Economics Letters 102 (2009) 192–194

Contents lists available at ScienceDirect

Economics Letters

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

The role of egalitarian motives in altruistic punishment☆

Tim Johnson [a,b], Christopher T. Dawes [c], James H. Fowler [c], Richard McElreath [d], Oleg Smirnov [e,*]

[a] Department of Political Science, Stanford University, Palo Alto, CA 94305, USA
[b] Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, Berlin 14195, Germany
[c] Department of Political Science, University of California, San Diego, CA 92093, USA
[d] Department of Anthropology, University of Utah, Salt Lake City, UT 84112, USA
[e] Department of Political Science, Social and Behavioral Sciences Building, Stony Brook University, Stony Brook, NY, 11794-4392, USA

ARTICLE INFO

Article history:
Received 18 May 2007
Received in revised form 19 November 2008
Accepted 5 January 2009
Available online 13 January 2009

Keywords:
Experiments
Public goods games
Altruistic punishment
Egalitarian motives

JEL:
C91
D31
D63
D64
H41

ABSTRACT

We conduct experiments in which subjects participate in both a game that measures preferences for income equality and a public goods game involving costly punishment. The results indicate that individuals who care about equality are those who are most willing to punish free-riders in public goods games.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Humans foster cooperation by sanctioning free-riders at a personal cost (Fehr and Fischbacher 2003; Henrich et al., 2006). Fehr and Gächter (2002) contend that this altruistic punishment results from anger toward norm violators. Fowler et al. (2005) propose that egalitarian motives inspire the destruction or transfer of resources in order to produce equal distributions of wealth (e.g., Dawes et al., 2007). Here we examine whether the same people who engage in egalitarian behavior also engage in altruistic punishment—a possibility suggesting that both behaviors result from a common disposition.

Ultimately, both motivations could trigger punishment, but separating them is difficult in scenarios where altruistic punishment occurs. Resource distributions in those games correlate perfectly with choices to cooperate or defect, thus efforts to reduce others' incomes punish free-riders and reduce inequality.

By eliminating this confound, past experiments offer insight into the mechanisms spurring altruistic punishment. For example, Falk et al. (2005) set punishment costs equal to the amount punishment reduces incomes, thereby preventing individuals from reducing inequality between themselves and those they punish. The fact that subjects continue to punish under these conditions suggests that norm enforcement can instigate punishment. Dawes et al. (2007), on the other hand, isolate egalitarian motives in an experiment without cooperation norms. Their results show that individuals incur costs reducing and augmenting others' incomes in order to create equal divisions of wealth. Thus, egalitarian motives might initiate punishment.

Given the plausibility of both motives, we examine whether those who engage in egalitarian behavior also engage in altruistic punishment. We perform an experiment in which subjects play both a *random income* game measuring inequality aversion and a modified public goods game with punishment. Game order is randomized and statistical analyses control for income differences between sessions. Our results suggest that those who exhibit stronger preferences for equality are more willing to punish free-riders in public goods games: both behaviors may result from a common disposition.

☆ The authors declare that they have no competing financial interests. The authors thank the Center for Adaptive Behavior and Cognition at the Max Planck Institute for Human Development and the UC Davis Institute of Government Affairs for generous research support.

* Corresponding author. Tel.: +1 786 566 1185; fax: +1 631 632 4116.
E-mail addresses: timj@stanford.edu (T. Johnson), cdawes@ucsd.edu (C.T. Dawes), jhfowler@ucsd.edu (J.H. Fowler), richard.mcelreath@anthro.utah.edu (R. McElreath), oleg.smirnov@stonybrook.edu (O. Smirnov).

