue of efficiency improvements ambiguous. Here we characterize the range of altruistic parameters for which welfare improvements are true for both, the individuals previously considered relatively selfish, as well as for the relatively more altruistic people they have become.

The other main question we ask, in the spirit of recent literature on evolution of preferences (see Henrich [7], Bergstrom and Stark [**?**] and [6]) is if altruism itself might be interpreted as a psychological mechanism, used somehow by selfish agents, in order to improve their welfare, given the inefficiencies derived from using alternative procedures. Specifically, the question could be posed as: Given an economy with public goods, under what conditions is it better for selfish individuals to become altruists and how do they do that? In a two step game, we show that there is a sense in which one could reasonably make such statement, and setup a model of preference changes, through a process of mutual courting that teaches each member of the relationship to go from selfishness towards reciprocal altruism. We thus provide a step towards a theoretical foundation for the phenomenon of altruism in an economy with public goods.

The issue of preference formation has basicaly two approaches: selection mechanisms through evolution, and teaching-learning mechanisms. Even though there are interesting evolution models that explain altruistic preference formation (for example H. Bester and W. Güth ([8])), Henrich [7] convincingly argues that altruism in humans are best explained as a result of "cultural evolution", or transmission through education mechanisms. Humans have developed brain capacity to learn some traits, among them altruism, that makes some groups of them take advantage of and compete favorably because of efficiency gains from altruistic behavior. Even though Henrich [7] talks about education, and learning capacity, he does not provide a specific model of how this could happen.

Rotemberg [17] presents a model that is an example of the teaching approach, even though elementary, of altruism, self teaching in this case. He presents a two stage model in which, at the first stage, selfish individuals choose their degree of altruism. At the second stage, agents maximize their preferences as chosen in the previous stage. It is shown that strategic complementarity in the second stage leads to the choice of some degree of altruism at the first stage. That is, agents benefit by changing their tastes towards altruism. In Rotemberg's model, one may view altruism as a form of commitment which, in the presence of strategic complementarities, can be rationalized.

But is difficult to justify that you affect your own preferences. If you endogenize the process there would be an indeterminacy problem of having a preference that acts as the objective function of the person, which somehow leads the person to change those preferences. On the other hand, allowing agents to manipulate the tastes of other people might seem more natural, and that is the approach we take, as in Stark and Falk [19]. The assumption that people influence the preferences of others in their own self-interest is justified by many real life situations. One can think of individuals who undertake actions (such as doing favors, inviting others to dine, bringing flowers, etc.) that modify the friendliness of those targeted. Parents, governments, religious institutions, couples, working partners and friends are examples of agents who take such kind of actions.

Stark and Falk [19] set up a model of uncertainty in which a well endowed individual in the first period donates to other, so that this one, by emphaty, will give a gift back to the donor if he is in trouble next period. Our model, on the other hand, uses the learning capacity of both individuals in the game in a context of public goods, and show that, in equilibrium, reciprocal altruism is formed. In the model, people may decide whether or not to spend some effort trying to alter the preferences of the people they interact with, and we show that there exists an equilibrium in which each individual spends a certain (strictly) positive amount of his or her resources in order to manipulate another person's feelings. And the welfare of everybody involved improves, both from their selfish as well as their altruistic point of view. The result hinges upon the fact of those welfare improvements, much in the way the results of Rottemberg and Bester and Güth depend on the existence of complementarities, and are easily extended to the specific gains of welfare due to different kind of market failure, like in economies with assymetric information, or incomplete markets, as described in Marhuenda and Pérez-Martí ([11]).

To carry out our modeling, we translate into a public goods economy standard altruistic preferences used in other fields. The felicity function of a given person is the sum of their own pleasure function (which is the standard utility function for selfish people) plus the pleasure function of the person he or she cares for, weighted by an altruistic parameter. This model is used in macroeconomic analyses dealing with intergenerational transfers and policy issues,[2] as well as in microeconomic models used to explain the behavior of relatives within a family.[3]

It is to point out that our model of altruistic preferences features a natural similarity with the one used in welfare analysis for characterizing, by means of the central planner approach, the set of Pareto optimal allocations in an economy with selfish agents. In this regard, our model could be thought of as an analysis of the relationships between *atomic local planners*. They are planners in the sense that they take into account the resources and preferences of the group in order to define the individually ideal allocations for the

[2]See, for example the seminal paper of Barro (1974, [3]).

[3]See, for example Becker (1974, [4]), Stark (1995, [18]).

whole society. They are atomic since, unlike the central planner, they do not have the power to carry out the required transformation of wishes into reality. And they are local because they care for their neighbors, but not necessarily for the whole society. Although the allocations preferred by individual planners are always efficient, the wills of different atomic planners in general do not agree with each other. Neither do they coincide with those of a central planner. This paper studies the set of their possible agreements in the way of showing the set of Pareto optima for various degrees of altruism.

The plan of the paper is as follows. We first show some results that allow us to make welfare comparisons between people displaying different degrees of altruism. We then show existence and uniqueness of Nash equilibrium for a given economy populated by individuals with a given degree of altruism. We completely characterize the set of altruistic equilibria, and show that the set goes in the right direction in terms of welfare. A corollary is that, indeed, an increase of altruism improves efficiency in the Pareto sense.[4] Then, we present the model of partner interactions to affect their preferences. In the last section we discuss the limitations of the model and need for further research, and present concluding remarks.

2 The Model

We consider an economy with two agents, one private and one public good. There is no initial endowment of the public good but each agent $i = 1, 2$ has ω^i units of initial endowment of the private good. We let $\omega = \omega^1 + \omega^2$ be the total endowment of private goods.

Assumption 1 *The preferences of the agents are represented by utility functions*

$$\begin{aligned} u^1(x_1, x_2, y) &= v^1(x_1, y) + \gamma_1 v^2(x_2, y) \\ u^2(x_1, x_2, y) &= v^2(x_2, y) + \gamma_2 v^1(x_1, y), \end{aligned} \tag{1}$$

where $\gamma_1, \gamma_2 \geq 0$, *and* $v^1, v^2 : \mathbb{R}^2_+ \to \mathbb{R}$, *are strictly concave functions, strictly increasing in both arguments and such that for* $i \neq j$,

$$v^k_{ij} = \frac{\partial^2 v^k}{\partial x_i \partial x_j}(x_1, x_2) \geq 0 \tag{2}$$

and

$$\lim_{x_k \to 0} \frac{\partial v^k}{\partial x_k}(x_1, x_2) = +\infty. \tag{3}$$

Here, x_i represents the consumption of the private good by individual $i = 1, 2$, and y is the output (and consumption) of the public good. Selfish preferences correspond to

[4]This last result complements the ones in paper ([11]), in which the case of incomplete markets is treated. In that paper it is shown that transfers among inter-relatives and inter-friends due to altruism are, in fact, a mechanism to “complete the markets” and improve welfare.

