CAES Working Paper Series

An Experimental Study of the Effect of Information Structure on the Possibility of Cooperation in the Prisoner's Dilemma

Ko Nishihara, Masato Kagihara, and Junichi Watanabe
Faculty of Economics
Fukuoka University

WP-2015-003

Center for Advanced Economic Study
Fukuoka University
(CAES)

8-19-1 Nanakuma, Jonan-ku, Fukuoka,
JAPAN 814-0180

An Experimental Study of the Effect of Information Structure on the Possibility of Cooperation in the Prisoner's Dilemma

Ko Nishihara * Masato Kagihara † Junichi Watanabe ‡
Faculty of Economics, Fukuoka University
8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan

May 31, 2015

Abstract

This paper experimentally investigates the effect of information structure on the possibility of cooperation in the Prisoner's Dilemma. We modify the information structure of the usual 2-person Prisoner's Dilemma as follows: Nature chooses the order of moves with fifty-fifty chance and only the first mover's *Defection* is observed by the second. When the payoffs of the modified Prisoner's Dilemma satisfy some conditions, there exists a Nash equilibrium in which both players take *Cooperation* as well as the noncooperative equilibrium (Nishihara (1997)). We find that this modification of information structure raises the possibility of cooperation when the cooperative equilibrium *risk-dominates* the noncooperative one.

Keywords: Prisoner's Dilemma, information structure, experimental study

JEL Classification Numbers: C72, C91

1 Introduction

Societal cooperation often fails when people make rational choices as individuals. *The Prisoner's Dilemma*, PD for short, has been studied by economists, sociologists, social psychologists etc. as a game that represents such a state. The game is used not only to depict social problems in the real world caused by individually rational actions but also to seek a way to solve these problems. In this paper, we experimentally investigate whether the possibility of cooperation will be enhanced when the information structure of PD is modified.

It has been pointed out that PD is too much simplified to describe the real-world condition leaving various factors of human interaction out of consideration. If some missing factors are added to the game, it may not be true that *Defection* is individually rational in the sense that it dominates *Cooperation*. And so, many theoretical studies have been done to make PD more practical by adding/appending such factors to it. Repetition of the game is a typical example of the factors.

*E-mail: nishi@fukuoka-u.ac.jp
†E-mail: kagihara@fukuoka-u.ac.jp
‡E-mail: junwata1@fukuoka-u.ac.jp

Nishihara (1997) noted the fact that noncooperative behavior can attract public attention in real life, and proposed a modified version of the n-person Prisoner's Dilemma. In the modified game, Nature chooses the order of moves with equal probability and a player can see a *Defection* taken before his move if any.

The followings are examples of this game. First, littering on the street, which corresponds to *Defection*, is easily found by other pedestrians, while no littering, *Cooperation*, is not. Second, in some village in savanna, people have traditionally kept cattle and made their living. But keeping goats in the commons is more profitable than keeping cattle for a single farmer, so he has incentive to raise his goats there. However, goats are among main causes of increasing desertification, since they eat plants right down to the ground without leaving a stub for the plant to regrow. Hence, keeping goats is *Defection*, while cattle *Cooperation*. Stocking the herd of cattle with goats is observable for other villagers.

When the payoffs of the game meet some condition, the modified game has a Nash equilibrium where each player takes *Cooperation*, which we call *the cooperative equilibrium*. The condition requires that *the incentive for Defection* is small enough. We call this modified game *PD with observable defection*, hereafter PDOD for short. However, PDOD always has a Nash equilibrium where each player takes *Defection*, which we call *the noncooperative equilibrium*. Therefore, it is not clear that which equilibrium occurs even when the existence of the cooperative equilibrium is guaranteed.

The purpose of this experimental study is to verify whether the ratio of people who take *Cooperation* is higher in PDOD than in PD. We concentrate on only 2-person games in this paper. One of the main reasons why we use 2-person games is to make the procedures of the experiment simple. In 2-person game, if a player knows that *Defection* was chosen, there is only one possibility: the opponent took it. Even in a 3-person game, a player must consider four cases when he knows that someone took *Defection* before his move: he may be the second mover, or he may be the third mover and the first mover and/or the second mover took *Defection*. We also make a questionnaire survey to find out what factor decides which action actually taken in the game.

