Behavioural Processes 87 (2011) 25–33

Contents lists available at ScienceDirect

Behavioural Processes

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

A behavioral analysis of altruism

Howard Rachlin*, Matthew Locey

Psychology Department, Stony Brook University, Stony Brook, NY 11794-2500, United States

ARTICLE INFO

Article history:
Received 20 August 2010
Received in revised form 8 December 2010
Accepted 8 December 2010

Keywords:
Altruism
Charity
Delay discounting
Evolution
Patterns of behavior
Prisoner's dilemma
Self-control
Social discounting
Soft commitment

ABSTRACT

Altruistic acts have been defined, in economic terms, as ". . .costly acts that confer economic benefits on other individuals" (Fehr and Fischbacher, 2003). In multi-player, one-shot prisoner's dilemma games, a significant number of players behave altruistically; their behavior benefits each of the *other* players but is costly to them. We consider three potential explanations for such altruism. The first explanation, following a suggestion by the philosopher Derek Parfit, assumes that players devise a strategy to avoid being free-loaders—and that in the present case this strategy dictates cooperation. The second explanation says that cooperators reject the one-shot aspect of the game and behave so as to maximize reward over a series of choices extending beyond the present situation (even though reward is not maximized in the present case). This explanation assumes that people may learn to extend the boundaries of their selves socially (beyond their own skin) as well as temporally (beyond the present moment). We propose a learning mechanism for such behavior analogous to the biological, evolutionary mechanism of group selection. The third explanation assumes that people's altruism is based on a straightforward balancing of undiscounted costs to themselves against discounted benefits to others (social discounting). The three proposed explanations of altruism complement each other.

The object of behavioral analysis is to identify reinforcers of acts. But an individual altruistic act apparently has no reinforcer; if it did, it would not be altruistic. Altruism thus seems to defy behavioral analysis. Altruistic acts have been defined, in economic terms, as ". . .costly acts that confer economic benefits on other individuals" (Fehr and Fischbacher, 2003). This definition does not say that the cost to the actor and the benefit to others must be equally valued. And, it does not say whether the "other individuals" are relatives of, friends of, or complete strangers to the actor. If you put a dollar in a machine and someone else gets (and eats) the candy bar, your act is altruistic according to the definition above. It may be that few of us would pay a dollar to give a perfect stranger a candy bar but we might very well pay a penny. Or, if the benefit to the stranger were very high (say, he was starving) we might pay a dollar – or even more than a dollar – to give him a candy bar. Or, if the beneficiary were not a stranger but our own child (and we were not at the moment worrying about cavities or obesity), many of us would pay the dollar. Such acts, fitting within Fehr and Fischbacher's (quite reasonable) definition of altruism, are extremely common in everyday life; behavioral analysis cannot just ignore them.

To illustrate how common altruistic behavior is, consider the multi-person prisoner's dilemma game that Rachlin has played with audiences over the last 15 years at public lectures. Let us call it, "the lecture game." At the start of the lecture game, blank index cards are handed out to 10 random members of the audience and the others are asked (as we ask the reader) to imagine that they had gotten a card. The task of the 10 players is to choose *X* or *Y* subject to the following rules (displayed on a slide):

(1) If you choose *Y* you will receive $100 times *Z*.
(2) If you choose *X* you will receive $100 times *Z* plus a bonus of $300.
(3) *Z* equals the number (of the 10 players) who choose *Y*.

The audience is told, regretfully, that the money is purely hypothetical; then several properties of the game are pointed out. First, for any particular player, it is always better to choose *X*. By choosing *X* a player subtracts 1 from *Z* and thereby loses $100 but more than makes up for that loss by the $300 bonus. The net gain for choosing *X* is therefore $200 *regardless of what anyone else chooses.* The point is then emphasized further by saying that any lawyer would advise choosing *X*.

It is noted that if everyone obeyed their lawyers and chose *X*, *Z* would equal zero and each person would earn just $300 while if everyone disobeyed their lawyers and chose *Y*, *Z* would equal 10 and each person would earn $1000. Hence the dilemma. It is then pointed out, that there is no right or wrong answer and that

* Corresponding author. Tel.: +1 212 996 3478.
E-mail addresses: howard.rachlin@sunysb.edu, hrachlin@notes.cc.sunysb.edu (H. Rachlin).

0376-6357/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.beproc.2010.12.004

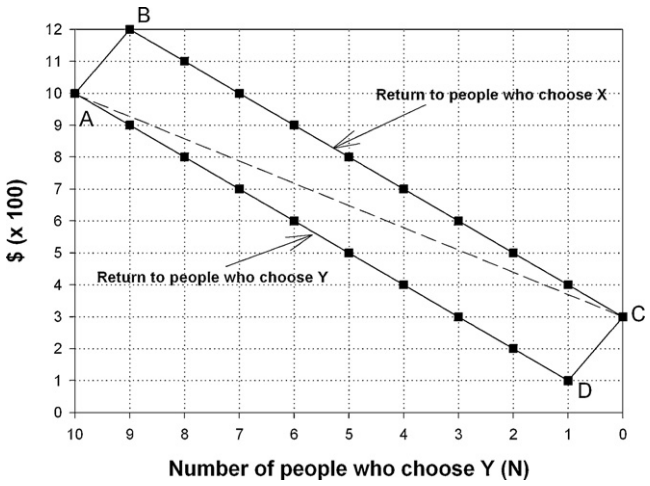

Fig. 1. Diagram of multi-person prisoner's dilemma game (lecture game). Lines A–D and B–C show returns to players who choose *Y* (cooperate) or choose *X* (defect) as a function of the decreasing number of players (out of 10) who choose *Y* (note reversed *x*-axis). The dashed line shows average earnings across the 10 players.

all choices will be forever anonymous. The 10 audience members with cards are then asked to mark them with *X* or *Y* as they would if the money were real, and the cards are collected.

