WORKING PAPER
DIPARTIMENTO DI ECONOMIA PUBBLICA

Working Paper n. 91

Marcello Basili and Maurizio Franzini

COOPERATION AND RECIPROCITY: A THEORETICAL APPROACH

Roma, Novembre 2005

UNIVERSITA' DEGLI STUDI DI ROMA
"LA SAPIENZA"

COOPERATION AND RECIPROCITY: A THEORETICAL APPROACH

Marcello Basili* and Maurizio Franzini**

Abstract

Cooperation among genetically unrelated agents occurs in many situations where economic theory would not expect it. A too narrow conception of self-interest is widely considered the culprit. In particular, relying on experimental evidence in plenty, we consider strong reciprocity rules of behaviour, according to which it is worth bearing the cost of punishing those who defect, and we give analytical foundation to such behaviour – and more generally to cooperation-proneness. The basic idea is that most agents may include self-esteem in their utility function and actually produce or destroy self-esteem through their effective behaviour. The latter amounts to introducing a moral system in individual behaviour in such a way to make it amenable to rational maximization. We also show how the presence of cooperation-prone agents may impact on the best contract in Principal-Agents situations by altering the convenience of gift giving and trust.

JEL Classification J41, D64

Keywords: agency, altruism, self-interest, punishment, reciprocity

* *University of Siena, Piazza san Francesco 7, 53100 Siena, basili@unisi.it*

** *University of Rome « La Sapienza », via Castro Laurenziano, maurizio.franzini@uniroma1.it*

Introduction

Cooperation among not genetically related agents is widely observed in behavioural experiments and also in everyday life, even when repeated interaction is absent. This evidence is very hard to reconcile with standard economic theory based on the assumption of self-interested agents. Recently, several economists have taken up the challenge of providing a general explanation of how cooperation can be established and maintained in a setting potentially open to free riding and opportunism. One of the most interesting strands of research is based on *strong reciprocity* (Bowles and Gintis 2003, Gintis et al. 2003, Gintis 2004) which represents a rather weak relaxation of the assumption that all agents are strictly self-interested. According to the strong reciprocity hypothesis many humans are ready to punish those who behave opportunistically even when this is costly to them. An adequate number of strong reciprocators may suffice to sustain cooperation when it would be impossible under more customary assumptions.

The strong reciprocity assumption has several appealing features. First of all it seems to fit in very well with empirical experiments both in laboratory and natural settings. Secondly, it is capable of sustaining a cooperative equilibrium even in the presence of a large number of standard self-interested agents: the prospect of being punished by few strong reciprocators may be enough to induce them to refrain from opportunistic behaviour.

However, we think that the analytical foundations of this hypothesis and its relation with the basic postulates of rationality might be generalized. The main purpose of this paper is to further develop the notion of strong reciprocity. More specifically we pursue three goals.

The first is to argue that a rational foundation for a more cooperative-prone behaviour can be provided by the twin assumptions that agents include self-esteem in their utility function and the amount of self-esteem depends on how they behave in social situations. Such a model may encompass the strong reciprocity hypothesis as a special case.

The second goal is to highlight those factors which make those agents who are responsive to self-esteem in our sense effective co-operators. In fact, rational agents, however prone to cooperation they may be, will not cooperate at any cost and in whatever setting. This is borne out by several pieces of evidence. We will show that other agents' behaviour may be crucial in this respect.

Related to this, the third goal is to show how the presence of a cooperation-prone Agent may impact on the best contract a Principal can offer. More specifically we will model a standard situation of adverse selection proving that a sort of *gift giving* contract, eliciting a self-esteem engendered

reciprocation on the part of the Agent, may lead to better results than contracts based on endogenous punishment or auditing.

The paper is organized as follows. In Section 2 we introduce and critically evaluate the strong reciprocity hypothesis. In Section 3 we present the utility function of an Agent who is cooperative-prone because she values self-esteem; we also illustrate how such utility function can lead either to cooperative or more traditional behaviour. In Section 4 we analyze a standard adverse selection problem by means of a Principal-Agent model with a cooperative-prone Agent and we compare the resulting solution based on gift giving with the different second best solutions obtained with auditing or endogenous punishment. Concluding remarks follow.

