Games and Economic Behavior ••• (••••) •••–•••

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Fostering cooperation through the enhancement of own vulnerability

Anita Kopányi-Peuker [a,b,*], Theo Offerman [a,b], Randolph Sloof [a,b]

[a] University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
[b] Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 5 May 2014
Available online xxxx

JEL classification:
D03
D81
D83

Keywords:
Prisoner's dilemma
Cooperation
Endogenous punishment

ABSTRACT

We consider the possibility that cooperation in a prisoner's dilemma is fostered by people's voluntary enhancement of their own vulnerability. The vulnerability of a player determines the effectiveness of possible punishment by the other. In the "Gradual" mechanism, players may condition their incremental enhancements of their vulnerability on the other's choices. In the "Leap" mechanism, they unconditionally choose their vulnerability. In our experiment, subjects only learn to cooperate when either one of these mechanisms is allowed. In agreement with theory, subjects aiming for cooperation choose higher vulnerability levels in Gradual than in Leap, which maps into higher mutual cooperation levels.

1. Introduction

In situations where private interests are at odds with the collective interest, the possibility to punish has proven to be an effective tool to support cooperation. In small groups, contributors are often willing to pay a small cost to punish free riders. This process helps to make free riders behave in agreement with the collective interest (see for instance Yamagishi, 1986; Ostrom et al., 1992; Fehr and Gächter, 2000, 2002). In practice, the possibility to punish free riders will often be limited though, because people are protected by property rights. For instance, a neighbor who refrains from contributing to the local public good can presumably be disciplined by damaging his car, but the person who inflicts the damage has to face the possibility that she will be prosecuted in court.

Here we examine a simple version of the conflict between public and private interest. We consider a prisoner's dilemma in which cooperation cannot be contracted and in which the possibility to punish may only endogenously become available. That is, players can only be punished if they voluntarily make themselves vulnerable in the first place, by giving the partner the possibility to punish them. They may do so to signal that they are interested in pursuing mutual cooperation. If the other is willing to punish a free rider at a small cost, then the player's signal to conditionally cooperate becomes credible. Theoretically and in an experiment, we investigate whether and how the possibility of enhancing one's own vulnerability may foster cooperation in the prisoner's dilemma. In particular, we want to know: (i) whether people voluntarily make themselves vulnerable if they have the possibility to do so, (ii) whether cooperation is enhanced when players have made themselves vulnerable and (iii) whether it matters if the trust-building process occurs gradually or in one single step.

* Corresponding author at: PO Box 15867, 1001 NJ Amsterdam, The Netherlands.
E-mail addresses: A.G.Kopanyi-Peuker@uva.nl (A. Kopányi-Peuker), T.J.S.Offerman@uva.nl (T. Offerman), R.Sloof@uva.nl (R. Sloof).

http://dx.doi.org/10.1016/j.geb.2015.10.001
0899-8256/

The trust-building process that we have in mind corresponds to how strangers are reported to build friendships. In his "Moralia", Plutarch already described how a reciprocal exchange of secrets may lead to a relationship in which there is a fear of loss of trust (Gambetta, 2009, p. 66; Plutarch, 1992). Strangers who have exchanged secrets before they interact in a prisoner's dilemma may refrain from free riding if they fear that this will trigger the other to publicly disclose the secret. More recent psychological research shows that feelings of intimacy develop in a dynamic process in which a person discloses personal information, thoughts and secrets and the partner responds in a likewise manner (Altman, 1973; Rotenberg, 1986; Dindia and Allen, 1992; Laurenceau et al., 1998). An important (but possibly unintended) by-product of such feelings may be that people have learned to trust each other when they subsequently interact in a prisoner's dilemma.[1] Interestingly, Derlega et al. (1976) report that reciprocal self-disclosure is especially observed among strangers, as hypothesized by Altman (1973). Once relationships have been established and friends have learned to trust each other, other arguments may take over. For instance, in long term relations the behavior of friends may be disciplined by repeated game considerations.

