European Journal of Political Economy 30 (2013) 15–34

Contents lists available at SciVerse ScienceDirect

European Journal of Political Economy

journal homepage: www.elsevier.com/locate/ejpe

On the evolution of conditional cooperation

Joel M. Guttman *

Department of Economics, Bar-Ilan University, 52900 Ramat-Gan, Israel

ARTICLE INFO

Article history:
Received 29 November 2010
Received in revised form 1 November 2012
Accepted 14 November 2012
Available online 27 November 2012

JEL classification:
D03
D64
H41

Keywords:
Reciprocity
Conditional cooperation
Public goods
Social preferences
Cooperation

ABSTRACT

A long series of laboratory and field experiments, as well as conventional empirical studies, has established that (1) individuals voluntarily provide themselves with public goods at levels exceeding those predicted by the Nash voluntary contributions mechanism, and (2) agents reciprocate increases in the contributions of their counterparts in such settings (conditional cooperation). This paper presents a simple model of the evolution of preferences for conditional cooperation in the presence of a public good, which explains these two empirical findings without employing reputational or group selection arguments. In this model, individuals inherit preferences to match other agents' contributions to the provision of a public good, at some specified "matching rate." Agents whose preferences induce them to be relatively successful – in material terms – increase in number, from one generation to the next. Under complete information and with randomly matched groups of N agents who have quasilinear preferences over the public good and a private good, the unique evolutionarily stable matching rate is 1, leading to Pareto optimal voluntary provision of the public good, regardless of group size N. The evolutionarily stable matching rate can be viewed as an endogenous social norm.

1. Introduction

A large body of laboratory and field experiments and other empirical evidence has accumulated, demonstrating that individuals can voluntarily provide themselves with public goods, even in groups of considerable size.[1] A second experimental literature (see, e.g., Guttman, 1986; Keser and van Winden, 2000; Fischbacher et al., 2001; Frey and Meier, 2004; Croson, et al., 2005; Croson, 2007) demonstrates that individuals typically *reciprocate* contributions of their counterparts in the voluntary provision of public goods. This evidence of "conditional cooperation" in the provision of public goods parallels extensive experimental evidence of reciprocity in two-person games, such as the ultimatum and investment games.[2]

This accumulation of evidence presents perhaps the central problem of public economics – explaining voluntary cooperation and reciprocity – just as the same problem, with regard to animals, has been termed the central problem of biology. Simply *assuming* reciprocal preferences[3] is only a first step, since almost any behavior can be explained by assuming the "right" preferences.[4]

The purpose of this paper is to provide a simple model capable of *predicting* which agent types – embodying varying degrees of preference for reciprocity – will survive in an evolutionary process.

* Corresponding author. Tel.: +972 52 762 3507; fax: +972 3 738 4034.
E-mail address: Joel.Guttman@biu.ac.il.

[1] See Ledyard (1995) for a survey of the experimental evidence, and Stewart (2009) for an interesting non-experimental empirical study.

[2] The fact that agents are conditionally cooperative has policy implications, for example in the case of policy to reduce tax evasion. See Traxler (2010).

[3] For models of reciprocal preferences, see Sugden (1984), Rabin (1993), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000).

[4] As Smith (2003) put it in his Nobel Lecture, "Technically, the issue is how most productively to model agent 'types' by extending game theory so that types are an integral part of its predictive content, rather than merely imported as an ex post technical explanation of experimental results."

0176-2680/$ – see front matter
http://dx.doi.org/10.1016/j.ejpoleco.2012.11.003

Previous theoretical work has shown that reciprocity can survive over the generations, when agents are assumed to be involved in bilateral strategic situations like the Prisoner's Dilemma.[5] Of particular interest to economists are models that do not assume irrationality of agents, but rather show that *reciprocal preferences* are evolutionarily stable. These "indirect" evolutionary models, which began in earnest with the work of Güth and Yaari (1992), do not reject the traditional assumption of individual rationality.[6] As in standard economic and game-theoretic models, agents optimally choose strategies given their preferences. The evolutionary process selects player types whose preferences induce them to choose strategies that yield relatively high (material) payoffs or "fitness". Thus evolutionary pressures operate on *preferences* and not on *strategies*, in contrast to the conventional sociobiological literature.

A limitation of results in this theoretical tradition, however, is that players must have some information of the type of their counterparts in order for the "reciprocator type" to survive.[7] If players know only the probability distribution of player types in an infinitely large population, a non-standard type (like the reciprocator type), whose preferences are not perfectly aligned with maximizing material payoff, cannot survive—at least not at a level that will support cooperation in the underlying two-person game.