$\gamma_1 = \gamma_2 = 0$ and a higher degree of altruism is associated with larger values of γ_1 or γ_2. Let the function $f : \mathbb{R} \to \mathbb{R}$, represent the technology, available to all members of the society, that transforms private goods into public goods.

Assumption 2 *The mapping $f : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing, strictly concave, C^1 and satisfies $f(0) = 0$.*

Thus, an economy is specified by a vector $(v^1, v^2; f; \gamma_1, \gamma_2; w^1, w^2)$. Since we always assume that v^1, v^2, f, w^1 and w^2 are constant, it is enough to specify an economy by the vector (γ_1, γ_2). The resource constraints for the economy are, then: $x_1 + x_2 + z = w^1 + w^2 = \omega$, and $y = f(z)$, where z is the amount of input of private goods used for the production of the public good.

An allocation (x_1, x_2) is feasible if $x_i \geq 0$, $i = 1, 2$, and $x_1 + x_2 \leq \omega$. So we define the set

$$\mathrm{T} = \{(x_1, x_2) \in \mathbb{R}^2_+ : x_1 + x_2 \leq \omega\}$$

as the set of feasible allocations. It is understood here that x_i is the consumption by agent $i = 1, 2$ of the private good and that an amount $f(\omega - x_1 - x_2)$ of the public good is provided along with the allocation (x_1, x_2). We now make our *atomic planner* assumption. Agents have not only knowledge of each other's pleasure function, but they are also aware of the private resources available to each other and of their consumption. Accordingly, since they also know the technology f, they can rationally deduce how much of the public good is produced. Hence their preferences may be summarized by means of the following utility functions

$$W^1(x_1, x_2; \gamma_1) = v^1(x_1, f(\omega - x_1 - x_2)) + \gamma_1 v^2(x_2, f(\omega - x_1 - x_2)) \quad (4)$$

$$W^2(x_1, x_2; \gamma_2) = v^2(x_2, f(\omega - x_1 - x_2)) + \gamma_2 v^1(x_1, f(\omega - x_1 - x_2)) \quad (5)$$

where now we only specify explicitly the consumption of the private good.

Definition 3 *A feasible allocation $(x_1, x_2) \in \mathrm{T}$ is Pareto Optimal if there is no other allocation $(r_1, r_2) \in \mathrm{T}$ such that for $i = 1, 2$ we have $W^i(r_1, r_2; \gamma_i) \geq W^i(x_1, x_2; \gamma_i)$ with some inequality strict.*

We let $\mathrm{P}(\gamma_1, \gamma_2)$ denote the set of Pareto optimal allocations in the economy (γ_1, γ_2). In Figure 1 we represent feasible allocations by points (x^1, x^2) such that $\omega - x^1 - x^2 \geq 0$. The amount y of public good produced can be inferred as $y = f(\omega - x^1 - x^2) \geq 0$. Any private consumption bundle (x^1, x^2) within the triangle is feasible, and z is the horizontal (or vertical) distance from the point to the diagonal, which represents a feasibility constraint. It follows that the graphical representation is useful, since in the canonical case, when f is the identity function, the allocation (x^1, x^2, y) can be visualized directly from the graph. Figure 1 also shows preferences and Pareto optimal allocations. It is straightforward to show that preferences represented by equations (4) and (5) are single-peaked. For the case $\gamma_1 = 0$, that peak is represented by the point X', where $x_2 = 0$. Similarly, for $\gamma_2 = 0$, the peak is represented by X''. The indifference curves of the selfish individiduals enclose those peaks. Two of the indifference curves drawn are tangent to each other at some feasible allocation. As pointed out in Figure 1, it can be shown that $P(0,0)$ is formed by the set of all such tangency points.[5]

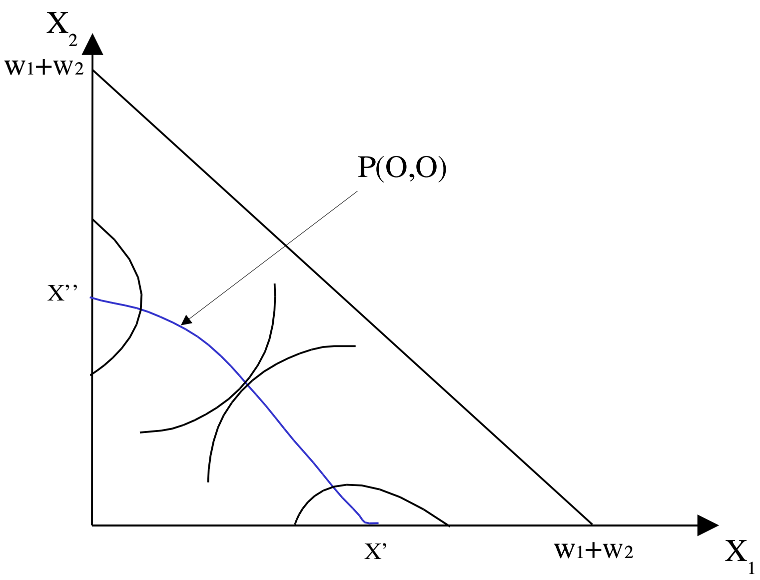

Figure 1: PO Allocations, $\gamma_1 = \gamma_2 = 0$ case

We shall restrict the values of the altruistic parameters. One may say that it is realistic not to expect large simultaneous values for γ_1, γ_2. In such cases, a conflict may be generated because of "too much" altruism: each person would prefer that the other one worked less for the common good than he or she wants to.To avoid such situations, let us focus our attention to the set $\Gamma = \{(\gamma_1, \gamma_2) \in \mathbb{R}^2_+ : \gamma_1\gamma_2 \leq 1\}$. We study first the

[5]We know that

$$P(0,0) = \left\{(x_1, x_2, y) : \frac{v_2^2(x_2, y)}{v_1^2(x_2, y)} + \frac{v_2^1(x_1, y)}{v_1^1(x_1, y)} = \frac{1}{f'(\omega - x_1 - x_2)}\right\},$$

which is Samuelson's efficiency condition. We are claiming here that

$$P(0,0) = \left\{(x_1, x_2) : \frac{W_2^2(x_1, x_2; 0)}{W_1^2(x_1, x_2; 0)} = \frac{W_2^1(x_1, x_2; 0)}{W_1^1(x_1, x_2; 0)}\right\}.$$

where

$$W_j^i = \frac{\partial W^i}{\partial x_j}$$

Pareto allocations P as a correspondence from Γ into T. The next result allows us to make preference comparisons.