2. Strong reciprocity: experimental evidence and theoretical foundations

Individuals cooperate in many situations in which economic theory would not predict cooperation. Gintis (2004) convincingly argues that it is not possible to offer a theoretical explanation of observed cooperation that fulfils some reasonable conditions[1] while retaining the assumption that agents are strictly self-interested. Indeed, the latter is to be relaxed. To this end Gintis and Bowles take a clear stand in favour of strong reciprocity, "cooperation is maintained because many humans have a predisposition to punish those who violate group-beneficial norms, even when this reduces their fitness relative to other group members" (Bowles and Gintis 2003). The resulting human behaviour is called strong reciprocity and it is defined as an altruistic behaviour "conferring group benefits by promoting cooperation, while imposing upon the reciprocator the cost of punishing shirkers" (Bowles and Gintis 2003).

The distinguishing behaviour of strong reciprocators is the following: when they detect a defector they impose a punishment on her despite the fact that such behaviour is costly to them. The model allows for the possibility of private and imperfect signals of defections and demonstrates that whatever the cost of the punishment to the reciprocator and to the defector self-interested agents may find it convenient to cooperate provided that the fraction of strong reciprocators in the population is adequate.

Strong reciprocators are considered altruistic people, in so far as the privately bear the cost of an action which is of advantage to the whole community. The empirical relevance of strong reciprocators seems to be documented by many experimental studies (Fehr *et al*. 1997, Fehr and Gachter 2000).

[1] These conditions, as Gintis calls them are: Incentive compatibility, Dynamic stability, Empirical Relevance, Plausible Informational Requirements and Plausible Discount Factors.

More in general, there is large evidence about the existence and importance of strong reciprocity in situations involving public goods, common pool resource and in ultimatum games (Yamagishi 1986, Ostrom, Walker and Gardner 1992, Fehr and Gächter 2002).

Being a matter of voluntary choice and given the implied costs, it is not a trivial task to reconcile strong reciprocity with rationality. Unfortunately we still lack a general model of the rational foundations of strong reciprocity or of cooperation-prone behavior.

To this end it can be useful to start from Sen more recent criticism of the traditional "rational model" of choice that is articulated in three steps (Sen 2002, p. 34-35). The first is related to a notion of welfare which is self-centred, whereby "a person's welfare depends only on her own consumptions and other features of the richness of her life". The second criticism concerns what Sen calls self-welfare goal, i.e. the assumption that welfare maximization is the individual's only goal. The last criticism points to self-goal choice, according to which a person's choice are exclusively geared to the pursuit of her own goals.

Sen clearly aims at enriching the traditional model by weakening, in particular, the assumption that people pursue other goals then a too narrowly defined notion of welfare. But of the three criticisms he levels against the conventional wisdom the less convincing is precisely the last one, essentially because we are practically left without an operating theory of choice. If people, as Sen argues, are maximizers but they care also about things different from their welfare, how do they solve their maximization problem?

Sen does not say much on this and it is actually quite difficult to figure out which solution could be given to this problem.

The solution we propose is much in line with Sen's approach but departs form it in the assumption that people do maximize their utility function which is enriched with an endogenously determined "moral" variable. More precisely, individuals are endowed with a "moral system" which transforms their actions in self-esteem. The latter, as determined by such system, enters their utility function and concurs to define their choice within a utility maximizing process. Therefore, self-esteem brings about utility but its "amount" is determined by a "moral system" which lies outside the preference system sustaining the utility function.

In our definition a moral individual has a high propensity to destroy self esteem when his actions are not consistent with his moral values. This will be reflected in his final utility given that self-esteem is positively related to utility. Therefore his actions in so far as they destroy his own self-esteem through the "moral value mechanism" are not determined by a too restricted notion of self-welfare. In

this respect we share Sen's approach. However, the inclusion of self-esteem in the utility function (which could very well be defined a goal-function) allows us to treat the choice problem as a typical maximization problem and give formal solution to it.

The explanation we offer seems capable of capturing the most important features of experimental behaviour. In particular, it fits in well with the observed attempts to induce cooperation through a sort of *gift giving* - as in the famous essay by Akerlof – and also with the apparent existence of limits to cooperative behaviour.

In fact, in a much quoted experiment Fehr, Gächter and Kirchsteiger (1997) divided subjects into two sets of employers and employees and considered their interaction in a Principal-Agent framework. First, the found that many employers offered generous wages and received reciprocating higher efforts with the result of increasing both their and the employee's payoff. Secondly, they noticed that there existed, however, a relevant difference between the level of effort agreed and the level of effort applied. They observed that this was not the behaviour of a small group of fraudulent individuals, because only 26%, i.e. a small minority, of individual honoured their announcement.

