Explaining Private Provision of Public Goods by Conditional Cooperation - An Evolutionary Approach - *

Maria Vittoria Levati[a]

November 2002

Abstract

We adopt an evolutionary approach to investigate whether and when conditional cooperation can explain the voluntary contribution phenomenon often observed in public goods experiments and in real life. Formally, conditional cooperation is captured by a regret parameter describing how much an individual regrets to contribute less than the average. We find that the evolutionary stability of conditional cooperation depends on what is known about the (individual) regret parameters of other group members.

Keywords: Public goods game, conditional cooperation, evolutionary stability, informational costs

JEL-Classification: A13, C72, D82, H41

* The author gratefully acknowledges helpful comments by Werner Güth and Axel Ockenfels.

a Max Planck Institute for Research into Economic Systems, Strategic Interaction Group, Kahlaische Str. 10, D-07745 Jena, Germany. Tel.: +49/3641/686-629, Fax: +49/3641/686-666, E-mail: levati@mpiew-jena.mpg.de

I. Introduction

A public good has two essential attributes: non-rivalry in consumption and non-excludability. The former characteristic refers to the possibility of simultaneous consumption by multiple consumers. The other characteristic means that it is difficult to prevent consumption of the good by those who fail to pay. Hence, when the public good must be financed by private arrangements, any individual faces a monetary incentive to free ride on the contributions of the others. Although, according to economic theory, this should lead to a Pareto inefficient provision of the public good, economists who study public goods in the laboratory observe that a significant number of people contribute to a degree greater than what would be implied by pure self-interest.[1]

A plausible explanation for the observed contribution behavior is the assumption that there are conditional cooperators, i.e. people who decide whether (and how much) to contribute being driven by a relevant reference contribution of the group rather than by opportunistically rational reasoning.[2] Such a relevant reference can be either the average contribution of the other group members or an extremal (i.e., the minimal or maximal) contribution of the group.

Numerous experiments support the idea that behavior is geared towards the average contribution of the other group members (see, for example, Keser, 1997; Croson, 1998; Sonnemans et al., 1999; Fischbacher et al., 2000; Keser and van Winden, 2000; and Brandts and Schram, 2001). On the other hand, caring to contribute at least as much as the minimal contribution of the others has been investigated experimentally by Levati and Neugebauer (2001), who use a clock mechanism to link an agent's contribution to the minimum contribution of the group. Although their mechanism should theoretically induce full contribution in groups of conditional cooperators, the presence of some opportunistic individuals (i.e., individuals only interested in own monetary reward) prevented the achievement of the efficient outcome.

1 See Davis and Holt (1993) and Ledyard (1995) for extensive surveys of experimental studies on voluntary contributions to public goods.

2 As Fischbacher et al. (2000, p. 397) suggest, conditional cooperation can be considered as a motivation in its own or be a consequence of some fairness preferences like altruism, warm-glow, inequity aversion or reciprocity. These "non standard-motivations" have received a lot of attention in the recent literature as explanation for cooperative behavior (e.g., Sugden, 1984; Andreoni, 1995; Palfrey and Prisbey, 1997; Anderson et al. 1998; Fehr and Schmidt, 1999; and Bolton and Ockenfels, 2000).

The observation that conditionally cooperative behavior plays an important role in explaining voluntary contribution in public goods settings (although of considerable interest in itself) does not reveal how it could evolve and prevail among human beings. Actually, ever since Darwin (1859) there exists a general suspicion that the fight for reproductive success must have driven out any non-competitive disposition. According to this view, biological competition should have educated human beings to behave competitively, in the sense of exploiting the cooperative behavior of their fellow humans. In other words, social cooperation should not be possible unless some external coercive power intervenes. Of course, such a (Hobbesian) view[3] contradicts not only the results of many controlled laboratory experiments[4] but also our ordinary life experiences. We observe daily that (besides those who exploit others) there are individuals who are willing to cooperate at least when they observe others' willingness to do so.

In this paper we will try to reach a deeper understanding of how and to what extent conditional cooperation might have evolved. In particular, we address the issue of evolutionary stability of conditionally cooperative behavior facilitating voluntary contributions among rational actors. Whereas in evolutionary biology (Maynard Smith, 1982) a genotype is assumed to determine behavior, we rely on an 'indirect evolutionary approach'. According to this approach, genetic evolution determines preferences and players behave rationally with respect to any given preferences (Güth and Yaari, 1992). Güth and Nitzan (1997) apply this methodology to a multi-person public goods game to investigate whether, also in the context of voluntary provision of public goods (but without the possibility of conditioning on others' behavior), moral objections against free riding are evolutionarily stable. Within a finite population framework, their main conclusion is that the emergence of an effective social conscience preventing free riding is generally not evolutionarily stable with just one exception, namely unanimity games (where the public good is produced only when everybody contributes, which is a strong form of conditional cooperation). Within an infinite population framework, Güth and Nitzan find that the evolutionarily stable probability that an individual develops social conscience is positive. Given these earlier mixed results, it is tempting to further explore public goods games and investigate whether and when conditional cooperation is evolutionarily stable.

3 Cf. Hobbes (1651). See also Parsons (1968).

4 Fehr and Gächter (2002), for instance, show experimentally that cooperation flourishes if altruistic punishment is possible, i.e. if individuals can punish free riders even though this is costly and yields no material benefits for the punishers.