0165-1765/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2009.01.003

2. Experimental design[1]

In the *random income* game, subjects are divided into groups of four anonymous members. Each player receives a sum of money randomly generated by a computer. Subjects see the payoffs of other group members for that round and are given an opportunity to reduce others' incomes by allocating "negative tokens." Each negative token reduces the purchaser's payoff by 1 monetary unit (MU) and decreases the payoff of a targeted individual by 3 MUs. Groups are randomized after each round to prevent reputation from influencing decisions; interactions are strictly anonymous and subjects know this. Since there is no normative rationale for income reduction in this experiment, we refrain from using the term "punishment" and, instead, use variations of the term "costly reduction." For replication purposes we also conducted an experiment in which subjects played a random income game that permitted the additional opportunity to pay to *increase* target incomes; it did not affect our substantive results (online Appendix).

In the *public goods* game with punishment, the same subjects are again divided into groups of four anonymous members. Each player receives an amount of money and is given the opportunity to contribute some or all of this endowment to a common pool. Once contributions to the common pool have been made, the pool's value is multiplied such that the group income is maximized when all contribute, but personal income is maximized by withholding contributions regardless of the behavior of other group members. After multiplication, the common pool is distributed equally among players. In order to distinguish egalitarian reductions from norm enforcement, the redistributed income from the common pool is *augmented randomly*, thereby allowing us to distinguish between the motives driving punishment. Subjects are then given the opportunity to reduce others' incomes by distributing "negative tokens." As in the random income game, each negative token reduces the purchaser's payoff by 1 monetary unit (MU) and decreases the payoff of a targeted individual by 3 MUs. Similarly, groups are randomized after each round and interactions between players are anonymous.

3. Results

In the random income experiment, income reduction was frequent: 62% of participants reduced others' incomes at least once; 31% did so five or more times (out of fifteen possible times). The total amount of negative tokens received increased with the relative income of the target (Fig. 1 a). Since reducing others' incomes was costly and yielded no material gain, we might expect income reduction to decline over time as subjects learn about the game. Period-specific purchases of negative tokens show no consistent pattern over time. The mean negative tokens received in period 5 (4.08 MU) is actually higher than that received in periods 1–4 (3.98 MU). Therefore, initial mistakes cannot explain individuals' willingness to reduce others' incomes. Similar to Fehr and Gächter (2002), income reduction occurred frequently in the public goods game. While the random income experiment allowed only for the costly reduction of group inequality, the public goods game also allowed for the punishment of non-contributors. The goal of our research design is to identify, using the random income experiment, the degree to which subjects prefer egalitarian outcomes and, then, to examine if those who exhibit egalitarian preferences punish at higher rates in the public goods game. When subjects who exhibit egalitarian preferences in the random income game punish low contributors in the public goods game, we interpret that behavior as evidence that