Nonetheless, this evidence "is compatible with the notion that the employers are purely self-interested, since their beneficent behaviour vis-à-vis their employees was effective in increasing employer profits" (Gintis, Bowles, Boyd and Fehr 2003, p. 157). Allowing the possibility of employer to reward and punish employees, Fehr, Gächter, and Kirchsteiger observe an increase up to 40% of the bet payoff of all subjects[2]. Gintis, Bowles, Boyd and Fehr comment is that "the subjects who assume the role of employee conform to internalized standards of reciprocity, even when they know there are no material repercussions from behaving in a self-interested manner. Moreover, subjects who assume the role of employer expect this behaviour and are rewarded for acting accordingly. Finally, employers draw upon the internalized norm of rewarding good and punishing bad behaviour when they are permitted to punish and employees expect this behaviour and adjust their own effort levels accordingly" (Gintis, Bowles, Boyd and Fehr 2003, p. 157).

The above situation can be represented in a Principal-Agent framework where: it is in the Principal's interest to induce reciprocal behaviour by the Agent, and the Agent may choose to cooperate – even independently from any material punishment - because she is a social being, feels part of a "community" (altruism) and, at least up to a certain extent, will loose self-esteem if she does not cooperate. However – and this is an important point in a rationality based approach – such mechanism

[2] Employers punish fraudulent employees (68%), reward employees that over-fulfil their contracts (70%) and reward employees that honour their contracts (Gintis, Bowles, Boyd and Fehr 2003, p. 157).

will not work in any case and independently from an accurate consideration of the relevant costs and benefits. The loss of self-esteem implied by lack of cooperation is not always high enough to ensure unlimited cooperation. In fact, as recalled above, experiments give support to the idea that there are limits to cooperative behaviour.

We show, within a unique theoretical framework, that altruistic individuals do not necessarily choose cooperative behaviour. Indeed it is remarked that "strong reciprocators are inclined to compromise their morality to some extent" (Gintis, Bowles, Boyd and Fehr 2003). The approach we suggest seems capable of explaining what determines this willingness to compromise: much depends on the characteristics of agents' moral system and how self-esteem enters their utility functions. The latter are captured by a function which relates their self-esteem to the degree to which they reciprocate other agents' gifts (or punish other people's defections). It seems appropriate to talk of moral system in this case because what is relevant is the behaviour in itself and not the results it leads to.

3. Cooperation-prone agents: Moral system and utility function

Our analysis is set in a Principal-Agent framework which is broad enough to encompass all the interesting cases. We focus on the Agent and model her behaviour as determined by the interaction of two functions representing respectively how she creates or destroys self-esteem and how self-esteem impacts on her utility.

Indeed, the key assumption is that the Agents' utility depends on the monetary transfers (m) from the Principal – considered in comparison to the effort delivered - and on an endogenous variable, i.e. Self-Esteem (E).

More precisely Self-Esteem positively depends on cooperative behaviour of which strong reciprocity is an important example and negatively depends on the gift which the Agent receives from the Principal – the gift can be defined as the excess of the remuneration over the reservation price of the effort, that is to say with respect to the minimum which would induce the Agent to offer that effort. The assumption is that, as the gift gets larger, the Agent will suffer a loss of Self-Esteem, if she refrains from making an adequate effort. The behaviour of the agent is, therefore, the product of two antagonist forces: altruism and self-interest. Moreover the agent can have too little or too much Self-Esteem.

Above a subjectively given amount Self-Esteem becomes a bad (excess of a good) since it induces subjection and exploitation.[3] Therefore there exists a satiation point, which we label $\bar{E}$.

The dynamic of Self-Esteem can be represented by the differential equation: $\frac{dE}{dt} = Q(E) - \alpha(1 - f(e))m$, with $\alpha \geq 0$ a parameter that depends on social cultural habits (e*thics*) and $f(e) \geq 0$ is the agent's effort parameterized to the monetary transfer; Q is a logistic function bounded above[4], i.e. $\exists D : Q(E) \leq D \ \forall E$, $Q(E)$ is a strictly concave function and twice differentiable, for $G \in R_+ \exists \ E \geq G$ such that $Q(E) = 0 \ \forall E \leq G$, threshold effect, and above a level $\bar{E}$ the logistic function may be decreasing, i.e. $Q_E < 0$ for $E \geq \bar{E}$.

Compensation and the level of productivity can influence the utility of the Agent through both the direct and traditional channel and the indirect channel represented by her Self-Esteem. On the basis of these assumptions, it is also possible to identify the characteristics and the costs of a transaction based on trust and to establish whether they are lower than those of a contract with some type of penalties. We will return to this problem in the next section.