The next section formally introduces the public goods game. Then, in Section III, this game is analyzed from an evolutionary point of view. It is shown that conditional cooperation can prevail in evolutionarily stable ways if players know a priori other players' types (III.1). On the other hand, conditional cooperation will not be evolutionarily stable if individuals, although knowing the distribution of types in the population, cannot identify the type of the other players (III.2). Afterwards our analysis of the two previous extreme cases is extended to an intermediate case in which specific information about the others' types is available at a cost (III.3). Some final observations and a general discussion of the results conclude the paper.

II. The model

The basic public goods game (as introduced by Isaac et al., 1984) relies on the player set: $N = \{1, \cdots, n\}$ with $n \in \mathbb{N}$, $n \geq 2$, and strategies/contributions c_i satisfying $0 \leq c_i \leq e$, where $e > 0$ is the initial endowment which is assumed to be identical for all $i \in N$.

For every contribution vector $\mathbf{c} = (c_1, \cdots, c_n)$ the monetary success of player $i \in N$ is defined as

$$v_i(\mathbf{c}) = e - c_i + \alpha\bar{c} \quad \text{with} \quad 0 < \alpha/n < 1 < \alpha$$

where $\bar{c}$ is the average contribution: $\bar{c} = \sum_{j=1}^{n} c_j/n$.

Due to $\alpha/n < 1$, the dominant strategy is to contribute nothing. Since $\alpha > 1$ the efficient outcome (maximizing the sum of $v_i(\mathbf{c})$ over $i \in N$) is, however, to contribute everything.

In this framework, a tendency to conditionally cooperate can be captured by regretting to contribute less than the average.[5] What motivates player i is then not only $v_i(\mathbf{c})$ but also this kind of regret. Along the lines of Fehr and Schmidt's (1999) and Bolton and Ockenfels' (2000) models, we formulate such a broader motivation as:

$$u_i(\mathbf{c}) = v_i(\mathbf{c}) - r_i \max\{0, \bar{c} - c_i\} \quad \text{with} \quad r_i \geq 0 \quad \text{for } i \in N.$$

5 An alternative reference point for a conditional cooperator may be the minimal contribution of his/her group. This has been suggested by Levati and Neugebauer (2001). To further strengthen the conclusions of the present study, the analysis of evolutionary stability presented here was also conducted using the others' minimum contribution (rather than the average) as relevant reference what, however, never yielded any substantial difference.

In the above formula, contributing more than the average is never optimal (regardless of r_i).[6] Furthermore, if

$$(1) \quad r_i > \frac{n-\alpha}{n-1},$$

also contributing less than the average is no longer optimal.[7] Specifically, every uniform contribution vector:

$$\mathbf{c} = \{k, \cdots, k\} \quad \text{with} \quad 0 \leq k \leq e \quad \forall i \in N$$

is a strict equilibrium[8] of the game with payoff functions $u_i(\cdot)$ satisfying (1) for all $i \in N$.

If, however, r_i is small, in the sense that

$$(2) \quad 0 \leq r_i < \frac{n-\alpha}{n-1} \quad \text{for } i \in N,$$

the old inefficient solution of free riding still holds: Whatever the others' contribution, player i's best response is $c_i^*(\mathbf{c}) = 0$. The degenerate case of $r_i = \frac{n-\alpha}{n-1}$ can be neglected since it will not matter for our evolutionary analysis.

In the following we will try to derive in an evolutionary set up whether and under which circumstances conditional cooperation, in the sense of behavior driven by (1), can be expected to prevail.

III. The evolutionary set up

In the tradition of indirect evolution (Güth and Yaari, 1992) we distinguish between:

- individual success $v_i(\cdot)$ determining the evolutionary selection of r_i-types over time, and
- individual payoff or utility $u_i(\cdot)$ guiding the choice of c_i in the public goods game.[9]

6 If $c_i > \bar{c}$, then $u_i(\mathbf{c}) = v_i(\mathbf{c})$ and i's dominant strategy is to free ride.

7 Suppose that $c_i < \bar{c}$. Then the derivative of $u_i(\mathbf{c})$ with respect to c_i is positive if and only if $-1 + \frac{\alpha}{n} - \frac{r_i}{n} + r_i > 0$. What implies $n - \alpha < r_i(n-1)$ or equivalently condition (1).

8 That is, a combination of strategies where every player suffers from a unilateral deviation.

9 The existence of different r_i-types is crucial in transforming the conventional direct into an indirect evolutionary approach. Different values of r_i, which in themselves do not measure differ-

Observing that the behavior of different r_i-types depends on whether $r_i > (n-\alpha)/(n-1)$ or $r_i < (n-\alpha)/(n-1)$, we consider an evolutionary game with mutant or strategy space:

$$M = \{\underline{r}, \bar{r}\} \quad \text{with} \quad 0 \leq \underline{r} < \tfrac{n-\alpha}{n-1} < \bar{r}.$$

Thus the mutants $\underline{r}$ and $\bar{r}$ can be described as an $\underline{r}$-type who prefers free riding since regret is too weak to prevent him/her from contributing less than the average and an $\bar{r}$-type who, regretting this, prefers contributing the average amount. We refer to the $\bar{r}$-types as the conditionally cooperative players.

Throughout our analysis we shall assume symmetry with respect to players of the same type. That is, we assume that all players i of the same r_i-type (with $r_i \in M$) will be making the same choices.

