E C O N O M I C S   B U L L E T I N

Can Altruism Hinder Cooperation?

Keisuke Nakao
College of Business and Economics, University of Hawaii, Hilo

Abstract

This note considers the effects of altruism on cooperation in the context of a repeated Prisoner's Dilemma. Altruism has two conflicting impacts on cooperation: One is to reduce the temptation for defection; the other is to make the future punishment ineffective. The total effect of altruism hinges on the shape of players' cost functions.

I am deeply grateful to Dilip Mookherjee for helpful discussions. I would also like to thank Barton Lipman, Zvika Neeman, Masatoshi Tsumagari, and participants at the Boston University micro theory workshop for helpful comments. All errors are my own. Summer research grant from the Institute of Economic Development, Boston University is greatly acknowledged.
Citation: Nakao, Keisuke, (2008) "Can Altruism Hinder Cooperation?." *Economics Bulletin,* Vol. 4, No. 26 pp. 1-6
Submitted: July 27, 2008. **Accepted:** August 15, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume4/EB-08D70029A.pdf

1. Introduction

It is broadly acknowledged that altruism can increase the benefits of social interactions by reducing the incentive for free riding (e.g., Axelrod 1984, 72; Andreoni 1990; Andreoni and Miller 1993; Fehr and Gachter 2002). These articles argue that altruism improves cooperative performances. In contrast, Bernheim and Stark (1988) described the detrimental effect of altruism which generates an enforcement problem; altruism may reduce the effectiveness of credible punishment.[1] Recent empirical findings from microcredit seem to confirm their prediction: a group of borrowers with strong social connectedness is more likely to default (Guinnane 1994; Ahlin and Townsend 2007).[2] How can these two conflicting forces of altruism coexist? Under what condition one of them dominates the other?

I address these questions by constructing a model of repeated game theory with a continuous version of Prisoner's Dilemma. Here are obtained results from the model. (a) Altruism may have two conflicting effects on cooperation. One is to moderate the temptation to betray others, as argued in the standard literature of altruism above. The other is to reduce the effectiveness of punishment because it is not altruist's interest to punish a deviant from cooperation. The total effect of altruism hinges on which effect dominates. (b) I further derive conditions for altruism to facilitate cooperation and to hinder it. The total effect of altruism depends on the shape of players' payoff functions. More precisely, if the marginal cost does not increase rapidly, altruism significantly reduces the effectiveness of punishment and thus creates a severe enforcement problem. (c) The effect of altruism might be non-monotonic. In such a case, social cooperation is most difficult in the middle range of altruism. Without altruism, people can make credible threats of punishment, and with sufficient altruism, there is no conflict of interests among people.

2. The Model
2.1 Stage Game

There are two players in the game. Each player $i \in \{1, 2\}$ chooses action $a_i \geq 0$. Player i's payoff in the stage game $\pi_i(a_1, a_2, \alpha)$ is a weighted average of two players' material payoffs:

$$\pi_i(a_1, a_2, \alpha) \equiv (1-\alpha)g_i(a_1, a_2) + \alpha g_j(a_1, a_2),$$

[1] Lindbeck and Weibull (1988) provided another (time-inconsistency) problem caused by altruism: Expecting endowment from altruistic parents in the future, a child may overconsume today, leading to a socially inefficient outcome.

[2] Guinnane (1994) compared financial markets of Germany and Ireland in the late nineteenth century and reported that a group lending system which was very successful in rural areas of Germany failed to be successfully transplanted to Ireland. He reasoned that if borrowers were socially closed to each other, their implicit threats to penalize defaulters would be weakened. Ahlin and Townsend (2007) found that the impact of altruism on loan repayment of group lending in financial markets of Thailand is negative and is statistically significant. These findings suggest that social cooperation might be less successful among people with altruistic concerns.

where $\alpha \in [0, 1/2]$ denotes the degree of players' altruism. When $\alpha = 0$, players are perfectly self-interested, and each player does not care about the other's interest at all. When $\alpha = 1/2$, they are perfectly altruistic, and their preferences exactly coincide. For simplicity, I consider only cases in which two players have the same degree of altruism (symmetric altruism).