Proposition 4 *The Pareto correspondence* P *is decreasing in* γ*, i.e. if* $(\gamma_1, \gamma_2), (\gamma_3, \gamma_4) \in \Gamma$ *are such that* $(\gamma_1, \gamma_2) \geq (\gamma_3, \gamma_4)$*, then* $\mathrm{P}(\gamma_1, \gamma_2) \subset \mathrm{P}(\gamma_3, \gamma_4)$*.*

Proof: The Proposition is trivial if $\gamma_1 = \gamma_2 = 0$. Let us assume, for example, that $\gamma_2, \gamma_4 > 0$. An allocation $(x_1, x_2) \in \mathrm{P}(\gamma_1, \gamma_2)$ iff there is an $\alpha \in [0, 1]$ such that (x_1, x_2) maximizes

$$\begin{aligned} & \alpha W^1(x_1, x_2; \gamma_1) + (1-\alpha) W^2(x_1, x_2; \gamma_2) \\ = \;& (\alpha + \gamma_2(1-\alpha)) v^1(x_1, f(\omega - x_1 - x_2)) + (\alpha\gamma_1 + 1 - \alpha) v^2(x_2, f(\omega - x_1 - x_2)). \end{aligned}$$

Let $a = \alpha + \gamma_2(1-\alpha)$, $b = \alpha\gamma_1 + 1 - \alpha$ and define β by the equation

$$\frac{a}{\beta + \gamma_4 - \gamma_4\beta} = \frac{b}{\beta\gamma_3 + 1 - \beta},$$

that is

$$\beta = \frac{a - b\gamma_4}{a - b\gamma_4 + b - a\gamma_3}.$$

Since $\gamma_4 \leq \gamma_2$ and $\gamma_1\gamma_2 \leq 1$, we have $b\gamma_4 \leq b\gamma_2 = \alpha\gamma_1\gamma_2 + \gamma_2 - \alpha\gamma_2 \leq \alpha + \gamma_2 - \alpha\gamma_2 = a$. Likewise, $a\gamma_3 \leq a\gamma_1 = \alpha\gamma_1 + \gamma_1\gamma_2(1-\alpha) \leq \alpha\gamma_1 + 1 - \alpha = b$. So $1 \geq \beta \geq 0$. Let

$$t = \frac{\alpha\gamma_1 + 1 - \alpha}{\beta\gamma_3 + 1 - \beta} \geq 0$$

Then, we see that

$$\alpha W^1(x_1, x_2; \gamma_1) + (1-\alpha) W^2(x_1, x_2; \gamma_2) = t\left(\beta W^1(x_1, x_2; \gamma_3) + (1-\beta) W^2(x_1, x_2; \gamma_4)\right)$$

so (x_1, x_2) is Pareto optimal in the economy $\gamma = (\gamma_3, \gamma_4)$ as well. **Q.E.D.**

Figure 2 shows the inclusion $\mathrm{P}(\gamma_1, \gamma_2) \subset \mathrm{P}(\gamma_3, \gamma_4) \subset P(0, 0)$. This result might be viewed as a more general approach to central planner welfare analysis. Here you have several atomic planners, and the problem is to coordinate their decisions, or aggregate their preferences in the social choice sense. In Figure 2, the γ_i correspond to the different levels of altruism of the proposition, and are placed in the set $P(0, 0)$ where the peaks of the functions (4) and (5) occur. It is easy to show that another way to find $P(0, 0)$ is to find the peaks of functions (4) and (5), varying the corresponding gamma's from zero to infinity as in standard welfare analysis. It also turns out that when $\gamma_1\gamma_2 = 1$, the peaks coincide, so that the solutions of the atomic planners are the same and are equal to the one proposed by the central planner. In this case there is no conflict at all, and the common preferred allocation is the unique Pareto optimal allocation of the economy.

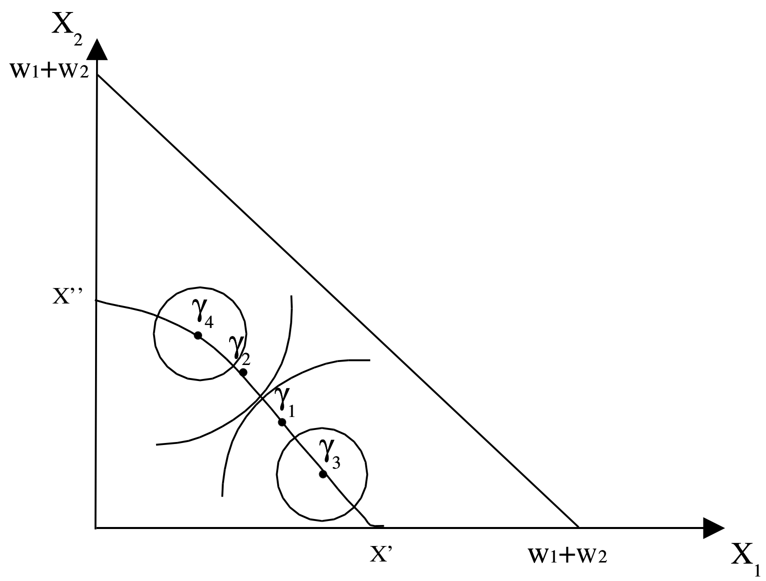

Figure 2: Efficiency Comparisons: P is decreasing

Next, we turn to the definition of equilibrium. We first remark that agent $i = 1, 2$, being in the possession of the initial endowment ω_i, can always guarantee for himself the utility

$$\max_{0 \le \xi \le \omega_i} W^i(\xi, \omega_j; \gamma_i) \tag{6}$$

where $j \ne i$. Let $\xi_i(\gamma_i)$ be the solution to the above problem. Now suppose agent $j \ne i$ could guarantee for himself the consumption x_j of the private good. Given this situation, agent i would choose the solution $g_i(x_j, \gamma_i)$ to the following maximization problem

$$\max_{0 \le \xi \le w - x_j} W^i(\xi, x_j; \gamma_i) \tag{7}$$

It is easy to verify that the mappings W^1, W^2 are strictly concave and so, given $x_j \in [0, \omega_j]$, there is a unique solution, $g_i(x_j, \gamma_i)$ to the above problem. Let $\varphi_i(x_j, \gamma_i) = \max\{g_i(x_j, \gamma_i), \xi_i(\gamma_i)\}$. The mapping $\varphi_i(x_j, \gamma_i)$ is agent i's best response (in terms of consumption of the private good) to the action x_j of agent j. Our notion of equilibrium coincides now essentially with the Nash solution.