For any constellation $\mathbf{r} = (r_1, \cdots, r_n)$ with $r_i \in M$ for $i = 1, \cdots, n$, the evolutionary success of type i is the individual success

$$v_i(\mathbf{c}^*) = e - c_i^*(\mathbf{c}^*|r_i) + \alpha\bar{c}^*$$

where $\mathbf{c}^*$ is the (symmetric) solution of the public goods game for the type vector $\mathbf{r}$. If there exist $\underline{r}$-type players in N, they will behave opportunistically and contribute zero. But then, for m $(< n)$ $\bar{r}$-type players contributing a positive amount, player i with $r_i = \bar{r}$ also prefers to contribute nothing since what (s)he would get by contributing one more unit (i.e. α/n) is less than what (s)he would lose by doing so (i.e. 1). What proves:

REMARK 1 *If $r_i = \underline{r}$ for at least one player $i \in N$, then the unique solution of the public goods game is $\mathbf{c}^* = (0, \cdots, 0)$ regardless of $\mathbf{r}$.*

From Remark 1 it follows that voluntary positive contributions require uniform groups of conditional cooperators in the sense of $r_i = \bar{r}$ for all $i \in N$. For such a group we assume that the solution is the payoff dominant strict equilibrium:

ASSUMPTION 1 *If $r_i = \bar{r}$ for all players $i \in N$, then $\mathbf{c}^* = (e, \cdots, e)$.*

ences in evolutionary success, *indirectly* influence it by inducing differences in behavior through differences in 'intrinsic' motivations.

Assumption 1 can be justified by the theory of equilibrium selection developed by Harsanyi and Selten (1988). This theory gives priority to strict over non-strict equilibria and also to payoff dominance among equilibria.[10]

Remark 1 also implies that it does not pay for an $\underline{r}$-type individual to invade[11] a uniform group of conditional cooperators if (s)he would be recognized as an $\underline{r}$-type. This would induce indeed all remaining $\bar{r}$-types to contribute zero. If, however, remaining unrecognized (the substitution is not noticed by the other conditional cooperators), an $\underline{r}$-type would greatly profit from this. This already indicates that the evolutionary stability of $\underline{r}$- or $\bar{r}$-types can depend crucially on what is known about the r_j-parameters of other group members. In the following we will discuss some possibilities concerning such information.

Let us generally assume within the framework of evolutionary game theory that an infinite population of individuals i with given r_i-types ($r_i \in M = \{\underline{r}, \bar{r}\}$) are randomly matched to infinitely many groups of size n to play public goods games. Under this general assumption, we will study the effects of different information conditions.

We start by supposing that the vector $\mathbf{r} = (r_1, \cdots, r_n) \in M^n$ is commonly known among all players in all groups. This leads to what we call the evolutionary public goods game with complete type information.

III.1. Complete type information

To study the evolutionary stability of the types $\underline{r}$ and $\bar{r}$, we use the concept of evolutionarily stable strategies.[12] Let us assume that an r_i-mutant is selected and confronted with $n-1$ other individuals of type r_j when investigating whether an r_j-monomorphism (namely a population with r_j-individuals only) can be invaded by r_i, with $r_i, r_j \in \{\underline{r}, \bar{r}\}$ and $r_i \neq r_j$. Let $v(r_i, r_j)$ measure a player's expected evolutionary success when (s)he is of the r_i-type while all other $n-1$ players are of type r_j. A strategy-type r_j is called (neutrally) evolutionarily stable if the following conditions are satisfied:

10 In a complete evolutionary approach, one should also justify why $\bar{r}$-type players' behavior converges towards the described unique solution. Güth and Yaari (1992) and Güth and Nitzan (1997) have discussed this in great detail. We therefore remit the reader to these papers for an evolutionary justification of Assumption 1.

11 Since we want to keep the group size n constant, this means to substitute one of the conditional cooperators.

12 See Maynard Smith and Price (1973), Maynard Smith (1982), Selten (1988), and van Damme (1991).

(i) $v(r_j, r_j) \geq v(r_i, r_j)$ for all $r_i \in M$, and

whenever $v(r_j, r_j) = v(r_i, r_j)$

(ii) $v(r_j, r_i) > v(r_i, r_i)$ or $(\hat{\text{ii}})$ $v(r_j, r_i) \geq v(r_i, r_i)$.

These conditions capture the idea that a monomorphic population of type r_j cannot be invaded by a small minority with deviant type r_i. According to condition (i), an evolutionarily stable strategy-type r_j is a best reply against itself: Any r_i-mutant invading a population of r_j-types cannot be more successful than the members r_j of the population. If in such a population some other r_i-mutants are equally successful, condition (ii) rules out that an alternative best reply $r_i \neq r_j$ can spread out in the population: Since r_j is better against r_i than r_i itself, r_i will be eliminated as soon as it becomes more frequent in the population. If, in the mixed population with r_j- and r_i-individuals, r_j earns at least as much as r_i (i.e., if condition $(\hat{\text{ii}})$ holds), then the strategy r_j can be characterized as neutrally evolutionarily stable.[13]

We first apply the ESS concept to the evolutionary public goods game with complete type information, in which we assume that players can identify the other players' types with certainty. In this case, an individual $\underline{r}$-mutant would end in an $\bar{r}$-monomorphic group as the only $\underline{r}$-type and this would be known by the remaining $\bar{r}$-types. According to Remark 1 this implies $\mathbf{c}^* = (0, \cdots, 0)$ for this group and the success $v(\cdot) = e$ whereas a group of conditional cooperators, all contributing e, earns $v(\cdot) = \alpha e > e$.