The material payoff function $g_i(a_1, a_2)$ is given as

$$g_i(a_1, a_2) \equiv a_j - c(a_i),$$

where $c(a)$ is the cost function which has the following properties: $c(0) = 0$, $c'(0) = 0$, $\lim_{a \to \infty} c'(a) = \infty$, $c'(a) > 0$ for $a > 0$ and $c''(a) > 0$ for $a > 0$ (strictly convex). One can imagine that each player i gives the amount a_i of gifts to the other player j at cost $c(a_i)$. Thus, the stage game has the Prisoner's Dilemma structure: unless players are perfectly altruistic ($\alpha = 1/2$), the pair of dominant strategies by the two players constitute a Pareto suboptimal outcome.

2.2 Repeated Game

Consider infinite interaction between the two players. Player i's average payoff in the repeated game is described as $(1 - \delta) \sum_{t=0}^{\infty} \delta^t \pi_{i,t}(a_{1,t}, a_{2,t}, \alpha)$, where δ is the rate at which future payoffs are discounted.

3. Results

3.1 Altruistic Cooperation without Punishments

In a Nash equilibrium of the stage game, each player maximizes his payoff, given the other player's action:

$$\max_{a_i} \{(1 - \alpha) [a_j - c(a_i)] + \alpha [a_i - c(a_j)]\} \text{ for } i = 1, 2.$$

By first-order condition,

$$c' \left(a^{NE}(\alpha)\right) = \frac{\alpha}{1 - \alpha},$$

where $a^{NE}(\alpha)$ is an action in the equilibrium.

Lemma 1 *$a^{NE}(\alpha)$ is the unique dominant strategy in the stage game.*

Proof. $a^{NE}(\alpha)$ is unique for given α because $c(a)$ is strictly convex and $c'(a)$ is unbounded from above. In addition, $a^{NE}(\alpha)$ is independent of the other player's action. Thus, $a^{NE}(\alpha)$ is the unique dominant strategy. ■

The pair of actions $\left(a^{NE}(\alpha), a^{NE}(\alpha)\right)$ forms the unique Nash equilibrium. Define the payoff in the equilibrium to be $P(\alpha) \equiv a^{NE}(\alpha) - c\left(a^{NE}(\alpha)\right)$.

Lemma 2 *$a^{NE}(\alpha)$ monotonically increases in α, and so does $P(\alpha)$. That is, altruism facilitates cooperation in the stage game.*

Proof. It is immediate that

$$\frac{\partial a^{NE}(\alpha)}{\partial \alpha} = \frac{1}{(1-\alpha)^2} \frac{1}{c''\left(a^{NE}(\alpha)\right)} > 0, \tag{1}$$

and

$$\frac{\partial P(\alpha)}{\partial \alpha} = \frac{1-2\alpha}{(1-\alpha)^3} \frac{1}{c''\left(a^{NE}(\alpha)\right)} > 0. \tag{2}$$

■

As players become more altruistic, they cooperate more in the sense that the level of actions in equilibrium becomes higher. When $\alpha = 1/2$, the conflict of interests between players disappears, and they cooperate most.

3.2 Altruistic Cooperation with Punishments

Following Bernheim and Stark (1988), I restrict attention to subgame perfect Nash equilibrium with Nash reversion in the repeated game. Suppose that both players are taking the same action $a^+ > a^{NE}(\alpha)$ and receiving the payoff $G(a^+) \equiv a^+ - c(a^+) > P(\alpha)$. If a player (say, player 1) cheats the other player (player 2) by choosing the dominant action $a^{NE}(\alpha)$ instead of a^+, then both players will take the action $a^{NE}(\alpha)$, ending up with lower payoff $P(\alpha)$ for all the subsequent periods. I define the (player 1's) temptation payoff by cheating the other to be $T(a^+, \alpha) \equiv \pi_1(a^{NE}(\alpha), a^+, \alpha)$. This temptation is affected by altruism as follows.