The experiment was conducted from 2009 through 2010 at Fukuoka University. We recruited undergraduate students of the university as subjects. The total number of the subjects was about 700. A pair of subjects was matched randomly and each subject, who was not informed of his opponent, played both PD and PDOD [1] by *the strategy method*. We use five kinds of payoff structures in conducting the experiment. Each payoff structure has the following features in common; an amount of the incentive for *Defection* is the same whatever the other player's choice is, and it is set for a cooperative equilibrium to exist theoretically. We find the following:

1. The ratio of people taking *Cooperation* is higher in PDOD than in PD at the significance level of 1% when the cooperative equilibrium *risk-dominates* the noncooperative one and the amount of the incentive for *Defection* is large enough.

2. A subject who believes that the other player would take *Cooperation* (resp. *Defection*) tends to take *Cooperation* (resp. *Defection*) in PDOD.

These results suggest that the observability of *Defection* plays an important role in increasing the ratio of people taking *Cooperation*, and that the subjects' behavior depends on the subjects' belief.

[1] In what follows, we use the terms PD and PDOD to describe 2-person PD and 2-person PDOD respectively unless otherwise noted.

This paper is organized as follows. The next section explains the theoretical back ground of this study and the hypothesis to test. Section 3 shows the results of experiment and gives some implications of them. Section 4 is devoted to conclusion.

2 Theoretical Background and Hypotheses

2.1 Theoretical Background

We consider PD given in Figure 1. Each player has two actions: X (*Cooperation*) and Y (*Defection*), which are pure strategies. Here, we call X *the cooperative strategy*. G stands for a gain when a player changes his strategy from X to Y whatever strategy the opponent chooses, which we call *the incentive for Defection*. In this experimental study, G is set at five levels.

	Cooperation (X)	Defection (Y)
Cooperation (X)	$500,\ 500$	$100,\ 500+G$
Defection (Y)	$500+G,\ 100$	$100+G,\ 100+G$

Note: $0 < G < 400$.

Figure 1: The Prisoner's Dilemma (¥)

PDOD, which is the 2-person case of the n-person game studied in Nishihara (1997), is the following game (See Figure 2).

1. Nature chooses the order of the moves with fifty-fifty chance.
2. Each player takes X or Y in the order chosen by Nature.
3. Each player has two information sets: *Information Set 0* is the one reached when either he is the first mover, or he is the second mover and the first mover takes X; *Information Set 1* reached when he is the second mover and the first mover takes Y.
4. Payoffs are determined by Figure 1, after the players take their actions.

In PDOD, each player has four pure strategies, XX, XY, YX, YY, where the first letter is an action taken in Information Set 0 and the second in Information set 1. We call XX and XY *the cooperative strategies* in PDOD because if both take these strategies then they will come to play X in the game. The following proposition is the 2-person case of a theorem shown in Nishihara (1997).

Proposition 1. *In PDOD with the payoff given in Figure 1, a pair of strategies (XY, XY) is a Nash equilibrium for $G \leq 200$, and (YY, YY) is a Nash equilibrium for any G.*

Note that in the equilibrium (XY, XY), both players choose X in the play since Information Set 1 is off the equilibrium path, while they choose Y in the equilibrium (YY, YY). The above proposition means that even if the equilibrium (XY, XY) exists, it is not clear whether players really choose the cooperative strategies.