For the sake of simplicity it is assumed that $m=c$ that is the price of the composite good c equals one. Formally, the utility function U of the Agent is bounded from above:

$U=U(m,E)$, such that U may have satiation in E or:

$U_m > 0$ always

$U_E \geq 0$ if $E \leq \bar{E}$

$U_E < 0$ if $E > \bar{E}$

The agent's decision problem can be represented as follows:

$$Max U(m, E) \qquad [1]$$

such that:

$$\frac{dE}{dt} = Q(E) - \alpha(1 - f(e))m, \text{ for } \alpha > 0 \text{ and } f(e) \in [0,1] \qquad [2]$$

[3] Self-Esteem derives from feeling part of a community or a social group. If the agent obtains too big a transfer she feels like betraying her social nature and this will induce shame and disapproval. If the Self-Esteem is too high and the transfer too low, the Agent feels silly.

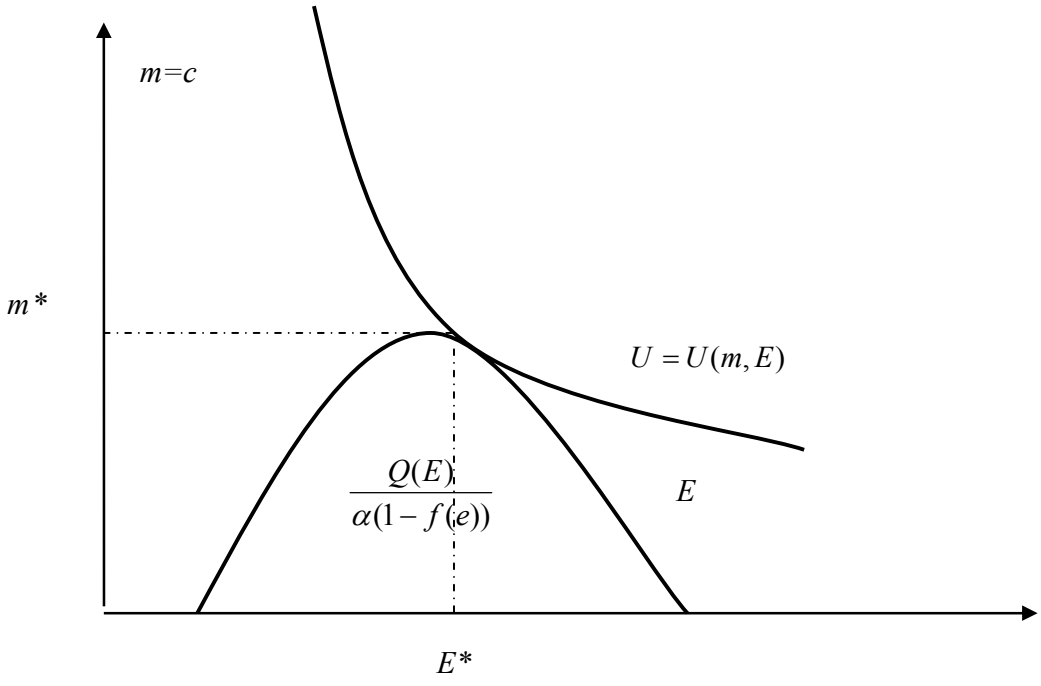
[4] Mathematically, the Self-Esteem is similar to a renewable factor (good) that reproduces itself by a Pearl and Verhulst function bounded above.

Considering the function $F = U(m,E) - \lambda[Q(E) - \alpha(1 - f(e))m]$, the first order conditions determine:

$$\frac{dF}{dm} = U_m + \lambda\alpha(1 - f(e)) = 0$$
$$\frac{dF}{dE} = U_E - \lambda Q_E = 0 \qquad [3]$$

Dividing the second equation for the first one, it is obtained $\frac{U_E}{Um} = -\frac{Q_E}{\alpha(1 - f(e))}$, that is the optimal solution for the Agent implies that her Marginal Rate of Substitution between monetary transfer and Self-Esteem equals her rate of (*technical*) transformation between them (*Fig. 1*). We are assuming that income can only increase by loosing Self-Esteem and monetary transfer is included into the agent's logistic function. As a result we obtain a function of re-production of Self-Esteem given its transformation in income and we summarize this relationship by a technical rate of transformation. Our problem is similar to the maximum long-run utility problem: $\max \lim_{t\to\infty} U(m_t, E_t)$, the solution of which requires finding the sustainable values of *m* and *E* and is characterized by the first order conditions $\frac{U_E}{Um} = -\frac{Q_E}{\alpha(1 - f(e))}$.

