Prisoners and their dilemma☆

Menusch Khadjavi, Andreas Lange*

Department of Economics, University of Hamburg, Von Melle Park 5, 20146 Hamburg, Germany

A R T I C L E   I N F O

A B S T R A C T

We report insights into the behavior of prisoners in dilemma situations that so famously carry their name. We compare female inmates and students in a simultaneous and a sequential Prisoner's Dilemma. In the simultaneous Prisoner's Dilemma, the cooperation rate among inmates exceeds the rate of cooperating students. Relative to the simultaneous dilemma, cooperation among first-movers in the sequential Prisoner's Dilemma increases for students, but not for inmates. Students and inmates behave identically as second movers. Hence, we find a similar and significant fraction of inmates and students to hold social preferences.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In 1950 Merrill M. Flood and Melvin Dresher conducted the first experiment of what they referred to as *A Non-cooperative Pair* (Flood, 1952, p. 17). Soon after this game was coined *Prisoner's Dilemma* by Albert W. Tucker based on the illustrious story of two suspects who are taken into custody and questioned separately (Luce and Raiffa, 1957; Roth, 1993). In this classic story of Albert W. Tucker cooperation means to keep quiet about the crime, while defection means to testify against the other suspect. Ever since virtually countless research articles and books have dealt with the Prisoner's Dilemma (e.g. Cooper et al., 1996; Ortmann and Tichy, 1999; Brandts and Charness, 2000; Clark and Sefton, 2001). It is one of the best known and most prominent examples used in almost all introductory game theory classes.

It is surprising, however, that in 60 years of research on the Prisoner's Dilemma and its applications,[1] to the best of our knowledge no study was published that reports the behavior of prisoners themselves in dilemma situations that

[1] Applications range from the arms race or oligopolistic competition for the two-player game, to voluntary public good provision reflected in an n-player Prisoner's Dilemma (e.g., Axelrod, 1984). Besides studying one-shot dilemma situations, repeated interaction has received much interest. For example, Axelrod (1984) compares the performance of different strategies in an iterated Prisoner's Dilemma with Anatol Rapoport's simple Tit-for-Tat strategy proving to be especially robust against defection, yet able to achieve cooperation if encountering other cooperative programs.

so famously carry their name. In this paper, we close this gap in the literature: we analyze the outcome of the original Prisoner's Dilemma story, i.e. we study how actual prisoners behave in a simultaneous and a sequential Prisoner's Dilemma. We compare their behavior to that of a conventional subject pool of university students in identical choice situations.

From a methodological point of view, our study combines a conventional laboratory experiment with an artefactual field experiment (Harrison and List, 2004), i.e. a lab-in-the-field experiment with a different subject pool. Such investigation of the robustness of findings from conventional laboratory experiments with students has recently received increasing interest (Levitt and List, 2007). Some studies do not find significant differences between students and subjects recruited from more general pools (e.g. Anderson et al., 2013; Cleave et al., 2010; Stoop, 2012). Other studies focus on specific groups to capture behavioral differences (e.g., financial market professionals in Alevy et al., 2007, real estate brokers and nurses in Jacobson et al., 2011). Our study meanwhile analyzes another special part of society: inmates.

Besides methodological reasons, we regard our study as useful in order to provide insights for behavioral law and economics. Jolls et al. (1998) and Korobkin and Ulen (2000) emphasize the importance of behavioral economic research on bounded rationality, bounded willpower and bounded self-interest for law and economics theory and policy. In this study we focus on bounded self-interest, i.e. the case when individuals hold other-regarding, social preferences. Meier (2007) discusses many settings in which these social preferences have been proven to play an important role for individual decision making. Rightfully convicted inmates are criminals who are addressed by specific policies such as conduct of the police force and courts, but also by prevention and reintegration programs. Hence, it is informative to examine how inmates react to incentives and whether they are similarly altruistic and cooperative when compared to a conventional subject pool. Naturally, a prison population differs along several dimensions from a students' subject pool. For instance, in prison specific social norms may exist and post-experiment interaction may be more frequent. We nevertheless regard it as insightful to study decisions of real prisoners in comparison to students, while we may not conclusively identify the source of differences.[2]

Institutionalized populations have been used for early experimental economic studies on token economies (e.g., Ayllon and Azrin, 1965; Phillips, 1968; Battalio et al., 1974) as such institutions appeared to provide particularly controlled environments for economic experimentation (see Kagel, 1972). Several recent studies also investigate the behavior of inmates: Farrington and Welsh (2005) and Petrosino et al. (2006) summarize a number of criminological experiments: often these experiments can be regarded as case study evaluations of specific prevention or correction programs. Few studies take experimental economic methods to this field. Notable exceptions are Chmura et al. (2010) who analyze dictator game giving of 68 male inmates from a South German prison and find no significant difference of behavior compared to a sample of non-prison subjects' results from a meta-study. Given that the dictator game is a good measure for altruistic behavior, they conclude that the prisoners in their study are not inherently more selfish than other subject pools. Birkeland et al. (2011) conduct experiments in Norway. They confirm dictator game findings of Chmura et al. and add evidence that this similarity does not depend on whether the game is played between prisoners, subjects from the general public or an interaction of the two groups. Hence, the two studies provide evidence that (male) prison subjects, students and subjects drawn from the general public are comparably altruistic in the dictator game.

In our experiment, we study social preferences beyond such altruism: Based on a simultaneous and a sequential version of a Prisoner's Dilemma game, we measure conditional and unconditional cooperativeness of inmates and students and provide a nuanced picture to the literature:[3] We find that in the simultaneous Prisoner's Dilemma only 37% of students choose to cooperate, relative to an individual cooperation rate of 56% among inmates. Inmates are therefore able to better solve *their* classical dilemma situation than students: on average one can expect inmates to mutually cooperate in 30% of cases, while only 13% of students' pairs fully cooperate.

In contrast, we find an equal share of about 60% of students and inmates to return cooperation in response to a cooperating first-mover in the sequential Prisoner's Dilemma. In both subject pools defection by the first mover will be answered with defection. We thereby obtain results on conditional cooperation consistent with Chmura et al. (2010) and Birkeland et al. (2011) results for dictator game giving. In this sequential Prisoner's Dilemma, 63% of first movers among students choose to cooperate, thereby significantly more than in the simultaneous choice situation. Conversely, among prisoners, the individual cooperation decision of first movers does not significantly differ from the individual cooperation decision in the simultaneous game. As such, collective cooperation rates, i.e. the percentage of cases in which both players cooperate, are significantly larger in the sequential than in the simultaneous game for students (39% vs. 13%), while no such difference is observed for inmates (27% vs. 30%).

The remainder of this paper is organized as follows: we discuss predictions and outline the experimental design and procedures in prison and at university in Section 2. We present and discuss the results of our study in Section 3, before drawing conclusions in Section 4.

---

[2] It would also be interesting to see if behavioral differences can be explained by the severity of the committed crime and to compare the behavior of female and male inmates.

[3] Note that differences may, for example, be driven by gender effects or by the different (and probably less severe) nature of committed crimes.