Definition 5 A feasible allocation $(x_1, x_2) \in \mathrm{T}$ is an Altruistic Agreement for $(\gamma_1, \gamma_2) \in \Gamma$, if for every $i, j = 1, 2$ with $i \ne j$, we have that $x_i = \varphi_i(x_j, \gamma_i)$.

In general, there is an inherent conflict among the agents of this economy, in spite of their altruism, as we said. We also remarked that only in the case $\gamma_1\gamma_2 = 1$, what each agent wants for himself of the private good coincides exactly with what the other agent wants for him or her (the peaks for equations (4) and (5) coincide for both agents), and they both agree on the preferred quality of the public good.[6] For the other cases, we postulate the Nash equilibrium solution concept as the outcome of their interaction.

When $\gamma_1\gamma_2 < 1$, a Nash equilibrium entails the kind of disagreement in which each agent, if he or she were to choose how to allocate the endowments for the entire economy, would prefer to curb his own work and have others work more than in the proposed equilibrium. When $\gamma_1\gamma_2 > 1$, on the contrary, a Nash equilibrium would entail an allocation whereby each agent would prefer to work more himself and less the other person than in the equilibrium. The case $\gamma_1\gamma_2 = 1$ generates an equilibrium that is a Pareto optimal allocation for both economies, the one with selfish, as well as the one with altruistic people. Furthermore, the Pareto allocation for the altruistic economy is, in this case, the only possible Pareto optimal allocation, as we said before.

Proposition 6 *Given $\gamma_1, \gamma_2 \geq 0$, $\gamma_1\gamma_2 \leq 1$, there is (at least) an Altruistic Agreement in the economy (γ_1, γ_2).*

Proof: By the Theorem of the Maximum, the functions $\varphi_i : [0, \omega] \to [0, \omega]$, $i = 1, 2$ are continuous. Hence, the function $\psi(x_1, x_2) = (\varphi_1(x_2), \varphi_2(x_1)) : [0, \omega] \times [0, \omega] \to [0, \omega] \times [0, \omega]$ is also continuous. By Brouwer's fixed point theorem, there is $(x_1, x_2) \in [0, \omega] \times [0, \omega]$ such that $(x_1, x_2) = (\varphi_1(x_2), \varphi_2(x_1))$. Thus, (x_1, x_2) is an Altruistic Agreement. **Q.E.D.**

We make the necessary assumptions so that the solutions to the problems 6 and 7 are determined by the first order conditions. Furthermore, for the sake of simplicity we also assume that $\varphi_i(x_j, \gamma_i) = g_i(x_j, \gamma_i)$. This is a realistic assumption for small values of the altruistic parameters. Hence, the mappings φ_1 and φ_2 defined above are defined implicitly by the equations

$$D_1W^1(\varphi_1(\theta, \gamma_1), \theta; \gamma_1) = 0, \quad D_2W^2(\theta, \varphi_2(\theta, \gamma_2); \gamma_2) = 0.$$

It follows that φ_1 and φ_2 are C^1 functions.

Lemma 7 1. *For $(\gamma_1, \gamma_2) \in \Gamma$ fixed, functions $\varphi_i(\theta, \gamma_i)$, $i = 1, 2$ are strictly decreasing in θ.*

2. *For θ fixed, $\varphi_1(\theta, \gamma_1)$ is strictly decreasing in γ_1, $\varphi_2(\theta, \gamma_2)$ is strictly decreasing in γ_2.*

Proof: By implicit differentiation we have now that $\varphi_1' = \partial\varphi_1/\partial\theta$ is derived from equation $D_{11}W^1(\varphi_1(\theta, \gamma_1), \theta)\varphi_1'(\theta) + D_{12}W^1(\varphi_1(\theta, \gamma_1), \theta) = 0$. Thus,

$$\varphi_1'(\theta, \gamma_1) = -\frac{D_{12}W^1(\varphi_1(\theta, \gamma_1), \theta)}{D_{11}W^1(\varphi_1(\theta, \gamma_1), \theta)}.$$

[6] In fact, if the maximization problems (6) and (7) were performed over the two first arguments, an internal single peak vector would be obtained as solution, given our assumptions. The result of Marhuenda and Pérez-Martí ([11]) that says that the peaks coincide when $\gamma_1\gamma_2 = 1$ apply here too.

A simple computation shows (using assumptions 2 and 1) that

$$D_{11}W^1 = v^1_{11} - 2v^1_{12}f' + v^1_{22}(f')^2 + v^1_2 f'' + \gamma_1(v^2_{22}(f')^2 + v^2_2 f'') < 0$$

and

$$D_{12}W^1 = -v^1_{12}f' + v^1_{22}(f')^2 + v^1_2 f'' - \gamma_1(v^2_{12}f' - v^2_{22}(f')^2 - v^2_2 f'') < 0$$

so $\varphi_1'(\theta, \gamma_1) < 0$ and φ_1 is decreasing in θ.

By differentiating now implicitly $D_1W^1(\varphi_1(\theta, \gamma_1), \theta; \gamma_1) = 0$ with respect to γ_1 we have

$$D_{11}W^1 \frac{\partial \varphi_1}{\partial \gamma_1} + \frac{\partial}{\partial \gamma_1} D_1 W^1 = 0$$

which is the same as

$$D_{11}W^1 \frac{\partial \varphi_1}{\partial \gamma_1} = v^2_2 f'$$

However, $D_{11}W^1 < 0$ and $v^2_2 f' > 0$, so $\frac{\partial \varphi_1}{\partial \gamma_1} < 0$. A similar argument applies to φ_2. **Q.E.D**.

Proposition 8 *Given $(\gamma_1, \gamma_2) \in \Gamma$, the Altruistic Agreement of Proposition 6 is unique.*

Proof: By Lemma 7, for each γ_1 fixed, the function φ_1 is monotonic in θ, and thus, it has an inverse $\psi_1 = \varphi_1^{-1}$. An altruistic Agreement is determined by a point θ such that $\psi_1(\theta, \gamma_1) = \varphi_2(\theta, \gamma_2)$.