Clearly, the extreme assumption that players can perfectly identify their peer's type (the so-called "green beard" assumption) is unrealistic. But the opposite (extreme) assumption that players have *no information* at all of their peer's type is also unrealistic. Frank (1988) has assembled persuasive empirical evidence that facial expressions, blushing, breathing rapidity, "body language" and the like provide signals, albeit noisy ones, about an agent's emotional status. Along the same lines, Caplan (2003, p. 393) cites studies by psychologists indicating that "even extremely brief and superficial contact leads to personality assessments measurably superior to random guessing." For example, Borkenau and Liebler (1993) report that "[r]atings of extraversion, conscientiousness, and intelligence by strangers having been exposed to a videotape of targets were significantly related to self-reports of these traits as well as to ratings by acquaintances."

Samuelson (2001) has posed the question, why would not types who successfully "fake" such signals, without actually being the (say) "nice" type, have higher fitness than agents who are truly nice? This question has been answered in at least one context, namely a finitely repeated Prisoner's Dilemma.[8] In an ongoing community where behavior is recalled from one generation to the next, agents' behavior in the last stage of the game, which reveals their type, can be recalled by agents of the following generation. Since personality characteristics are inherited with positive probability,[9] this last-stage behavior automatically provides a noisy signal of the type of the agent's offspring.

Consider an arbitrarily small weakening of the zero-information assumption. A far more realistic, intermediate assumption – that agents can identify their fellow agents' type with some *arbitrarily small* but positive probability – is sufficient to permit an evolutionary explanation of reciprocal preferences in certain contexts. For example, Güth (1995) assumes only that players can correctly identify their opponent's type with some positive probability, in the context of a trust game.[10] More generally, Heifetz et al. (2007) show that imperfect identification of an opponent's type can support the evolutionary survival of non-standard types. In the present paper, the model is first set up under the assumption of complete information, and then this assumption is relaxed in the manner of Güth (1995).

The above-cited models examine the evolutionary stability of the reciprocator type only in the context of *two-person* games. Previous theoretical work (particularly Boyd and Richerson, 1988) casts doubt on the evolutionary stability of the reciprocator type in games involving more than two players, such as in the provision of public goods.[11] Yet, aside from the experimental studies cited above, anthropological studies (e.g., Lee, 1979; Kaplan et al., 1984) provide evidence that in many food foraging groups, game is shared equally among group members, regardless of who makes the kill. This equal-sharing rule presents a serious puzzle to economists, since it would induce free-riding by group members with standard (non-reciprocal) preferences. These anthropological studies suggest that reciprocal preferences developed very early in human history, in societies where the voluntary provision of public goods is a cornerstone of individuals' livelihood.

The present paper develops a simple model of the evolution of reciprocal preferences in groups of arbitrary size which must voluntarily provide a public good. It is shown that if players have identical, quasilinear preferences over the public good and a

[5] For excellent surveys of the literature on social preferences, see Sethi and Somanathan (2003), and Sobel (2005).

[6] The work of Güth and Yaari (1992) has been applied and extended by a large literature, including Güth and Kliemt (1994), Güth (1995), Huck and Oechssler (1999), Güth et al. (2000), Guttman (2000, 2003), Possajennikov (2000), Ely and Yilankaya (2001), Güth and Peleg (2001), Ok and Vega-Redondo (2001), Sethi and Somanathan (2001), Bowles and Gintis (2004), Friedman and Singh (2004a, 2004b, 2009), Güth and Pull (2004), Heifetz and Segev (2004), Dekel et al. (2007), and Heifetz et al. (2007).

[7] Ely and Yilankaya (2001), Güth and Peleg (2001), Ok and Vega-Redondo (2001), Samuelson (2001), Sobel (2005), and Heifetz et al. (2007), among others, have analyzed this problem. The model of Huck and Oechssler (1999) is an exception, in that their model predicts the survival of a type who makes fair allocations in the ultimatum game, without assuming that players can observe the type of their opponents. But they reach this result by assuming a finite (and not too large) population and that agents know the population distribution of types, or that agents interact in small groups, and the agents know the group composition of types.

[8] See Guttman (2003, p. 650).

[9] For citations to the abundant evidence on this point, see Caplan (2003).

[10] Similarly, Guttman (2003) assumes that agents emit a noisy signal correlated with their type. This correlation can be arbitrarily small in order to support an evolutionary explanation of preferences for reciprocity when agents play a finitely repeated Prisoner's Dilemma game, provided that the number of stages in the game is sufficiently large.

[11] Three papers that have applied the *indirect* evolutionary approach to the voluntary provision of public goods are Güth and Nitzan (1997), Levati (2006), and Traxler and Spichtig (2011). Güth and Nitzan (1997) obtain generally negative results regarding the evolutionary stability of "moral objections to free riding," at least when population size is finite. Levati (2006), assuming a different technology of providing the public good, obtains the result that if information of other agents' types is complete, there exists an evolutionary equilibrium in which all agents are non-reciprocators, as well as an evolutionary equilibrium in which all agents are reciprocators. Traxler and Spichtig (2011) explain the survival of conditional cooperation by introducing an environment in which there are multiple equilibria in the public goods game. Conditional cooperators have an evolutionary advantage over unconditional cooperators and pure free-riders because of their flexible responses to different equilibria.