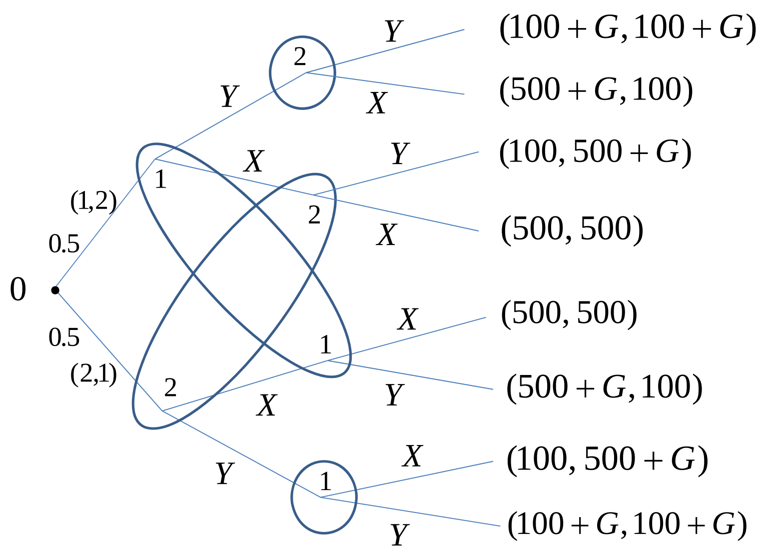

Figure 2: The Prisoner's Dilemma with Observable Defection

2.2 Hypotheses to Test

Although Proposition 1 shows that there is a cooperative equilibrium in PDOD, the existence of two equilibria implies that the possibility of cooperation is not necessarily enhanced. This paper experimentally investigates whether the ratio of people taking the cooperative strategy is higher in PDOD than in PD. Note that the difference between PD and PDOD is in the information structure of these games: *No Information* (hereafter NI) in PD and *Observable Defection* (hereafter OD) in PDOD. Hence, the purpose of this experimental study is to examine whether the modification of the information structure of PD, from NI to OD, raises the possibility of cooperation in PD. In the experiment, *the outcome* is player's strategy, a cooperative strategy or not, and *the treatment* is information structure of the games, NI or OD. Thus, we estimate *an average treatment effect* of information structure on the choice of cooperative strategy in PD.

For the purpose stated above, let us consider a group of people as a population. Let α be the ratio of people in the population who choose the cooperative strategy X when they play PD. And, let p, q, r and s be the ratios of people who take XY, XX, YY and YX respectively if they play PDOD. Note that $p + r$ is the ratio of people who take the cooperative strategy in PDOD. In this setting, the average treatment effect to be estimated is $(p + r) - \alpha$. If the experimental reslult suggests that $(p + r) - \alpha > 0$, then we can conclude that the possibility of cooperation is enhanced by modifying the information structure of PD, from NI to OD. Hence, the null-hypothesis H_0 and the alternative-hypothesis H_1 are:

$$H_0 : \alpha = p + r, \qquad H_1 : \alpha < p + r. \tag{1}$$

3 Experimental Results

3.1 The Experiment

The experiment was conducted at Fukuoka University in 2009 and 2010. The subjects were recruited from undergraduate students of the university.[2] In the experiment the recruitment was made 7 times, and in each recruitment the subjects were assigned to either session of the two. Each session consisted of two subsessions, Subsession 1 and 2. Hence, the experiment consisted of 14 sessions and 28 subsessions. In each subsession subjects played PD or PDOD [3]; each subject played two kinds of games in the session. About 50 subjects participated in each session, and the total number of 704 students participated in the experiment. The game played in each subsession was conducted by the strategy method, i.e., the subjects were asked to make their choices at every information set in the game.

Table 1: List of Sessions

Recruitment	G	Session (Date)	Information structure (SS1, SS2)	# of subjects (SS1, SS2)
1	30	1 (11/18/2009)	(NI, OD)	(58, 58)
		2 (11/19/2009)	(OD, NI)	(42, 42)
2	50	3 (12/02/2009)	(NI, OD)	(51, 51)
		4 (12/03/2009)	(OD, NI)	(46, 46)
3	50	5 (12/16/2009)	(NI, OD)	(69, 69)
		6 (12/17/2009)	(OD, NI)	(43, 43)
4	100	7 (11/17/2010)	(NI, OD)	(49, 48)
		8 (11/19/2010)	(OD, NI)	(23, 23)
5	100	9 (12/01/2010)	(PI, OD)	(68, 67)
		10 (12/02/2010)	(OD, PI)	(60, 60)
6	200	11 (12/08/2010)	(NI, OD)	(56, 53)
		12 (12/09/2010)	(OD, NI)	(51, 51)
7	150	13 (12/15/2010)	(NI, OD)	(38, 36)
		14 (12/16/2010)	(OD, NI)	(36, 36)

Note 1. SS1 and SS2 stand for Subsession 1 and 2, respectively.
Note 2. NI (No Information), OD (Observable Defection) and PI (Perfect Information) stand for the information structures of PD, PDOD and PDPI, respectively.