We claim now that $\psi_1'(\theta, \gamma_1) < \varphi_2'(\theta, \gamma_2)$ for $0 < \gamma_1\gamma_2 \le 1$ and all θ in the domain of ψ_1. To see this fix γ_1, γ_2 such that $0 < \gamma_1\gamma_2 \le 1$. Let $\gamma_1' = 1/\gamma_2$. Note that $\gamma_1 \le \gamma_1'$. After some computation we have that

$$\begin{aligned}
\varphi_2'(\theta, \gamma_2) &= -\frac{D_{12}W^2(\varphi_1(\theta, \gamma_1), \theta)}{D_{22}W^2(\varphi_1(\theta, \gamma_1), \theta)} \\
&= \frac{v^2_{12}f' - v^2_{22}(f')^2 - v^2_2 f'' + \gamma_2(v^1_{12}f' - v^1_{22}(f')^2 - v^1_2 f'')}{v^2_{11} - 2v^2_{12}f' + v^2_{22}(f')^2 + v^2_2 f'' + \gamma_2(v^1_{22}(f')^2 + v^1_2 f'')}
\end{aligned}$$

Since W^1 is strictly concave, the determinant

$$\begin{vmatrix} D_{11}W^1 & D_{12}W^1 \\ D_{12}W^1 & D_{22}W^1 \end{vmatrix} = D_{11}W^1 D_{22}W^1 - (D_{12}W^1)^2$$

is strictly positive. Thus,

$$\begin{aligned}
\psi_1'(\theta, \gamma_1) &= -\frac{D_{11}W^1(\varphi_1(\theta, \gamma_1), \theta)}{D_{12}W^1(\varphi_1(\theta, \gamma_1), \theta)} \\
&< -\frac{D_{12}W^1(\varphi_1(\theta, \gamma_1), \theta)}{D_{22}W^1(\varphi_1(\theta, \gamma_1), \theta)} \\
&= -\frac{-v^1_{12}f' + v^1_{22}(f')^2 + v^1_2 f'' - \gamma_1(v^2_{12}f' - v^2_{22}(f')^2 - v^2_2 f'')}{v^1_{22}(f')^2 + v^1_2 f'' + \gamma_1(v^2_{11} - 2v^2_{12}f' + v^2_{22}(f')^2 + v^2_2 f'')}
\end{aligned}$$

$$
\begin{aligned}
&< \frac{v_{12}^1 f' - v_{22}^1 (f')^2 - v_2^1 f'' + \gamma_1' (v_{12}^2 f' - v_{22}^2 (f')^2 - v_2^2 f'')}{v_{22}^1 (f')^2 + v_2^1 f'' + \gamma_1' (v_{11}^2 - 2v_{12}^2 f' + v_{22}^2 (f')^2 + v_2^2 f'')} \\
&= \frac{v_{12}^2 f' - v_{22}^2 (f')^2 - v_2^2 f'' + \gamma_2 (v_{12}^1 f' - v_{22}^1 (f')^2 - v_2^1 f'')}{v_{11}^2 - 2v_{12}^2 f' + v_{22}^2 (f')^2 + v_2^2 f'' + \gamma_2 (v_{22}^1 (f')^2 + v_2^1 f'')} \\
&= \varphi_2'(\theta, \gamma_2)
\end{aligned}
$$

where the last inequality follows because $v_{12}^2 f' - v_{22}^2 (f')^2 - v_2^2 f'' > 0$, $v_{11}^2 - 2v_{12}^2 f' + v_{22}^2 (f')^2 + v_2^2 f'' < 0$ and $\gamma_1 \leq \gamma_1'$. The claim is now proved.

Suppose now that there are two Altruistic Agreements $\theta_1 < \theta_2$ such that $\psi_1(\theta_1) = \varphi_2(\theta_1)$ and $\psi_1(\theta_2) = \varphi_2(\theta_2)$. Then $\psi_1(\theta) - \varphi_2(\theta, \gamma_2)$ has two zeros in the interval $[\theta_1, \theta_2]$, so there is $\theta_3 \in (\theta_1, \theta_2)$ such that $\psi_1'(\theta_3) = \varphi_2'(\theta_3)$ which contradicts the claim. **Q.E.D.**

In Figure 3 Altruistic Agreements coincide with the intersection of the reaction curves $\psi_1 = \varphi_1^{-1}$ and φ_2, for different profiles $\gamma \in \Gamma$. At the crossing points, $\psi_1' < \varphi_2' < 0$. When $\gamma_1\gamma_2 = 1$, the graphs of $\psi_1(\ ,\gamma_1)$ and $\varphi_2(\ ,\gamma_2)$ cross exactly at the singleton $\mathrm{P}(\gamma_1, \gamma_2) \subset \mathrm{P}(0,0)$. We have also represented the case of $(\gamma_1, \gamma_{2'})$, with $\gamma_{2'} < \gamma_2$, so that $\gamma_1\gamma_{2'} < 1$. Here $\psi(\ ,\gamma_1)$ and $\varphi(\ ,\gamma_{2'})$ cross at an inefficient allocation, above $\mathrm{P}(0,0)$. In this case, $\mathrm{P}(\gamma_1, \gamma_{2'})$ is the segment in the Pareto boundary $\mathrm{P}(0,0)$ that goes from the point where $\varphi(\ ,\gamma_{2'})$ crosses it (which is, precisely, the peak of the function in (5), and is depicted in the figure by $\gamma_{2'}$) to $\mathrm{P}(\gamma_1, \gamma_2)$.