On the other hand, in the $\underline{r}$-monomorphism a rare, e.g. single $\bar{r}$-mutant does not change the result of $\mathbf{c}^* = (0, \cdots, 0)$. Thus, the $\underline{r}$-monomorphism is neutrally stable.[14] What proves:

REMARK 2 *When the type-vector* $\mathbf{r}$ *is commonly known, both monomorphic population distributions, in the sense of only* $\bar{r}$*-, resp.* $\underline{r}$*-, types in the population, are evolutionarily stable. Whereas the* $\bar{r}$*-monomorphism as a strict equilibrium of the evolutionary game is evolutionarily stable, the* $\underline{r}$*-monomorphism is only neutrally stable.*

13 Cf., Maynard Smith (1982).

14 Random mutation in the sense of Kandori et al. (1993) would render the $\underline{r}$-monomorphism as instable. In an evolutionary model with mutations, every player $i \in N$ changes his/her r_i-type at random with ϵ-probability ($\epsilon > 0$). What implies that in an $\underline{r}$-monomorphism an $\bar{r}$-type has ϵ^{n-1}-chance to end up in a group of $\bar{r}$-types only.

Remark 2 amounts to an evolutionary explanation of why conditionally cooperative behavior can evolve under ideal information conditions. We now turn to the polar case in which the identification beforehand of the others' types is not possible and each player knows only his/her own type. We refer to this case as the evolutionary public goods game with private type information.

III.2. Private type information

Let us assume in the tradition of Güth (1995) that, if only individual i knows his/her own type r_i, the share of $\bar{r}$-, resp $\underline{r}$-, types in the entire population is still commonly known. Assume that an $\bar{r}$-monomorphic population is invaded by a rare $\underline{r}$-mutant. Since all expect to end in an $\bar{r}$-monomorphic group with probability 1 (a single or finitely many $\underline{r}$-types do not question this because the population is infinite), one has $c_{\bar{r}}(\mathbf{c}) = e$ except for the rare $\underline{r}$-type mutant(s) who chooses $c_{\underline{r}}(\mathbf{c}) = 0$ and earns the success

$$v_{\underline{r}}(\cdot) = e + \alpha \frac{n-1}{n} e$$

exceeding the success

$$v_{\bar{r}}(\cdot) = \alpha \frac{n-1}{n} e$$

of the $\bar{r}$-types. So invasion is successful and an $\bar{r}$-monomorphism cannot be stable (as long as monotonic dynamics are assumed).[15]

For an $\underline{r}$-monomorphic population a rare $\bar{r}$-type does not question the behavior $\mathbf{c}^* = (0, \cdots, 0)$. In case of an $\underline{r}$-monomorphism even rare $\bar{r}$-mutants will choose to contribute zero since they expect to end up with $n-1$ (non cooperative) $\underline{r}$-types with probability 1. Thus, an $\underline{r}$-monomorphism is neutrally stable even when regret types are private information.[16] What proves:

Remark 3 *With private type information only the $\underline{r}$-monomorphism is evolutionarily stable (in the sense of neutral stability).*

Comparing Remarks 2 and 3, the importance of type information becomes obvious. While conditional cooperation is evolutionarily stable if players can identify their

15 Condition (i) in the definition of ESS is violated.

16 Even random mutation in the sense of Kandori et al. (1993) will not change this.

partners' types before they play a public goods game, this is not longer true if players merely know the distribution of types in the population.

The arguments above were related to pure, extreme cases. As such, they give us a clue to what might be crucial factors of the evolutionary process in which conditional cooperation emerges. Namely, villages or small communities, where everyone knows everyone else, would favor the evolution of conditionally cooperative behavior whereas cities or big metropolitan areas, where knowing (the attitudes of) the other people is not easy, would not.

However, it is quite improbable that the ideal conditions of the two polar cases could prevail during human evolution. It seems more plausible that these conditions can be characterized by intermediate informational assumptions that locate the interaction somewhere in-between the two extreme cases analyzed so far. We shall turn now our attention to an analysis of such an intermediate case.

III.3. Costly type information

As the comparison of the two polar cases analyzed above shows, it would be in the interest of the conditional cooperators to learn about the other players' types. However, an $\underline{r}$-type has an incentive to avoid this and try to question the reliability of r-type signals,[17] or the signals themselves may be imperfect (as in Frank, 1987). Since the biological aspects of these issues are not easily manageable, we will simply suppose (following Güth and Kliemt, 1994) that a perfectly reliable detection technology is available at a cost C.

Specifically, let us assume that individuals are first randomly split up in infinitely many groups of size n, and can then learn about the others' types by investing into r-type detection at some positive monetary cost C.

Players must decide about the acquisition of costly information being ignorant of the other players' types but, as in Section III.2, knowing the population shares of both r-types. When deciding whether to invest in type detection or not they are, of course, not aware of the others' investment choices nor of the others' contribution decisions. This means that they must make that decision before playing the public goods game.

17 It has been, however, suggested that individuals cannot convincingly lie because they cannot avoid involuntary signalling inherent in their behavior (cf., Gauthier, 1978). Several papers in evolutionary game theory are based on this assertion (see, for instance, Frank, 1987, 1988; and Robson, 1990), which has been empirically confuted by Ockenfels and Selten (2000) in a two-person bargaining experiment.

Observe first that an $\underline{r}$-type's optimal contribution decision does not depend on his/her information status: Whatever the other players' types, an $\underline{r}$-type will rationally choose to free ride. Such a type will therefore never invest in type detection. On the contrary, an $\bar{r}$-type's contribution decision depends on the others' types. If (s)he ends up in a group with other $n-1$ $\bar{r}$-types, (s)he will (in equilibrium) choose $c_i(\cdot) = e$. Otherwise, (s)he will choose $c_i(\cdot) = 0$. Thus, our subsequent analysis of when and under which conditions it is convenient to acquire costly information will focus only on $\bar{r}$-types.

