ELSEVIER

Journal of Economic Behavior & Organization
Vol. 65 (2008) 489–505

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

The power of conventions: A theory of social preferences

Jing Li*

Department of Economics, Duke University, 213 Social Sciences Building, Durham, NC 27705, United States

Received 7 February 2005; accepted 16 February 2006
Available online 12 October 2006

Abstract

People often act as if they care about others' welfare as well as their own (i.e. have "social preferences"). One plausible assumption is that people have preferences for social implications of their actions, determined by exogenous "conventions", in addition to the material consequences of actions. I construct games with conventions using the psychological games framework developed in Geanakoplos et al. [Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games and Economic Behavior 1, 60–79]. With a notion of distributional convention combining efficiency and fairness, I show equilibrium behavior reflects social preferences. The model yields tight and testable predictions consistent with a large body of experimental results, is parsimonious, and is suggestive of further studies, both experimentally and theoretically.

JEL classification: A14; D63; D64; C72

Keywords: Convention; Social preferences; Coordination; Fairness; Psychological games

1. Introduction

Social preferences refer to the phenomena that people seem to care about certain "social" goals, such as the well-being of other individuals or a "fair" allocation among members in society, in addition to their own material benefits. The evidence is ample; Camerer (2003), Kahneman and Tversky (2000) and Sobel (2005) all contain extensive accounts of both real life examples and experimental results.

* Corresponding author. Tel.: +1 919 660 1864; fax: +1 919 684 8974.
E-mail address: jingli@econ.duke.edu (J. Li).

0167-2681/$ – see front matter
doi:10.1016/j.jebo.2006.02.005

Depending on the fine details of the environment, social preferences exhibit many patterns: sometimes people reciprocate, rewarding kindness and punishing unkindness; sometimes people show unmotivated altruism; sometimes people act in the entire group's interest, even if it hurts some individuals in the group. The following experimental results are illustrative of the variety of the patterns of social preferences.

1. In an experiment of the dictator game,[1] subjects choose between pairs of (self, other) allocations. About 50 percent of the subjects choose (375, 750) over (400, 400) (Charness and Rabin, 2002).
2. Subjects first play a dictator game, choosing between (self, other) allocations of (US$ 10, US$ 10) and (US$ 18, US$ 2).[2] Then some choices were randomly selected and realized. Finally, those subjects whose decisions were not realized were given the choice of evenly splitting US$ 12 with a person whose first offer was (18, 2) or evenly splitting US$ 10 with a person whose first offer was (10, 10). The one who was not chosen for the interaction receives 0. About 74 percent of the subjects chose the latter (Kahneman et al., 1986).
3. Two players sequentially make private contributions to a public good, which is supplied either at the maximum of the two contributions (the best-shot game) or at the sum of them (the summation game). The first-mover has a smaller marginal-willingness-to-pay than the second-mover.[3] Subjects behave very differently in experiments of these two games: the first-mover typically does free ride in the best-shot game, but not in the summation game; in addition, when the first-mover contributes 0, the second-mover responds by contributing 0 almost three times more often in the summation game than in the best-shot game (Andreoni et al., 2002).
4. The ultimatum game is another famous example where theoretical prediction fails.[4] In laboratory experiments, it is rarely observed that the proposer demands the entire sum, and offers of 20–30 percent are frequently rejected. Offers of 50/50 split are observed in all experiments, often being the mode. With stakes between US$ 5 and US$ 20 and as high as US$ 100, the average offer is around 40 percent of the sum. Moreover, the rejection rate seems to depend on possible offers the proposer *did not* make. For instance, when the proposer chooses between offering 20 or 75 percent, an offer of 20 percent is rejected 33 percent of the time; however, when the proposer's choice set is changed to (20 percent, 87.5 percent), the rejection rate for an offer of 20 percent drops to 16 percent (Brandts and Sola, 2001; Camerer and Thaler, 1995; Charness and Rabin, 2002; Thaler, 1988).

It turns out to be a challenging task to explain *all* these complex patterns in a parsimonious model. The existing literature on social preferences includes two main classes of models, the distributional preferences models and the reciprocal preferences models.

Distributional preferences models assume players have preferences over final payoff *allocations*. For example, Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) explain the

[1] The dictator game is a one-player game in which the player allocates a sum of money between himself and another inactive player. Obviously, traditional game theory predicts the player takes everything himself.

[2] Consistent with the previous result, in this experiment 122 out of 161 subjects chose (10, 10) over (18, 2).

[3] Subgame perfect equilibrium predicts the same outcome in both games: the first-mover free rides and the second-mover provides the entire public good.

[4] In the ultimatum game, two players split a sum of money. The proposer moves first and makes an offer to the responder. If the responder accepts the offer then the money is divided as such. If the responder rejects it, then both players get nothing. In all subgame perfect equilibria of this game, the proposer makes an offer of no greater than the minimum share he can offer, and the responder accepts whatever she is offered.