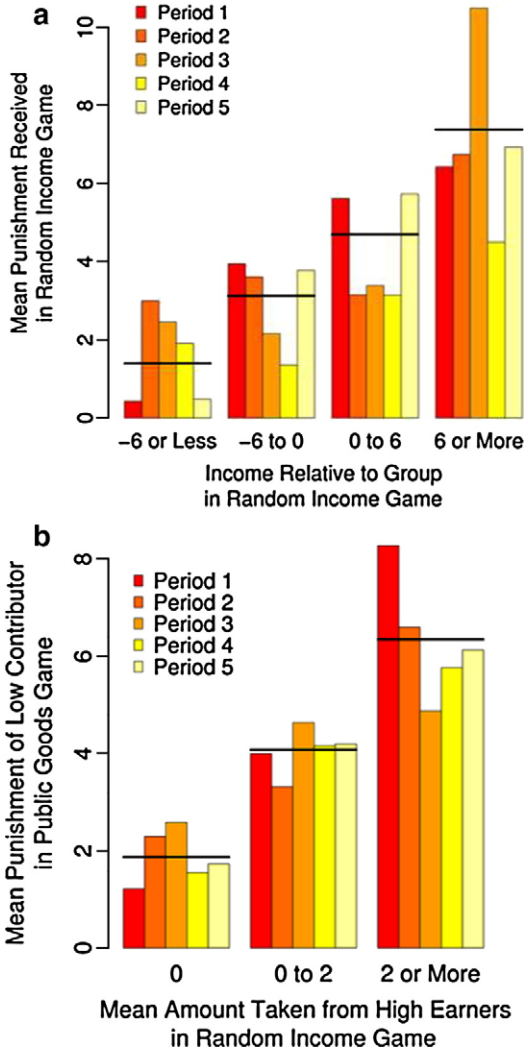

Fig. 1. a) Mean income reduction received in each period as a function of relative income. Relative income is target income minus mean income of other three group members. b) On average, people who engage in the costly reduction of high earners' incomes (high earners = those earning more than the group average) in the random income game send more punishment to the lowest contributor in a group in the public goods game with random payoff. In both panels, solid horizontal lines indicate the average punishment across all periods.

altruistic punishment and egalitarian behavior could result from a common source (cf Fehr and Schmidt, 1999).

Fig. 1b shows that the same subjects who assign negative tokens to high earners in the random income experiment also spend significantly more to punish low contributors in the public goods game. This figure, however, does not account for subjects who might be willing to give negative tokens to others regardless of others' incomes or contributions. Ideally, we would like to have a measure for the willingness to assign negative tokens as a *function of the target's relative income*. To create such a measure, we use the 15 observations for each subject (subjects play a total of 5 rounds and in each round they make decisions concerning 3 other group members) and plot the subjects' purchases of negative tokens against the target's relative income. We then fit a line to the data produced by each subject (Fig. 2a). The intercept indicates the subject's general willingness to reduce others' incomes, while the slope represents the degree to which the subject is motivated to reduce the income of those who earn the most and to refrain from reducing the income of those who earn the least. We denote the slope *income sensitivity*; it serves as a simple measure of egalitarian preferences. We show in the online

[1] Detailed discussion of experimental procedures, including the exact wording of participant instructions, is available in our online appendix at <http://jhfowler.ucsd.edu>.

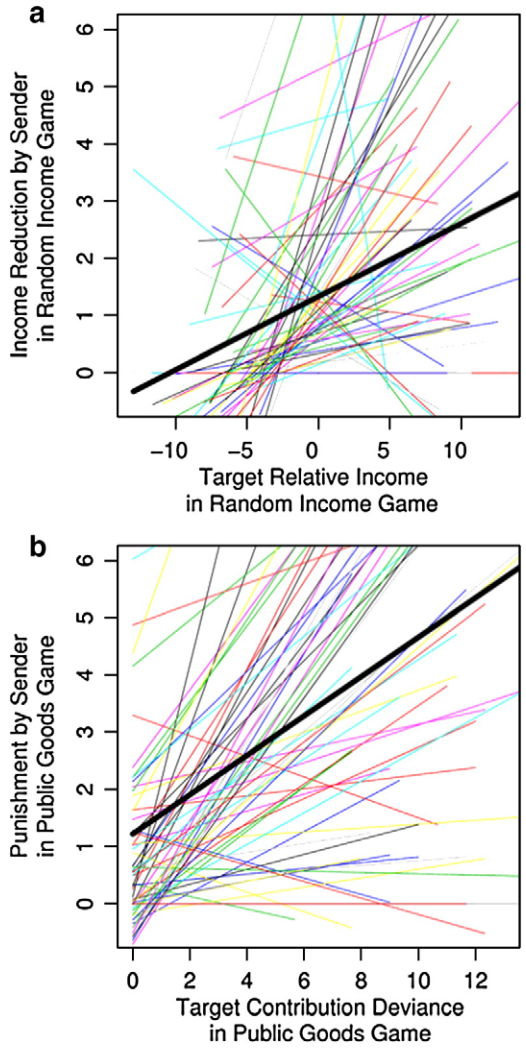

Fig. 2. a) Best fitting lines for each subject that describe the effect of target relative income on net income reduction in the random income game. b) Best fitting lines for each subject that describe the effect of target (negative) contribution deviance on net income reduction in the public goods game with random payoff. Most lines have a positive slope, indicating a general desire to reduce the income of the wealthy and to punish free-riders. Only a handful of subjects have negative income sensitivity (9%) or negative contribution sensitivity (8%). The solid line in both panels shows the average sensitivity for all subjects.

appendix that income sensitivity correlates significantly with emotional responses to inequality.