Assume that the population share of $\bar{r}$-types is $q \in [0, 1]$ while that of $\underline{r}$-types is $1 - q$, what is common knowledge among the players. Let m be the number of $\bar{r}$-type co-players in a group ($0 \leq m \leq n-1$). Then the probability, p, of being in a group with other m individuals of type $\bar{r}$ is binomially distributed with parameters m and q, i.e. $p = \binom{n-1}{m} q^m (1-q)^{n-1-m}$.

An $\bar{r}$-type will invest into information (incurring a sunk cost of C) only if this makes him/her better off than remaining uninformed. Assume that (s)he had invested. (S)he then gets to know the types of all his/her co-players. Therefore, if all $n-1$ co-players j are of type $\bar{r}$ and (due to Assumption 1) all contribute $c_j(\cdot) = e$, (s)he will contribute the entire endowment, whereas when facing at least one $\underline{r}$-type opponent (s)he will contribute nothing.

Had the $\bar{r}$-type chosen not to invest in type detection, (s)he does not have access to specific type information. Therefore, (s)he must rely on the prior probability p of meeting other $\bar{r}$-type individuals and will choose his/her level of contribution on the basis of his/her expected payoffs. In order to suffer no regret, the uninformed $\bar{r}$-type must contribute exactly the expected average contribution, which is contingent on m (i.e., the number of $\bar{r}$-type co-players in the group). Let us suppose that there exists a contribution level, c^+, that *a priori* matches the average amount. This means that c^+ must be such that

$$(3) \quad c^+ = f(c^+) := \sum_{m=0}^{n-1} \binom{n-1}{m} q^m (1-q)^{n-1-m} \frac{(m+1)}{n} c^+.$$

In words, c^+ must be a fix point of $f(c^+)$. For $0 \leq q < 1$ the unique c^+ complying with this requirement is $c^+ = 0$.[18] Hence, a tiny doubt on the conditionally

18 c^+ cancels out from both sides of (3), which (after algebric manipulations) becomes $n - 1 = \sum_{m=0}^{n-1} \binom{n-1}{m} q^m (1-q)^{n-1-m} m = E(m) = (n-1)q$, where $E(m)$ is the expected value of m. For $0 \leq q < 1$, $(n-1)$ is always smaller than $(n-1)q$ and the unique c^+ satisfying (3) is $c^+ = 0$. Only for $q = 1$, the equality of interest can be satisfied. In this case, all $0 \leq c^+ \leq e$ can be sustained as fix points (so that, only by payoff dominance, one can assume $c^+ = e$).

cooperative attitude of the co-players suffices to drive the uninformed $\bar{r}$-types' contribution decision to zero. This proves:

REMARK 4 *For $0 < q < 1$, all uninformed $\bar{r}$-types (who have not invested in information) will contribute zero.*

We can now find out when it is optimal for $\bar{r}$-types to invest into information. Whether investing into information makes the $\bar{r}$-types better off than not investing depends on the value of the information to which they gain access. Let us therefore look at the value of the information by considering the payoff expectations of the informed $\bar{r}$-types as opposed to those of the uninformed $\bar{r}$-types.

If the $\bar{r}$-types invest into information, they expect (before playing the public goods game) $\alpha e - C$ with probability q^{n-1} and $e - C$ with complementary probability $(1 - q^{n-1})$.[19] If instead all $\bar{r}$-types do not invest into information, due to Remark 4, they would expect e. We get therefore the following expected payoffs for informed and uninformed $\bar{r}$-types:[20]

$$u_I(\cdot) = [q^{n-1}\alpha e + (1 - q^{n-1})e] - C$$

and

$$u_U(\cdot) = e$$

where the subscripts I and U stand for informed, resp. uninformed, $\bar{r}$-types.

For $u_I(\cdot) > u_U(\cdot)$ to be fulfilled it is necessary that

$$[\alpha q^{n-1} + (1 - q^{n-1})]e - C > e,$$

which entails:

$$(4) \quad q > \hat{q}(C) := \Big[\frac{C}{(\alpha - 1)e}\Big]^{\frac{1}{n-1}}$$

19 q^{n-1} refers to the probability to meet $(n-1)$ $\bar{r}$-type co-players who, due to symmetry, are all informed and contributing e; $1 - q^{n-1}$ refers to the probability to face at least one $\underline{r}$-type opponent; in this case, the informed $\bar{r}$-types contribute 0 and earn e minus the sunk cost C.

20 Note that payoff expectations do not include regret since the $\bar{r}$-types always meet the average contribution.

and, due to $q \in [0,1]$:

(5) $\quad 0 < C \leq \bar{C} := (\alpha - 1)e.$

Inequalities (4) and (5) determine the value of the population composition q, resp. of the detection cost C, for which investing in type detection is optimal. The expected gain from investment, $q^{n-1}(\alpha-1)e - C$, is zero for $q = 0$, increases in q, and is $(\alpha-1)e - C$ for $q = 1$. The latter implies that for any $C > (\alpha-1)e$ a non-investment situation prevails: The cost is prohibitively high to make it convenient acquire information. Within the range $(\alpha-1)e \geq C > 0$, the result depends on the population composition q as illustrated in Figure 1 for the detection cost $C' = 5$.[21]