Each session proceeded as follows. At the beginning of each session, subjects were given general instructions of the session, which were then read aloud by the experimenter. Subsession 1 and 2 were conducted with the following procedure. The instructions of the game were handed out to the subjects. The subjects were asked to read them in silence. The instructions said the following: The subjects would be paired at random with each other; they would not know who they were paired with; their monetary reward would depend on their choices. After all the subjects finished reading

[2] Announcements of the recruitment were made on the university electronic bulletin board and in class. The announcements informed that the participants would obtain some monetary rewards and that they were allowed to attend the experiment only once.

[3] In the 9th and 10th sessions, subjects played PD with perfect information (hereafter PDPI) instead of PD. The purpose of these sessions will be explained in Section 3.2.

the instructions, they were given a quiz to test their understanding. Only those who answered the quiz correctly were permitted to join the game. At the end of the session, the subjects were asked to answer the questionnaire about their decision in PDOD. Finally, the subjects were paid some monetary rewards, which were the sum of ¥500 show-up fee and the payoffs earned in the two subsessions.

In this experiment, PD and PDOD were played for five different Gs ($G = 30, 50, 100, 150, 200$). In addition, PDPI was conducted only for $G = 100$. Hence, a total of 11 kinds of games were played in the experiment. The numbers of subjects in these games are listed in Table 2.

Table 2: Number of Subjects

Information Structure\ G	30	50	100	150	200	Total
NI (SS1, SS2)	$(58, 42)$	$(120, 89)$	$(49, 23)$	$(38, 36)$	$(56, 51)$	$(321, 241)$
OD(SS1, SS2)	$(42, 58)$	$(89, 120)$	$(83, 115)$	$(36, 36)$	$(51, 53)$	$(301, 382)$
PI (SS1, SS2)			$(68, 60)$			$(68, 60)$
Total (SS1, SS2)	$(100, 100)$	$(209, 209)$	$(200, 198)$	$(74, 72)$	$(107, 104)$	$(690, 683)$

Note 1. SS1 and SS2 stand for Subsession 1 and 2, respectively.
Note 2. NI, OD and PI stand for the information structures of PD, PDOD and PDPI, respectively.
Note 3. Although total number of participants was 704, 690 and 683 students joined SS1 and SS2 as subjects respectively because 14 students failed the quiz in SS1 and 7 in SS2.

3.2 The Comparison: No Information and Observable Defection

The data obtained in the experiment are categorized into three types: the data from Subsession 1, the data from Subsession 2 and the pooled data from both the subsessions. This section reports the results obtained from the former two types of experimental data: the data from Subsession 1 and the data from Subsession 2, because there arises no need to consider *order effect* when using these types of data. The data from Subsession 1 reflects the fact that the subjects made their decisions without any prospect. On the other hand, in Subsession 2 the subjects could make their decisions more stably because they had already learnt how to play the game in Subsession 1.[4]

Table 3 and Table 4 show the strategies taken by the subjects in PDOD based on the data from Subsession 1 and from Subsession 2 respectively. In PDOD, most of the subjects took XY or YY, which means that they chose Y in reaction to Y taken by the opponent. First, we report the results based on the data from Subsession 1. Table 5 shows the ratios of subjects taking cooperative strategies, X under NI and XX or XY under OD. When G is 50 or 100, the ratio is higher by about 20% under OD than under NI. The difference is statistically significant at the 1% level.[5] However, this is not true when G is 30, 150 or 200. In these Gs, the ratio is lower under OD than under NI, although the difference is not statistically significant even at the 25% level. Next, the results based

[4]Note that the subjects knew nothing about their opponents' decisions and any results in Subsession 1 because we used the strategy method and the monetary rewards were paid after all the subsession finished.

