KOCHI UNIVERSITY OF TECHNOLOGY

Social Design Engineering Series *SDES-2015-12*

The approval mechanism experiment: A solution to prisoners dilemma

Tatsuyoshi Saijo
School of Economics and Management, Kochi University of Technology
Research Center for Social Design Engineering, Kochi University of Technology
Institute of Economic Research, Hitotsubashi University
Urban Institute, Kyusyu University

Yoshitaka Okano
School of Economics and Management, Kochi University of Technology
Research Center for Social Design Engineering, Kochi University of Technology

Takafumi Yamakawa
Osaka University

24th May, 2016

School of Economics and Management
Research Center for Social Design Engineering
Kochi University of Technology

May 23, 2016
Not for circulation!

The Approval Mechanism Experiment: A Solution to Prisoner's Dilemma†

January 2010/Revised August 2012

Tatsuyoshi Saijo, Yoshitaka Okano and Takafumi Yamakawa
Osaka University

Abstract

Players can approve or reject the other choice of the strategy after announcing the choices in a prisoner's dilemma game. If both approve the other choice, the outcome is what they choose, and if either one rejects the other choice, it is the outcome when both defect, which is called the mate choice mechanism. Theoretically, this mechanism implements cooperation in backward elimination of weakly dominated strategies (*BEWDS*) assuming that players are payoff maximizers, reciprocaters, inequality averters or the mixture of them, but it does not implement cooperation in Nash equilibria (*NE*) or subgame perfect equilibria (*SPE*). Although the coverage of behaviors shrinks, it also implements cooperation in neutrally stable strategies (*NSS*). Experimentally, we observe that the cooperation rate with the mechanism is 90% in round 1 and it is 93.2% through 19 rounds, and that subjects' behavior is consistent with *BEWDS* rather than *NE*, *SPE* or *NSS* behavior using questionnaire analysis. Utilizing off equilibrium path data, we find that payoff maximizers or reciprocators are 88-90%, inequality averters are 10-11%, and utilitarians are 0-1%.

JEL Classification Numbers: C72, C73, C92, D74, P43

† This research was supported by the Suntory Foundation, the Joint Usage/Research Center at ISER, Osaka University and "Experimental Social Sciences: Toward Experimentally-based New Social Sciences for the 21st Century" that is a project in the Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science and Culture of Japan. We thank the Economics Department at Osaka University who kindly allowed us to use the computer lab. We also thank comments from participants of seminars at Okinawa, UCLA, Hokkaido, Kyoto Sangyo, UCSB, UCSD, Tohoku, Waseda, Keio, VCASI, Seoul National, Hawaii, UMass and 2011 Japanese Economic Association Meeting at Tsukuba. We thank Jiro Akita, Jim Andreoni, Masahiko Aoki, Kenemi Ban, Ted Bergstrom, Tim Cason, Gary Charness, Youngsub Chun, Takako Fujiwara-Greve, Yukihiko Funaki, Yoichi Hizen, Eiji Hosoda, Tatsuya Kameda, Michihiro Kandori, Kazunari Kainou, Shunsuke Managi, Takehito Masuda, Yuko Morimoto, Mayuko Nakamaru, Yusuke Narita, Ichiro Obara, Masao Ogaki, Cheng-Zhong Qin, Junyi Shen, Kazumi Shimizu, Hideo Shinagawa, Martin Shubik, John Stranlund, John Spraggon, Masanori Takaoka, Masanori Takezawa, Toshio Yamagishi, Takehiko Yamato and Bill Zame for their valuable comments.

1. Introduction

Dresher and Flood conducted the first experiment on Prisoner's Dilemma game at RAND in January of 1950[1], and after their work, numerous papers on its theory and experiments have been published in not only economics but also many fields such as mathematics, computer science, biology, psychology, sociology, political science, management science and so on[2]. There are at least three approaches in order to tame the dilemma.

The first possible direction is to introduce the repetition of the game. David M. Kreps, Paul Milgrom, John Roberts, and Robert Wilson (1982) investigated possible cooperation in *finitely* repeated prisoner's dilemma game. The source of cooperation was some asymmetries of types of players. James Andreoni and John H. Miller (1993) conducted a series of prisoner's dilemma experiments to confirm the prediction by Kreps et al. (1982), and found that subjects' beliefs of their opponent altruism increased reputation building and therefore they were more cooperative than subjects in a repeated single-shot game[3]. However, the average cooperation scarcely exceeded more than 60%. Yoella Bereby-Meyer and Alvin E. Roth (2006) reported that noisy payoffs reduced cooperation in repeated game although they increased cooperation in one-shot game.

The second approach is related with biology and ecology. Genetic relationship between participants changes the payoff matrix structure called kin selection due to W.D. Hamilton (1964). This idea has been extended to direct, indirect or network reciprocity and group selection (see Michael Doebeli and Christoph Hauert (2005) and Martin A. Nowak (2006) for the review).[4] Economists such as Robert Sugden (1984), Matthew Rabin (1993) and the followers also have been pursuing this avenue. Rachel T.A. Croson (2007) found that reciprocity plays a key role in linear public good experiments compared with commitment and altruism although it is not good enough to attain the Pareto efficient allocation.

---

[1] According to William Poundstone (1992), " the prisoner's dilemma was 'discovered' in 1950, just as nuclear proliferation and arms races became serious concerns" (page 9). See also Merrill M. Flood (1958) and Chapter 6 of Poundstone (1992). Of course, Dresher and Flood were not the first to notice this dilemma problem. David Hume (1739), for example, noticed sequential prisoner's dilemma: "Your corn is ripe to-day; mine will be so to-morrow. 'Tis profitable for us both, that I shou'd labour with you to-day, and that you shou'd aid me to-morrow. I have no kindness for you, and know you have as little for me. I will not, therefore, take any pains upon your account; and shou'd I labour with you upon my own account, in expectation of a return, I know I shou'd be disappointed, and that I shou'd in vain depend upon your gratitude. Here then I leave you to labour alone: You treat me in the same manner. The seasons change; and both of us lose our harvests for want of mutual confidence and security." in Book 3.2.5. See also Chapter 3 of Alex Abella (2008).

[2] See Alvin E. Roth (1995) for an overview of the experiments.

[3] Simon Gächter and Christian Thöni (2005) confirmed that knowing other subjects who are cooperative made subjects cooperative in a public good provision experiment.

[4] One of their modeling tools is evolutionary dynamics. For example, Christoph Hauert, Arne Traulsen, Hannelore Brandt, Martin A. Nowak, and Karl Sigmund (2007) found how altruistic punishment evolved in the model.

The third approach is to introduce one more stage to the dilemma game in order to *implement* the cooperative outcome.[5] James Andreoni and Hal Varian (1999) and Gary Charness, Guillaume R. Fréchette and Cheng-Zhong Qin (2007) set up a stage where subjects can reward the other subject conditional upon cooperation before the prisoner's dilemma game stage, called the compensation mechanism. The cooperation rate was about 40-70% in this design. Jeffrey S. Banks, Charles R. Plott and David P. Porter (1988) introduced a voting stage after a public good provision stage as Shubik (2011) suggested, and observed that unanimity reduced efficiency. Although costly punishment does not implement cooperation in a traditionally rational model, following Toshio Yamagishi (1986), Ernst Fehr and Simon Gächter (2000) introduced it in a public good provision experiment, and observed that the average contribution rate was 57.5% with the punishment and 18.5% without it under the stranger matching.[6]

Our approach belongs to the third one. As Elinor Ostrom (1990) showed, many successful examples of the commons usually have some devices before and/or after the strategic decisions of obtaining benefits from the commons, and leaving the dilemma in the commons alone without introducing any devices is extremely rare.[7] Therefore, our goal is to find a "minimum" reasonable device or mechanism to make players cooperate *theoretically* and *experimentally* in environments as stark as possible. For this end, we first assume the behavioral principle that appeared in Hume's quotation in Footnote 1, i.e., all players are absolutely selfish. In addition to that, our challenge is to design a mechanism that is also compatible with non-selfish behaviors such as *reciprocal norm* following the finding by Croson (2007), *inequality aversion* and/or *utilitarianism*.[8]

In order to accomplish this task, we restrict ourselves to the class of mechanisms satisfying the following two stringent conditions. First, they must be "*onto*". That is, the four possible outcomes of prisoner's dilemma game are exactly the same as the outcomes of the

---

[5] Martin Shubik (2011) emphasized the need a stage *after* prisoner's dilemma game. "Instead of switching to the cooperative game *per se* if the gap were large enough the agents could construct a mechanism in the form of a second stage to the game that provides coordination, signaling and possibly some other forms of control on the original matrix game in such a way that the players can pay for the administrative costs and still all be able to benefit from its existence."

[6] In addition to this observation, Fehr and Gächter (2000) observed that the average contribution rate was 85% with the punishment and 37.5% without it under the *partner* matching.

[7] Broadly speaking, our approach is one of "*the Game 5 Ways*" proposed by Ostrom (1990) who set up a contract stage before the dilemma stage called Game 5. This game based upon *empirical* findings is quite different from traditional games that utilize central authority or private property rights.

[8] For example, the incentive of a costly punisher with a norm ("if you do not cooperate, I will punish you") is the opposite to the incentive of a payoff maximizer. That is, under costly punishment with two players where one is a costly punisher and the other is a payoff maximizer, the punisher chooses defection and the maximizer chooses cooperation.

mechanisms, and hence the outcomes other than these four should not be used. This implies that they do not accept any payoff flow from or to the outside. In this sense, they must be budget-balanced.[9] For example, a mechanism that gives some monetary payoff to a player who chooses cooperation from the outside is not onto. Furthermore, we impose not using *direct* punishment (or reward) since *personal* punishment (or bribe) is usually prohibited in our modern societies or legal systems.[10] Second, the mechanisms must be "*voluntary*". Any player who chooses defection should not be forced to change the decision to cooperation.

Under the above constraints, we introduce the *approval* stage after the prisoner's dilemma as Adam Smith (1759) suggested.[11] After the dilemma stage, each subject can approve ("*yes*") or disapprove ("*no*") the other choice of the strategy in the first stage. Although there are many ways to design the rule or the mechanism, we employ the following simple one called the *mate choice mechanism*: if both approve the other strategy, the outcome is the one with which both choose in the first stage, and if either one disapproves it, the outcome is the one with which both defect in the first stage[12]. Apparently, the mate choice mechanism satisfies the onto and voluntary conditions. Furthermore, this mechanism satisfies *forthrightness* saying that the outcome must be what players choose whenever both approve the other choices. We also show that the mate choice mechanism is *unique* satisfying forthrightness and several other conditions.

Our basic behavioral principle is that each player is a payoff maximizer. We also consider three types of a player who is a *reciprocator,* an *inequality averter* or a *utilitarian*. A reciprocator chooses "*yes*" if the other player chooses to cooperate, and the player chooses "*no*" if not.[13] An inequality averter prefers equal payoff pairs to unequal payoff pairs. A utilitarian cares the sum of payoff of the two. We assume that they are a payoff maximize in the following sense: they maximize the payoff as far as they follow the behavioral principles. Therefore, our research question is whether or not the mate choice mechanism can align incentives of players who are payoff maximizers, reciprocators, inequality averters and/or utilitarians theoretically and experimentally. Of course, the behavior depends upon an equilibrium concept employed.

---

[9] This is a standard condition in mechanism design. See, for example, Chapter 23 of Andreu Mas-Colell, Michael D. Whinston and Jerry R. Green (1995).

[10] Costly punishment does not satisfy the onto condition. Francesco Guala (2010) surveyed literature including ethnology, anthropology and biology, and concluded that costly punishment was rare.

[11] Smith (1759) stressed the importance of approval in the following manner: "We either approve or disapprove of the conduct of another man according as we feel that, when we bring his case home to ourselves, we either can or cannot entirely sympathize with the sentiments and motives which directed it."

[12] Although Raúl López-Pérez and Marc Vorsatz (2010) also investigated the approval stage after the prisoner's dilemma game, their design at the stage did not affect the final outcomes, and the cooperation rates were 22-38%.

[13] The typical definition of reciprocity is "if the other cooperate, then I will cooperate" (see, for example, Robert Sugden (1984), Matthew Rabin (1993) and Rachel Croson (2007) among others). Since our game has two stages, we take advantage of this structure, and regard approval as a norm of reciprocity.

We prepare five possible equilibrium concepts: Nash equilibrium (*NE*), subgame perfect equilibrium (*SPE*), evolutionarily stable strategies (*ESS*), neutrally stable strategies (*NSS*), and backward elimination of weakly dominated strategies (*BEWDS*).[14] *NE* or *SPE* includes the equilibrium paths where at least one player chooses defection (*D*) and hence (*C,C*), i.e., the outcome where both choose cooperation is not always attained whatever behavioral principles players have. Under *ESS*, (*C,C*) is attained when both are reciprocators. Other than that, no *ESS* exists or the definition is not applicable due to asymmetry. Under *NSS*, (*C,C*) is attained when both are payoff maximizers, reciprocators, inequality averters or utilitarian. Other than that, the definition is not applicable due to asymmetry. Under *BEWDS*, (*C,C*) is attained for all cases except for the case where one player is either a payoff maximize or a reciprocator assuming that the other is a utilitarian. That is, the coverage of *BEWDS* among these behavioral principles is the broadest, and the one of *NSS* is the second.

Our experimental task is to find how subjects cooperate and to identify which equilibrium concepts they choose. In our experimental design, we aim at constructing the environment as bleak as possible against cooperation. In order to avoid possible learning or building-up reputation, no subject ever met another subject more than once, called the complete stranger design.[15] Furthermore, each subject could not identify where the other subject was located in the lab. As usual in this type of experiment, no talking was allowed.

Our observation is rather striking. We observe 93.2% cooperation in the session of prisoner's dilemma game with the mate choice mechanism in 19 periods, and 7.9% cooperation in the session of the game without the mechanism.

We also check the robustness of the mate choice mechanism with two additional and slightly different sessions. The first one is prisoner's dilemma game with *unanimous voting* where we change the wording for the mate choice mechanism, but keep the game-theoretical structure. After prisoner's dilemma game decision, each subject votes "*yes*" or "*no*" to the choice pair in the first stage. Whenever both choose "*yes*", the choice pair is finalized. If either one says "*no*", the outcome when both choose defection is selected. This mechanism is mathematically equivalent to the mate choice mechanism and we observe that the average cooperation rate of three sessions is 95.8%.