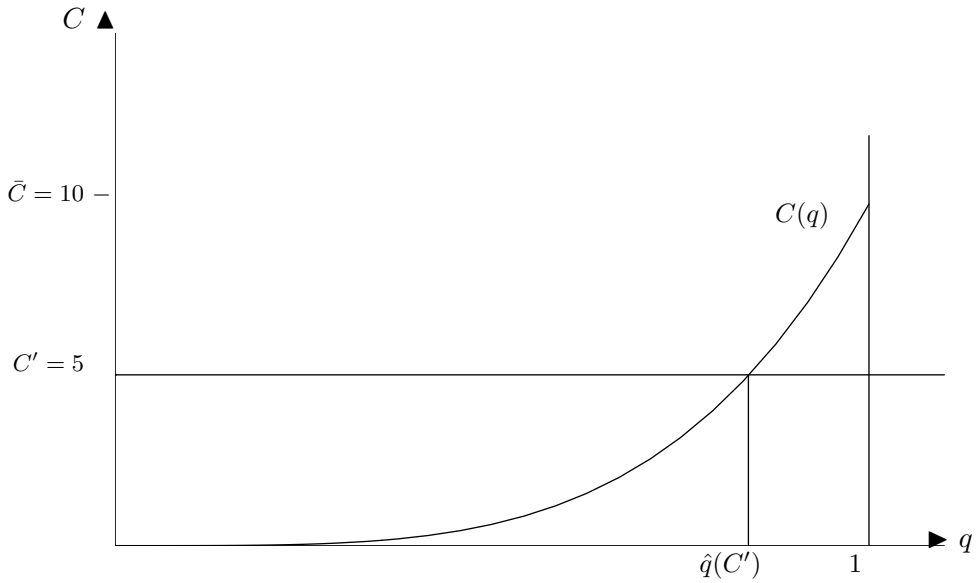

Figure 1: C-q relationship.

For all q with $\hat{q}(C') < q < 1$ expected gains from a positive investment decision exceed their costs C', whereas for any $q < \hat{q}(C')$ the opposite holds. Thus investing in type detection is worthwhile if, for $C = C'$, the actual population composition complies with $q > \hat{q}(C')$. More generally we can state:

REMARK 5 *For all cost parameters $C \in (0, (\alpha-1)e]$, investing in type detection is rational when the population share q of $\bar{r}$-types fulfils $\hat{q}(C) < q < 1$*

21 The figure relies on the parameters: $n = 5$, $e = 20$ and $\alpha = 1.5$. These values imply $\bar{C} = 10$. Choosing $C' = 5$, we get $\hat{q}(C') = 0.84$.

Remark 5 deals merely with non-degenerate cases. Degenerate cases emerge whenever one of the players is indifferent between alternative moves. Given optimal play this may happen only for $q = \hat{q}(C)$. It shall become obvious subsequently that this value of q refers to transitory states of evolutionary dynamics and thus can be neglected.

Before turning to the issue of evolutionary stability itself it seems helpful to take a closer look at the relationships between group size, n, marginal return from the public good, α, and optimal investment behavior. For cost levels C for which investing in type detection is rational (i.e., $C < (\alpha - 1)e$), the fraction $C/[(\alpha - 1)e]$ in the definition of $\hat{q}(\cdot)$ is smaller than 1 (see Inequality (4)). This implies that $\hat{q}(\cdot) := \left[\frac{C}{(\alpha-1)e}\right]^{\frac{1}{n-1}}$ is positively related to n. Hence, reducing the group size decreases $\hat{q}(\cdot)$ (that is, the threshold or the dividing line between investment and non-investment decision) and increases the range $\hat{q}(\cdot) < q < 1$ of population compositions for which investing in information is optimal.

From Inequality (4) one can infer as well that $\hat{q}(\cdot)$ is negatively related to α: A decrease in the marginal return from the public good causes a rise in $\hat{q}(\cdot)$ and, hence, reduces the range of q for which investing in type detection is rational.

Keeping in mind these observations about the relationships between n, α and $\hat{q}(\cdot)$, we can now approach the issue of evolutionary dynamics and investigate how q would evolve if n-person public goods games are played among randomly matched players chosen from an infinite population of r_i-type individuals with $r_i \in M = \{\underline{r}, \bar{r}\}$. Our basic assumption is that q increases (decreases) through time as long as the $\bar{r}$-types are more (less) successful than the $\underline{r}$-types. The latter strictly depends on whether the $\bar{r}$-types invest in type detection or not.

Assume initially that the cost C of being informed and the population composition are such that the relations defined in Remark 5 cannot be fulfilled: There are no values of C and q rendering a positive investment decision optimal. All $\bar{r}$-types remain uninformed so that knowledge of r_i-types is strictly private. Players only know their own type and the distribution of types in the population. Therefore Remark 3 directly applies. The population share q of $\bar{r}$-types converges to $q = 0$; i.e., the $\bar{r}$-types are driven out of the population.

Suppose now that Remark 5 is satisfied and, hence, the $\bar{r}$-types do invest into information. To study the dynamics of q, we must compare the relative success of the informed $\bar{r}$-types with that of the $\underline{r}$-types. The informed $\bar{r}$-types expect the

success:

$$v_{\bar{r}}(\cdot) = [q^{n-1}\alpha + (1 - q^{n-1})]e - C$$

while the $\underline{r}$-types (due to the information status of the $\bar{r}$-types) expect:

$$v_{\underline{r}}(\cdot) = e.$$

For $v_{\bar{r}}(\cdot) > v_{\underline{r}}(\cdot)$ to hold good it is necessary that:

$$q > \hat{q}(C),$$

where $\hat{q}(C)$ is determined by the left side of (4). Thus, an increase of q is to be expected from the mere fact that the $\bar{r}$-types acquire information. This increase in q will continue until q hits its upper bound. At $q = 1$ the $\bar{r}$-types have spread out in the population defeating completely the $\underline{r}$-types. This proves:

REMARK 6 *In the evolutionary public goods game with perfect but costly detection technology, an $\bar{r}$-monomorphic population with $q = 1$ is evolutionarily stable if $C \in (0, (\alpha - 1)e]$.*

The other boundary of the interval $[\hat{q}(C), 1]$ does not satisfy the conditions for evolutionary stability. Since according to Remark 5 $\bar{r}$-types do not invest into information for $q < \hat{q}(C)$ (what implies a decrease of q) and they do invest for $q > \hat{q}(C)$ (what induces an increase of q), the population composition $q = \hat{q}(C)$ is highly unstable. Any slight disturbance of q will lead away from $q = \hat{q}(C)$.