[5]In this paper we use *a difference in sample means* as a test statistics, as shown in Table 5. For the difference in sample means to be a consistent estimator of the average treatment effect, the assignment of the subjects to the treatment must be *mean independent* of the outcome. Since we assigned each subject to a session at his convenience without informing him about the order of games played, the assignment is considered to have no influence on the outcome although this is not a random assignment. Therefore, the average treatment effect is consistently estimated by the difference in sample means.

on the data from Subsession 2 are shown in Table 6. The ratio is higher under OD than under NI for all Gs except when $G = 150$. The difference is statistically significant at the 1% or 5% level.

Table 3: Ratios of Strategies Taken by Subjects in PDOD: SS1 Data

G (# of subjects)	XX	XY	YX	YY
$G = 30$ (42)	0.119	0.357	0.024	0.500
$G = 50$ (89)	0.067	0.427	0.022	0.483
$G = 100$ (83)	0.048	0.422	0.012	0.518
$G = 150$ (36)	0.000	0.167	0.000	0.833
$G = 200$ (51)	0.020	0.118	0.020	0.843

Table 4: Ratios of Strategies Taken by Subjects in PDOD: SS2 Data

G (# of subjects)	XX	XY	YX	YY
$G = 30$ (58)	0.086	0.552	0.000	0.362
$G = 50$ (120)	0.017	0.542	0.008	0.433
$G = 100$ (115)	0.043	0.400	0.026	0.530
$G = 150$ (36)	0.000	0.194	0.028	0.778
$G = 200$ (53)	0.038	0.189	0.000	0.774

Table 5: Ratios of Subjects Taking Cooperative Strategies: SS1 Data

G (# of subjects: NI, OD)	NI	OD	Difference
$G = 30$ (58, 42)	0.534 (0.065)	0.476 (0.077)	−0.058 (0.101)
$G = 50$ (120, 89)	0.308 (0.042)	0.494 (0.053)	0.186* (0.068)
$G = 100$ (49, 83)	0.224 (0.060)	0.470 (0.055)	0.245* (0.087)
$G = 150$ (38, 36)	0.211 (0.066)	0.167 (0.062)	−0.044 (0.091)
$G = 200$ (56, 51)	0.161 (0.049)	0.137 (0.048)	−0.023 (0.069)

Note 1. Standard deviations in parentheses.
Note 2. NI and OD stand for no information and observable defection respectively.
Note 3. * Significance at the 1% level.

Based on the data from Subsession 1, our hypothesis is statistically supported only for $G = 50$ and 100 as stated above. We could explain this fact as follows. When $G = 30$, the incentive for *Defection* is too small for the subjects to recognize Y as a dominant strategy.[6] When $G = 200$, the equilibrium condition of (XY, XY) holds with equality, hence, XY is less preferable for the subjects to YY under the uncertainty of the opponents' strategy. Therefore, it is meaningful to compare the information structure only for $G = 50$, 100 and 150. Finally, applying the equilibrium selection theory of Harsanyi and Selten (1988), the equilibrium (YY, YY) risk-dominates (XY, XY) when G is greater than 133 (See Appendix A). Hence, the subjects prefer YY when $G = 150$.

In this experiment, we asked each subject to fill out a questionnaire about the reason why he chose his strategy in PDOD. Table 7 gives a brief summary of the answers to the question B-4. It shows that

[6]Because of altruism and/or inequality aversion, subjects often feel "sense of guilty" for taking Y in PD. Our questionnaire survey also confirmed this fact. Hence, the game the subject has in his mind may not be PD when G is small enough.

Table 6: Ratios of Subjects Taking Cooperative Strategies: SS2 Data

G (# of subjects: NI, OD)	NI	OD	Difference
$G = 30$ (42, 58)	0.310 (0.071)	0.638 (0.063)	0.328* (0.101)
$G = 50$ (89, 120)	0.326 (0.050)	0.558 (0.045)	0.232* (0.070)
$G = 100$ (23, 115)	0.217 (0.086)	0.443 (0.046)	0.226** (0.112)
$G = 150$ (36, 36)	0.194 (0.066)	0.194 (0.066)	0.000 (0.093)
$G = 200$ (51, 53)	0.078 (0.038)	0.226 (0.057)	0.148** (0.071)

Note 1. Standard deviations in parentheses.
Note 2. NI and OD stand for no information and observable defection respectively.
Note 3. * Significance at the 1% level, ** Significance at the 5% level.

the subjects who took cooperative strategies expected their opponents to take cooperative strategies, and that those who took non-cooperative strategies non-cooperative strategies. In practice, the hypothesis that two ditributions of answers shown in Table 7 are same is rejected by χ^2 test with p-value 0.000 ($\chi^2 = 30.50$).