The second one is the prisoner's dilemma game with the mate choice mechanism *without*

[14] R. Selten (1975) is the initiator who used the idea of *BEWDS* in game theory, and later Ehud Kalai (1981) used *BEWDS* in the *PD* Game and Banks, Plott and Porter (1988) used it in the provision of a public good in implementing cooperation.

[15] John Duffy and Jack Ochs (2009) reported that random matching treatment in a repeated prisoner's dilemma game failed to generate cooperative norm contrary to a theoretical prediction by Michihiro Kandori (1992).

repetition. All subjects of ten pairs chose cooperation in the first stage, and then chose approval in the second stage. The other session is the prisoner's dilemma game only without repetition and observed that two subjects out of twenty chose cooperation. These sessions show that the mate choice mechanism is robust enough to attain almost full cooperation.

Which equilibrium concept is compatible with the data? The equilibrium paths of *NE* and *SPE* always contain the paths where at least one player chooses *D*. On the other hand, the set of equilibria based upon *BEWDS* is a proper subset of the set of them based upon *NSS* as far as *NSS* equilibria exits. Comparing the off equilibrium paths of *BEWDS* and *NSS*, we find that *BEWDS* is most suitable to explain the data.

The mate choice mechanism reduces cognitive burden of subjects under *BEWDS*. Subjects who use *BEWDS* must compare two dimensional vectors at each subgame after the choice of either cooperation or defection in the prisoner's dilemma game stage. Notice that a payoff vector (u,v) weakly dominates (x,y) if $u \geq x$ and $v \geq y$ and at least one strict inequality. If either one disapproves the other choice in the second stage, then all three payoff vectors out of four at the subgame are the same, that is called the *mate choice flat*. Therefore, subjects must compare just two numbers u and x, not two vectors since $v = y$ due to the flat. Furthermore, this made subjects think backwardly easier than the situation without the flat. That is, we have strong evidence where subjects considered the two stage game backwardly together with the eliminations.

In order to compare the above results, we used the compensation mechanism by Andreoni and Varian (1999) *before* the prisoner's dilemma game that has the same symmetric payoff table in the above experiments although they used an asymmetric payoff table.[16] The outcome when both cooperate is the unique *SPE* in the prisoner's dilemma game with the compensation mechanism although all possible combinations of *C* and *D* are the outcomes of *BEWDS* assuming that both are payoff maximizers. Our finding with 19 rounds was that the average cooperation rate of three sessions is 75.2% that is higher than that in their experiment, but it is significantly different from the rate of the mate choice mechanism.

We observed 760 (10 pairs x 19 rounds x 4 sessions) pairs in mate choice or unanimous voting sessions. An interesting question is what type of behavioral principles the subjects used. The clues are in the off equilibrium path data since off equilibrium choices are different from each other depending upon the combinations of behavioral principles. Utilizing the path data, we estimate that the ratios of payoff maximizers or reciprocators, inequality averters, and utilitarians are 88.13-89.64%, 10.24-10.86% and 0.1-1.02% respectively. This is partially justified by a coder of

[16] See also Hal R. Varian (1994).

the questionnaires.

A good example of the mate choice mechanism is so called *MAD* (Mutually Assured Destruction) that led the earth to the avoidance of nuclear disaster around the last half of the twentieth century.[17] Even though superpower *A* attacks the other superpower *S* using nuclear weapons, superpower *S* can monitor the attack and then has enough time to mount the counterattack. In other words, this is a two stage game where the first stage is a *PD* game, and the second stage is a special case of the approval stage. The approval in the second stage is "No (Further) Attack" and the non-approval is "(Counter) Attack." If a superpower decides to choose "Attack" in the *PD* game, she must choose "No (Further) Attack" automatically in the second stage since she has already chosen "Attack" in the first stage. Then each chooses "No Attack" or "No Action" in the first stage, and then chooses "No (Further) Attack" in the second stage is the unique *BEWDS* path.[18] Notice that the second stage mechanism is not by man made one such as convention, but by an evolving mechanism due to technological constraints including the monitoring accuracy and the time lag between the discharge and explosion that are called second-strike capability by Bruce Russett, Harvey Starr and David Kinsella (page 237, 2009). The technological progresses were due to the battle of holding hegemony over the other superpower.

There are many other examples of the mate choice mechanism.[19] Consider a merger or a joint project of two companies. They must propose plans (the contents of cooperation) in the first stage, and then each faces the approval decision in the second stage. In order to resolve the conflicts such as prisoner's dilemma, interested parties usually form a committee consisting of representatives of the parties. Consider two companies facing confrontation on the standardizations of some product. Each company chooses cooperation (or compromise) or defection (or advocating of the own standard), and then the committee consisting of two company members and/or bureaucrats gives the approval. Another example is the two party system. Each party chooses either cooperation (or compromise) or defection (or insistence of policy for the own party), and then diet (or national assembly) plays a role of approval. The bicameral system also has two stages. One chamber decides a policy (or compromise) and the other chamber plays a role of approval. The negotiation process at United Nations also has this structure. Negotiators among relevant countries get together to find compromise, i.e., the content

---

[17] We thank Toshio Yamagishi who pointed out this example.

[18] Robert J. Aumann (2006) in his Nobel Lecture described *MAD* as an outcome of infinitely repeated games in order to maintain cooperation. The idea of approval mechanism is not to use infinite periods but to consider the game in two stages. Notice also that (*Attack*, *Attack*) is a part of *SPE*, but not a part of the *BEWDS* path. That is, it was fortunate that the decision makers of the superpowers did not follow this path.

[19] We thank Kazunari Kainou who provided some of the examples in this paragraph.

of cooperation in the first week and then high ranked officials such as presidents and prime ministers get together to approve or disapprove it in the second week. Adding the second stage in resolving conflicts has been used widely in our societies.

The organization of the paper is as follows. Section 2 explains the mate choice mechanism as a special case of the approval mechanism. Sections 3, 4, 5 and 6 are for theoretical properties of the mate choice mechanism under *NE, SPE*, *ESS*, *NSS,* or *BEWDS* with payoff maximization, reciprocity, inequality aversion and/or utilitarianism. Section 7 takes care of various possibilities of implementation. Section 8 describes experimental procedures and section 9 is for experimental results. Section 10 explains why *BEWDS* works well and then characterizes the mate choice mechanism. Section 11 is for further research agenda.

2. Prisoner's Dilemma with Approval Mechanism

The prisoner's dilemma (*PD*) game with approval mechanism consists of two stages. In the first stage, players 1 and 2 face a usual *PD* game such as Figure 1. In each cell, the first number is the payoff for player 1 and the second is for player 2. Both players must choose either cooperation (*C*) or defection (*D*) simultaneously. There might be many ways to interpret the matrix in Figure 1, but a typical interpretation in public economics is the payoff matrix of the voluntary contribution mechanism in the provision of a public good. Each player has ten dollars (or initial endowment w) at the beginning, and (s)he must decide whether (s)he contributes all ten dollars (or cooperates) or nothing (or defects). The sum of the contribution is multiplied by $\alpha \in (0.5,1)$, that is 0.7 in the following example, and the benefit goes to both of them, which expresses non-rivalness of the public good. If both contribute, then the benefit of each player is $(10+10)\times 0.7=14$. If either one of them contributes, contributor's benefit is $10\times 0.7=7$, and non-contributor's benefit is 10+7=17 since (s)he has 10 dollars at hand. Therefore, the payoff matrix in Figure 1 keeps this linear structure. Of course, non-contribution (*D*) is the dominant strategy.[20] Bold and italic numbers in the lower right cell show the equilibrium payoff in Figure 1.

		Player 2	
		C	*D*
Player 1	*C*	14,14	7,17
	D	17,7	***10,10***

Figure 1. Prisoner's dilemma game.

Consider now the second stage as in Figure 2. Knowing the strategy pair of the first

[20] In the experiment, we used payoff numbers that are 100 times of the numbers in Figures 1 and 2 due to the exchange rate.

stage, each player must either approve the strategy choice of the other (y) or disapprove it (n) simultaneously. Ellipses show the information sets. Since each set has two alternatives, and there are ten information sets, the total number of possible strategy profiles is 1024 ($=2^{10}$). The upper (lower) number at the bottom of the game tree show player 1's (2's) payoff respectively.

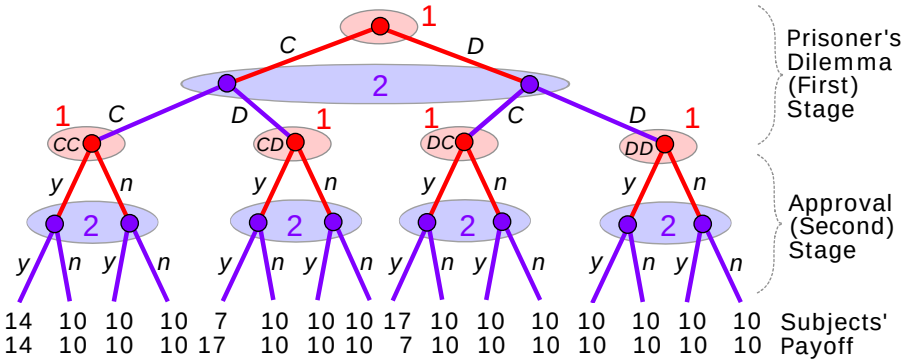


Figure 2. Prisoner's dilemma game with the mate choice mechanism.

Although there are many ways to connect the approval decisions to the strategy choices in the first stage, we choose the following simple way since this procedure has a special feature on the uniqueness of the approval mechanism that will be discussed later: if both approve the other choice in the first stage, then the payoff (or outcome) is what they choose in the *PD* stage. Otherwise, the payoff is (10,10) that corresponds to (*D*,*D*) in the first stage. In the context of public good provision, when either one of them disapproves the choice of the other, the public good will not be provided and hence the money is simply return to the contributors.[21]

Another interpretation of the above specification of the approval mechanism is *mate choice*. A male and a female meet together. If both approve the other, then they can make a mate. If either one of them disapproves the other choice, then they must stay at the status quo since they cannot be a mate. We call this specification of approval mechanism the *mate choice mechanism* (*MCM*).[22] The *PD* game with *MCM* in Figure 2 is abbreviated as *PDMC*.[23]

We will consider several equilibrium concepts whose equilibrium outcomes are different for the next four sections. Sections 3, 4, and 5 are for payoff maximizers, and section 6 is

[21] This specification is different from so called the money back guarantee mechanism. Consider the mechanism if either one of the two chooses *C* but not both, then the 10 contribution is returned to the cooperator. This mechanism cannot generate (7,17) where (*C*,*D*) in the first stage and both choose *y* in the second stage in Figure 2.
[22] There are six approval rules whose outcome is exactly the same as the mate choice rule. See Tatsuyoshi Saijo and Yoshitaka Okano (2009).
[23] The definition of mate choice in biology is much broader than our usage: according to T. R. Halliday (1983), "Mate choice may be operationally defined as any pattern of behaviour, shown by members of one sex, that leads to their being more likely to mate with certain members of the opposite sex than with others."

for reciprocators, inequality averters or the mixture of them including payoff maximizers.

3. Nash and Subgame Perfect Equilibria of the *PDMC*

An equilibrium concept that has been widely used in analyzing the two stage game is subgame perfect equilibrium (*SPE*). Consider four subgames in the second stage in Figure 3. Bold and italic numbers in a cell show that the pair is Nash equilibrium (*NE*) and some of them are not black but gray that are eliminated under *BEWDS* in the next section. Subgame *CC* has two *NE*s ((*y*,*y*) and (*n*,*n*)). Similarly, subgame *CD* has two *NE*s ((*n*, *y*) and (*n*, *n*)), subgame *DC* has two *NE*s ((*y*,*n*) and (*n*,*n*)), and subgame *DD* has four *NE*s ((*y*, *y*), (*y*, *n*), (*n*,*y*) and (*n*,*n*)).

The four subgames have a point in common: the payoff at (*y*,*n*), (*n*,*y*) and (*n*,*n*) is (10,10) and hence it is *flat,* that we call the *mate choice flat*, in the matrices due to the *MCM*. Any payoff that is lower than the status quo payoff, i.e., 10, would never be an *NE*. That is, (7,17) or (17,7) would not be chosen and the mechanism prevents free-riding. In this sense, the mechanism is a device for *survival* not to end up at payoff 7. Another point in common is that (*n*,*n*) is always an *NE* due to the mate choice flat. This makes Pareto inferior payoff vector (10,10) to (14,14) survive as an equilibrium, and hence this needs an equilibrium refinement or different equilibrium concepts to exclude (10,10).

Player 2

Player 1		Subgame *CC*		Subgame *CD*		Subgame *DC*		Subgame *DD*	
		y	*n*	*y*	*n*	*y*	*n*	*y*	*n*
y		***14,14***	10,10	7,17	10,10	17,7	***10,10***	***10,10***	***10,10***
n		10,10	***10,10***	***10,10***	***10,10***	10,10	***10,10***	***10,10***	***10,10***

Figure 3. Four subgames in *PDMC*.

Given the outcomes of four subgames, we can construct the reduced normal form games. Since the payoff of all *NE*s in subgames *CD, DC* and *DD* is (10,10), consider two cases (*y*,*y*) and (*n*,*n*) in subgame *CC*. If it is (*y*,*y*) in subgame *CC*, there are two *NE*s (*C*,*C*) and (*D*,*D*) in the reduced normal form game (see Figure 4-(i)). Since each equilibrium has 16 cases (i.e., two *NE*s in subgame *CD*, two in subgame *DC*, and four in subgame *DD*), there are 32 *SPE*s. On the other hand, if it is (*n*,*n*) in subgame *CC*, there are four *NE*s (*C*,*C*) , (*C*,*D*), (*D*,*C*) and (*D*,*D*) (see Figure 4-(ii)). Since each equilibrium has 16 cases, we have 64 *SPE*s. In total, there are 96 *SPE*s. Notice also that two games in Figure 4 have the mate choice flat.