Lemma 3 *$T(a^+, \alpha)$ monotonically decreases in α; i.e., altruism reduces the temptation to cheat the other.*

Proof. By taking the derivative of $T(a^+, \alpha)$, we have

$$\frac{\partial T(a^+, \alpha)}{\partial \alpha} = -(g_1(a^{NE}(\alpha), a^+) - g_2(a^{NE}(\alpha), a^+)) < 0. \tag{3}$$

■

The incentive constraint for the action profile (a^+, a^+) to be sustained as subgame perfect Nash equilibrium with "Nash-threat" punishments is:

$$G(a^+) \geq (1-\delta)T(a^+, \alpha) + \delta P(\alpha). \tag{4}$$

Putting Condition (4) in another way, the discount factor must be large enough such that:

$$\delta \geq \overline{\delta}(a^+, \alpha) \equiv \frac{T(a^+, \alpha) - G(a^+)}{T(a^+, \alpha) - P(\alpha)}.$$

This implies that players must be patient enough that they are afraid of future punishments.

3.3 Conflicting Effects of Altruism

Lemmas 2 and 3 with Condition (4) imply that altruism has two conflicting effects on cooperation: One is to make the future punishment ineffective; the other is to reduce the temptation to defect. The total effect of altruism depends on which force is dominant to the other.

Proposition 1 *Altruism makes cooperation less sustainable (i.e., $\overline{\delta}(a^+,\alpha)$ increases in α) if and only if its effect on punishment dominates its effect on temptation in the sense that*

$$\frac{\frac{\partial P(\alpha)}{\partial \alpha}}{G(a^+)-P(\alpha)} > \frac{-\frac{\partial T(a^+,\alpha)}{\partial \alpha}}{T(a^+,\alpha)-G(a^+)}. \tag{5}$$

Proof. By taking the derivative of $\overline{\delta}(a^+,\alpha)$ with respect to α, we have

$$\frac{\partial \overline{\delta}(a^+,\alpha)}{\partial \alpha} = \frac{\frac{\partial T(a^+,\alpha)}{\partial \alpha}(T(a^+,\alpha)-P(\alpha)) - (T(a^+,\alpha)-G(a^+))\left(\frac{\partial T(a^+,\alpha)}{\partial \alpha} - \frac{\partial P(\alpha)}{\partial \alpha}\right)}{(T(a^+,\alpha)-P(\alpha))^2},$$

which is positive if and only if Condition (5) holds. ■

In Condition (5), $G(a^+)-P(\alpha)$ and $T(a^+,\alpha)-G(a^+)$ denote the strength of punishment and that of temptation, respectively. Thus, Condition (5) compares the percentage change in the effectiveness of punishment with the percentage change in the incentive for temptation.

The following corollary shows that the shape of cost function $c(a)$ matters for the effect of altruism on cooperation.

Corollary 1 *A small rise in α from $\alpha = 0$ increases $\overline{\delta}(a^+,\alpha)$ if*

$$\lim_{a \to +0} c''(a) = 0. \tag{6}$$

Proof. In Condition (5), $G(a^+)-P(\alpha)$, $T(a^+,\alpha)-G(a^+)$ and $\frac{\partial T(a^+,\alpha)}{\partial \alpha}$ are all finite (see Equation (3)), but $\frac{\partial P(\alpha)}{\partial \alpha}$ is not necessarily so: $\frac{\partial P(\alpha)}{\partial \alpha} = \frac{1-2\alpha}{(1-\alpha)^3}\frac{1}{c''(a^{NE}(\alpha))}$ (Equation 2). Since $a^{NE}(\alpha)$ for small α, if the term $c''(a)$ is small enough with small a, the left hand side of Condition (5) becomes very large, and the condition surely holds. ■

Condition (6) means that the marginal cost rises very slowly or the *slope* of marginal cost is very flat for low levels of actions. If the marginal cost increases in the level of action very slowly, then altruism substantively deprives players of hard punishment, and as a result, players are more induced to deviate from cooperation.

The next example is the case that the cost function is the power function.

