*Chapter 65*

# PAYOFF UNCERTAINTY AND COOPERATION IN FINITELY-REPEATED PRISONER'S DILEMMA GAMES[1]

LAWRENCE M. KAHN

*Cornell University, 264 Ives Hall, Ithaca, NY 14853, USA*

J. KEITH MURNIGHAN

*Northwestern University*

It is well known that for the finitely repeated prisoner's dilemma game, the unique Nash equilibrium strategy for rational players is non-cooperation at each stage of the game. The argument rests on the backwards induction argument described by Luce and Raiffa (1957): neither player should be motivated to choose cooperatively on the last trial. If neither expects the other to choose cooperatively on the last trial, the next to last trial becomes important. And so on, back to the current trial. The logic of the last trial motivates non-cooperative choices back to the point when the players believe that the game is finite; after that point, rational players should never choose cooperatively (Friedman, 1986).

This expectation, however, does not fit the behavior observed in many experiments including those by Rapoport and Chammah (1965), Axelrod (1984), Selten and Stoecker (1986), Roth (1988), and Andreoni and Miller (1993), among others.[2]

Kreps et al. (1982) suggested that uncertainty might be a factor that contributes to this disparity in theoretical predictions and observed behavior. They formulated two different models: Model 1 assumes that each player thinks that his/her opponent might play tit-for-tat instead of persistent non-cooperation; Model 2 assumes that players may be uncertain about their opponent's payoffs, particularly whether they have a dominant strategy and what that strategy is. The central feature of their analysis is that players' uncertainties about their counterparts' strategies or payoffs leads to uncertainty in their expectations about their opponents' likely strategies, which in turn influences their own strategic choices and may lead to cooperation when it would not otherwise be theoretically expected. In this chapter, we summarize the part of our previous research

[1] Portions of this chapter were written while the second author was a Fellow at the Center for Advanced Study in the Behavioral Sciences.

[2] In contrast to finitely repeated PD, cooperative equilibria may exist in games with low probabilities of termination when the future gains from current cooperation (i.e., future Pareto-optimal outcomes) outweigh the gains from current defection. Roth and Murnighan (1983) have shown that a PD with a given probability $p$ of continuation to at least the next round is analytically equivalent to an infinitely repeated game with $p$ as discount factor. Murnighan and Roth (1983) presented data from a dozen different games that supported the basic predictions of their theory.

*Handbook of Experimental Economics Results, Volume 1*

*DOI: 10.1016/S1574-0722(07)00065-0*

(Kahn and Murnighan, 1993) that investigates their second model, testing whether payoff uncertainty can lead to cooperation in finitely-repeated games,[3] one of which was a prisoner's dilemma.

The players in our experiment were either certain or uncertain about their counterpart's payoffs in a finitely-repeated game; theoretically, this uncertainty provided the basis for cooperative equilibria during the game's early play in some conditions but not in others. For all of the games, even for those with a cooperative equilibrium, the non-cooperative outcome was also an equilibrium. Thus, cooperative behavior was frequently not predicted, even given the players' uncertainties about their counterparts' outcomes.

## 1. Methods

The participants were 154 undergraduate volunteers, primarily enrolled in a management class, at the University of Illinois at Urbana-Champaign. Participating gave them a small amount of extra credit in one of their courses (a common practice) and the chance for one of four monetary prizes (of $200, $100, $100, and $100).

## 2. The Experimental Design

We manipulated two basic factors: the strength of the players' roles and their uncertainty about their counterpart's payoffs. We labeled players Strong if they had a dominant strategy; we labeled them Weak if they did not.

Players played one of two games, either Game 1 or Game 2 in Table 1a. In some conditions, they were completely uncertain (i.e., chances were 50–50) whether they were playing one of these two games or one of Games 3 and 4 (also shown in Table 1a). The uncertainty conditions included neither or both players being uncertain of the game they were playing, or one or the other of the Row and Column players being uncertain. As noted in Table 1b, uncertainty for Row was essentially immaterial in a one shot game since they always had a dominant single-period strategy, even when they were uncertain about which game their counterpart was playing. Of course, in the repeated games we study here, even players with a dominant one-period strategy may gain by cooperating

[3] While not based on PD, Camerer and Weigelt's (1988) experiments also bear on the rationales for cooperation identified by Kreps et al. They established a lending game in which a "banker" (B) chooses whether to "lend" to an entrepreneur (E). B is uncertain about E's preferences to repay the loan or renege. Each E played eight rounds of this game against a series of lenders. The results suggested that lenders had "homemade" prior beliefs of the probability that E would prefer to repay the loan even when E gained more by reneging. Specifically, the E players repaid the loan more often and later in the game than predicted by a rational model. Also, see Kahn and Murnighan (1993) for some results on Model 1 and further detail on the findings we report here.