		Player 2	
		C	*D*
Player 1	*C*	***14,14***	10,10
	D	10,10	***10,10***

(i) (*y*, *y*) in subgame *CC*

		Player 2	
		C	*D*
Player 1	*C*	***10,10***	***10,10***
	D	***10,10***	***10,10***

(ii) (*n*, *n*) in subgame *CC*

Figure 4. The prisoner's dilemma stage in the backward induction.

Consider the *SPE* paths in Figure 2.[24] First, fix (*y*,*y*) at subgame *CC*. Then they are (*C*,*C*,*y*,*y*) (16 cases), (*D*,*D*,*y*,*y*) (4 cases), (*D*,*D*,*y*,*n*) (4 cases), (*D*,*D*,*n*,*y*) (4 cases), and (*D*,*D*,*n*,*n*) (4 cases). Second, fix (*n*,*n*) at subgame *CC*. Then they are (*C*,*C*,*n*,*n*) (16 cases), (*C*,*D*,*n*,*y*) (8 cases), (*C*,*D*,*n*,*n*) (8 cases), (*D*,*C*,*y*,*n*) (8 cases), (*D*,*C*,*n*,*n*) (8 cases), (*D*,*D*,*y*,*y*) (4 cases), (*D*,*D*,*y*,*n*) (4 cases), (*D*,*D*,*n*,*y*) (4 cases), and (*D*,*D*,*n*,*n*) (4 cases). In what follows, for example, we use *CDny* instead of (*C*,*D*,*n*,*y*).

Consider *NE*s of the game in Figure 2. Since each player has 32 strategies, we can construct a 32 by 32 payoff matrix. Finding the intersection of best responses of two players, we have 416 *NE*s with equilibrium paths of *CCyy* (16 cases), *DDyy* (64 cases), *DDyn* (64 cases), *DDny* (64 cases), *DDnn* (64 cases), *CCnn* (16 cases), *CDny* (32 cases), *CDnn* (32 cases), *DCyn* (32 cases), and *DCnn* (32 cases). Summarizing these, we have,

Property 1. *In the PDMC, we have*
(i) *416 NEs and 96 SPEs out of 1024 possible strategy profiles*;
(ii) *the NE paths and the SPE paths are the same and they are CCyy*, *DDyy*, *DDyn*, *DDny*, *DDnn*, *CCnn*, *CDny*, *CDnn*, *DCyn*, *and DCnn*; *and*
(iii) *the payoff of 16 NEs and 16 SPEs is* (*14,14*) *on the path CCyy and the payoff of the rest is* (*10,10*).

Let us define player *i*'s strategy of the two stage game as $s_i = (E_i, s_i^{CC}, s_i^{CD}, s_i^{DC}, s_i^{DD})$ where E_i is *i*'s choice between *C* and *D* in the *PD* stage, and s_i^{AB} is *i*'s choice between *y* and *n* in the *MCM* when player *i* chooses *A* and player *j* chooses *B* in the *PD* stage.[25] Then we have,

Property 2. *16 strategy profiles where the outcome of SPE is* (*C*,*C*) *are*
$(s_1, s_2) = ((C, y, n, \cdot, \cdot), (C, y, n, \cdot, \cdot))$ *where "* $\cdot$ *" indicates either y or n.*

4. Neutrally Stable Strategies of *PDMC*

[24] A *path* is (subject 1's choice between *C* and *D*, subject 2' choice between *C* and *D*, subject 1's choice between *y* and *n*, subject 2's choice between *y* and *n*).
[25] For example, consider $s_1 = s_2 = (C, y, n, y, y)$. This indicates that player 1 chooses *n* and player 2 chooses *y* at subgame *CD*, and player 1 chooses *y* and player 2 chooses *n* at subgame *DC*.

This section presents neutrally stable strategies (*NSS*), which is a refinement of *NE*, of the *PDMC*. We will show that all *NSS* paths are *CCyy*, and the game does not have evolutionarily stable strategy. Since players 1 and 2 are symmetric, let us abbreviate player's subscript and let $v(s,t)$ be the payoff of player 1 when the strategy profile is (s,t). Due to payoff symmetry, player 2's payoff at (s,t) is $v(t,s)$.

Definition 1. A strategy t is a *neutrally stable strategy* if and only if for all $t' \neq t$,
(i) $v(t,t) \geq v(t',t)$ and (ii) $v(t,t) = v(t',t)$ implies $v(t,t') \geq v(t',t')$.

If the weak inequality in (ii) becomes strict, then t is called an *evolutionarily stable strategy* (*ESS*).

Property 3. *A strategy t is an NSS if and only if* $t = (C, y, n, \cdot, \cdot)$ *where "* $\cdot$ *" indicates either y or n.*

Proof. See Appendix.

Combining Properties 2 and 3, we have,

Property 4. *All 16 NSS profiles are exactly the same as the 16 SPE profiles whose payoff is* (14,14).

Regarding *ESS*, we have,

Property 5. *There is no ESS.*

Proof. See Appendix.

5. Backward Elimination of Weakly Dominated Strategies of *PDMC*

Another equilibrium concept whose outcome exactly coincides with (C,C) is backward elimination of weakly dominated strategies (*BEWDS*) which is also adopted, for example, in Ehud Kalai (1981). This requires two properties. The first is subgame perfection and the second is that players do not choose weakly dominated strategies in each subgame and the reduced normal form game.

Let us take a look at the subgame whose starting node is *CC* in Figure 2. Notice that (14,10) corresponds to *y* and (10,10) to *n* for both players. We say strategy α with (u,v) *weakly dominates* strategy β with (x,y) if $u \geq x$ and $v \geq y$ with at least one strict inequality or $(u,v) \geq (x,y)$. That is, since *y* weakly dominates *n*, *n* should not be chosen. Therefore, (y,y) is realized at subgame *CC*. Similarly, (n,y) at subgame *CD* and (y,n) at subgame *DC* are realized. In subgame

DD, since no weakly dominated strategy exists, (*y*,*y*), (*y*,*n*), (*n*,*y*) and (*n*,*n*) are realized. Given the realized strategies in all subgames, we have the reduced normal form game in Figure 4-(i). In this game, *C* weakly dominates *D* for both players and hence, (*C*,*C*) is the realized outcome. Since there are four realized pairs in subgame *DD*, there are four realized predictions in the two stage game. Notice that the order of elimination in each stage does not change the final outcome.

Property 6. *Using BEWDS in PDMC, we have*
(i) *four realized predictions; and*
(ii) *the unique prediction path is CCyy with the payoff of* (14,14).

6. Reciprocators, Inequality Avertors and Utilitarians

Thus far, the payoff maximization is the objective of each player. Following Croson (2007), we impose the following reciprocal norm: if the other chooses *C*, then I will approve it, and if not, then I will disapprove it. We call a payoff maximizer who has this norm a *reciprocator* (*R*). Maximizing behavior depends on equilibrium concepts. Consider the following two cases: both are reciprocators, and one of them is a reciprocator and the other is payoff maximizer.[26]

Let us consider the case where both are reciprocators. Then the reciprocal norm implies $s_i = (E_i, s_i^{CC}, s_i^{CD}, s_i^{DC}, s_i^{DD}) = (\cdot, y, n, y, n)$. The payoff maximization behavior and equilibrium concepts employed determine the first element of s_i. Suppose first that both are *R*s. Then, the reduced normal form game is exactly the same as the payoff matrix of Figure 4-(i). Therefore, the choices of *NE* and *SPE* are (*C*,*C*) and (*D*,*D*). By definition, the choice of *NSS, ESS,* and *BEWDS* is (*C*,*C*).

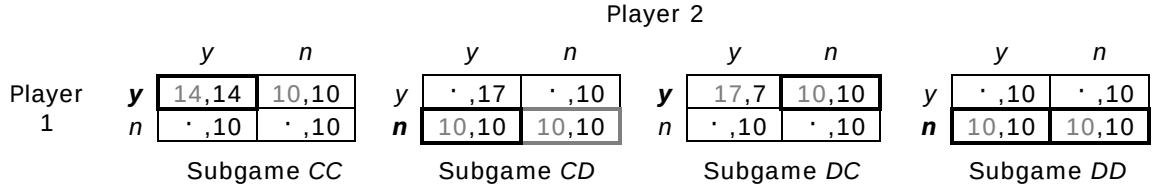


Figure 5. Subgames when player 1 is a reciprocator and player 2 a payoff maximizer.

Consider the case where player 1 is an *R* and player 2 is a payoff maximizer. Then strategies for player 1 are *C* and *D*, i.e., $E_1 = C$ or D since the choices of subgames are determined by the norm, i.e., $(s_1^{CC}, s_1^{CD}, s_1^{DC}, s_1^{DD}) = (y, n, y, n)$, and player 2 has $2^5 = 32$ strategies. Then there are 24 *NE*s with equilibrium paths of *CCyy*, *DCyn*, *DDny* and *DDnn*. The cells with bold square in

[26] We implicitly assume that each player knows the other type of behavior in this section. However, this assumption will be relaxed in the next sections.

Figure 5 show the outcomes of each subgame, and hence we can obtain the matrix of Figure 4-(i) and 8 *SPE*s with equilibrium paths of *CCyy*, *DDny* and *DDnn*. Consider *BEWDS*. Since player 2 chooses *y* at subgame *CD* in Figure 5, and *y* and *n* are indifferent in subgame *DD*, $s_2 = (E_2, s_2^{CC}, s_2^{CD}, s_2^{DC}, s_2^{DD}) = (C, y, n, y, \cdot)$. That is, we have two realized predictions in the normal form game, and the unique prediction path is *CCyy*.

We introduce two more behavioral principles: an *inequality averter* who prefers (10,10) to (7,17) or (17,7) and prefers (14,14) to (10,10) and a *utilitarian* who cares the sum of payoffs. That is, a utilitarian prefers (14,14) to (7,17) or (17,7) and prefers (7,17) or (17,7) to (10,10). Since we have four behavioral principles and five equilibria, we must consider 50 combinations. Table 1 summarizes all possible equilibrium paths of the combinations.[27]

(Player 1, Player 2)	*NE*	*SPE*	*NSS*	*ESS*	*BEWDS*
MM	*CCyy, CCnn, CDny, CDnn, DCyn, DCnn, DD··*	*CCyy, CCnn, CDny, CDnn, DCyn, DCnn, DD··*	*CCyy*	*No ESS*	*CCyy*
RR	*CCyy, DDnn*	*CCyy, DDnn*	*CCyy*	*CCyy*	*CCyy*
II	*CCyy, CCnn, CDyn, CDny, CDnn, DCyn, DCny, DCnn, DD··*	*CCyy, CCnn, CDny, CDyn, CDnn, DCyn, DCny, DCnn, DD··*	*CCyy*	*No ESS*	*CCyy*
UU	*CCyy, CCnn, CDyy, CDnn, DCyy, DCnn, DD··*	*CCyy, CCnn, CDyy, CDnn, DCyy, DCnn, DD··*	*CCyy*	*No ESS*	*CCyy*
MR	*CCyy, CDny, DDyn, DDnn*	*CCyy, DDyn, DDnn*	*n.a.*	*n.a.*	*CCyy*
MI	*CCyy, CCnn, CDyn, CDny, CDnn, DCyn, DCnn, DD··*	*CCyy, CCnn, CDyn, CDny, CDnn, DCyn, DCnn, DD··*	*n.a.*	*n.a.*	*CCyy*
MU	*CCyy, CCnn, CDny, CDnn, DCyy, DCnn, DD··*	*CCyy, CCnn, CDny, CDnn, DCyy, DCnn, DD··*	*n.a.*	*n.a.*	*DCyy*
RI	*CCyy, DCyn, DDny, DDnn*	*CCyy, DDny, DDnn*	*n.a.*	*n.a.*	*CCyy*
RU	*CCyy, DCyy*	*DCyy*	*n.a.*	*n.a.*	*DCyy*
IU	*CCyy, CCnn, CDny, CDnn, DCny, DCnn, DD··*	*CCyy, CCnn, CDny, CDnn, DCny, DCnn, DD··*	*n.a.*	*n.a.*	*CCyy*

M: Payoff Maximizer, *R*: Reciprocator, *I*: Inequality Averter, *U*: Utilitarian
n.a.: not applicable due to asymmetry of the game.
"··" in *DD··* indicates *yy, yn, ny or nn*.

Table 1. Equilibrium Path of *PDMC* under Four Behavioral Principles and Five Equilibrium Concepts.

[27] See Okano (2012) for the proofs.

7. Implementability

We will consider implementability of the *MCM* in an economic environment with a public good.[28] Let $u_i(x_i, y) = x_i + \alpha_i y$ be a utility function defined on R_+^2 where x_i is a private good, y is a pubic good, and let $\boldsymbol{U} = \{(u_1, u_2) : u_i = x_i + \alpha_i y \text{ for some } \alpha_i \in (0.5,1)\}$. Let $y = h(x) = \sum t_i$ be a production function of the public good where $t_i = w_i - x_i$ and w_i is player i's initial endowment. Then let $\boldsymbol{A} = \{((x_1, y),(x_2, y)) \in R_+^4 : y = \sum (w_i - x_i)\}$ be the set of feasible allocations. Define a social choice correspondence $f : \boldsymbol{U} \to\!\!\!\!\to \boldsymbol{A}$ by $f(u) =$ the set of maximizers of $\sum u_i(x_i, y)$ on $\boldsymbol{A}$. Apparently, this correspondence is a function in our setting, and the optimal level of the public good is $w_1 + w_2$. Let g: $\boldsymbol{S} \to \boldsymbol{A}$ be a game form (or mechanism) where $\boldsymbol{S}$ is the set of strategy profiles, and let $E_g : \boldsymbol{U} \to\!\!\!\!\to \boldsymbol{S}$ be the equilibrium correspondence based upon equilibrium concept E. Then we say that mechanism g *implements* f in E if $f(u) = g \cdot E_g(u)$ for all u. In our special case, we set $w_1 = w_2 = 10$ and $\alpha_1 = \alpha_2 = 0.7$. We also regard $t_1 = t_2 = 10$ as *C*, and $t_1 = t_2 = 0$ as *D*. Furthermore, we do not allow any number between 0 and 10.[29]

Consider first that both are payoff maximizers. In the *PD* case, the strategy space for each player is {0,10} and the game form is $h(t_1, t_2) = ((x_1, y),(x_2, y))$ with $y = \sum t_i$. Let $\boldsymbol{D_h}$ be the set of dominant strategy equilibria. Then since $h \cdot \boldsymbol{D_h}(u) = ((w_1, 0),(w_2, 0))$, h cannot implement f in dominant strategy equilibria.