Figure 1. Agent Equilibrium



Graphically, the solution of the Agent's optimization problem is the tangency between her indifference curve and the *re-production function* of her Self-Esteem.

The interpretation of this condition is therefore the following: a maximizing individual will take her own moral values into account when making a choice. The moral values determine what has just been called the technical rate of transformation between self-esteem and monetary advantages, while the marginal rate of substitution represents how the individual is ready to trade off monetary benefits against higher self-esteem. Therefore the choice is the result of the working of the moral and pleasure mechanisms. Moral values dominate the self-esteem producing mechanism while pleasure or welfare mechanisms set the rate at which the two goals can be substituted for each other. It is important to stress that the moral mechanism endogenizes self-esteem and allows us to understand that a moral individual is not only she who gets pleasure from self esteem but also – and especially - she who has to behave properly in order to reproduce self-esteem.

Individuals are different from a moral point of view because they attach a different marginal utility to self-esteem or because they transform bad behaviour in a grater or smaller amount of lost self-esteem. Our model allows taking both aspects into account.

3. Trust and reciprocity in a Principal-Agent model with adverse selection

When a cooperation-prone individual enters a Principal-Agent relationship playing the role of the Agent, the Principal may rationally consider the possibility of turning this proneness to his own advantage by devising a contract that transforms it into an effective cooperative behavior. In order to achieve this result the Principal has to bear a cost (much of the gift-type envisaged by Akerlof 1982) which is borne in the expectations that the Agent will reciprocate. This may be taken as the cost of an implicit contract based on trust. In this sense, trust which creates cooperation is costly and endogenous. It is worth stressing that cooperation-proneness is not the same as effective cooperation. Unlike other approaches, ours draws a clear distinction between propensity to cooperation (that may be understood as a form of altruism) and effective cooperation.

In a previous paper (Basili-Duranti-Franzini, 2004) we have developed a model that allowed establishing under what conditions a contract based on trust may yield the Principal a higher return than alternative arrangements, like endogenous punishment or auditing. Building on that model we now consider how a cooperation-prone Agent may interfere with the choice of the best contract and how it could make the cooperative solution less costly. Our assumption on the utility function of the Agent

and the relevance of self-esteem has, therefore, an impact on traditional Principal-Agent models and may alter the relative convenience of different contractual arrangements.

Consider a Principal-Agent model in which the information asymmetry concerns the productivity of the Agent that could be high or low (efficient or inefficient Agent), giving rise to low or high marginal costs, respectively.[5]

Let θ_H be the constant marginal cost of the efficient Agent and θ_L the constant marginal cost of the inefficient Agent. Since the Principal cannot observe θ, he cannot equalise the marginal value of each Agent's production, $S'(q)$, to its marginal cost.

If he were to offer a contract calling for different compensation levels on the basis of the quantity produced and equal to the respective marginal benefit, the efficient Agent could simulate being inefficient (producing less) with a view to pocketing the information rent. The latter is equal to the difference between the two marginal costs at the low production levels or $\Delta\theta q_L$ and it is impossible to write a first best contract.

The Principal has to establish compensation levels by disregarding the equality between marginal benefit and marginal cost, or he has to define incentive and punishment mechanisms. In both cases he has to bear an additional cost with respect to the first best solution and, consequently, he will choose the less costly solution.

The problem of the Principal is that of maximising profit that is the difference between the value of production and the associated costs. Profit is assumed to be a linear function of the quantity produced q.

Let: $S(q_H)$ and $S(q_L)$ be the value of production obtained with the efficient and inefficient Agents;

m_H and m_L the transfer to the efficient and inefficient Agents;

θ_H and θ_L the marginal cost of the efficient and inefficient Agents;

$\Delta\theta$ the information rent;

$\Delta\theta q_L$ the value of the information rent;

v and $(1-v)$ the probability to come across an efficient or inefficient Agent, respectively;

s_L the probability of discovering the inefficient Agent's deception;

$\alpha(1-f(e_H))$ and $\alpha(1-f(e_L))$ indicate the rate of transformation of Self-Esteem into money for the efficient or inefficient Agent, respectively;

$c(s_L)$ the cost of auditing the inefficient Agent;

[5] To analyze this problem we rely on a standard situation of adverse selection, as modeled in Laffont-Martimort (2002).