Table 1

(a) The four possible games

Game 1

| | | Column A | Column B |
|---|---|---|---|
| Row | A | (18, 18) | (5, 20) |
| | B | (24, 5) | (12, 12) |

Both players have dominant, non-cooperative strategies; a prisoner's dilemma game

Game 2

| | | Column A | Column B |
|---|---|---|---|
| Row | A | (18, 18) | (5, 10) |
| | B | (24, 5) | (12, 12) |

Row has a dominant, non-cooperative strategy; Column has no dominant strategy

Game 3

| | | Column A | Column B |
|---|---|---|---|
| Row | A | (24, 18) | (15, 20) |
| | B | (18, 5) | (12, 12) |

Both players have dominant strategies: cooperation for Row and non-cooperation for Column

Game 4

| | | Column A | Column B |
|---|---|---|---|
| Row | A | (24, 18) | (15, 10) |
| | B | (18, 5) | (12, 12) |

Row has a dominant, cooperative strategy; Column has no dominant strategy

(b) Uncertainty conditions: The game the players felt they were playing with one-shot game predicted outcomes

| | Strong–Strong | Predicted Outcome(s) | Strong–Weak | Predicted Outcome(s) |
|---|---|---|---|---|
| Neither Uncertain | | | | |
| For Row: | Game 1 | (12, 12) | Game 2 | (12, 12) |
| For Column: | Game 1 | (12, 12) | Game 2 | (12, 12) |
| Row Uncertain | | | | |
| For Row: | Game 1 or 2 | (12, 12) | Game 2 or 1 | (12, 12) |
| For Column: | Game 1 | (12, 12) | Game 2 | (12, 12) |
| Column Uncertain | | | | |
| For Row: | Game 1 | (12, 12) | Game 2 | (12, 12) |
| For Column: | Game 1 or 3 | (12, 12) or (15, 20) | Game 2 or 4 | (12, 12) or (24, 18) |
| Both Uncertain: | | | | |
| For Row: | Game 1 or 2 | (12, 12) | Game 2 or 1 | (12, 12) |
| For Column: | Game 1 or 3 | (12, 12) or (15, 20) | Game 2 or 4 | (12, 12) or (24, 18) |

(c) Predicted periods of mutual cooperation, 20 trial repeated game

| | Strong–Strong | Strong–Weak |
|---|---|---|
| Neither Uncertain | 0 | 0 |
| Row Uncertain | 0 | 0 |
| Column Uncertain | 0 | 19 |
| Both Uncertain | $\geqslant 2$ | $\geqslant 14$ |

over portions of the repeated game. Column players, on the other hand, often had no dominant single-period strategy (they were Weak in our terms) and they therefore were considerably affected by their uncertainty even in the one-shot game.

We refer to Game 1 as Strong–Strong since both players have dominant strategies, in this case non-cooperative (Game 1 is a prisoner's dilemma). Game 2 is Strong–Weak since only the Row player has a dominant strategy; this is not a prisoner's dilemma game. Uncertain players were told that one of two of the games were equally likely (see Table 1b). Certainty or uncertainty was common knowledge: both players knew who was certain or uncertain and what they were certain or uncertain about.

The design resulted in eight cells (Table 1b) created by crossing the four possible uncertainty combinations with the two actual payoff matrices. With player rationality, only three of the eight cells in the design admitted cooperative equilibria for some portion of the 20-round game: Strong–Weak (SW), Column Uncertain; Strong–Strong (SS), Both Uncertain; and SW, Both Uncertain (Table 1c). The proof is available from the authors upon request.

Intuitively, the clearest case for a cooperative equilibrium is the Strong–Weak, Column Uncertain condition. Here, Column perceives a 50% chance that Row's dominant strategy is A, to cooperate. Since Column does not gain even temporarily by playing B in response to Row's A play, Column has a strong motivation to cooperate. Row's motivation to cooperate comes from the idea that cooperation may build Row's reputation for having a dominant cooperative strategy, inducing Column to choose A rather than B. Further, Row knows that Column does not gain by playing B in response to Row's A play – Column's only temptation to choose B and defect is to protect against the chance that they are playing Game 2. The theory, then, suggests that Column should cooperate for all 20 periods and Row should cooperate until the last period. These strategies provide Row with clearly preferable payoffs of 18 each trial and 24 on the last trial rather than payoffs of 12 on every trial after the first.