Let g be the *MCM*. Write $\boldsymbol{BEWDS}_g(u)$ as the set of realized predictions by *BEWDS* using g under u. Then $g \cdot \boldsymbol{BEWDS}_g(u) = f(u)$ and hence g implements f in *BEWDS*. Similarly, writing $\boldsymbol{NSS}_g$ as the set of *NSS* pairs, we have $g \cdot \boldsymbol{NSS}_g(u) = f(u)$ for all possible u. That is, g implements f in *NSS*[30]. Let $\boldsymbol{NE}_g(u)$ ($\boldsymbol{SPE}_g(u)$) be the set of *NE*s (*SPE*s) using g under u. Then $g \cdot \boldsymbol{NE}_g(u) \neq f(u)$ ($\neq g \cdot \boldsymbol{SPE}_g(u)$), and hence g cannot implement f in *NE* (or *SPE*). Obviously, g cannot implement f in *ESS* since no *ESS* exists. For the other cases, using Table 1, we have,

Property 7.

(1) *Cooperation cannot be attained in the PD game in dominant strategy;*

(2) *MCM cannot implement cooperation of PD in either NE or SPE;*

(3) *MCM implements cooperation of PD in NSS if both are payoff maximizers, reciprocators, inequality avertors or utilitarians;*

---

[28] See Eric Maskin (1999) and Saijo (1988) for Nash implementation, John Moore and Rafael Repullo (1988) for *SPE* implementation, and Matthew O. Jackson (2001) for a general survey.

[29] As Takehito Masuda, Okano and Saijo (2011) shows, if the number of strategies is more than two, the *MCM* fails to implement the social choice correspondence in *BEWDS*.

[30] Sho Sekiguchi (2012) show that the prisoner's dilemma game with approval stage implements the Pareto optimal outcome in an evolutional dynamics model.

(4) *MCM implements cooperation of PD in ESS if both are reciprocators; and*
(5) *MCM implements cooperation of PD in BEWDS if both players are payoff maximizers, reciprocators, inequality averters, or the mixture of them except for the case when one player is either a payoff maximize or a reciprocator when the other player is a utilitarian.*

MCM implements cooperation with four behavioral principles including some of the mixture of them under *BEWDS*. In particular, the mechanism implements a social goal under *an* equilibrium concept even there are *three* types of players such as payoff maximizers, reciprocators and inequality avertors, and the mixture of them. In this sense, we name this *tripartite* or *multipartite* implementation. Under this notion, it is not necessary to know that a player must know which type of the other player is. Notice that this is different from *double* or *triple* implementation in the literature. Given *one* behavioral principle such as payoff maximizing behavior, a mechanism *doubly* or *triply* implements some social goal with *two* or *three* equilibria.[31]

8. Experimental Procedures

We conducted the experiments in November 2009, March and November 2010, and October, November and December 2011 at Osaka University. We had, in total, ten experimental sessions. The *PDMC* experiment had three sessions and the *PD* experiment had one session. The experiment of *PD* with unanimous voting (*PDUV*) had one session and the experiment of *PD* with compensation mechanism (*CMPD*) had three sessions, which will be explained later. In these sessions, subjects played the game nineteen rounds. The *PDMC** and *PD** experiments, where "*" indicates no repetition of the game, had one session.

Twenty subjects participated in each session, and hence the total number of subjects was 200. No subjects participated in more than one session. We recruited these subjects by campus-wide advertisement. They were told that there would be an opportunity to earn money in a research experiment. Communication among the subjects was prohibited, and we declared that the experiment would be stopped if it was observed. This never happened. The subjects' information and durations of experimental session were summarized in Table A1 in the appendix.

The experimental procedure is as follows. We made ten pairs out of twenty subjects seated at computer terminals in each session[32]. The pairings were anonymous and were

[31] For the recent development of multiple (not multipartite) implementation and its experiment, see Saijo, Tomas Sjöström, and Yamato (2007) and Timothy Cason, Saijo, Sjöström and Yamato (2006). As for tripartite implementation, the *MCM* is the first mechanism that explicitly deals with multiple behavioral principles under one equilibrium concept to the best of our knowledge.
[32] We used the z-Tree program developed by Urs Fischbacher (2007).

determined in advance so as not to pair the same two subjects more than once in sessions with repetitions. Since most of the previous studies such as Andreoni and Varian (1999) (Charness, Fréchette and Qin (2007)) employed random matching among 4 to 8 subjects (2 to 4 groups)[33], the repetition necessarily entails of pairings of the same two subjects. Therefore, compared to the previous experiments, this "complete" strangers design might reduce possibility of cooperation among subjects.[34] Each subject received instruction sheet and record sheet. The instruction was read loudly by the same experimenter.

Let us explain the *PDMC* experiment. Before the real periods started, we allowed the subjects five minutes to examine the payoff table and to consider their strategies. When the period started, each subject selected either *A* (defection) or *B* (cooperation) in the choice (or *PD*) stage, and then inputted the choice into a computer and also filled in it on the record sheet. After that, each subject wrote the choice reason in a small box on the record sheet by hand. Then the next was the decision (or approval) stage. Knowing the other's choice, each subject chose to either "accept" or "reject" the other's choice, and then inputted the decision into a computer and also filled in it on the record sheet. After that, each subject wrote the reason in a small box by hand. Once every subject finished the task, each subject could see "your decision," "the other's decision," "your choice," "the other's choice," "your points," and "the other's points" on the computer screen. However, neither the choices nor the decisions in pairs other than "your" own were shown on the computer screen. This ended one period. The experiment without the decision stage became the *PD* experiment. After finishing all nineteen periods, every subject filled in questionnaire sheets. The *PDMC** and *PD** experiments were exactly the same as the *PDMC* and *PD* experiments without repetition, respectively.

In order to examine the robustness of the mate choice mechanism and to understand the framing effect, we also conducted the *PD* game with unanimous voting (*PDUV*) experiment. The experimental procedure is exactly the same as in the *PDMC* experiment except for the unanimous voting stage. Each subject must vote for the outcome of the *PD* stage. If both affirm the strategy choices in the *PD* stage, then the outcome is what they choose in the *PD* stage. Otherwise, the outcome is (10,10). That is, *PDMC* and *PDUV* are mathematically equivalent, but not *cognitively*. For example, suppose that (*C,D*) (or (*B,A*) in the experiment) is observed in the *PD* stage. In the *PDMC*, subject 1 is asked to choose either approve or disapprove subject 2's choice *D*, but in the *PDUV*, subject 1 is asked to vote on the outcome (*C,D*). In this sense, comparing the *PDMC* with

[33] Charness et al. (2007) partitioned 16 subjects in one session into four separate groups, with the 4 subjects in each group interacting only with each other over the course of the session.
[34] An exception is Cooper et al. (1996) who employed the complete stranger matching.

the *PDUV* is to understand the framing effect.

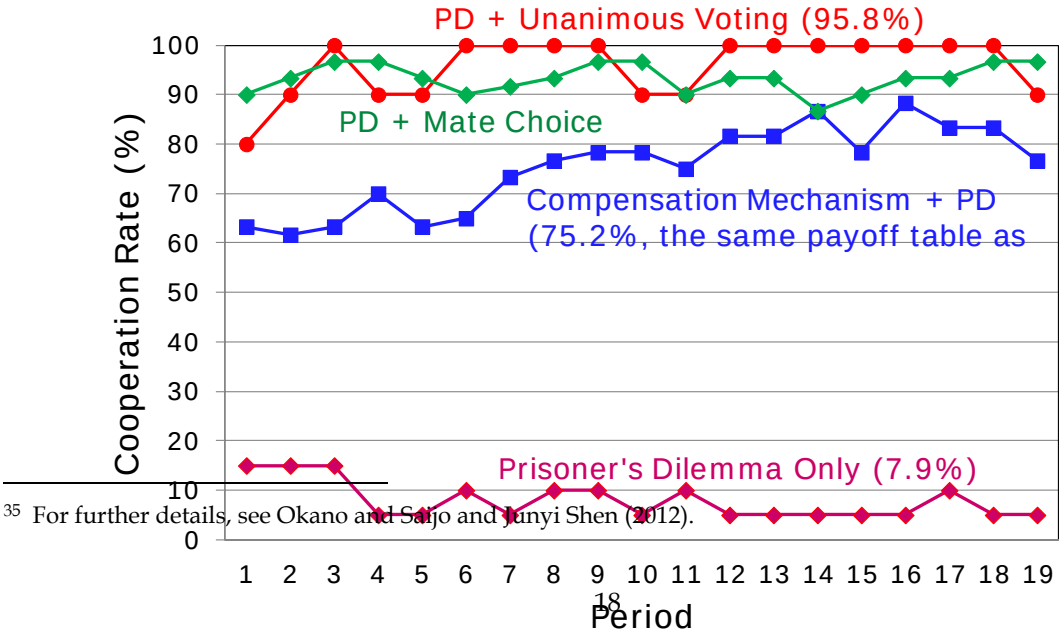
We also compare our results with two-stage game experiments introduced by Andreoni and Varian (1999). They added a stage called the *compensation mechanism* (*CM*) where each subject could offer to pay the other subject to cooperate *before* the *PD* stage. Then they showed that the unique *SPE* outcome was Pareto efficient in their asymmetric payoff table although all possible combinations of *C* and *D* are the outcomes of *BEWDS* assuming that both are payoff maximizer.[35] Eight subjects in a group formed four groups and the matching was random. They played a usual *PD* game for the first 15 periods, and then played the two stage game from 16 to 40 periods. The cooperation rate of the former was 25.8% and the latter was 50.5%.

We used the *PD* game in Figure 1 rather than their asymmetric game. We refer to this experiment as the *CMPD* experiment. The unique *SPE* outcome is that both offer three in the compensation mechanism, and choose cooperation in the *PD* stage. Due to discreteness of strategies, the equilibrium offers are either three or four in the compensation mechanism. On the other hand, all possible combinations such as (*C*,*C*), (*C*,*D*), (*D*,*C*) and (*D*,*D*) fall into *BEWDS* with equilibrium offer of three. When the offer takes discrete value, the *BEWDS* outcomes are 33*CC*, 33*CD*, 33*DC*, 33*DD*, 34*CC*, 34*CD*, 43*CC*, 43*DC* and 44*CC*.

9. Experimental Results

9.1. The Effect of the Mate Choice Mechanism

Figure 6 shows the cooperation rates of the *PDMC*, the *PDUV*, the *CMPD* and the *PD* experiments per period. We use *ex post* cooperation rate in *PDMC* and *PDUV* experiments. For example, if both chose *C* in the choice stage and one of the subjects disapproved the other choice in the decision stage, then we did not count their choices as cooperation.



[35] For further details, see Okano and Saijo and Junyi Shen (2012).

Figure 6. Cooperation Rates of Four Experiments.

The *PDMC* experiment achieves high cooperation rate from periods 1 to 19. The cooperation rate is more than or equal to 90 percent in all periods except for period 14. Overall, the cooperation rate is 93.2 percent[36]. On the other hand, the *PD* experiment exhibits 7.9 percent cooperation rate, 11 percent for the first five periods declining to 6 percent for the last five periods. No (*C*,*C*) was observed among 190 pairs of choices. The cooperation rate in our experiment is slightly lower than the previous experiments. For example, Alvin E. Roth and J. Keith Murnighan (1978) find 10.1% cooperation, Russell Cooper et al. (1996) find 20%, and Andreoni and Miller (1993) find 18%. Hence, our subjects are more in line with game-theoretic logic such as adopting dominant or Nash equilibrium strategy. The difference of cooperation rates between *PDMC* and *PD* experiments is statstically significant (p-value < 0.001, Wilcoxon rank-sum test[37]). Hence, *MCM* has strong effect making subjects more cooperative.

Takaoka, Okano and Saijo (2012) conducted a series of experiment with a session in which 22 subjects played the *PDMC* for the first ten periods and then the *PD* for the last ten periods, and a session in which another 22 subjects played the *PD* for the first ten period and then the *PDMC* for the last ten periods. They found the similar results described above. Using the data in the first ten periods of these sessions, *PDMC* achieves 91.8 percent of the cooperation rate and *PD* achieves 7.7 percent of the cooperation rate. This difference is statistically significant (p-value < 0.001, Wilcoxon rank-sum test)[38].

Observation 1.

(i) *In the PD game with the mate choice mechanism experiment, the cooperation rate is 93.2 percent and more than or equal to 90 percent in all periods except for period 14.*

(ii) *In the PD game only experiment, the cooperation rate is 7.9 percent and no* (*C*,*C*) *was observed among*

[36] 96.9 percent of choices are *C* in the choice (first) stage.

[37] Performing Wilcoxon rank-sum test, we first calculate average cooperation rate of each subject across periods, and then calculate the test statistic using the averages in order to eliminate correlation across periods.

[38] Using the data in the last ten periods, Takaoka, Okano and Saijo (2011) found that *PDMC* and *PD* achieve 90.9% and 9.1% of the cooperation rates respectively. The difference is statistically significant (p-value < 0.001, Wilcoxon rank-sum test).

190 pairs of choices.
(iii) *The cooperation rate in the PD game with the mate choice mechanism experiment is significantly different from that in the PD game only experiment.*

9.2. The Robustness of the Mate Choice Mechanism

The *PDUV* experiment also achieves high cooperation rate, more than or equal to 90 percent in all periods except for period 1. The overall cooperation rate is 95.8 percent[39]. The difference of cooperation rates between the *PDUV* and the *PD* experiments is statistically significant (p-value < 0.001, Wilcoxon rank-sum test). Hence, the unanimous voting also has strong effect making subjects more cooperative. The difference of cooperation rates between the *PDUV* and the *PDMC* experiments is not statistically significant (p-value = 0.657, Wilcoxon rank-sum test). This indicates that the difference of the wordings for the mechanism does not have a significant effect on the subjects' behavior.