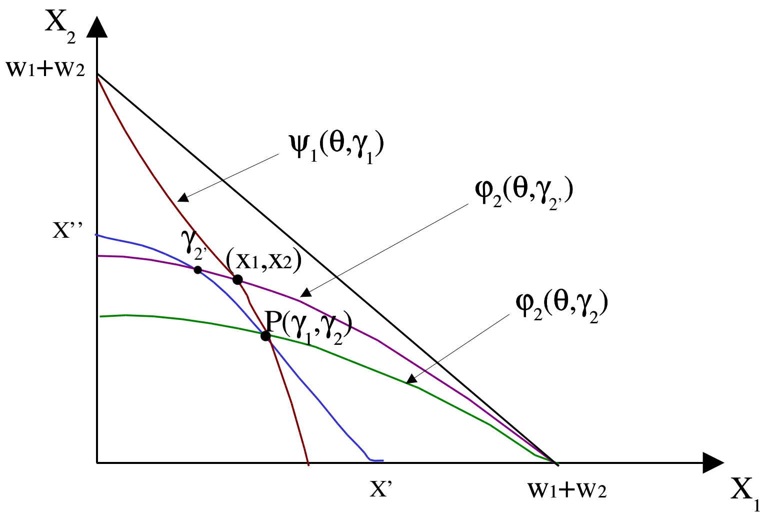

Figure 3: Examples of Altruistic Equilibria

In order to fully characterize the set of all possible Altruistic Agreements as (γ_1, γ_2) vary in Γ, we analyze the structure of the mapping $\mathrm{R} : \Gamma \to \mathrm{T} \times \mathrm{T}$ which assigns to every $(\gamma_1, \gamma_2) \in \Gamma$ its unique Altruistic Agreement $\mathrm{R}(\gamma_1, \gamma_2)$.

First note that $\mathrm{R}(\gamma_1, \gamma_2)$ is given implicitly by the system of equations

$$D_1 W^1 = D_2 W^2 = 0.$$

The Jacobian of this system is $D_{11}W^1 D_{22}W^2 - D_{12}W^1 D_{12}W^2 > 0$, whenever $(\gamma_1, \gamma_2) \in \Gamma$. Hence, by the implicit function Theorem, $\mathrm{R}(\gamma_1, \gamma_2)$ is a C^1 function of (γ_1, γ_2).

By Proposition 7, the function R defines a diffeomorphism from the open set $A = \{(\gamma_1, \gamma_2) \in \mathbb{R}^2 : \gamma_1, \gamma_2 > 0, \gamma_1\gamma_2 < 1\}$ into the set of feasible allocations which extends to the boundary $\partial A = \{(\gamma_1, \gamma_2) \in \mathbb{R}^2_+ : \gamma_1\gamma_2 = 0\} \cup \{(\gamma_1, \gamma_2) \in \mathbb{R}^2_+ : \gamma_1\gamma_2 = 1\}$. Let $B = \{(\gamma_1, \gamma_2) \in \mathbb{R}^2_+ : \gamma_1\gamma_2 = 1\}$.

Proposition 9 $\mathrm{P}(0,0) = \mathrm{R}(B) \cup \{(\bar{x}_1, 0)\} \cup \{(0, \bar{x}_2)\}$, *where, for* $i = 1, 2$, $\bar{x}_i$ *maximizes*

$$\max_{0 \leq z \leq w} v_i(z, f(\omega - z)).$$

It follows that $\mathrm{R}(A)$ is a region of $\mathbb{R}^2$ with a non–empty interior whose boundary is determined by $\mathrm{P}(0,0)$ and the graphs of the mappings $\varphi_1(\theta, 0)$ and $\varphi_2(\theta, 0)$. The set $\mathrm{R}(A)$ is represented in Figure 4 as the shaded area. In the context of Figure 4 we see also that, in general, a market equilibrium underproduces public goods, and generates allocations above $\mathrm{P}(0,0)$. A straightforward consequence of the characterization of $\mathrm{R}(A)$ is that, given a market allocation that is, say, inefficient, there are altruistic parameters which improve upon that allocation in the Pareto sense. This result is stated formally as

Proposition 10 *Let* (x_1, x_2) *be an inefficient allocation for the society* $(\gamma_1, \gamma_2) \in \Gamma$. *Then there exists an economy* $(\gamma_3, \gamma_4) \in \Gamma$ *with* $(\gamma_3, \gamma_4) \geq (\gamma_1, \gamma_2)$ *such that, if* (x_3, x_4) *is the Altruistic Equilibrium that corresponds to the economy* (γ_3, γ_4), *this allocation is Pareto preferred to* (x_1, x_2) *in both economies.*

This proposition implies that once people increase their degree of altruism by means of some mechanism, they do not have any incentives to reverse the change, as a group. In other words, individuals have an incentive to change their preferences, as a group, from selfish to altruistic. Once they have become altruistic, they lack incentives to change back their preferences towards selfishness.

3 Preference formation through group interaction

In the previous section, we have seen that individuals can benefit by changing their preferences towards altruism. How can this happen? We extend now the previous model to propose a mechanism by which a rational agent i will have incentives to manipulate the

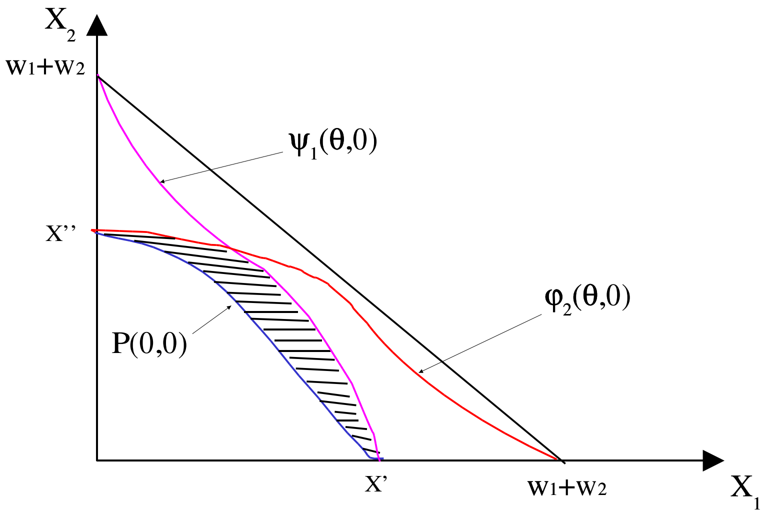

Figure 4: Set of Altruistic Equilibria, for different (γ_1, γ_2) profiles

preferences of the other player $j \neq i$ so that j becomes more altruistic towards i. With this, we provide an approach to endogenous preference formation.

Here, agents play a two-stage game. At the first stage, each agent $i = 1, 2$ decides the amount of effort he spends on "being nice" to the other agent $j \neq i$. By spending this effort she could make agent j more "sympathetic" towards her. We model this phenomenon by postulating that the effort agent i spends on j affects the parameter of altruism γ_j of agent j.

More formally, if agent i spends effort t_i on agent $j \neq i$, then the preferences of agent j in the second stage of the game are given by

$$u^j(x_1, x_2, y) = v^j(x_j, y) + \gamma_j(t_i)v^i(x_i, y),$$

as in the preceding model. But, with the important difference that now $\gamma_j = \gamma_j(t_i)$ depends on the effort of the other agent i[7].