The two other conditions with cooperative equilibria are less compelling. In SW, Both Uncertain, both parties have an incentive to build a cooperative reputation. However, Row no longer knows whether Column can achieve single-period gains by choosing B. If Row acts defensively to protect against such behavior, cooperation should theoretically break down some time after 14 periods. In SS, Both Uncertain, reputation building for part of the repeated game is again a sequential equilibrium. However, Row may realize that Column may be able to reap a single-period gain by playing B in response to Row's A play. This should lead to an earlier breakdown in cooperation, theoretically after 2 periods.

The five other conditions in Table 1 do not admit any cooperative sequential equilibria: if Column knows that Row is Strong, then cooperation should unravel from the last period to the first. Thus, there are no cooperative equilibria for any game in which Column is certain about Row's payoffs. In the one remaining condition, SS, Column Uncertain, Row knows that Column has a dominant, non-cooperative strategy; therefore, the backwards induction argument applies here as well.

Theoretically, then, the uncertainty in three of the eight conditions provided an opportunity for the players to make cooperative choices, if only to suggest to their counterparts that their payoffs favor cooperative action. Uncertainty in the other games still did not create cooperative equilibria.

Each game was played for 20 rounds; following each round, players received feedback on the results of the prior round. Everyone participated in two sessions: each pair's first bargaining session was randomly chosen from among the eight cells in the design; the second session either duplicated the first, with a different opponent, or switched to either the Both or Neither Uncertain Conditions, also with a different opponent. Data analysis indicated that order had no significant impact on the results.

## 3. Results

This large experiment produced several results. In particular:

1. The two conditions that generated theoretical cooperative equilibria for many trials of the game (SW, Column Uncertain and SW, Both Uncertain) led to significantly more cooperative choices than the other conditions. Participants in these conditions made an average of 13.8 cooperative choices, compared to 9.5 cooperative choices in all of the other conditions. These data suggest that the theoretical analysis of Kreps et al. (1982) has behavioral consequences.
2. The five conditions that did not provide the conditions for a cooperative equilibrium nevertheless led to considerable cooperative behavior. Without a cooperative equilibrium, participants still chose cooperatively almost half the time. This suggests a serious problem for the theory.
3. People seemed to pay no attention to the fact that the games were finite – there was no end game play – until the 17th trial. Figure 1 shows the pattern of cooperative choices across the 20 trials: after the first few trials, cooperative choices were relatively stable for much of the game, even though all of the players knew that the game was finite.
4. The Strong–Weak conditions generated more cooperative choices (average = 13.5) than the Strong–Strong (average = 7.7).
5. When both parties were certain about the payoffs, whether they were SS or SW, they tended to make more cooperative choices (on average, 12.7 versus 9.6). In addition, for both the Strong–Strong and the Strong–Weak game, there was more first round cooperation when both parties were certain about their payoffs than when either or both parties were uncertain about them.

The figure provides an important picture of all of the data; it displays the mean frequency of cooperative choices in each condition across trials, pooled over Row and Column players, who behaved similarly in each condition. After a primarily cooperative first round, the data reveal an almost immediate and substantial drop in cooperation, particularly in the conditions that were less cooperative on the first round. Then, cooperation rates leveled off for most of the rest of the game. Over most of the 20 trials, the