P_H and P_L the amount of the endogenous punishment for the efficient and inefficient Agents. Given information asymmetry and adverse selection, the Principal's profit maximization problem can be written as follows:

$$\max_{\{q_L, q_H\}} \{v[S(q_H) - \theta_H q_H]\} + \{(1-v)[S(q_L) - \theta_L q_L]\} - v\Delta\theta q_L \quad [4]$$

such that:

(i) $m_H - \theta_H q_H \geq m_L - \theta_H q_L$

(ii) $m_L - \theta_L q_L \geq m_H - \theta_L q_H$

(iii) $U_H \geq 0$

(iv) $U_L \geq 0$

The solution of this problem induces the same production as first-best for the efficient Agent but a reduction with respect to first-best production for the inefficient Agent, with $S'(q_L) = \theta_L + \frac{v}{1-v}\Delta\theta$

Introducing an audit mechanism with an endogenous punishment (P) in the previous framework, other things being equal, the maximization problem of the Principal can be written as follows:

$$\max_{\{q_L, q_H, P_H, P_L\}} \{v[S(q_H) - \theta_H q_H - \Delta\theta q_L + s_L P_H]\} + \{(1-v)[S(q_L) - \theta_L q_L - c(s_L)]\} \quad [5]$$

such that:

(v) $P_H \leq \Delta\theta q_L$

The solution with endogenous punishment implies: the same production as first-best for the efficient Agent; a reduction with respect to first-best production for the inefficient Agent, with $S'(q_L^S) = \theta_L + \frac{v}{1-v}(1-s_L)\Delta\theta$. Crucially, only the inefficient Agent is monitored with a strictly positive probability.

It is worth noting that the loss of efficiency for the Principal is lower in the case of auditing with punishment, since he obtains the following:

$$S'(q_L^S) = \theta_L + \frac{v}{1-v}(1-s_L)\Delta\theta \leq S'(q_L) = \theta_L + \frac{v}{1-v}\Delta\theta \quad [6]$$

Consider the possibility of resorting to endogenous cooperation (i.e. cooperative behaviour or strong reciprocity) in order to induce the Agent to refrain from attempting to obtain all the information

rent. The Principal pays a *gift*, or incentive cost, equal to G_H and G_L for the efficient or inefficient Agent, that are cooperation-prone.

The Principal's maximization problem can be written as follows:

$$\max_{\{U_L, q_L, U_H, q_H, G_H, G_L\}} \{v[S(q_H) - \theta_H q_H - U_H]\} + \{(1-v)[S(q_L) - \theta_L q_L - U_L]\} \qquad [7]$$

such that:

(vi) $U_H = m_H - \theta_H q_H \geq m_L - \theta_H q_L - \alpha_H \Delta\theta q_L$

(vii) $U_L = m_L - \theta_L q_L \geq m_H - \theta_L q_H - \alpha_L \Delta\theta q_L$

(viii) $G_H \leq m_H - \theta_H q_H$

(ix) $G_L \leq m_L - \theta_L q_L$

(x) $U_H \geq 0$

(xi) $U_L \geq 0$

Inequalities (vi)-(vii) represent the incentive constraints for the high and low productivity Agent, respectively, inequalities (x)-(xi) are participation constraints, while (viii) and (ix) imply a non-negative gift for the two Agents.

If constrains (vi)-(xi) are both binding, $U_H = \Delta\theta q_L - \alpha(1 - f(e_H))\Delta\theta q_L$ for $\alpha(1 - f(e_H)) \geq 0$, and then (viii) can be re-written as follows:

(xii) $G_H \leq \Delta\theta q_L$

The problem [7] becomes:

$$\max_{\{q_L, q_H, G_H, G_L\}} \{v[S(q_H) - \theta_H q_H - (1 + (-\alpha(1 - f(e_H)))\Delta\theta q_L]\} + \{(1-v)[S(q_L) - \theta_L q_L]\} \qquad [8]$$

such that (xii) is binding.

The optimal contract implies that there is:

- no distortion with respect to the first-best solution for the efficient Agent;
- a downwards distortion is determined with respect to the first-best solution for the less efficient Agent, such that: $S'(q_L^A) = \theta_L + \frac{v}{1-v}(1 - \alpha(1 - f(e_H)))\Delta\theta$

The solution with induced cooperation has, naturally, second best characteristics due to the cost which it generates and only contemplates a gift for the efficient agent. Given the informative rent and the probability of crossing each kind of agent, such a cost depends on the Agent's rate of transformation of Self-Esteem into money.