Figure 2 summarizes the basic insights of the above dynamics in the q, C-plane. As indicated by the arrows, the share q of $\bar{r}$-type individuals increases inside the area formed by the $C(q)$-curve, the $q = 1$-(vertical) line and the q-axis whereas q decreases outside this realm.

Remark 6 provides some hope that conditionally cooperative behavior can survive. If a perfectly reliable type detection technology is available at a cost C, the $\bar{r}$-monomorphism is evolutionarily stable provided that the $\bar{r}$-types invest into information. For any not prohibitively high value of C, small groups and high marginal returns from the public good favor the evolution of the $\bar{r}$-types by inducing an increase in the range of q for which the investment decision is rational. If the cost of the detection technology is too high or the population share q of $\bar{r}$-types falls below the threshold $\hat{q}(C)$, conditionally cooperative individuals

will be driven out of a population and the other monomorphic population (with $q = 0$) will be evolutionarily stable. Thus, under certain values of the parameters, different stable population compositions are viable.

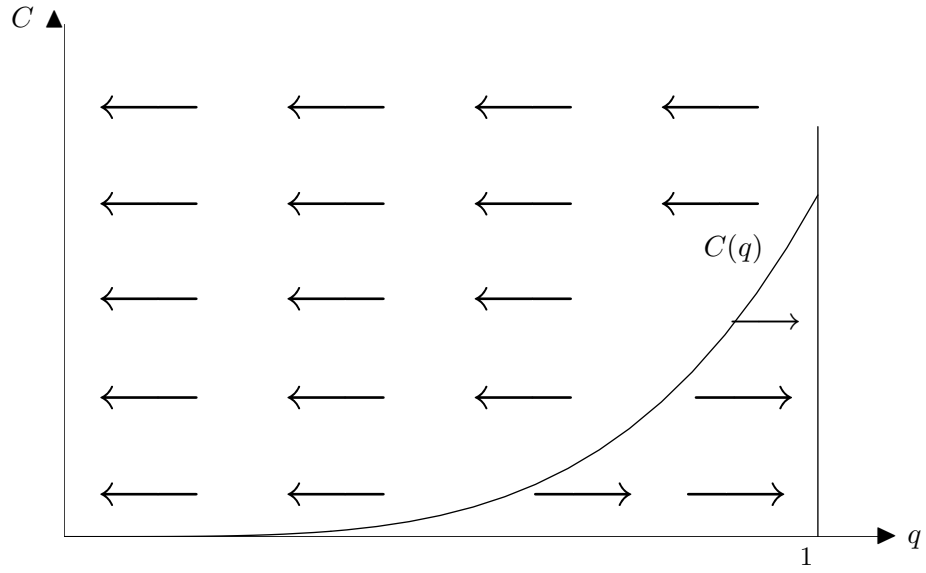

Figure 2: q-dynamics

IV. Concluding remarks

The preceding exercise aimed at pointing out the circumstances under which conditionally cooperative behavior (facilitating voluntary contributions among rational actors in public goods settings) might evolve.

The main finding of our study is that evolutionarily stable conditional cooperation crucially depends on the individuals' information conditions. If the type (opportunistically rational or conditionally cooperative) of the other group members is commonly known, a monomorphic society of conditional cooperators is evolutionarily stable. But with private type information, if a free riding mutant appears (s)he will be treated as a conditional cooperator (since each conditional cooperator will consider the probability of interacting with the mutant as negligible) and (s)he will earn a higher success than his/her conditionally cooperative counterparts. As a result, conditional cooperation will become vulnerable to free riding mutants and will not survive.

Our analysis, therefore, suggests that conditional cooperation is more likely to emerge in small societies (like villages or even restricted districts of big cities)

where individuals are not anonymous. This corroborates the elementary yet extremely important insight that the stability of cooperation depends not only on the own cooperative disposition but also on the possibility to discriminate between cooperative and opportunistic partners.

Nonetheless, even when the others' type is not directly known, our model indicates that conditional cooperation can evolve if a perfectly reliable type detection technology is available at a (non prohibitively high) monetary cost C. In this case, there exists a stable monomorphism containing only conditionally cooperative types.

The results of Section III.3 justify at least some optimism that, as long as people invest in type detection, conditional cooperation can survive and eliminate free riding once conditionally cooperative types have somehow gained a sufficient share $\hat{q}(C)$ in a population. According to our analysis, for any given value of C, small groups and high efficiency gains from the public good reduce the threshold $\hat{q}(C)$ and, therefore, ameliorate the conditions for the survival of conditional cooperation in evolutionary competition. From Figure 2, the region

$$\{(q, C) : \hat{q}(C) < q < 1, \quad 0 < C \leq \bar{C}\}$$

forms the attraction set of $q = 1$ while the region complementary to this area forms the source of $q = 0$ (disregarding degenerate cases with a starting point $q_0 = \hat{q}(C)$). According to Figure 2, the starting point of the dynamic process is essential: If $0 < C < \bar{C}$, it depends on $q_0 \in [0, 1]$ whether the population composition will converge to $q = 0$ or to $q = 1$. Thus (like in earlier evolutionary studies), only if we have a "good start" will there be a happy ending; otherwise we are stuck with a "competitive" world in which only exploitative behavior is known.