We also create an index of altruistic punishment. In Fig. 2b we plot a line that describes, for each subject, the relationship between a target's contribution behavior and the subject's willingness to punish the target. We label the slope *contribution sensitivity*.

To determine the relationship between income and contribution sensitivity, we construct a model that jointly estimates income and contribution sensitivity for each subject (online Appendix). The results from our model indicate that the relationship between income sensitivity and contribution sensitivity is significantly positive ($p=0.005$) at 0.50 with a 95% confidence interval: [0.13, 0.97]. Each unit increase in egalitarian behavior in the random income game, as reflected in our income sensitivity measure, is associated with a half unit increase in the desire to engage in altruistic punishment in the public goods game.

If a common disposition underlies both behaviors, income sensitivity (estimated from punishment in the random income game) should be a strong predictor of punishment in the public goods game for these subjects. We calculate predicted values for subject punishment in the public goods game by applying income sensitivities estimated in the random income game to the incomes observed in the public goods game. The correlation between predicted and observed punishment in this out-of-sample test is significant (rho = 0.59, Spearman's rank test, $p<0.0001$). These and other sensitivity tests (*including a full-scale replication on a different sample* — see online Appendix) show that subjects act consistently across both games, supporting the argument that a common disposition underlies egalitarian preferences and altruistic punishment.

The evidence here is the first to demonstrate empirically a link between egalitarian motives and altruistic punishment. As predicted by Fehr and Schmidt's (1999) formal model of egalitarian behavior, a concern for equality can yield cooperation in groups where decentralized punishment is a possibility, even if that punishment is individually costly.

Although we interpret our results as indicating that a common disposition might underlie both behaviors, we rule out three interpretations of this disposition. One is that some individuals like to reduce others' incomes regardless of the context, so they will appear to be consistent across treatments. When we measured income and contribution sensitivities, we accounted for this alternative possibility by controlling for the average tendency to reduce others' incomes (the intercepts on lines in Fig. 2). A second possibility is that people in our experiment were motivated to engage in spiteful behavior to "win" by obtaining a higher relative income. When we estimated income sensitivities based on income relative to the self rather than the group, the alternative sensitivity measure does a much poorer job predicting behavior in the public goods game. The relationship between income sensitivity and contribution sensitivity ceases to be significant, even when we estimate contribution sensitivity with respect to the self as well ($p>0.10$ for all specifications). A third possibility is that consistent behavior in both games is spurious: some people may have a tendency to be "moral" or susceptible to social pressures stipulating that they should contribute to "fair" outcomes. This criticism, however, raises the question of what it means to be "moral" or "fair." Our experiments show a distinctive content to this moral behavior: *the same people who pay to enforce equality are those who are likely to punish free-riders*. The argument that some people are moral and others are not thus offers no explanation for why these two particular behavioral tendencies would be related. We argue that the behaviors share a common source because egalitarian motives help to drive altruistic punishment, which significantly effects cooperation.

References

Dawes, C.T., Fowler, J.H., Johnson, T., McElreath, R., Smirnov, O., 2007. Egalitarian motives in humans. Nature 446, 794–796.
Falk, A., Fehr, E., Fischbacher, U., 2005. Driving forces behind informal sanctions. Econometrica 73, 2017–2030.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition and cooperation. Quarterly Journal of Economics 114, 817–868.
Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415, 137–140.
Fehr, E., Fischbacher, U., 2003. The nature of human altruism. Nature 425, 785–791.
Fowler, J.H., Johnson, T., Smirnov, O., 2005. Egalitarian motive and altruistic punishment. Nature 433 10.1038/nature03256.
Henrich, J., et al., 2006. Costly punishment across human societies. Science 312, 1767–1770.