Comparing the second-best solution for the less efficient Agent in the case of cooperation with the second-best solutions obtained with contracts without punishments [9] and with endogenous punishment in the event of the discovery of deception [10] and [11], it can be observed that:

$$S'(q_L^A) = \theta_L + \frac{v}{1-v}(1-\alpha(1-f(e_H))\Delta\theta \leq S'(q_L) = \theta_L + \frac{v}{1-v}\Delta\theta \qquad \textit{Always} \qquad [9]$$

$$S'(q_L^A) = \theta_L + \frac{v}{1-v}(1-\alpha(1-f(e_H))\Delta\theta \leq S'(q_L^S) = \theta_L + \frac{v}{1-v}(1-s_L)\Delta\theta$$

$$\textit{if } (1-\alpha(1-f(e_H)) \leq (1-s_L) \qquad [10]$$

$$S'(q_L^A) = \theta_L + \frac{v}{1-v}(1-\alpha(1-f(e_H))\Delta\theta > S'(q_L^S) = \theta_L + \frac{v}{1-v}(1-s_L)\Delta\theta$$

$$\textit{if } (1-\alpha(1-f(e_H)) > (1-s_L) \qquad [11]$$

Therefore cooperation is chosen (rejected) if the necessary cost of activating altruism $(1-\alpha(1-f(e_H))$ of the more efficient Agent is lower (greater) than the probability $(1-s_L)$ of the loss connected to the non-punishment (exposure) of the less efficient Agent in the event of fraudulent behaviour.

Finally, consider both altruism and endogenous punishment *P* and assume that the probability of auditing Agents behaviour depends on their rate of transformation of Self-Esteem into money, that is $s_{\alpha H} = s(\alpha(1-f(e_H))$ and $s_{\alpha L} = s(\alpha(1-f(e_L))$. Since the cost of auditing *c* depends on probability of discovering a fraudulent behaviour $s_{\alpha i}$, for *i=L,H* the problem [8] becomes, other things being equal:

$$\max_{\{q_L, q_H, G_H, G_L\}} \{v[S(q_H) - \theta_H q_H - (1+(-\alpha(1-f(e_H)))\Delta\theta q_L + s_{\alpha L} P_H]\} + \{(1-v)[S(q_L) - \theta_L q_L - c(s_{\alpha L})]\} \qquad [12]$$

Such that (v) and (xii) are binding. The optimal contract implies that there is:

- no distortion with respect to the first-best solution for the efficient Agent;
- a downwards distortion is determined with respect to the first-best solution for the less efficient Agent, such that: $S'(q_L^A) = \theta_L + \frac{v}{1-v}[(1-a(1-f(e_H)) - s_{aL}]\Delta\theta$

This model implies that if the Agent's rate of transformation of Self-Esteem into money is large (small), transfers should be cheap (expensive). It is reasonable (i.e. Gintis, Bowles, Boyd and Fehr 2003) to assume that induced cooperative behaviour reduces the inefficient Agent's attitude to deflect from agreement and increases the probability of discovering her fraudulent behaviour, that is $(s_{aL})>(s_L)$ with a lower cost, by effect of Self-Esteem, that is $c(s_{aL})<c(s_L)$. In any case even assuming

the same probability of discovering the fraudulent behaviour of the less efficient Agent is the same in the case of endogenous punishment with or without trust, that is $s_{aL} = s_L$ and $c(s_{aL})=c(s_L)$, it will be that:

$$S'(q_L^A) = \theta_L + \frac{v}{1-v}[(1-\alpha(1-f(e_H)) - s_{\alpha L}]\Delta\theta \leq S'(q_L^A) = \theta_L + \frac{v}{1-v}(1-s_L)\Delta\theta \quad Always \qquad [13]$$

$$S'(q_L^A) = \theta_L + \frac{v}{1-v}[(1-\alpha(1-f(e_H)) - s_{\alpha L}]\Delta\theta \leq S'(q_L^A) = \theta_L + \frac{v}{1-v}(1-\alpha(1-f(e_H))\Delta\theta$$

$$if \quad s_{\alpha L} \geq 0 \qquad [14]$$

The existence of strong reciprocity and endogenous punishment permits the Principal to implement trust and cooperation at the lowest cost.