Example 1 *Assume $c(a) = a^l$ with $l > 1$. If $l > 2$, then a small rise in α increases $\overline{\delta}(a^+,\alpha)$.*

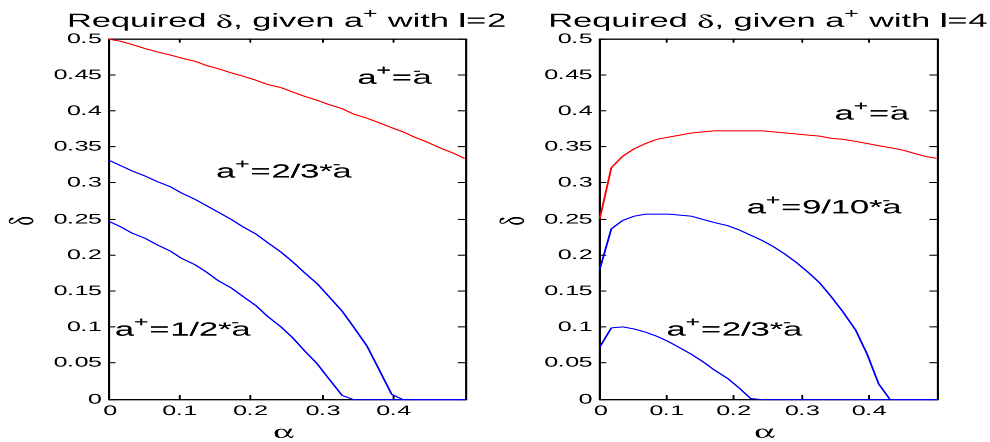

Figure 1: Required discount factors for various levels of action and altruism.

Put differently, when $l > 2$, Condition 5 is satisfied, and altruistic preferences require the larger δ to enforce cooperation. Figure 1 shows the relationship between required discount factors $\overline{\delta}(a^+, \alpha)$ and altruism α for $l = 2, 4$, given levels of actions a^+. It can be seen that with $l = 2$, altruism monotonically lowers $\overline{\delta}(a^+, \alpha)$ and it makes cooperation easier. In contrast, with $l = 4$, the effect of altruism α on required discount factor $\overline{\delta}(a^+, \alpha)$ is non-monotonic. The figure implies that cooperation is most difficult with a middle range of altruism. Cooperation is relatively easy for small α, because players can create effective punishment. It is also easy for sufficiently large α because there is no conflict of interests. This example confirms that altruism may make cooperation more difficult to enforce in a certain circumstance.

4. Conclusion

This note shows two conflicting effects of altruism on cooperation: one is to moderate the temptation for cheating, and the other to alleviate the effectiveness of punishment on a defector. The total effect of altruism hinges on the shape of cost functions. If the marginal cost of cooperation is sufficiently low, altruism significantly reduces the effectiveness of punishment, and altruism can hinder cooperation.

References

Ahlin, C. and R. Townsend (2007) "Using Repayment Data to Test Across Models of Joint Liability Lending," *Economic Journal* **117**: F11-F51.

Andreoni, J. (1990) "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *Economic Journal*, **100**: 464-477.

Andreoni, J. and J. H. Miller. (1993) "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma," *Economic Journal*, **103**: 570-585.

Axelrod, R. (1984) *The Evolution of Cooperation*, Basic books: New York.

Bernheim, B. D. and O. Stark (1988) "Altruism with the Family Reconsidered: Do Nice Guys Finish Last?" *American Economic Review* **78**: 1034-1045.

Fehr, E., and S. Gachter (2002) "Altruistic Punishment in Humans," *Nature*, **415**: 137-140.

Guinnane, T. W. (1994) "A Failed Institutional Transplant: Raiffeisen's Credit Cooperatives in Ireland, 1894-1914," *Explorations in Economic History* **31**: 38-61.

Lindbeck, A. and J. W. Weibull (1988) "Altruism and Time Consistency: The Economics of Fait Accompli," *Journal of Political Economy* **96**: 1165-1182.