At the second stage, agents play the altruistic public good provition game of the preceding section and consumption is realized. The payoffs are those obtained in the altruistic agreement of that game. However, if agent $i = 1, 2$ has spent effort t_i during

[7]This is without loss of generallity, since t_i could also enter directly into the utility function of j, more like a private good gift or donation; we focus the analysis on the distinctive feature of the model, having in mind that the gift itself is what could have motivated in practice the chance in preferences, and taking into account that the gift does not alter the results of the model.

stage 1, then his initial endowment of the private good has been reduced accordingly to $\omega_i - t_i$.

Hence, given that the amount of effort (t_1, t_2) has been decided during the first period, at the second stage, agent 1 has the following utility function on consumption bundles x_1, x_2 and $y = f(w - t_1 - t_2 - x_1 - x_2)$,

$$W^1(t_1, t_2; x_1, x_2) = v^1(x_1, f(w - t_1 - t_2 - x_1 - x_2)) + \gamma_1(t_2) v^2(x_2, f(w - t_1 - t_2 - x_1 - x_2)).$$

Similarly, for agent 2,

$$W^2(t_1, t_2; x_1, x_2) = v^2(x_2, f(w - t_1 - t_2 - x_1 - x_2)) + \gamma_2(t_1) v^1(x_1, f(w - t_1 - t_2 - x_1 - x_2)).$$

Abusing notation, we use the same W^i for utility functions as in the previous section, even though, formally, preferences now depend additionally on t_1 and t_2. It is natural to assume that, if no effort is spent on the other agent, then the preferences remain unaltered. We also assume that the greater the effort, the greater the effect on the other agent. Formally,

Assumption 11 *1. $\gamma_i(0) = 0$, for $i = 1, 2$.*

2. $\gamma_i'(t) > 0$, for all $t \in (0, \omega_i)$ and $i = 1, 2$.

Consider the mapping Φ assigning to each $(t_1, t_2) \in [0, \omega_1] \times [0, \omega_2]$ the unique altruistic equilibrium $\Phi(t_1, t_2) = R(\gamma_1(t_2), \gamma_2(t_1)) = (x_1(t_1, t_2), x_2(t_1, t_2))$ of the induced game at the second stage with preferences $W^1(t_1, t_2, x_1, x_2)$ and $W^2(t_1, t_2; x_1, x_2)$.

Lemma 12 *The mapping Φ is differentiable. Furthermore,*

$$\frac{\partial x_i}{\partial t_j}(t_1, t_2) \geq 0 \quad \text{if } i = j \tag{8}$$

$$\frac{\partial x_i}{\partial t_j}(t_1, t_2) \leq 0 \quad \text{if } i \neq j \tag{9}$$

The proof is straightforward if we take into account Lemma 7, the discussions preceding it, and assumption 11. The lemma states that the effort spent by, say, agent 1, during the first stage has the effect of increasing his consumption of the private good at the second stage. Correspondingly, it reduces the amount of private good consumed by the other agent at that stage. Our definition of equilibrium is essentially the notion of subgame perfect equilibrium.

Definition 13 An allocation $(t_1^*, t_2^*; x_1^*, x_2^*) \geq 0$ is an equilibrium provided the following conditions hold:

1. $\Phi(t_1^*, t_2^*) = (x_1^*, x_2^*)$.

2. $W^1(t_1^*, t_2^*; x_1^*, x_2^*) \geq W^1(t_1, t_2^*; \Phi(t_1, t_2^*))$ for any other $t_1 \in [0, \omega_1]$ and $W^2(t_1^*, t_2^*; x_1^*, x_2^*) \geq W^2(t_1^*, t_2; \Phi(x_1^*, x_2))$ for any other $t_2 \in [0, \omega_2]$.

In particular, agents anticipate their future altruistic preferences at the first stage and evaluate the outcome in view of their preferences when consumption is realized, i.e., at the second stage. In view of the results of section 2, the altruistic outcome at the second stage would also be preferred from the point of view of selfish preferences.

Proposition 14 *There exists an equilibrium* $(t_1^*, t_2^*; x_1^*, x_2^*)$.

Proof: Let $i = 1, 2$ and for fixed $t_j \in [0, \omega_j]$, with $j \neq i$, consider the mapping $t_i \longmapsto W^i(\Phi(t_1, t_2))$ with $t_i \in [0, \omega_i]$. Since, it is continuous it attains a maximum say $\Lambda_i(t_j)$ which is agent i's best response to the action t_j by agent j. An equilibrium corresponds to a point (t_1^*, t_2^*) such that $\Lambda_i(\Lambda_j(t_i^*)) = t_i^*$ for $i = 1, 2$, $i \neq j$. So consider the mappings $\Lambda_i \circ \Lambda_j : [0, \omega_i] \rightarrow [0, \omega_i]$ and $\Lambda_j \circ \Lambda_i : [0, \omega_j] \rightarrow [0, \omega_j]$. By Brouwer's fixed point theorem there are fixed points $(t_1^*, t_2^*) \in [0, \omega_1] \times [0, \omega_2]$ such that $\Lambda_i(\Lambda_j(t_i^*)) = t_i^*$ for $i = 1, 2$, $i \neq j$. **Q.E.D**.

Proposition 15 *Let* $(t_1^*, t_2^*; x_1^*, x_2^*)$ *be an equilibrium with* $t_i^* \neq \omega_i$ *for* $i = 1, 2$. *Then,* $t_1^*, t_2^* > 0$.

Proof: In equilibrium, the allocations $(x_1^*(t_1^*, t_2^*), x_2^*(t_1^*, t_2^*))$ are defined implicitly by the equations

$$\begin{array}{rcccl} D_1 W^1 & = & v_1^1 - v_2^1 f' - \gamma_1 v_2^2 f' & = & 0 \\ D_2 W^2 & = & v_1^2 - v_2^2 f' - \gamma_2 v_2^1 f' & = & 0 \end{array}$$

so differentiating implicitly the first equation with respect to t_1 and rearranging terms we obtain that, at $t_2 = 0$, the derivatives $D_1 x_1$and $D_1 x_2$ satisfy the equation

$$(1 + D_1 x_1 + D_1 x_2)\left(v_{12}^1 f' - v_{22}^1 (f')^2 - v_2^1 f''\right) = D_1 x_1 (v_{11}^1 - v_{12}^1 f')$$

and, in view of the signs of $D_1 x_j$ and v_{kl}^i, we see that, whenever $t_2 = 0$, we must have that $1 + D_1 x_1 + D_1 x_2 < 0$. On the other hand, if $t_2 = 0$ then,

$$\frac{dW^1}{dt_1} = v_1^1 D_1 x_1 - v_2^1 f' \left(1 + D_1 x_1 + D_1 x_2\right) > 0$$

which shows that when $t_2 = 0$, agent 1 will want to increase t_1 up to the boundary ω_1. Thus, $t_2 = 0$ cannot be an equilibrium of the game. **Q.E.D**.