Table 7: Answers to Question B-4: $G = 100$

Types of Subjects (# of subjects)\Answers	1	2	3	4	5	Total
Cooperative strategy: XX or XY (39)	0.538	0.231	0.205	0.026	0.000	1.00
Non-cooperative strategy: YY or YX (44)	0.068	0.727	0.136	0.000	0.068	1.00

Question B-4 You are asked about your expectation on the opponent's choice when you receive the message, "Select your action". Choose the most adequate one among the following 1 through 5.

1. The opponent would probably choose X.
2. The opponent would probably choose Y.
3. The opponent would choose X or Y with almost equal probability.
4. I could not imagine the opponent's choice.
5. I did not expect on the opponent's choice.

Note: Subjects were instructed that they would receive the messages "Select your action", and "The opponent chose Y. Select your action," in Information Sets 0 and 1, respectively.

3.3 The Comparison: No Information, Observable Defection and Perfect Information

Cooperative equilibrium does not exist when the information structure of PD is either too coarse or too fine. The extremes are NI for the former case and PI for the latter.[7] Actually, Nishihara (1997) shows that the cooperative equilibrium does not exist under any information structure if it

[7]Note that there are five kinds of information structure in PD. In addition to NI, OD and PI, there are two other information structures: *Observable Cooperation* and *Observable Order of Moves.*

does not exist under OD. It means that the "adequate" coarseness of OD is needed to generate the cooperative equilibrium. In order to see this point, we also compare PI with NI and OD.

Consider again the population stated in Section 2.2. Under PI, each player has three information sets: one is reached when he is the first mover; the next one when he is the second mover and the opponent chose cooperative action X; the last one when he is the second mover and the opponent chose non-cooperative action Y. Each player's pure strategy is denoted by a 3-tuple abc, where a, b and c are the actions taken in the first, second and third information sets, respectively. Hence, each player has 8 pure strategies: XXX, XXY, XYX, XYY, YXX, YXY, YYX, and YYY. Let $p_1, p_2, \ldots, p_8$ be the ratios of the people in the population who take these strategies, respectively. We call XXX and XXY cooperative strategies since if both take these strategies then they will come to play X. The ratio of the people who take the cooperative strategies is $p_1 + p_2$. If it is statistically significant that $p_1 + p_2 < p + r$, then we can conclude that possibility of cooperation is higher in PDOD than in PDPI. The null and alternative hypotheses are:

$$H_0 : p_1 + p_2 = p + r, \quad H_1 : p_1 + p_2 < p + r.s$$

Similarly, we also compare NI and PI. The null and alternative hypotheses are:

$$H_0 : \alpha = p_1 + p_2, \qquad H_1 : \alpha < p_1 + p_2.$$

The strategies taken by the subjects in PDPI are shown in Table 8 and Table 9. Table 10 and Table 11 show the ratios of subjects taking cooperative strategies, X under NI, XX or XY under OD, and XXX or XXY under PI, based on the data from Subsession 1 and from Subsession 2 respectively. The data show that the ratios of the subjects who took the cooperative strategies were about 45% under OD. The ratios are higher by about 20% under OD than under NI or PI. The differences are statistically significant at the 1% or 5% level. However, the differences are not statistically significant between NI and PI. Thus, for $G = 100$, among the above three information structures, the ratio of the subjects who take the cooperative strategies is significantly increased under OD. It implies that the imperfectness of OD plays an important role in increasing the possibility of the cooperative strategies.