Next, we describe the results of the experiments without repetition. All twenty subjects chose *C* in the choice stage, and then approved the other choice in the decision stage in the *PDMC** experiment. Two subjects chose *C* and eighteen chose *D* in the *PD** experiment (the cooperation rate is 10 percent), which is similar to the data of *PDMC* and *PD* experiments. The difference of cooperation rates between the *PDMC** and the *PD** experiments is statistically significant (p-value < 0.001, chi-square test[40]). When we compare the cooperation rates between the *PDMC* experiment of the first period and the *PDMC** experiments, the *PDMC* experiment of the last period and the *PDMC** experiment, the *PD* experiment of the first period and the *PD** experiment and the *PD* experiment of the last period and the *PD** experiment, they are not statistically significant (p-values > 0.1 in all tests, chi-square test). Hence, the number of repetition does not have the effect on the performance of the mechanism. Summarizing the results, we have,

Observation 2.
(i) *In the PD game with unanimous voting experiment, the cooperation rate is 95.8 percent.*
(ii) *The cooperation rate in the PD game with unanimous voting is significantly different from the rate in the PD game only experiment and not significantly different from the rate in the PDMC experiment.*
(iii) *In the PD game with the mate choice mechanism experiment without repetition, all twenty subjects*

[39] 98.2 percent of choices are C in the choice (first) stage.
[40] In the chi-square test, the true cooperation rate in each session is replaced by its maximum likelihood estimate. Test statistic is distributed asymptotically as a chi-square with 1 degree of freedom under the null hypothesis.

chose the cooperative strategy in the dilemma stage, and then approved the other's choice. In the PD game only experiment without repetition, on the other hand, two out of twenty subjects (10%) chose the cooperative strategy.

(iv) *The cooperation rate in the PD game with the mate choice mechanism without repetition is significantly different from the rate in the PD game only experiment without repetition.*

(v) *The cooperation rate in the PD game with the mate choice mechanism experiment without repetition is not statistically different from the rate in the first and the last period in the PD game with the mate choice mechanism experiment with repetition.*

(vi) *The cooperation rate in the PD game only experiment without repetition is not statistically different from the rate in the first and the last period in the PD game only experiment with repetition.*

9.3. The Comparison of the Mate Choice Mechanism with the Compensation Mechanism

We will compare the performances between the mate choice mechanism and the compensation mechanism. Overall, 75.2 percent of choices are cooperative in the *CMPD* experiment. The cooperation rate is significantly different from that in the *PD* experiment (p-value < 0.001, Wilcoxon rank-sum test), indicating that the compensation mechanism has the effect of making subjects more cooperative. The cooperation rate is also significantly different from those in *PDMC* and *PDUV* experiments (p-value < 0.001 for both tests, Wilcoxon rank-sum test), indicating that the mate choice mechansm outperforms the compensation mechanism.

Though the cooperation rate does not reach more than 90 percent at the end of the experiment, it increases over time, 64.3 percent for the first five periods increasing to 82.0 percent for the last five periods. In order to examine whether the cooperation rate increases as periods proceed, we ran a simple random effect probit model. The dependent variable takes the value of 1 if the subject chooses *C* and 0 otherwise. The independent variables are the period number and the constant. The result indicates that the coefficient of the period is significantly greater than zero at the 1 percent significance level.

Table 2 reports the p-values of the chi-square tests for equality of cooperation rates for each period between *PDMC* and *CMPD* experiments, between the *PDUV* and the *CMPD* experiments, between combination of *PDMC* and *PDUV* experiments and the *CMPD* experiment. We evaluate the test at the 5 percent significance level. The p-values in bold face indicate that we cannot reject the null hypothesis that the cooperation rate is the same between two experiments. From periods 1 to 10, the difference of cooperation rate is not significant in only periods 1, 4 and 10 in the tests of *PDUV* vs. *CMPD*. This indicates that the compensation mechanism cannot achieve high cooperation rate from ealry periods while the mate choice mechanism can do. From

periods 11 to 19, on the other hand, there are many periods in which the difference of cooperation is not significant. This indicates that the compensation mechanism can acheives high cooperation so that it is not statistically different from the mate choice mechanism though it needs the repetition.

Period	1	2	3	4	5	6	7	8	9	10
PDMC vs. *CMPD*	0.001	0.000	0.000	0.000	0.000	0.001	0.008	0.011	0.002	0.002
PDUV vs. *CMPD*	**0.168**	0.018	0.001	**0.074**	0.024	0.002	0.010	0.017	0.023	**0.247**
(*PDMC* & *PDUV*) vs. *CMPD*	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003

Period	11	12	13	14	15	16	17	18	19
PDMC vs. *CMPD*	0.031	**0.053**	**0.053**	**1.000**	**0.080**	**0.343**	**0.088**	0.015	0.001
PDUV vs. *CMPD*	0.156	0.039	0.039	**0.085**	0.023	**0.110**	**0.051**	**0.051**	**0.197**
(*PDMC* & *PDUV*) vs. *CMPD*	0.018	0.012	0.012	**0.540**	0.015	**0.147**	0.023	0.003	0.001

Table 2: The *p*-values of Chi-square Test for Each Period

In the *SPE* and *BEWDS*, players should offer 3 or 4 (300 or 400 in the experiment). We find that the actual behavior is consistent with this prediction on average. Overall, the average side payment is 349.47. In each of 19 periods, subjects offer their side payment between 300 and 400 on average. The minimum average side payment is 316.67 in period 5, and the maximum average side payment is 366.67 in period 7.

Observation 3.

(i) *In the CMPD experiment, 75.2 percent of choices are cooperative: 64.3 percent for the first five periods and 82.0 percent for the last five periods, and the cooperation rate significantly increases as periods proceed.*

(ii) *The overall cooperation rate in the CMPD experiment is significantly different from the rate in the PD game only experiment, the PDMC experiment and the PDUV experiment.*

(iii) *The average side payment is 349.47 with 316.67 as the minimum and 366.67 as the maximum where the expected side payment is between 300 and 400 under the SPE and BEWDS.*

9.4. Performance of Equilibria in *PDMC* and *PDUV* Experiments

Table 3 reports the frequencies of pairs with which *NE*, *SPE*, *NSS*, *BEWDS* and the other paths occurred in *PDMC* and *PDUV* experiments. In total, 99.1 percent of all pairs fall into *NE* and *SPE* paths. However, *NE* and *SPE* have many paths where almost no pairs are. In particular, no pairs are observed on *CCnn*, *DDyn* and *DDnn* paths. In this sense, *NE* and *SPE* poorly describe subjects' behavior. On the other hand, since the unique *NSS* and *BEWDS* path covers 93.8 percent

of all pairs, it seems that *NSS* and *BEWDS* have evidential strength of explanation on subjects' behavior in *PDMC* and *PDUV* experiments.

We cannot distinguish *NSS* and *BEWDS* from the viewpoint of the equilibrium path because both predict *CCyy*. However, *NSS* and *BEWDS* predict different off equilibrium paths on subgame *CD*. Consider first *NSS* where the strategy pair is ((*C,y,n,* ·,·),(*C,y,n,* ·,·)). A subject who chooses *D* has freedom to choose either *y* or *n* since the fourth component of his strategy is either *y* or *n*, while a subject who chooses *C* must choose *n* due to the third component. Therefore, *CDny* and *CDnn* are the off equilibrium paths. On the other hand, *BEWDS* predicts only *ny* on subgame *CD* since the strategy pair is ((*C,y,n,y,*·),(*C,y,n,y,*·)). In *PDMC* and *PDUV* experiments, we observed 40 pairs who chose *CD* in the first stage and 39 pairs (97.5 percent) are either *ny* or *nn* as *NSS* predicts. However, only 4 pairs (10 percent) are *nn*. On the other hand, 35 pairs (87.5 percent) are *ny*. Hence, *BEWDS* has more descriptive power than *NSS* has in *PDMC* and *PDUV* experiments.

	BEWDS, NSS	*NE, SPE*								
Path	*CCyy*	*CCnn*	*CDny*	*CDnn*	*DDyy*	*DDyn*	*DDnn*	*CCyn*	*CDyy*	*CDyn*
PDMC	531	0	28	4	1	0	0	5	1	0
PDUV	182	0	7	0	0	0	0	1	0	0
Total	713	0	35	4	1	0	0	6	1	0
	713 (93.8%)	40 (5.3%)						7 (0.9%)		

Table 3. Frequencies of *NE*, *SPE, NSS, BEWDS* and the Other Paths in *PDMC* and *PDUV* Experiments

Observation 4.

(i) *The pairs on the unique NSS and BEWDS path, i.e., CCyy, is 93.8% and the pairs on NE and SPE paths other than CCyy is 5.3% of all pairs.*

(ii) *The pairs on the BEWDS path are 87.5% and the pairs on the NSS path other than the BEWDS path is 10% of all pairs at subgame CD.*

9.5. Performance of *SPE* and *BEWDS* in the *CMPD* Experiment

Table 4 reports the frequencies of pairs with which *BEWDS*, *SPE* and the other paths occurred in the *CMPD* experiment. *SPE* paths account for 42.3 percent and *BEWDS* paths have additional 15.4% of all pairs. Although 57.7% is the *BEWDS* paths, 42.3% is the non-*BEWDS* paths. Comparing Table 4 to Table 3, the descriptive power of both *SPE* and *BEWDS* is relatively weak.

These results may indicate that there is an affinity between the structure of the game and the descriptive power of the equilibrium concepts.

	SPE			*BEWDS*			
Path	33*CC*	34*CC*	44*CC*	33*CD*	33*DD*	34*CD*	Others
CMPD	48	58	135	5	0	83	241
Total	241 (42.3%)			88 (15.4%)			241 (42.3%)

Table 4. Frequencies of *BEWDS*, *SPE* and the Other Paths in *PDMC* and *PDUV* Experiments

Observation 5.

In the CMPD experiment, SPE explains 42.3 percent of the experimental data. In addition to that, BEWDS explains another 15.4 percent of the data.

9.6. The Ratios of Payoff Maximizer, Reciprocator, Inequality Averter and Utilitarian

We will develop a method of estimating ratios of four types of choice behaviors using the path data with *BEWDS* as an equilibrium concept in this section. Consider, for example, a pair with a reciprocator (*R*) and an inequality averter (*I*). The equilibrium path is *CCyy*, the choice is *nn* at subgame *CD*, and it is *yn* at subgame *DC* under *BEWDS*. That is, the off equilibrium paths under *BEWDS* are *CDnn* and *DCyn*. The entries at *RI* in Table 5 are *CDnn* and *CDny* since *DCyn* is exactly the same as *CDny*. We will not consider subgame *DD* since the payoff table is flat for all cases. The differences among the entries make the estimation possible.

	M	*R*	*I*	*U*
M	*2CDny*	*2CDny*	*CDnn,CDny*	*CDny,CCyy*
R	*2CDny*	*2CDny*	*CDnn,CDny*	*CDny,CCyy*
I	*CDnn,CDny*	*CDnn,CDny*	*2CDnn*	*CDny,CDyn*
U	*CDny,CCyy*	*CDny,CCyy*	*CDny,CDyn*	*2CDyy*

Table 5. Off equilibrium paths under *BEWDS*.

Notice that the columns and rows of *M* and *R* in Table 5 are identical. This indicates that the behaviors of *M* and *R* are indistinguishable using the off equilibrium path data. Hence, deleting the first row and column, we obtain Table 6, and name the first row and column *M or R*, or simply *MR*. Notice that Table 6 has the equilibrium paths with bold face.

	MR	*I*	*U*
MR	***CCyy***,*2CDny*	***CCyy***,*CDnn*,*CDny*	***CDyy***,*CDny*,*CCyy*
I	***CCyy***,*CDnn*,*CDny*	***CCyy***,*2CDnn*	***CCyy***,*CDny*,*CDyn*
U	***CDyy***,*CDny*,*CCyy*	***CCyy***,*CDny*,*CDyn*	***CCyy***,*2CDyy*

Table 6. Equilibrium and off equilibrium paths among *M* or *R*, *I* and *U*.

Let (*a*,*b*,*c*,*d*,*e*) be the numbers of paired data of (*CDny*,*CDnn*,*CCyy*,*CDyn*,*CDyy*) in our experiment. Then (*a*,*b*,*c*,*d*,*e*)=(28,4,531,0,1) in the *PDMC* experiments. Notice that off equilibrium paths *CDyy* at *UU* and *CCyy* at *MRU* in Table 6 are the equilibrium paths at *MRU* and all pairs except for *MRU* respectively. That is, for example, the number of the *CDyy* data comes from either *MRU* or *UU*. Similarly, the number of the *CCyy* data comes from the equilibrium path data at all pairs except for *MRU*, and the off equilibrium path at *MRU*.

The total number of pairs in *PDMC* is 190x3=570, and there are 5 pairs with *CCny* and one pair with *DDyy*. Since *CCny* is not a result of the elimination of weakly dominated strategies at *CC* for all pairs, we take them out of our consideration. Since *DDyy* is a path of subgame *DD*, we also take it out of the data. That is, the total number of pairs in consideration is 564.

Let the ratios among three cases where both participants are at the equilibrium path, an off equilibrium path, and the other off equilibrium path be 1:*q*:*q* where $q > 0$. We assume that all pairs have the same q which can be interpreted as a mistake ratio.[41] Let w be the number of all data, i.e., 564, and let p_{MR}, p_I and p_U be the ratios of *MR*, *I* and *U* among the participants and we assume that they are independent.[42] Then the number of *CDny, CDnn, CCyy, CDyn,* and *CDyy* are

(1) $\frac{2qw}{1+2q}(p_{MR}p_{MR}+p_{MR}p_I+p_{MR}p_U+p_Ip_U)=28$, (2) $\frac{2qw}{1+2q}(p_{MR}p_I+p_Ip_I)=4$,

(3) $\frac{w}{1+2q}(p_{MR}p_{MR}+2p_{MR}p_I+p_Ip_I+2p_Ip_U+p_Up_U+2p_{MR}p_Uq)=531$,

(4) $\frac{2qw}{1+2q}p_Ip_U=0$ and (5) $\frac{2w}{1+2q}(p_{MR}p_U+p_Up_Uq)=1$ respectively.

Since $p_{MR}+p_I+p_U=1$, the number of equations is 6 and the number of variables is 4, and hence

[41] Another way is to introduce that each player has a mistake rate. Since both methods give almost the same result, we employed the "*q*" method to avoid complication.