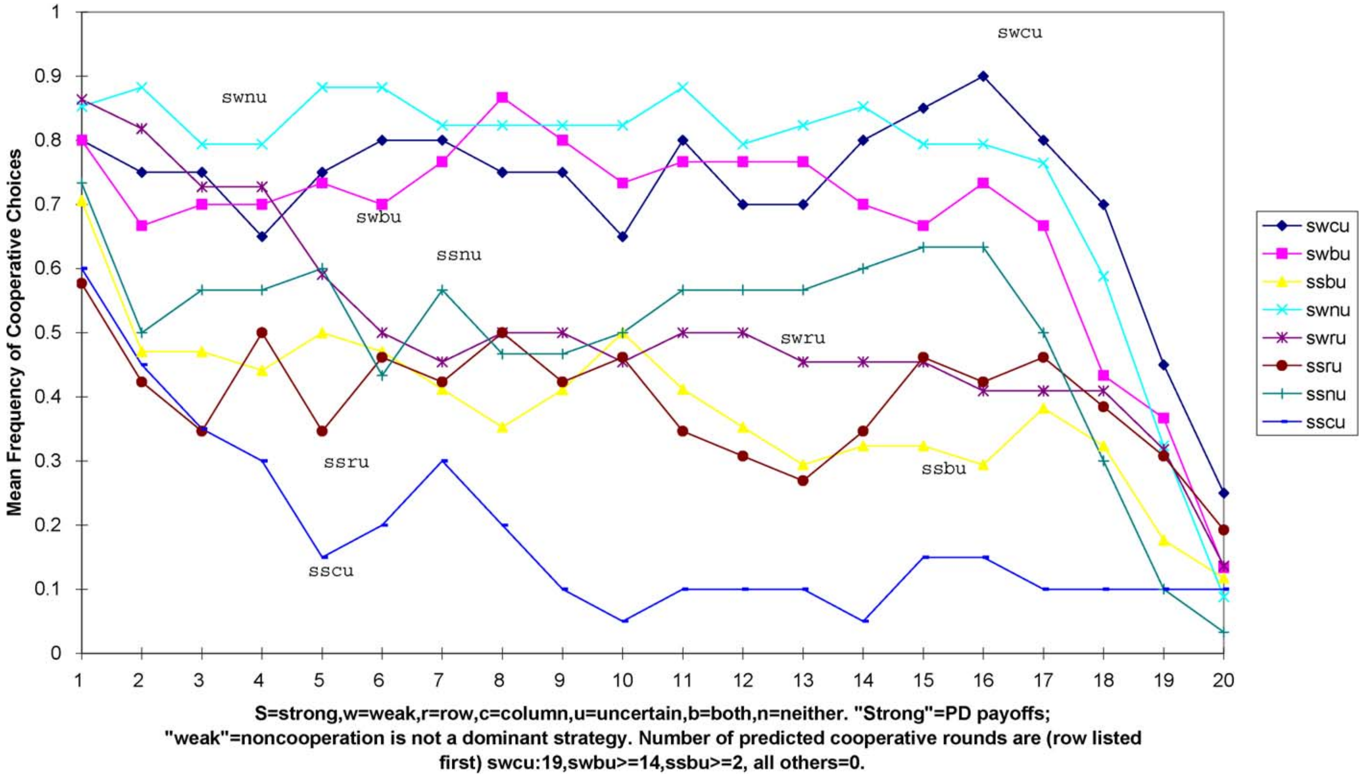


Figure 1. Mean frequencies of cooperation over trials.

eight conditions seemed to fall into three distinct sets: the three cooperative SW conditions, a set of four moderately cooperative conditions, and the primarily non-cooperative SS, Column Uncertain condition. Among the moderately cooperative group, the SS, Neither Uncertain condition was the most cooperative and, after the early plunge in cooperation rates, seemed to evidence a gradual increase in cooperation from trials 8 to 16. This is particularly surprising given the fact that non-cooperation is a dominant strategy for both players and they have no uncertainty about it.

Finally, the data reveal a huge drop in cooperation rates in all of the conditions starting on the 17th of the 20 trials. We liken the image to a ball falling off a table. In this case, however, the ball falls a few trials before the end of the game rather than at the very end. This steep drop in cooperation shows awareness of the impending end game among players who had otherwise displayed high levels of cooperation.

Further, Figure 1 again illustrates that the conditions when neither of the two parties had any uncertainty about each others' payoffs generally led to more cooperation throughout the repeated trials than any of the Row, Column, or Both Uncertain conditions within a Strong–Strong or Strong–Weak game.

Needless to say, these data provide only minimal evidence to suggest that uncertainty about a counterpart's payoffs explains why they are cooperative in finite prisoner's dilemma games. Other models are clearly necessary to explain these anomalies. The extensive nature of the data in this study, however, provide some clues.

## 4. Discussion and Conclusions

Several of the results reported here broadly support theories of sequential equilibria. In particular, at least four pieces of data indicate that game theory's structurally-oriented models predict some of the reasons for individuals' play in the prisoner's dilemma game: (1) more cooperation resulted when a cooperative equilibrium existed; (2) players cooperated more overall and more in the first round in SW than in SS; (3) Column was more cooperative than Row on the first trial; and (4) the SW Column Uncertain condition produced the most end game cooperation.