References

Anderson, S. P., Goeree, J. K. & Holt, C. (1998), 'A Theoretical Analysis of Altruism and Decision Error in Public Goods Games', *Journal of Public Economics* **70**, 297–323.

Andreoni, J. (1995), 'Cooperation in Public Goods Experiments: Kindness or Confusion?', *American Economic Review* **85**(4), 891–904.

Bolton, G. E. & Ockenfels, A. (2000), 'ERC: A Theory of Equity, Reciprocity and Competition', *American Economic Review* **90**, 166–193.

Brandts, J. & Schram, A. (2001), 'Cooperation and Noise in Public Goods Experiments: Applying the Contribution Function Approach', *Journal of Public Economics* **79**, 399–427.

Croson, R. (2000), 'Theories of Altruism and Reciprocity: Evidence from Linear Public Goods Games', Discussion Paper, Wharton School, University of Pennsylvania, Philadelphia.

Darwin, C. (1859), *The Origin of Species*, Greg Suriano (reprinted in 1998).

Davis, D. D. & Holt, C. A. (1993), *Experimental Economics*, Princeton, NJ: Princeton University Press.

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

Fehr, E. & Schmidt, K. M. (1999), 'A Theory of Fairness, Competition, and Cooperation', *Quarterly Journal of Economics* **114**, 817–868.

Fischbacher, U., Gächter, S. & Fehr, E. (2000), 'Are People Conditionally Cooperative? Evidence from a Public Goods Experiment', *Economics Letters* **71**, 397–404.

Frank, R. H. (1987), 'If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?', *American Economic Review* **77**, 593–604.

Frank, R. H. (1988), *Passions Within Reason*, Norton, New York.

Gauthier, D. (1978), *Morals by Agreement*, Clarendon Press, Oxford.

Güth, W. (1995), 'An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives', *International Journal of Game Theory* **24**, 323–344.

Güth, W. & Kliemt, H. (1994), 'Competition or Co-operation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes', *Metroeconomica* **45**, 155–187.

Güth, W. & Nitzan, S. (1997), 'The Evolutionary Stability of Moral Objections to Free Riding', *Economics and Politics* **9**, 133–149.

Güth, W. & Yaari, M. (1992), Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach, *in* U. Witt, ed., 'Explaining Process and Change. Approaches to Evolutionary Economics', The University of Michigan Press, Ann Arbor, pp. 23–34.

Harsanyi, J. C. & Selten, R. (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge, MA: Harvard University Press.

Hobbes, T. (1651), *Leviathan*, Harmondsworth: Penguin (reprinted in 1968).

Isaac, R. M., Walker, J. & Thomas, S. (1984), 'Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations', *Public Choice* **43**(1), 113–149.

Kandori, M., Mailath, G. J. & Rob, R. (1993), 'Learning, Mutation, and Long Run Equilibria in Games', *Econometrica* **61**(1), 29–56.

Keser, C. (1997), SUPER: Strategies Used in Public Goods Experimentation Rounds. Working Paper 97/24, Sonderforschungsbereich 504, University of Mannheim.

Keser, C. & van Winden, F. (2000), 'Conditional Cooperation and Voluntary Contributions to Public Goods', *Scandinavian Journal of Economics* **102**, 23–39.

Ledyard, J. O. (1995), Public Goods: A Survey of Experimental Research, *in* J. H. Kagel & A. E. Roth, eds, 'The Handbook of Experimental Economics', Princeton, NJ: Princeton University Press.

Levati, M. V. & Neugebauer, T. (2001), 'An Application of the English Clock Market Mechanism to Public Goods Games', Discussion Paper No. 4, Papers on Strategic Interaction, Max Planck Institute for Research into Economic Systems, Jena, Germany.

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

Maynard Smith, J. & Price, G. R. (1973), 'The Logic of Animal Conflict', *Nature* **246**, 15–18.

Ockenfels, A. & Selten, R. (2000), 'An Experiment on the Hypothesis of Involuntary Trust-Signalling in Bargaining', *Games and Economic Behavior* **33**, 90–116.

Palfrey, T. R. & Prisbrey, J. E. (1997), 'Anomalous Behavior in Public Goods Experiments: How Much and Why?', *American Economic Review* **87**, 829–846.

Parsons, T. (1968), Utilitarianism. Sociological Thought, *in* 'International Encyclopedia of Social Sciences', New York and London.

Robson, A. E. (1990), 'Efficiency in Evolutionary Games: Darwin, Nash and the Secret Hand-Shake', *Journal of Theoretical Biology* **144**, 379–396.

Selten, R. (1988), 'Evolutionary Stability in Extensive Two-Person Games - Correction and Further Developments', *Mathematical Social Sciences* **16**, 223–266.

Sonnemans, J., Schram, A. & Offerman, T. (1999), 'Strategic Behavior in Public Good Games: When Partners Drift Apart', *Economics Letters* **62**, 35–41.

Sugden, R. (1984), 'Reciprocity: The Supply of Public Goods Through Voluntary Contributions', *Economic Journal* **94**, 772–787.

van Damme, E. (1991), *Stability and Perfection of Nash Equilibria*, Berlin: Springer Verlag.