[42] For example, the number of pairs at *IU* is $2p_Ip_U(1+q+q)w/(1+2q)$. "1" corresponds to *CCyy* and the first "*q*" corresponds to *CDny* and the second "*q*" corresponds to *CDyn*. "2" before p_Ip_U is for *UI*.

the system of non-linear equations normally does not have the solution. In order to find q and (p_{MR}, p_I, p_U) as the starting value in the following numerical simulation, we solve the equations choosing three out of (1)-(5) using Mathematica.[43] The subscripts in Table 7 show the choices of the five equations. The first row in each cell shows $(p_{MR}, p_I, p_U; q)$, and the second shows (*c,e*) after taking into account of the number of the equilibrium paths. For example, consider the 145 case in Table 7. Then solving equations with (1), (4), (5) and $p_{MR} + p_I + p_U = 1,$ we obtain $(p_{MR}, p_I, p_U; q)$ =(0.9991, 0, 0.0009; 0.0261). Using this information, we have that the number of off equilibrium paths at *CCyy* must be $\hat{c} = 2p_{MR}p_U qw/(1+2q) = 0.0261$. Similarly, $\hat{e} = 1 - 2p_{MR}p_U w/(1+2q) = 0.0009.$

Using the data as the initial values, we will estimate (p_{MR}, p_I, p_U) minimizing the sum of the square of the difference between the probability and data in the following manner. Using Table 6, we obtain the probabilities of off equilibrium paths.

$p_{CDny} = p_{MR}p_{MR} + p_{MR}p_I + p_{MR}p_U + p_I p_U$, $p_{CDnn} = p_{MR}p_I + p_I p_I$, $p_{CCyy} = p_{MR}p_U$, $p_{CDyn} = p_I p_U$ and $p_{CDyy} = p_U p_U$.[44] Let $\hat{s} = a + b + \hat{c} + d + \hat{e}$ where $\hat{c}$ and $\hat{e}$ are obtained in the above procedure. Define g by

$$g(p_{MR}, p_I, p_U) = (p_{CDny} - a/\hat{s})^2 + (p_{CDnn} - b/\hat{s})^2 + (p_{CCyy} - \hat{c}/\hat{s})^2 + (p_{CDyn} - d/\hat{s})^2 + (p_{CDyy} - \hat{e}/\hat{s})^2.$$

Let us now consider the following minimization problem: Find $(p_{MR}, p_I, p_U) \geq 0$ that satisfies

Minimize $g(p_{MR}, p_I, p_U)$ subject to $p_{MR} + p_I + p_U = 1.$

For example, consider again the 145 case. Using the Newton method with the initial values (p_{MR}, p_I, p_U) =(0.9991, 0, 0.0009) and $(a, b, \hat{c}, d, \hat{e})$ = (28,4,0.0261,0,0.0009) in Mathematica,[45] we obtain (p_{MR}, p_I, p_U) =(0.8739, 0.1249, 0.0012). The last number in the last row in each cell in Table 7 shows the minimized value of g, and it is 1.6156x10^{-7}. Table 7 shows that the initial solutions are relatively wide spread, but the discrepancies among the solutions go down after the minimization. We name a series of procedures including solving equations using the data of *BEWDS* paths and minimizing g the *path data analysis*. Appendix provides the supporting

[43] We used "NSolve" in Mathematica 8.

[44] Taking out the bold face entries in Table 6, we obtain a table consisting of all "mistake" entries. Consider *CDny*. Since all entries at (*MR,MR*) are *CDny*, the probability of *CDny* at (*MR,MR*) is $p_{MR}p_{MR}$. Since one of two entries at (*MR,I*) is *CDny*, the probability of *CDny* at (*MR,I*) is $(1/2)p_{MR}p_I$. Due to the symmetry of the table, the probability at (*I,MR*) is also $(1/2)p_{MR}p_I$. That is, the probability of *CDny* when one player is *MR* and the other is *I* is $p_{MR}p_I$. Repeating the same procedure at (*MR,U*) and (*I,U*), we obtain p_{CDny}.

[45] We used "FindMinimum" in Mathematica 8.

evidence of path data analysis using the data in *PDMC* and *PDUV* experiments.

An important remark is the interpretation of the ratio of four types of behavior. Even a subject may behave like a utilitarian in some periods, but an inequality averter in some other periods. That is, the viewpoint of this analysis is not based upon specific subject, but behavior of a period. Keeping this point in mind, we have,

Property 8. *Using the path data analysis, we have*
(*i*) *the ratios of payoff maximizing or reciprocating, inequality averting, and utilitarianizing behaviors in the PDMC experiments are 85.55-87.39%, 12.48-13.17% and 0.12-1.29% respectively; and*
(ii) *the ratios of payoff maximizing or reciprocating, inequality averting, and utilitarianizing behavior in the PDMC and PDUV experiments are 88.13-89.64%, 10.24-10.86% and 0.1-1.02% respectively.*

$(p_{MR}, p_I, p_U; q)$: solution of simultaneous equations $(\hat{c}, \hat{e})$ derived from the equations $(p_{MR}, p_I, p_U; g)$: solution of minimization	
$(0.8739, 0.1250, 0.0011; 0.0301)_{123}$ $(0.0300, 0.0043)_{123}$ $(0.8737, 0.1249, 0.0014; 2.094\text{x}10^{-7})_{123}$	$(0.8750, 0.1250, 0; 0.0301)_{124}$ $(0, 1)_{124}$ $(0.8555, 0.1317, 0.0129; 0.0012)_{124}$
$(0.8739, 0.1250, 0.0011, 0.0301)_{125}$ $(0.0301, 0.0012)_{125}$ $(0.8738, 0.1248, 0.0014; 2.019\text{x}10^{-7})_{125}$	$(0.8485, 0.1515, 0; 0.0311)_{134a}$ $(0,1)_{134a}$ $(0.8555, 0.1317, 0.0129; 0.013)_{134a}$
$(0.9953, 0, 0.0047; 0.0262)_{134b}$ $(0.1312, 0)_{134b}$ $(0.8707, 0.1248, 0.0045; 4.311\text{x}10^{-7})_{134b}$	$(0.8740, 0.1249, 0.0011; 0.0301)_{135}$ $(0.0301, 0.0012)_{135}$ $(0.8738, 0.1248, 0.0014; 2.0260\text{x}10^{-7})_{135}$
$(0.9991, 0, 0.0009; 0.0261)_{145}$ $(0.0261, 0.0009)_{145}$ $(0.8739, 0.1249, 0.0012; 1.6156\text{x}10^{-7})_{145}$	$(0.8788, 0.1212, 0; 0.0311)_{234}$ $(0,1)_{234}$ $(0.8555, 0.1317, 0.0128; 0.0012)_{234}$
$(0.8739, 0.1250, 0.0011; 0.0301)_{235}$ $(0.0301, 0.0012)_{235}$ $(0.8738, 0.1248, 0.0014; 2.0183\text{x}10^{-7})_{235}$	$(0.9991, 0, 0.0009; 0.0301)_{345}$ $(0.0301, 0.0009)_{345}$ $(0.8738, 0.1249, 0.0013; 1.364\text{x}10^{-7})_{345}$

Remark: 134 and 145 have two solutions. Since $\hat{e}$ in 134b is negative, we used $\hat{e} = 0$ in this case. If $p_U = 0$ (i.e., cases 124, 134a and 234), both the number of equilibrium path data and the off equilibrium path data at *CDyy* must be zero (see (5)) although the number of *CDyy* in the experiment is one. If this "one" is regarded as the datum from the equilibrium path, $\hat{e}$ must be close to zero, and the simulation results are similar to other cases. However, we regard $\hat{e}$ =1 on these inconsistency cases since a subject who chose *y* to *D* at *CDyy* showed apparent utilitarian motive in the record sheet. Since *q*=535.474 in one of the solutions of 145 and Mathematica did not return the answer in 245, we exclude them in the table.

Table 7. The ratios of maximizers or reciprocators, inequality averters and utilitarians.[46]

[46] It seems that the reason why the ratio of utilitarian is not zero comes from one observation of *CDyy* in Table 3. That is, a subject who chose *C* approved the other subject who chose *D*.

We recruited a coder to determine if the descriptions of the record sheet during the session and the questionnaire after the session were based on the idea of maximizer, reciprocator, inequality averter or utilitarian (or none of the above). She did not major in economics, and did not know the content of this project at all. This maintains objectivity for counting. She received a written instruction for coding in which the content of the experiment was written, which is the same as that the subjects were received in the experimental session. She was requested to determine one judgment for the description of record sheet for each period and determine one or possibly more than one judgments for the whole description of questionnaire after the session. If she could not determine the type of behavioral principle, she can choose "none of the above." We did not provide her the decision criteria for coding and asked her to set it up by herself. She felt that the judgment of the record sheet during the session was hard because the description was extremely short, and her judgment would lack credibility, while the description of questionnaire after the session was relatively easy to judge.

Table 8 reports frequencies and rates of maximizer, reciprocator, inequality averter and utilitarian judged from the description of record sheets during the session. Table 9 is their frequencies and rates judged from the description of questionnaire after the session. In these tables, we eliminated the data which the coder judged "none of the above."[47] Moreover, the frequencies and rates of maximizer and reciprocator are combined for the comparison of the rates in table 7. In table 9, when answers were multiple, we divided one by the number of answers in order to keep the weight across subjects. In both tables, the rate of maximizer was highest, followed by inequality averter, while the rates of reciprocator and utilitarian were little. The rate of maximizer is larger in table 9 than in table 8 while the rate of inequality averter is larger than in table 8 than in table 9. The chi-square test reveals that the rates are significantly different between two tables (p-value < 0.05)[48].

	Maximizer and Reciprocator	Inequality Averter	Utilitarian
Frequencies	1037 (M: 999, R: 38)	363	0
Rates (%)	74.07 (M: 71.36, R: 2.71)	25.93	0.00

Table 8. Frequencies and Rates of Maximizer, Reciprocator, Inequality Averter and Utilitarian Judged from the Description of Record Sheets during the Session.

[47] The number of judgments of "none of the above" is 120 out of 1520 (7.89 percent) descriptions in the record sheet and 4 out of 80 (5 percent) descriptions in the questionnaire after experiment.
[48] Since the rate of utilitarian was zero in both tables, we eliminated this category when we calculate the test statistic.

	Maximizer and Reciprocator	Inequality Averter	Utilitarian
Frequencies	65.5 (M: 65.5, R: 0)	10.5	0
Rates (%)	86.18 (M: 86.18, R: 0.00)	13.82	0.00

Table 9. Frequencies and Rates of Maximizer, Reciprocator, Inequality Averter and Utilitarian Judged from the Description of Questionnaire after the Session.

In order to compare the rates of tables 8 and 9 with each of ten estimated rates in table 7, we pool the rate of maximizer and reciprocator, and then perform the chi-square tests. The rates in table 8 is significantly different from those in table 7 (p-values < 0.00001 in all tests), while the rates in table 9 is not significantly different from those in table 7 (p-values > 0.6 in all tests). Hence, the rates in table 9 (description of questionnaire after the session) are consistent with the estimated rates in table 7, while the rates in table 8 (description of record sheets during the session) are not.

10. *BEWDS* and the Role of Mate Choice Flat

As shown in Table 3, *BEWDS* explains the subjects' behavior well in *PDMC* and *PDUV* experiments. In the first half of this section, we will consider why *BEWDS* has predictive power on subjects' behavior when the *MCM* is employed. A possible reason is that there are simple heuristics whose strategies are equivalent to *BEWDS*. Furthermore, questionnaire analysis finds that most subjects were likely to adopt them. In the second half of this section, we will argue that, with subjects behaving in line with *BEWDS*, the *MCM* has the uniqueness property under some axioms.

The mate choice flat derives from *MCM*, and it has several nice features. Although the following property is trivial, the mate choice flat alleviates cognitive burden of subjects from two dimensional to one dimensional comparison.

Property 9. *Under the mate choice flat, strategy* α *with the payoff vector* (u,v) *weakly dominates strategy* β *with* (x,y) *if and only if* $u > x$.

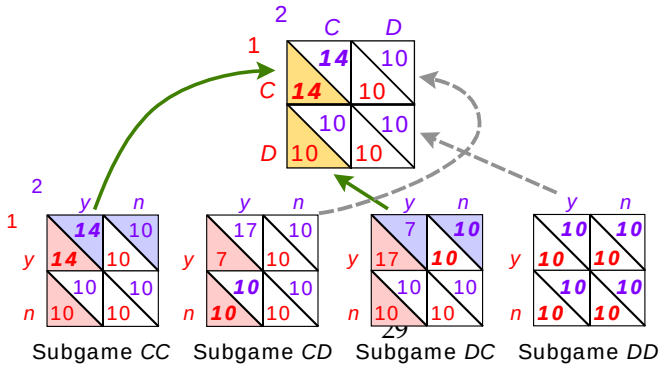


Figure 7. The Triangles that Subjects must Consider and Backwardability.

Since subjects can instantaneously understand that each subgame in the second stage has the mate choice flat, and can identify (10,10) as the outcome of subgame *DD* in Figure 7, the cells or triangles that subjects must see or consider are dark parts in subgames *CC, CD* and *DC* in the second stage under *BEWDS*. Subject 1's own possible outcomes to be compared are the two lower left triangles of the left column. Subject 2's outcomes to be compared are the two upper right triangles of the upper row. Hence, the number of triangles that each subject must see is four in each of subgames *CC, CD* and *DC*. Hence, the remaining triangles are unnecessary for their decision making including the lower right cell.