Economic applications of this process include the formation of cartels and bribery operations. That both sides may benefit from voluntarily enhanced vulnerability was well understood in the Dutch Construction Cartel uncovered in a parliamentary inquiry in 2002 (VARA/Zembla, 2009). It was revealed that representatives of Dutch construction companies slowly but surely built good relationships with civil servants of the Ministry of Transport. Gradually civil servants were buttered up by the companies, starting small with a good bottle of wine or a fancy dinner. If the civil servant took the bait, the gifts increased, from a refurbished home garden to the building of a fireplace or dormer windows. Sometimes the civil servant could even be seduced to a visit to the exclusive brothel Yab Yum, or was treated on golf trips abroad traveling with the company's private Jet. After the civil servant had opened up to blackmail, the construction company knew that it could easily follow up by suspicious overruns on projects supervised by the civil servant. That the relationship was risky was proven in the aftermath. In total 1300 construction companies were fined for a total of 306 million euro and paid an additional 100 million euro in settlements.[2]

In this paper, we model the trust-building process in a stylized three stage game. In the first stage, the two players voluntarily decide upon the extent to which they make themselves vulnerable for punishment. In the second stage, after having been informed of each other's own possible punishment level, the two partners decide whether or not to cooperate in a prisoner's dilemma. Based on the observed outcome, each player decides in the third stage whether or not to punish the partner at a small cost. If a player decides to punish, the partner loses an amount equal to her own possible punishment level chosen in the first stage.

We consider two different trust-building mechanisms. In the "Gradual" variant, players may build trust in small steps, while observing the partner's willingness to go along in this process. This variant has the advantage that a player can condition the own possible punishment level on the partner's possible punishment level. It agrees with the empirical observation that trust is often formed in small incremental steps. The possibility of "starting small" can be advantageous, as has been shown by Andreoni and Samuelson (2006) for a repeated prisoner's dilemma and by Weber (2006) in a team production game with Pareto ranked equilibria.[3]

In the "Leap" variant, players decide whether or not to take a leap in the dark by simultaneously choosing an own possible punishment level without the possibility of conditioning it on the partner's level. This variant may correspond to situations where players do not have the time to build trust in small steps, or where it is too costly to engage in a slow gradual process. Theoretically, using equilibrium refinements, we show that the Gradual treatment may support higher levels of vulnerability and cooperation than the Leap treatment. In essence, the possibility to condition one's own vulnerability on the partner's vulnerability in the Gradual variant allows players to turn themselves into unconditional cooperators, provided that the partner does the same. This possibility does not exist in the Leap variant.

We test the performance of the mechanisms in an experiment where in each round subjects are rematched with a different partner in their matching group. In a Control treatment where players do not have the possibility to make themselves vulnerable, cooperation is not sustained. In the treatments with a trust-building mechanism, after some initial aversion a significant fraction of the subjects actively employs the mechanism to achieve cooperative outcomes. The two trust-building treatments do not differ in the extent to which subjects use the mechanism. Instead, conditional on the mechanism being

[1] Gossip may also serve as a form of self-disclosure that promotes trust through the enhancement of vulnerability. If a worker communicates damaging information about a superior to a colleague, he faces the risk that the colleague reveals this to the superior. If instead the colleague reciprocates with another negative story, a bond may be formed which may help the workers to solve free rider incentives in the work place. See Sommerfeld et al. (2007) for a discussion of other functions of gossip.

[2] Other applications of the mechanisms studied in this paper include the use of hostages and trust-building among criminals. Since Roman times, hostages have been exchanged to enforce truces and treaties. In some cases hostages were voluntarily exchanged (Schelling, 1960, pp. 135–137; Lee, 1991; Herrmann and Palmieri, 2005). Posting a hostage does not bind an agent to a specific action, but creates a stick that the partner can use later on. The use of the hostage cannot be contracted either, so the partner can punish after any action of the agent. Williamson (1983) discusses how the exchange of hostages can be used to support trade in contractual relations hampered by holdup threats; he argues that "... the use of hostages to support exchange is widespread and economically important" (p. 537). Gambetta (2009) describes some examples where trust-building mechanisms are used to support criminal activities. For instance, mafia bosses have been reported to bring their wives to potentially explosive dinners to signal their own willingness not to start a shooting. Pedophiles are often asked to share compromising photos before they get access to a child-pornography website.

[3] The mechanism behind the result in those papers is quite different though. In Andreoni and Samuelson (2006), players differ in their taste for cooperation when they participate in a twice-played prisoner's dilemma. If the stakes are larger in the second stage game, players have a larger willingness to invest in the first stage to achieve cooperation in the second. In Weber (2006), coordination on the efficient equilibrium is facilitated if teams start small and new members observe the history of the team before they enter.

Please cite this article in press as: Kopányi-Peuker, A., et al. Fostering cooperation through the enhancement of own vulnerability. Games Econ. Behav. (2015), http://dx.doi.org/10.1016/j.geb.2015.10.001