Unfortunately, these effects only explain part of the picture. In essence, even after considering the effects of uncertainty, the original issue still remains: considerable cooperation resulted when no cooperative equilibria exist. It is our impression that the most important findings in these data are that: (1) the players were more cooperative when they were certain of each other's payoffs; and (2) they only began to react to the presence of the endgame with four of the 20 trials remaining. Rather than being best explained by rational models, it appears that what seems to be unexpected cooperation may depend on psychological factors, including: (1) the personal discomfort of uncertainty; (2) myopia with respect to the endgame; (3) the desire to avoid losses; and (4) the fact that only two choices are available in these games. Points 1, 3, and 4 combine to suggest that, because the non-cooperative equilibrium exists in all conditions of both games, players may have reacted to uncertainty by choosing non-cooperatively. A systematic psychological discontinuity between the certain and uncertain conditions may have driven them to choose non-cooperatively without non-cooperative, anti-opponent intent. Indeed, this reflects a criticism of PD games, that the person choosing non-cooperation may perceive it as defensive rather than offensive (see, for example, Apfelbaum, 1974 or Nemeth, 1972]. In economic terms, the existence of payoff uncertainty may raise the transaction costs involved in reaching a cooperative outcome: with payoff uncertainty, it may be difficult for players to identify the efficiency gains from cooperating. Thus, the net effect of payoff uncertainty may be negative – the negative effects on the ability of parties to negotiate cooperative agreements may outweigh the positive effects of the existence of additional cooperative equilibria.

Points 1, 2, and 3, above, may have led people to depend more on their counterparts' previous choices than on any calculation of the underlying structural contingencies of the game. In particular, participants may have focused on reaping as many mutually cooperative outcomes as they could but, as the end approached, tried to avoid being the last person to shift to non-cooperation. This compounds a player's uncertainty. Yet, almost all of the conditions, regardless of their structural uncertainty, show that cooperation starts to break down on trial 17. This suggests that payoff uncertainty and the uncertainty of one's counterpart's reactions to the endgame are qualitatively different, one requiring a wait-and-see strategy and the other requiring a preemptive strategy. This suggests that further modeling of the play in prisoner's dilemma games may need to be even more complex.

## Acknowledgements

We are grateful for the financial support provided by the National Science Foundation (#BNS87-00864 and SES88-15566) and the Russell Sage Foundation. The authors thank Patty DeForrest and Felice Herbin for serving as experimenters.

## References

Andreoni, J., Miller, J. (1993). "Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence". Economic Journal 103, 570–585.

Apfelbaum, E. (1974). "On conflicts and bargaining". Advances in Experimental Social Psychology 7, 103–156.

Axelrod, R. (1984). "The Evolution of Cooperation". Free Press, New York.

Camerer, C., Weigelt, K. (1988). "Experimental tests of a sequential equilibrium reputation model". Econometrica 56, 1–36.

Friedman, J.W. (1986). "Game Theory with Applications to Economics". Oxford University Press, New York.

Kahn, L.M., Murnighan, J.K. (1993). "Conjecture, uncertainty, and cooperation in prisoner's dilemma games". Journal of Economic Behavior and Organization 22, 91–117.

Kreps, D., Milgrom, P., Roberts, J., Wilson, R. (1982). "Rational cooperation in the finitely repeated prisoner's dilemma". Journal of Economic Theory 27, 245–252.

Luce, R.D., Raiffa, H. (1957). "Games and Decisions; Introduction and Critical Survey". Wiley, New York.

Murnighan, J.K., Roth, A.E. (1983). "Expecting continued play in prisoner's dilemma games". Journal of Conflict Resolution 27, 279–300.

Nemeth, C. (1972). "A critical analysis of research utilizing the prisoner's dilemma paradigm for the study of bargaining". In: Berkowitz, L. (Ed.), Advances in Experimental Social Psychology, vol. 6. Academic Press, New York.

Rapoport, A., Chammah, A. (1965). "Prisoner's Dilemma". University of Michigan Press, Ann Arbor.

Roth, A.E. (1988). "Laboratory experimentation in economics: A methodological overview". Economic Journal 98, 974–1031.

Roth, A.E., Murnighan, J.K. (1983). "Equilibrium behavior and repeated play of the prisoner's dilemma". Journal of Mathematical Psychology 17, 189–198.

Selten, R., Stoecker, R. (1986). "End behavior in sequences of finite prisoner's dilemma supergames". Journal of Economic Behavior and Organization 7, 47–70.