Let us consider the minimum informational or intellectual requirement to achieve *CCyy* under *BEWDS*. Consider subject 1. The information of the two lower left triangles of the left column in each of subgames *CC, CD*, and *DD* is enough to solve or choose either *y* or *n* in each subgame, but this is not enough to solve the two stage game since subject 1 cannot identify which cell would be realized without having the information of the two upper right triangles of the upper row at subgames *CC* and *DC*. In other words, subject 1 must use *theory of mind* to understand which strategy subject 2 chooses. If this is successful, subject 1 can construct the reduced normal form game out of two stages shown at the top in Figure 7. During this construction, subject 1 understands that (s)he really needs to know for the decision making is the two lower left triangles of the left column in the reduced normal form game, i.e., subject 1's outcomes of subgames *CC* and *DC*. Using this information, subject 1 chooses *C*. In this sense, subject 1 must have *backwardability* that identifies chosen cells in subgames *CC, CD* and *DD*, and then finds his or her own two triangles corresponding to subgames *CC* and *DC* in the reduced normal form game. Finally, subject 1 also uses a simple heuristic: "*the other subject thinks the same way as I think.*" For example, subject 1 who understands the outcome of subgame *DC* can find the outcome of subgame *CD* using this heuristic. These two simplified methods mitigate subjects' burden considerably, and we found many subjects actually employ the methods in the following.[49] We can apply the same procedure to the cases when a subject is an inequality

[49] Player 1 must compare two numbers six times in order to decide (*C,y*) in Figure 7: two comparisons (i.e., my own 14 and 10, and the other's 14 and 10) in subgame *CC*, one comparison (i.e., my own 10 and 7, and the other's choice does not matter since my own outcome is 10 regardless the choice of the other) in subgame *CD*, two comparisons in subgame *DC*, and one comparison between 14 and 10 in the reduced normal form game. This is

averter or a utilitarian. For example, if a subject is a reciprocator, since the decisions in subgames have already determined, the cognitive burden is just to compare 14 with 10 in the reduced normal form game.

From the record sheets during the session and questionnaires after the session, we will detect whether subjects adopt backwardability, whether subjects adopt the idea of weak dominance or its equivalence (Property 9) and whether subjects adopt simple heuristic that "*the other subject thinks the same way as I think*". We recruited an economics graduate student who did not know the contents of this project at all in order to maintain objectivity. He received a written instruction for coding in which the content of the experiment was written, which is the same as that the subjects were received in the experimental session, and he was asked to count the number of subjects who seemed to have in mind the concept of backwardability, weak dominance or its equivalence and a heuristic that "*the other subject thinks the same way as I think*". He was not provided the decision criteria for counting and was asked to set up it by himself.

	Backwardability	Weak Dominance or Its Equivalence	The Same as I Think
Complete Description	49	43	39
Partial Description	9	15	13
No Description	2	2	8

Table 6: Questionnaire Analysis of the *PDMC* experiment

	Backwardability	Weak Dominance or Its Equivalence	The Same as I Think
Complete Description	20	17	14
Partial Description	0	3	5
No Description	0	0	1

Table 7: Questionnaire Analysis of the *PDUV* experiment

Tables 6 and 7 show the results. In both *PDMC* and *PDUV* experiments, many subjects described that they had the idea of backwardability, weak dominance or its equivalence and the simple heuristic that "*the other subject thinks the same way as I think*."[50] In every component, more

quite a contrast when we find Nash equilibria of the two stage game. Since the number of information set is 5, each player has 2^5 strategies. Player 1 must compare (2^5-1) numbers to find best responses for any given strategies of player 2. This indicates that player 1 must compare two numbers $(2^5-1)\times 2^5$ times. In order to find the Nash equilibria, player 1 must find the best responses of player 2, and hence must compare two numbers $(2^5-1)\times 2^5$ times. That is, the number of comparisons is $2\times(2^5-1)\times 2^5 = 1984$ which is more than 300 times of 6, which might trigger qualitative difference between *NE* and *BEWDS*.

[50] This shows that cases such as *MU* and *RU* in Table 1 scarcely happened from the view point of each player.

than 65 percent of subjects describe completely or partially the corresponding idea in both experiments. Furthermore, 33 out of 60 subjects describe completely all three components in the *PDMC* experiment, and 13 out of 20 in the *PDUV* experiment. Note that our questionnaire was free description. We did not restrict subjects such that they would pay attention to these components. This analysis indicates that most subjects seem to behave in line with *BEWDS* under the *MCM*.

Concerning the weakly dominated strategies, we wrote the following in the instruction for counting,

> Suppose that the own choice is A (defection) and the other choice is B (cooperation). … The following answer is an example where a subject is in mind of weakly dominated strategies. "The payoff vector is (1700,1000) when I choose acceptance and it is (1000,1000) when I choose rejection. So, acceptance is better." In addition to the comparison of the payoff vectors, please count the answer of comparison of two numbers, 1700 and 1000.

This instruction assumes that the subject is a payoff maximizer, not reciprocator, inequality averter or utilitarian. Hence, the questionnaire analysis about the weak dominance or its equivalence described above should be considered that subjects are assumed being payoff maximizers.

MCM that we employed has the uniqueness property under *BEWDS*. That is, any approval mechanisms satisfying Axioms 1, 2 and 3 in the following must be *MCM.* We say that an approval mechanism satisfies *forthrightness* if both choose *y* in the second stage after the choice of a strategy pair in a *PD* game, then the outcome of the approval mechanism is the outcome of the *PD* game with the strategy pair.[51] For example, suppose that subjects 1 and 2 choose (*C*,*D*) and both choose *y* in the approval mechanism. Then forthrightness requires that the outcome must be (*C*,*D*). In order to limit the class of approval mechanisms, we introduce the following axioms.

Axiom 1 (*Onto*): An approval mechanism satisfies the *onto* condition if every outcome of a *PD* game is an outcome of the *PDMC* and every outcome of *PDMC* must be an outcome of the *PD*

Even though in case *MU*, for example, player *M* thought that this case was *MM* and player *U* thought that the case was *UU* since players assumed that the other player behaved as they behaved .

[51] This definition is slightly different from forthrightness introduced by Tatsuyoshi Saijo, Yoshikatsu Tatamitani and Takehiko Yamato (1996).

game.

The onto condition requires that the set of outcomes of subgames *CC, CD, DC* and *DD* must be {(14,14),(7,17),(17,7),(10,10)}. Consider a (non-approval) mechanism that gives 5 if a subject chooses "*C*" in Figure 1 as reward for cooperation. Then the outcomes become (14+5,14+5),(7+5,17),(17,7+5), and (10,10). That is, (14,14),(7,17),(17,7) would never be realized with this mechanism. Therefore, this reward mechanism with the *PD* game does not maintain the outcomes of the *PD* game. In this sense, many reward or punishment mechanisms are not "onto". The onto condition also excludes mechanisms that are not budget balanced. That is, the reward mechanism above needs outside money to maintain the mechanism, and the onto condition does not allow this budget deficit. Furthermore, punishment occasionally reduces total payoff, and hence it is not efficient.

Axiom 2 (*Mate choice flat at the approval stage*): An approval mechanism satisfies *mate choice flat at the approval stage* if either subject chooses *n*, the outcome of these strategy pairs must be the same for each subgame.

Axiom 2 allows that the flat outcome in subgame *CC* can be different from the one of subgame *CD,* for example.

Axiom 3 (*Mate choice flat at the reduced normal form stage*): A *PD* game with an approval mechanism satisfies *mate choice flat at the reduced normal form stage* if either subject says "*D*" in the normal form game derived from the two stage game, the outcome of these strategy pairs must be the same.

This axiom requires that the reduced normal form game of the two stage game must also have a mate choice flat if either subject chooses *D*, but does not require that the same outcome is related to the outcomes in the second stage.

We say that an approval mechanism with a *PD* game is *natural* if the outcome is (10,10) when either one of two subjects chooses *n*, and it is *voluntary* if any subject who chooses *D* should not be forced to change from *D* to *C*. Clearly, if it is natural, it should be voluntary since a defector is not forced to contribute $10. The forthright and natural mechanism is exactly *MCM* by its construction. It is straightforward to see that the forthright and natural mechanism satisfies Axioms 1, 2 and 3. On the other hand, the following property guarantees that a forthright mechanism satisfying Axioms 1, 2 and 3 must be natural, and hence it is voluntary.

Property 10. *Suppose that the unique equilibrium path of a PD game with an approval mechanism is CCyy under BEWDS and suppose Axioms 1, 2 and 3. Then the approval mechanism satisfying forthrightness is natural.*

Proof. See Appendix.

Since an approval mechanism that is natural and forthright must be *MCM*, Property 10 shows the uniqueness of the mechanism under Axioms 1, 2 and 3.

11. Concluding Remarks

The *MCM* implements cooperation in the *PD* game in *BEWDS*, and it works well among maximizers, reciprocators, inequality averters, utilitalians and the mixture of them except the combinations of either a maximizer or reciprocator when the other player is a utilitarian. We also found that the *MCM* promotes cooperation significantly in the *PD* game with human subjects. This experimental evidence is most compatible with the behavioral principle based upon BEWDS. Using the path data analysis, we found that the ratios of payoff maximizing or reciprocating, inequality averting, and utilitarianizing behaviors are 86-90%, 10-13% and 0-1% respectively, which is partially consistent with the classification by a coder. The elimination of weakly dominated strategies can be done by comparison of two numbers, but not two vectors due to the mate choice flat. It seems that the flat reduced subjects' cognitive burden, and made them easily consider backwardly. We noticed that the *MCM* is unique with several axioms. We also found that the cooperation rate in the *PDMC* is significantly higher than that in the *CMPD*.

Of course, the *MCM* does not always solve all prisoner's dilemma. First, the participants must agree upon using the mechanism as mechanism designers of all fields in economics implicitly presume. Second, the mechanism might need monitoring devices and/or enforcing power. Otherwise, a participant might not conduct the deed described in "*C*" even after two participants choose "*C*" and "*y*". Third, we cannot apply the mechanism if the contents of "*C*" have not been settled down before applying it. Many researchers have been using global warming as an example of *PD*. Although countries and parties have been negotiating the substance of coping with it for over twenty years under the United Nations Framework Convention on Climate Change (*UNFCCC*), they have not reached what exactly "*C*" should be.

The *PD* game has two participants and two strategies. We will consider the directions of our further research agenda based upon these numbers. First, fix the number of participants two, and then consider the number of strategies is at least three. This is nothing but a voluntary contribution mechanism for the provision of a public good with two participants. Masuda,

Okano and Saijo (2012) show that the *MCM* with *BEWDS* cannot implement the Pareto outcome when both have the same linear utility function. Then they designed the minimum mate choice mechanism that is based upon the spirit of the *MCM*, and found that it implements the Pareto outcome theoretically and experimentally. The contribution rates of several sessions exceeded 95%.

Second, fix the number of strategies two, and then consider the number of participants is at least three. This is nothing but a social dilemma situation. As Banks, Plott and Porter (1988) found, Okano, Masuda and Saijo (2012) also show that the *MCM* with *BEWDS* cannot implement the Pareto outcome. Then they design new mechanisms utilizing the idea of the *MCM* that implement the Pareto outcome. However, the first five round cooperation rates are 70-80% and then they go beyond the 90% cooperation rate. This is due to the fact that the mechanisms implementing the Pareto outcome necessarily contain *PD* games with two participants. For example, consider the case where one player chooses *D*, and the other two players choose *C*. In the second stage, two players with choice *C* face a *PD* game in the mechanism. Two players should not cooperate theoretically in order to attain full cooperation (i.e., the three choose *C*), but they occasionally choose *C* experimentally. In other words, cooperation of two players is a major stumbling block against full cooperation that is a fundamental difficulty in designing workable mechanisms in social dilemma without any repetition. Although many researchers do not see any differences between two and more than two participants, they find a deep fissure between them.

Third, consider that both are at least three. This environment is a wide open area. Of course, there are quite a number of papers with this environment (see, for example, related papers in Plott and Vernon L. Smith (2008)), but the crack between theory and experiment has not been filled up.

Mechanism designers have not been considering *comfortability* of mechanisms. Although it is still early stage of research, Hideo Shinagawa, Masao Nagatsuka, Okano and Saijo (2012) find that subjects facing the *MCM* did not show any significant activation of anterior prefrontal cortex (*PFC*) in the processing of decision making using a near-infrared spectroscopy (*NIRS*)-based system. This finding suggests the possibility that subjects with the *MCM* made decision at ease. On the other hand, subjects playing the *PD* game showed significant activation of right *PFC* and left orbitofrontal cortex that are related to unpleasant emotion.[52]

Takaoka, Okano and Saijo (2012) compare the *MCM* with costly punishment measuring

[52] See Yoko Hoshi, Jinghua Huang, Shunji Kohri, Yoshinobu Iguchi, Masayuki Naya, Takahiro Okamoto, and Shuji Ono (2011).

salivary alpha-amylase (s*AA*) of subjects. *SAA* has been proposed as a sensitive biomarker for stress-related changes in the body that reflect the activity of the sympathetic nervous system. They find that subjects who experienced the *MCM* reduced the level of *sAA* and subjects who experienced costly punishment increased the level of *sAA*. This indicates that the *MCM* is a mechanism that is relatively stress free.

Appendix

	PDMC			*PD*	*PDUV*	*CMPD*			*PDMC**	*PD**
	1	2	3			1	2	3		
Letters	0	0	1	2	0	2	1	3	1	0
Human Sciences	1	0	2	0	1	2	2	2	2	2
Foreign Studies	1	2	2	1	2	2	3	2	1	0
Law	4	2	0	1	3	0	0	1	2	1
Economics	3	1	0	2	2	1	2	2	1	1
Science	0	1	2	1	1	5	1	2	0	1
Medicine	1	1	0	0	2	1	0	0	0	1
Dentistry	0	0	0	0	1	0	0	0	0	0
Pharmaceutical Sciences	1	0	1	1	0	1	0	0	0	0
Engineering	5	10	10	4	4	2	9	8	11	12
Engineering Science	4	2	2	7	4	4	1	0	2	2
Information Science and Technology	0	1	0	0	0	0	1	0	0	0
Frontier Biosciences	0	0	0	1	0	0	0	0	0	0
# of Females	4	5	3	3	9	6	3	7	2	3
Average Age	22.1	21.95	21.65	22.6	21.4	21.4	21.35	21.5	21.8	23
Average Earning ($)	61.5	67.2	66.6	45.3	59.2	55.9	64.1	63.6	65.1	48.0
Maximum Earning ($)	61.5	68.9	68.9	48.8	59.9	58.6	67.7	66.6	65.1	78.5
Minimum Earning ($)	61.5	54.1	56.5	30.7	55.4	52.9	59.2	57.0	65.1	32.3
Duration of Session (min.)	115	120	114	75	98	132	129	131	80	72

1) Numbers of divisions in rows of affiliation show the numbers of participants.
2) No repetition in "*" sessions.
3) $1=86.55 yen for *PDMC-1* and *PD*, $1=77.23 yen for *PDMC-2* and *PDMC-3*, $1=88.8 yen for *PDUV*, *CMPD-1*, $1=77.66 yen for *CMPD*-2, $1=76.8199 yen for *CMPD*-3, $1=81.7099 yen for *PDMC** and $1=81.7099 yen for *PD**.