Table 8: Ratios of Strategies Taken by Subjects in PDPI ($G = 100$): SS1 Data

G (# subjects)	XXX	XXY	XYX	XYY	YXX	YXY	YYX	YYY
$G = 100$ (68)	0.059	0.235	0.000	0.118	0.044	0.147	0.000	0.397

Table 9: Ratios of Strategies Taken by Subjects in PDPI ($G = 100$): SS2 Data

G (# of subjects)	XXX	XXY	XYX	XYY	YXX	YXY	YYX	YYY
$G = 100$ (198)	0.033	0.100	0.000	0.083	0.033	0.133	0.017	0.600

4 Summary

We experimentally studied whether the information structure OD increases the possibility of cooperation in PD. The experiments were conducted for $0 < G \leq 200$, where a cooperative Nash

Table 10: Ratios of Subjects Taking Cooperative Strategies ($G = 100$): SS1 Data

Information Structure (# of Subjects)	Ratios of Subjects Taking Cooperative Strategies	Differences (vs NI)	Differences (vs PI)
NI (49)	0.224 (0.060)		−0.070 (0.083)
OD (83)	0.470 (0.055)	0.245* (0.087)	0.176** (0.080)
PI (68)	0.294 (0.055)	0.070 (0.083)	

Note 1. Standard deviations in parentheses.
Note 2. NI, OD, and PI stand for no information, observable defection, and perfect information, respectively.
Note 3. * Significance at the 1% level, ** Significance at the 5% level.

Table 11: Ratios of Subjects Taking Cooperative Strategies ($G = 100$): SS2 Data

Information Structure (# of Subjects)	Ratios of Subjects Taking Cooperative Strategies	Differences (vs NI)	Differences (vs PI)
NI (23)	0.217 (0.086)		0.084 (0.089)
OD (115)	0.443 (0.046)	0.226** (0.112)	0.310* (0.075)
PI (60)	0.133 (0.044)	−0.084 (0.089)	

Note 1. Standard deviations in parentheses.
Note 2. NI, OD, and PI stand for no information, observable defection, and perfect information, respectively.
Note 3. * Significance at the 1% level, ** Significance at the 5% level.

equilibrium exists in PDOD. When $G = 50, 100$, OD significantly increases the ratio of taking cooperative strategies in comparison to NI. On the other hand, at other levels of G ($G = 30, 150, 200$), the differences between the ratios are not significant. The reason why not at $G = 150$ could be explained by the notion of risk-dominance. Finally, our questionnaire survey finds that the subjects who believe that their opponents are cooperative take cooperative strategies, and that those who believe that the opponents are noncooperative take noncooperative strategies.

Appendix

A Risk Dominance

Consider a two person strategic form game in which the players can choose XY or YY which are defined in Section 2.1. Fix a player arbitrarily, and let p be the probability that the opponent will choose XY. The expected payoff of taking XY is the following:

$$EXY(p) = 500p + \left(100 + \frac{G}{2}\right)(1-p) = \left(100 + \frac{G}{2}\right) + \left(400 - \frac{G}{2}\right)p.$$

The expected payoff of taking YY is:

$$EYY(p) = (300 + G)p + (100 + G)(1-p) = (100 + G) + 200p. \tag{2}$$

Hence,

$$EXY(p) > EYY(p) \iff \left(200 - \frac{G}{2}\right) > \frac{G}{2} \iff p > \frac{G}{400 - G}. \tag{3}$$

Therefore, if $\frac{G}{400-G} < \frac{1}{2}$, namely, $G < \frac{400}{3} = 133.33..$,, then the range of p in which XY is the best response is larger than that in which YY is the best response. Thus, (XY, XY) risk-dominates (YY, YY) when $G < 133.33....$

References

Funaki, Y. and J. Vyrastekova (2009), "Cooperation in a sequential N-person prisoners' dilemma game: the role of information and reciprocity," *mimeo.*

Hamberger, H., M. Guyer and J. Fox (1975), "Group size and cooperation," *Journal of Conflict Resolution*, **19**, pp.503-531.

Harsanyi, J. C. and R. Selten (1988) *A General Theory of Equilibrium Selection in Games*, MIT Press.

Kagel, J. and A. Roth eds.(1995) *The Handbook of Experimental Economics*, Princeton University Press.

Nishihara, K. (1997) "A resolution of N-person prisoners' dilemma," *Economic Theory*, **10**, pp.531–540.