Cooperation by Indirect Revelation Through Strategic Behavior

ROBERT LAPSON
KGSM, Northwestern University, Evanston, IL 60208, USA

Abstract: The paper deals with a one-shot prisoners' dilemma when the players have an option to go to court but cannot verify their testimonies. To solve the problem a second stage is added to a game. At the first stage the players are involved in the prisoners' dilemma and at the second stage they play another game in which their actions are verifiable. In such a setup the information about the actions chosen at the prisoners' dilemma stage can be revealed through strategic behavior of the players during second stage. A mechanism for such revelation in the extended game is described. It provides an existence of a unique sequential equilibrium, which may be obtained by an iterative elimination of dominated strategies and has a number of desirable properties.

Prisoners' dilemma, a non-cooperative game whose nickname is attributed to A. W. Tucker, is probably the most notorious example in the game theory. It is discussed at length in such monographs as Luce and Raiffa (1957), Rapoport and Chammah (1970), Kreps (1990), Fudenberg and Tirole (1991), Myerson (1991).

The prisoners' dilemma represents a classical example of a contradiction between individual and collective interests, and may serve as a quite realistic description of many economic and social conflicts. In economics, however, cooperation between agents may lead towards inefficiency to prevent which many countries introduce antitrust laws. Nevertheless, even such legislation does not rule out certain types of agreements.

Since the first presentation of the prisoners' dilemma there have been many attempts to "solve" it. Most of these endeavors are summarized by Tsebelis (1990). They can be split into four groups.

The first group of arguments relies on belief that there is no such thing as the prisoners' dilemma in real life. For example, Stinchcombe (1980) asserts that people do not try to maximize their goals, and therefore do not follow rationality assumptions adopted by game theory. Similar views are shared by many proponents of evolutionary biology (see Maynard Smith, 1982). Another branch within the first group can be represented by Howard's (1970) metagames theory.

Approaches of the second group suggest introduction of communication with monitoring and binding contracts (see, for example, Myerson, 1991). The latter can be ensured by existence of an authority which can monitor players' actions and impose obligatory sanctions when the deviations from the terms of the agreement occur.

1 I wish to thank Itzhak Gilboa and Morton Kamien for helpful comments, as well as Leo Hurwicz and Roger Myerson, whose courses in J. L. Kellogg School of Management inspired this study.
The third way to solve the dilemma is to admit that the game is played repeatedly. If the game is played infinitely many times or information is incomplete (i.e., there is a small probability that the game will stop after each round) there are conditions under which always to cooperate for both players is an equilibrium.

Finally, Brams (1975) introduces asymmetry in informational structure of the game, making one of the players a principal and another a follower, who adapts to the principal's choice. If the players can almost always correctly predict rivals' choices, then there is a place for cooperation.

In this paper we use an approach that does not precisely fit any of the groups sketched above. We deal with a one-shot prisoners' dilemma in which the players have an option to go to court but cannot verify their testimonies. The question is what kind of a mechanism can be designed to enforce cooperation in such a game. In search for an answer the original game is extended into a two-stage game, where at the first stage the players are involved in the prisoners' dilemma and at the second stage they play another game in which their actions are verifiable. The payoff at both stages are publicly known, and each player knows before choosing a strategy for the second stage, what his opponent's choice was at the first stage. In such a setup the information about the actions chosen at the prisoners' dilemma stage can be revealed through strategic behavior of the players during the second stage. A mechanism for such revelation in the extended game is described. It provides an existence of a unique sequential equilibrium, which may be obtained by an iterative elimination of dominated strategies and has a number of desirable properties.

The paper does not suggest a way to resolve the dilemma once and for all. Our point is, rather, that in certain economic situations, in which participants are involved in prisoners' dilemma and cooperation is both legitimate and socially desirable, it might be achieved by traditional social institutions even without perfect monitoring, incomplete information or infinite repetition. Instead, we require that the players be able to think "one period" ahead and that there was some relationship between the prisoners' dilemma played today and the game that will be played tomorrow. A remarkable fact is that even a very slight connection between the stages provides extensive opportunities for cooperation if the court and the players behave strategically.

Our model is presented in section 2. Section 3 describes mechanism design. The intuition behind the result is sketched in section 4.

# 1 Setup of the Problem

We assume that two firms (players): I and II, are involved in a two stage game, the first stage of which is a prisoners' dilemma. The firms sign a contract to cooperate. Assume that the firms have endowments big enough for their assets to remain positive after the first stage, so that the utility levels, with which they enter the second stage are given by the matrix