Over the years this game has been played dozens of times: with college students, with philosophers, with economists (American-capitalist and Italian-socialist), and with psychologists (American, Japanese and Polish). The median response is 5 *Y*'s to 5 *X*'s. If there is any bias it is usually in the direction of more *Y*'s than *X*'s. The chart in Fig. 1 is then shown to the audience and used to look up the earnings of the *X* and *Y* choosers. For example, if 5 of the 10 participants chose *Y*, *Z* = 5 and each *Y*-chooser would get \$500 while each *X*-chooser would get \$800. In keeping with standard prisoner's dilemma terminology, an *X*-choice is called a "defection" and a *Y*-choice is called a "cooperation." Although the money earned in the lecture game is hypothetical, laboratory experiments with real money (albeit in lesser amounts) have found significant numbers of cooperations in one-shot, multi-person prisoner's dilemma games such as this one (Camerer, 2003).

Because a *Y*-choice always earns \$200 less than an *X*-choice, choosing *Y* is a "costly act." Because a *Y*-choice increases *Z* by 1, and each of the 9 other players earns \$100 more than she would have otherwise, choosing *Y* "confers economic benefits on other individuals." Because the choices are completely anonymous, it cannot be claimed that a player's reputation would be enhanced by choosing *Y*. Thus, according to Fehr and Fischbacher's definition, *Y*-choices in the lecture game are altruistic. Altruism toward one's family members may be explained in terms of genetic selection, but it is unlikely that any lecture-game players were related. Although there is indeed a bias in altruistic behavior toward relatives over non-relatives (Jones and Rachlin, 2008) numerous instances of altruism in everyday life are directed toward friends, acquaintances and even complete strangers.

The philosopher, Parfit (1984, pp. 61–62) listed a sample of situations from everyday life modeled by multi-person prisoner's dilemma games such as the lecture game:

- Commuters: Each goes faster if he drives, but if all drive each goes slower than if all take busses;
- Soldiers: Each will be safer if he turns and runs, but if all do more will be killed than if none do;
- Fishermen: When the sea is overfished, it can be better for each if he tries to catch more, worse for each if all do;
- Peasants: When the land is overcrowded, it can be better for each if he or she has more children, worse for each if all do. . ..
- There are countless other cases. It can be better for each if he adds to pollution, uses more energy, jumps queues, and breaks agreements; but if all do these things, that can be worse for each than if none do. . .. In most of these cases the following is true. If each rather than none does what will be better for himself, or his family, or those he loves, this will be worse for everyone.[1]

Some of the situations Parfit cites (fishermen, peasants) may be described in terms of Hardin's (1968) "tragedy of the commons"—overuse by individuals of a common resource; all are multi-person prisoner's dilemma games.

Sometimes, depending on circumstances, the lecture game is followed by a discussion (without revealing what anyone actually chose) of why a person should choose *Y* in this anonymous game. One reason people often give for choosing *Y* is that they believe that many or most of the other players will choose *Y*. This rationale is common but not strictly rational. Since the cost of a *Y*-choice is constant at \$200, regardless of what anyone else chooses, what anyone else chooses should, in theory, not influence any person's individual choice. What others choose in this game will indeed have a strong effect on a player's earnings. But, within this one-shot, anonymous game, no player can have any direct influence on another player's choice. It is true that if a person chooses *Y* she will be better off if all other players chose *Y*, and she earns \$1000, than if any other player or players had chosen *X*. But, given that all other players chose *Y*, she would be still better off if she alone had chosen *X* and earned \$1200.

Parfit (1984, pp. 100–101) considers choosing *Y* because you believe that others will choose *Y* as a rationale for cooperating and makes an appealing suggestion for quantifying this rationale. First, he suggests, determine how much you would earn if all players, including you, defected (chose *X*); in the lecture game, that would be \$300. Then make your best guess of how many other players will probably cooperate (how many would choose *Y*). Then determine how much you would earn if you also cooperated. If the amount you earned by cooperating is at least as great as the amount you would earn if all players including you defected then, according to Parfit, you should cooperate. In order to earn \$300 or more by cooperating in the lecture game, at least 2 *other* players would also have to cooperate. If you were a player in the lecture game, knowing that about half of the 10 players usually cooperate, and you wanted to follow Parfit's suggestion, you would cooperate. However, as Parfit admits, his suggestion does not resolve the dilemma. If two other lecture-game players chose *Y* and you also chose *Y*, you would earn \$300 whereas, if you had chosen *X*, you would have earned \$500. Why should you not have chosen *X*? Although Parfit's suggestion has intuitive appeal, and corresponds to people's verbal rationalizations, it is not a good explanation, in itself, of why people behave altruistically.

A second reason that people may give for choosing *Y* is simply that choosing *Y* is altruistic (or good or generous or public-spirited) and they see themselves as altruistic (or good or generous or public-spirited) people. Choosing *X* would not fit in with their self-image and would make them feel bad about themselves. We believe that this seemingly vacuous reason for cooperating is actually a good reason and worth discussing in some detail. But it is first necessary to outline a behavioral concept of the self—to answer the question: Where does one person end and another begin? Skinner (1969) claimed that the skin is not important as a boundary. We agree with this claim, but for reasons different from Skinner's. Skinner meant

[1] See a review of Parfit's interesting and valuable book by Rachlin (2010a).