4 Summary and Final remarks

We have shown that it is possible to interpret altruism in an economy with public goods as a mechanism used by members of society in order to improve their welfare. In fact, individuals might have an incentive to deviate and stay selfish if they know the other members of society will evolve towards altruism. Nevertheless, societies (including in this concept families, groups of friends, countries) with an appropriately higher degree of altruism might be more successful than others with a lesser degree of altruism, due to efficiency gains. Our results give some insight regarding the existence of a grup fitness function that makes possible an evolutionary process that benefits groups of agents with altruistic prefences, since, if the group produces more of a local public good, like defense or technology, the group is more fit than others to survive. As for the individual incentives to free ride within the group, we present a model of group interaction in which people affect the preferences of others, and reciprocal altruism arises as a consequence of investment in education, depending on the well documented capacity of humans to learn reciprocal altruism (see Henrich [7]).

In our model of preference formation, even though rational selfish agents would prefer to stay selfish in order to "use" the other person, who becomes altruistic through their influence, they realize that their own preferences might change in the process of courting, and they decide to engage in the game, anyway, gaining more as a reciprocated altruist than in the previous selfish interaction with the same person. Since this process might lead to a Pareto improvement, it is reasonable to expect that it will be favored by natural selection, which would explain the "weakness" of the selfish people that allows them to be able to learn to true love others.

The local interactions model shows how an economy populated by altruists, may reach complete harmony with complete efficiency. The failures of the model add the need of a specific combination in the degrees of altruism to the failures of the central benevolent planner mechanism, since each agent has to know all the citicens, and all re resources of society, as well as being able to calculate the optimal plan. in fact, an extension of the model to n individuals require that each person cares for every other person in society, and the assumed level of rationality regarding the amount of information and computational ability becomes unrealistically large. Such a model would suffer from the same criticism (at even a larger scale) of impracticality as the central planner proposal. One alternative would be to require that any individual knows and cares only about his neighbor. It turns out that evolution towards altruism has been shown[8] to be robust only in the cases of complete information. Individualistic preferences become evolutively robust in the presence of incomplete information. Thus, altruism is more likely to emerge in societies where individuals are rather well informed about each other. This extension, natural in the model we present here, in which people care for their neighbors, would then have two advantages: it would be realistic, and it would make the model evolutionarily stable.

The paper also opens a window to consider the copyright laws under a new light. It might be inefficient to enforce a copyright law implying exclusion of a public good when it

[8] See Ok and Vega-Redondo (1999, [16]).

has been produced by a spontaneous non–profit movement [9]. Accordingly, property rights of law-excludable public goods, like technology, which are today crucial for any growth model, should be reconsidered, in at least some cases, in light of efficiency considerations. In addition, public policy considerations regarding economic growth, due to technological development, might include the alternative of preference formation and support to altruistic groups, such as the Free Software movement, that produce such law-excludable public goods.

References

[1] Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence", *Journal of Political Economy*, 1989, v.97.

[2] Andreoni, James, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving", *The Economic Journal*, 1990, v.100, pp. 464-477.

[3] J. R. Barro, "Are government bonds net wealth?", *Journal of Political Economy*, 1095–1117, 1974.

[4] G. S. Becker, "A theory of social interactions", *Journal of Political Economy*, 1063–1093, 1974.

[5] Bersgstrom, Ted; Stark, Oded, "How Altruism Can Prevail in an Evolutionary Environment", *American Economic Review*, May 1993, v. 83, iss. 2, pp. 149-55

[6] Ted Berstrom and Oded Stark, "How Altruism Can Prevail Under Natural Selection", Working paper, University of Michigan and Harvard University, 2002.

[7] Joseph Henrich, "Cultural group selection, coevolutionary processes and large-scale cooperation", *Journal of Economic Behavior & Organization*, 53, 3-35, 2004.

[8] Bester, Helmut and Güth, Werner. "Is altruism evolutionary stable?", *Journal of Economic Behavior and Organization* 34, 193-209, 1998.

[9] Border, K. C. *Fixed Point Theorems with Applications to Economics and Game Theory.* Cambridge University Press, 1992.

[10] Cornes, Richard, and Sandler, Todd. "Easy Riders, Joint Production, and Public Goods" *Economic Journal* 94, 1984: 580-98

[11] Marhuenda, Francisco and Felipe Pérez-Martí, "Altruism and Incomplete Markets", IESA, Caracas, Avance de Investigación (working paper), May 1998.

[12] Marhuenda, Farncisco and Felipe Pérez-Martí, "On altruism, Efficiency and Public Goods", IESA, Caracas, Avance de Investigación (working paper), 1998.

[13] Marhuenda, Francisco and Felipe Pérez-Martí, "Consistency of interdependent preferences", Universidad Carlos III de Madrid working paper. September 1999.

[9] Think of the case of UNIX, produced originally by what should well be considered an altruistic group, subsequently appropriated by others, retarding the Linux phenomenon, which is quite succesfull right now.

[14] Mas-Colell, A., Whinston, M. and J. G. *Microeconomic Theory* Oxford University Press, 1995.

[15] Mulligan, Casey B, *Parental Priorities and Economic Inequality*, Chicago Press, 1997.

[16] Ok, Efe and Vega-Redondo, Fernando, "On the Evolution of Individualistic Preferences: Complete versus Incomplete Information Scenarios". University of Alicante working paper, March 1999.

[17] Rotemberg, Julio J., "Human relations in the workplace", *Journal of Political Economy* 102 (4), 684–717, 1994.

[18] Stark, O. *Altruism and Beyond.* Cambridge University Press, 1995.

[19] Stark, O. and Ita Falk, "Tranfers, Empathy Formation, and Reverse Transfers", *The American Economic Review.* Vol. 88, No. 2, Papers and Proceedings of the Hundred and Thenth Annual Meeting of the American Economic Association (May, 1998).