4. Concluding remarks

Standard economic theory is undeniably too pessimistic as to the possibility of cooperation among *strangers* (Seabright 2004). Genetic relatedness is not the only condition for cooperation to develop in situations where self-interest would make destructive opportunism the best course of action. A huge bulk of evidence can be invoked to this end. In particular, as Bowles and Gintis have argued, many humans seem to adhere to a strong reciprocity rule of behaviour that implies the bearing of a personal cost in order to punish those members of the community who defect from cooperation. However, the analytical foundation of this type of cooperation-prone behaviour and how it relates to rationality have not been yet spelled out. In this paper we have advanced our own explanation relying on the notion of self-esteem and modelling cooperation-prone agents in terms both of a *moral* function transforming cooperation into self-esteem and of a utility function which includes self-esteem in its argument. On the basis of this model we have drawn a clear distinction between propensity to cooperation, on the one hand, and effective cooperation, on the other, which are too often muddled. We have also shown the impact of our hypothesis on the best contract a Principal can offer in a typical Principal-Agent situation and how it can help understand the role that gift giving and endogenous trust can play to counter opportunism.

Our results support to the idea that the presence of altruists may lead to cooperative solution because it can make the eliciting of cooperation the best strategy for standard self-interested agents. Moreover there are good reasons to believe that altruists will not be dominated by self-interested people in the evolutionary game.

Finally, in our definition a moral individual has a high propensity to destroy self-esteem when her actions are not consistent with her moral values. This will be reflected in her final utility given that self-esteem is positively related to utility. Therefore her actions in so far as they destroy her own self-esteem through the moral value mechanism are not determined by a too restricted notion of self-welfare.

Interesting enough our approach is coherent with the most recent Sen's criticism of the standard rational model of choice based on the notion of self-centred welfare, that is a system in which: a person's welfare depends only on her own consumptions and other features of the richness of her life, the welfare maximization is the individual's only goal and an individual choices are exclusively geared to the pursuit of selfish goals (Sen 2002, p. 34-35).

However, differently from Sen we retain an operating theory of choice that makes people able to behave as maximizers, particularly with respect to endogenously determined moral variable. More precisely, individuals are endowed with a *moral system* which transforms their actions into self-esteem. The latter, as determined by such system, enters their utility function and concurs to define their choice within a utility maximizing process. Therefore, self-esteem brings about utility but its amount is determined by a moral system which *lies outside* individual preference system. Eventually, the inclusion of self-esteem in the utility function (which could very well be defined a goal-function) allows us to treat the choice problem as a typical maximization problem and give formal solution to it.

References

Akerlof, G. (1982), Labour Contracts as Partial Gift Exchange, *Quarterly Journal of Economics* 97, pp. 543-69.

Basili, M., Duranti, C., Franzini, M.(2004), Networks, trust and institutional complementarities, *Rivista di Politica Economica* 1-2, 159-180.

Bowles, S., Gintis, H. (2003), The evolution of strong reciprocity: cooperation in heterogeneous populations, *Theoretical Population Biology* (forthcoming).

Fehr, E., Gächter, S., Kirchsteiger, G. (1997), Reciprocity as a contract enforcement device: experimental evidence, *Econometrica* 65, 833–860.

Fehr, E., Gächter, S. (2000), Cooperation and punishment, *American Economic Review* 90, 980–994.

Fehr, E., Gächter, S. (2002), Altruistic punishment in Humans, *Nature* 415, 137–140.

Gintis, H., Bowles, S., Boyd, R., Fehr, E. (2003), Explaining altruistic behavior in humans, *Evolution and Human Behavior* 24, 153-172.

Gintis, H., 2004. Modelling cooperation among self-interested agents: a critique. mimeo, December.

Laffont, J.J., and Martimort, D. (2002), *The theory of Incentives*, University Press, Princeton.

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

Seabright, P. (2004), *The Company of Strangers. A Natural History of Economic Life*, University Press, Princeton

Sen, A. (2002), *Rationality and freedom*, Belknap Press, Cambridge Ma.

Yamagishi, T. (1986), The provision of a sanctioning system as a public good, *Journal of Personality and Social Psychology* 51, 110–116.

Pubblicato in proprio
Dipartimento di Economia Pubblica
Facoltà di Economia
Università degli Studi di Roma “La Sapienza”
Via del Castro Laurenziano 9 – 00161 Roma

Working Paper del Dipartimento di Economia Pubblica
Università degli studi di Roma "La Sapienza"
Via del Castro Laurenziano 9 – 00161 Roma

I Working Paper vengono pubblicati per favorire la tempestiva divulgazione, in forma provvisoria o definitiva, dei risultati delle ricerche sulla teoria e la politica economica.
La pubblicazione dei lavori è soggetta all'approvazione del Comitato Scientifico, sentito il parere di un referee.

I Working Paper del Dipartimento di Economia Pubblica ottemperano agli obblighi previsti dall'art. 1 del D.L.: 31.8.45 n. 660.