Available online at www.sciencedirect.com
ScienceDirect
Games and Economic Behavior 60 (2007) 287–306
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Endogenous transfers in the Prisoner's Dilemma game: An experimental test of cooperation and coordination

Gary Charness [a,*], Guillaume R. Fréchette [b], Cheng-Zhong Qin [a]

[a] *Department of Economics, University of California at Santa Barbara, Santa Barbara, CA 93106-9210, USA*
[b] *New York University, Department of Economics, 269 Mercer Street, 7th floor, New York, NY 10003, USA*

Received 21 February 2005

Available online 29 January 2007

Abstract

We test a two-stage compensation mechanism for promoting cooperation in Prisoner's Dilemma games. Players first simultaneously choose binding non-negative amounts to pay their counterparts for cooperating, and then play the induced game knowing these amounts. In our games, all payment pairs *consistent* with mutual cooperation in subgame-perfect equilibrium transform these games into *coordination games*, with *both* mutual cooperation and mutual defection as Nash equilibria in the second stage. When endogenous transfer payments are not permitted, cooperation is much less likely. Mutual cooperation is most likely when the (sufficient) payments are identical, and it is also substantially more likely with payment pairs that bring the mutual-cooperation payoffs closer together. Both the Fehr–Schmidt and Charness–Rabin models predict that transfers that make final payoffs closer are preferred; however, they do not explain why equal transfers are particularly effective. Transfers are also effective in sustaining cooperation even when they are imposed and not chosen.

JEL classification: A13; B49; C72; C78; C91; K12

Keywords: Prisoner's dilemma; Endogenous transfer payments; Compensation mechanism; Coase theorem; Coordination games; Equilibrium selection

* Corresponding author. Fax: +1 805 893 8830.
E-mail addresses: charness@econ.ucsb.edu (G. Charness), frechette@nyu.edu (G.R. Fréchette), qin@econ.ucsb.edu (C.-Z. Qin).
URLs: http://www.econ.ucsb.edu/~charness (G. Charness), homepages.nyu.edu/~gf35 (G.R. Fréchette), http://www.econ.ucsb.edu/faculty/~qin/ (C.-Z. Qin).

doi:10.1016/j.geb.2006.10.010

1. Introduction

The prisoner's dilemma is by far the most famous example of a game with a unique Pareto-inefficient Nash equilibrium. The chief characteristic of this game is that while there are substantial gains that could be attained through cooperation, non-cooperation (*defection*) is dominant for each player. The theoretical result is that all players defect, even though joint defection leaves each player with less than he or she could have obtained through mutual cooperation. A multitude of experiments have been conducted on the Prisoners' Dilemma (see Rapoport and Chammah, 1965; Dawes, 1980, and Roth, 1988 for surveys of these experiments). The central finding in these studies is that mutual cooperation is indeed rather rare in the Prisoner's Dilemma. Since players always do better with respect to their individual payoffs by defecting, few people elect to cooperate in this environment, leading to poor social outcomes. It is thus desirable to design mechanisms that will implement the efficient outcome.

Coase (1960) presents an example involving a rancher and a farmer, in which the rancher's cattle stray onto the farmer's property and damage the crops beyond the benefit to the rancher. Coase argues that even if the rancher's cattle are legally allowed to trespass, the efficient outcome, as in the case where the rancher's cattle are legally prohibited from trespassing, will still result because the farmer would then have an incentive to pay the rancher to cooperate (reducing the number of straying cattle). That is, with well-defined property rights, no transaction costs, and fully symmetric information, efficiency is neutral to the assignment of responsibilities for damages; this result has come to be called the *Coase theorem.*

Varian (1994) presents a general two-stage compensation mechanism that can be seen as being complementary to the Coasian approach.[1,2] It implements efficient outcomes through subgame-perfect equilibria in a wide range of environments with externalities, including Prisoner's Dilemma games with certain specifications of the payoffs.[3] The mechanism provides a formalization of bargaining involved in the Coase theorem, and it does not involve a regulator or central planner mandating taxes or transfer payments; instead it relies upon the parties to design transfer payments that leads to the efficient outcome. In essence, the Prisoner's Dilemma can be seen as an environment with a two-sided externality.

Applying this mechanism to a Prisoner's Dilemma game, each party would make a binding pre-play offer to pay the other for cooperating in stage 1; upon observing these offers, each party then chooses to cooperate or to defect in the Prisoner's Dilemma game in stage 2. A natural solution concept is subgame-perfect equilibrium (henceforth SPE); while one wishes to offer enough to induce the other to cooperate, it is best to offer the minimum amount that is required to achieve this goal.[4] Qin (2002) characterizes the conditions on payment pairs that are necessary and sufficient to "induce the players to cooperate" (to be defined shortly).[5]

---

[1] In fact, Varian presents two mechanisms, one of which is general and one of which only works for certain Prisoner's Dilemma games.

[2] Varian (1995) explicitly spells out the connection with Coase.

[3] Moore and Repullo (1988) show that, given certain assumptions, almost all choice rules can be implemented by subgame-perfect equilibria. The compensation mechanism seems to be considerably simpler than the examples provided in Moore and Repullo (1988).

[4] See also Ziss (1997) where it is shown that the efficient outcome is not among the set of possible SPE outcomes for certain Prisoner's Dilemma games.

[5] Jackson and Wilkie (2003) consider more general strategy-dependent transfer payments that players may offer to each other before playing a game in strategic form. For example, with a Prisoner's Dilemma game, players can offer among permissible transfer payments those that will be carried out only when both players defect or only when Player 1