Table A1. Subjects' Information

Proof of Property 3. Let $P(s_1, s_2)$ be the path with (s_1, s_2). The following four cases cover all strategies.

Case 1. $t = (D, \cdot, \cdot, \cdot, \cdot)$ is not an *NSS*.

Let $t' = (C, y, n, n, \cdot)$. Since $P(t,t) = (D, D, \cdot, \cdot)$, $P(t',t) = (C, D, n, \cdot)$, $P(t,t') = (D, C, \cdot, n)$ and $P(t',t') = (C, C, y, y)$, we have $v(t,t) = v(t',t) = v(t,t') = 10$ and $v(t',t') = 14$. Therefore, t is not an

NSS.

Case 2. $t=(C,n,\cdot,\cdot,\cdot)$ is not an *NSS*.

Let $t'=(C,y,\cdot,\cdot,\cdot)$. Since $P(t,t)=(C,C,n,n)$, $P(t',t)=(C,C,y,n)$, $P(t,t')=(C,C,n,y)$ and $P(t',t')=(C,C,y,y)$, we have $v(t,t)=v(t',t)=v(t,t')=10$ and $v(t',t')=14$. Therefore, *t* is not an *NSS*.

Case 3. $t=(C,y,y,\cdot,\cdot)$ is not an *NSS*.

Let $t'=(D,\cdot,\cdot,y,\cdot)$. Since $P(t,t)=(C,C,y,y)$ and $P(t',t)=(D,C,y,y)$, we have $v(t,t)=14<v(t',t)=17$. Therefore, *t* is not an *NSS*.

Case 4. $t=(C,y,n,\cdot,\cdot)$ is an *NSS*.

Since $P(t,t)=(C,C,y,y)$, $v(t,t)=14$.

(i) If $t'=(D,\cdot,\cdot,\cdot,\cdot)$, $v(t',t)=10$ because $P(t',t)=(D,C,\cdot,n)$. Therefore, $v(t,t)=14>v(t',t)=10$.

(ii) If $t'=(C,n,\cdot,\cdot,\cdot)$, $v(t',t)=10$ because $P(t',t)=(C,C,n,y)$. Therefore, $v(t,t)=14>v(t',t)=10$.

(iii) If $t'=(C,y,\cdot,\cdot,\cdot)$ with $t'\neq t$, $P(t',t)=P(t,t')=P(t',t')=(C,C,y,y)$. Therefore, $v(t',t)=v(t,t')=v(t',t')=14$. Hence, $v(t,t)=v(t',t)$ and $v(t,t')=v(t',t')$.

These cases show that *t* is an *NSS*. ∎

Proof of Property 5. By definition, if *t* is not an *NSS*, then *t* is not an *ESS*. Therefore, by the proof of Property 4, the strategies other than (*C,y,n,·,·*) are not *ESS*. We will show that *t* = (*C,y,n,·,·*) is not an *ESS*. Let *t′* = (*C,y,·,·,·*) with *t′* ≠ *t*. Since, *P*(*t,t*) = *P*(*t′,t*) = *P*(*t,t′*) = *P*(*t′,t′*) = *CCyy*, *v*(*t,t*) = *v*(*t′,t*) = *v*(*t,t′*) = *v*(*t′,t′*) = 14. Therefore, *t* is not an *ESS*. Hence, there is no *ESS* in this game. ∎

Proof of Property 10. We will show a slightly general proof allowing asymmetry of the *PD* game matrix. Consider the following *PD* game in Figure A1. Each cell represents (player 1's payoff, player 2's payoff). We assume that *c* > *a* > *d* > *b* and *x* > *w* > *z* > *y*.

	C	*D*
C	(*a,w*)	(*b,x*)
D	(*c,y*)	(*d,z*)

Figure A1. Prisoner's dilemma game.

First, consider the second stage, i.e., the approval mechanism stage. In Figures A2, A3, A4 and A5, the upper choice is for *y* and the lower is for *n* for player 1 and the left is for *y* and the right is for *n* for player 2. Consider subgame *CC*. Then the upper left cell must be (*a,w*) by forthrightness in Figure A2, and there are four possibilities of the flat by Axioms 1 and 2.

(*a,w*)	(*a,w*)
(*a,w*)	(*a,w*)

(1)

(*a,w*)	(*b,x*)
(*b,x*)	(*b,x*)

(2)

(*a,w*)	(*c,y*)
(*c,y*)	(*c,y*)

(3)

(*a,w*)	(*d,z*)
(*d,z*)	(*d,z*)

(4)

Figure A2. Four Possible Cases at subgame *CC*.

Shaded areas show the remaining outcomes using elimination of weakly dominated strategies. Since (*C,C,y,y*) is the unique path, (4) must be the case among the four possibilities.

Applying the same procedure for subgames *CD*, *DC* and *DD*, we have the following figures.

Figure A3. Four Possible Cases at subgame *CD*.

Figure A4. Four Possible Cases at subgame *DC*.

Figure A5. Four Possible Cases at subgame *DD*.

Second, consider the reduced normal form game stage. The outcome of the upper left cell with (*C,C*) must be (*a,w*) in each game of Figure A6, and the cells of the rest must have the same outcome in each game due to Axioms 1 and 3. The shaded areas in Figures A6 show the outcomes of elimination of weakly dominated strategies.

Figure A6. Four Possible Mate Choice Flats in the Reduced Normal Form Game.

Finally, since (*C,C,y,y*) is the unique path, (4) is the only possible case in Figure A6. That is, the outcome of mate choice flat is (*d,z*) in the reduced normal form game. Let us go back to Figures A3, A4 and A5 where the outcome (*d,z*) must be chosen (i.e., (4) in Figure A3, (4) in Figure A4 and (1) in Figure A5). Then (*d,z*) is the outcome of mate choice flat for each case. Since the outcome (*d,z*) is also the mate choice flat in Figure A2, all four cases have the common mate choice flat (*d,z*) in the approval mechanism. Hence, this approval mechanism must be natural. ∎

Supporting Evidence of Path Data Analysis

Three *PDMC* sessions + one *PDUV* session

$(p_{MR}, p_{I}, p_{U}; q)$: solution of simultaneous equations (c,e) derived from the equations $(p_{MR}, p_{I}, p_{U}; g)$: solution of minimization	
$(0.8967, 0.1026, 0.0008; 0.0273)_{123}$ $(0.0272, 0.0031)_{123}$ $(0.8964, 0.1024, 0.0013; 3.337\text{x}10^{-7})_{123}$	$(0.8974, 0.1026, 0; 0.0273)_{124}$ $(0, 1)_{124}$ $(0.8813, 0.1086, 0.0101; 0.0008)_{124}$
$(0.8967, 0.1026, 0.0008, 0.0273)_{125}$ $(0.0273, 0.0008)_{125}$ $(0.8964, 0.1024, 0.0013; 3.375\text{x}10^{-7})_{125}$	$(0.8750, 0.1250, 0; 0.0281)_{134a}$ $(0,1)_{134a}$ $(0.8813, 0.1086, 0.0102; 0.0008)_{134a}$

Remark: 134 and 145 have two solutions. Since e in 134b is negative, we used $e = 0$ in this case. Since q=528.467 in one of the solutions of 145 and Mathematica did not return the answer in 245, we exclude them in the table. We excluded *CCny* in the *PDUV* session when we applied the path data analysis. Total number of pairs = 753 = 564 + 189.

References

Abella, Alex. 2008. *Soldiers of Reason: The RAND Corporation and the Rise of the American Empire*, Houghton Mifflin Harcourt Publishing.

Andreoni, James, and John H. Miller. 1993. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence." *Economic Journal*, 103(418): 570-85.

Andreoni, James, and Hal Varian. 1999. "Preplay Contracting in the Prisoners' Dilemma." *Proceedings of the National Academy of Sciences,* 96(19): 10933-8.

Aumann, Robert J. 2006. "War and Peace." *Proceedings of the National Academy of Sciences,* 103(46): 17075-78.

Banks, Jeffrey S., Charles R. Plott, and David P. Porter. 1988. "An Experimental Analysis of Unanimity in Public Goods Provision Mechanisms." *Review of Economic Studies*, 55(2): 301-22.

Bereby-Meyer, Yoella, and Alvin E. Roth. 2006. "The Speed of Learning in Noisy Games: Partial Reinforcement and the Sustainability of Cooperation." *American Economic Review*, 96(4): 1029-42.

Cason, Timothy N., Tatsuyoshi Saijo, and Tomas Sjöström, and Takehiko Yamato. 2006. "Secure Implementation Experiments: Do Strategy-proof Mechanisms Really Work?" *Games and Economic Behavior*, 57(2): 206-235.

Charness, Gary, Guillaume R. Fréchette, and Cheng-Zhong Qin. 2007. "Endogenous Transfers in the Prisoner's Dilemma Game: An Experimental Test of Cooperation and Coordination." *Games and Economic Behavior*, 60(2): 287-306.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. 1996. "Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games." *Games and Economic Behavior,* 12(2): 187–218.

Croson, Rachel T.A. 2007. "Theories of Commitment, Altruism and Reciprocity: Evidence from Linear Public Goods Games." *Economic Inquiry*, 45(2): 199–216.

Doebeli, Michael, and Christoph Hauert. 2005. ""Models of Cooperation Based on the Prisoner's Dilemma and the Snowdrift Game." *Ecology Letters*, 8(7): 748–766.

Duffy, John, and Jack Ochs. 2009. "Cooperative Behavior and the Frequency of Social Interaction." *Games and Economic Behavior,* 66(2): 785-812.

Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review,* 90(4): 980-94.

Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10(2): 171-8.

Flood, Merrill M. 1958. "Some Experimental Games." *Management Science*, 5(1): 5-26.

Gächter, Simon, and Christian Thöni. 2005. "Social Learning and Voluntary Cooperation among Like-Minded People." *Journal of the European Economic Association*, 3(2–3): 303–14.

Guala, Francesco. 2010. "Reciprocity: Weak or Strong? What Punishment Experiments Do (and Do Not) Demonstrate." University of Milan Department of Economics, Business and Statistics Working Paper No. 2010-23. Forthcoming in *Brain and Behavioral Sciences*.

Halliday, T. R. 1983. "The Study of Mate Choice." In *Mate Choice*, ed. Patrick Bateson, 3-32. New York: Cambridge University Press.

Hamilton, W. D. 1964. "The Genetical Evolution of Social Behaviour." *Journal of Theoretical Biology*, 7(1): 1-16.

Hauert, Christoph, Arne Traulsen, Hannelore Brandt, Martin A. Nowak, and Karl Sigmund. 2007. "Via Freedom to Coercion: The Emergence of Costly Punishment." *Science*, 316(5833): 1905-7.

Hoshi, Yoko, Jinghua Huang, Shunji Kohri, Yoshinobu Iguchi, Masayuki Naya, Takahiro Okamoto, and Shuji Ono. 2011. "Recognition of Human Emotions from Cerebral Blood Flow Changes in the Frontal Region: A Study with Event-Related Near-Infrared Spectroscopy." *Journal of Neuroimaging*, 21(2): 94-101.

Hume, David. 1739 (1874 Edition). *A Treatise of Human Nature Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects and Dialogues Concerning Natural Religion.* London: Longmans, Green, and Co.

Jackson, Matthew O. 2001. "A Crash Course in Implementation Theory." *Social Choice and Welfare*, 18(4): 655-708.

Kalai, Ehud. 1981. "Preplay Negotiations and the Prisoner's Dilemma." *Mathematical Social Sciences*, 1(4): 375-9.

Kandori, Michihiro. 1992. "Social Norms and Community Enforcement." *Review of Economic Studies,* 59(1): 63–80.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory*, 27(2): 245-52.

López-Pérez, Raúl, and Marc Vorsatz. 2010. "On Approval and Disapproval: Theory and Experiments." *Journal of Economic Psychology,* 31(4): 527–41.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. Oxford: Oxford University Press.

Maskin, Eric. 1999. "Nash Equilibrium and Welfare Optimality." *Review of Economic Studies,* 66(1): 23–38.

Masuda, Takehito, Yoshitaka Okano and Tatsuyoshi Saijo. (2012). "The Minimum Approval Mechanism implements Pareto Efficient Outcome Theoretically and Experimentally." In Preparation.

Moore, John, and Rafael Repullo. 1988. "Subgame Perfect Implementation." *Econometrica,* 56(5): 1191-1220.

Nowak, Martin A. 2006. "Five Rules for the Evolution of Cooperation." *Science*, 314(5805):1560-1563.

Okano, Yoshitaka, Tatsuyoshi Saijo, and Junyi Shen. 2011. "Backward Elimination of Weakly Dominated Strategies is Compatible with Experimental Data Compared with Subgame Perfection: Approval and Compensation Mechanisms." In Preparation.

Okano, Yoshistaka. 2012. "The Equilibrium paths of the Mate Choice Mechanism with Prisoner's Dilemma Game with Four Behavioral Principles and Five Equilibrium Concepts." Mimeo.

Ostrom, Elinor. 1990. Governing the Commons. Cambridge: Cambridge University Press.

Plott, Charles R. and Vernon L. Smith. 2008. *Handbook of Experimental Economics Results*. Amsterdam: North-Holland.

Poundstone, William. 1992. *Prisoner's Dilemma*. New York: Anchor.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281-1302.

Roth, Alvin E. 1995. "Introduction to Experimental Economics." In *The Handbook of Experimental Economics*, ed. John H. Kagel and Alvin E. Roth, 3-109. Princeton: Princeton University Press.

Roth, Alvin E. and J. Keith Murnighan. 1978. "Equilibrium Behavior and Repeated Play of the Prisoner's Dilemma." *Journal of Mathematical Psychology*, 17(2): 189-98.

Russett, Bruce, Harvey Starr and David Kinsella. 2009. *World Politics: The Menu for Choice*, Wadsworth Publishing. 9th edition.

Saijo, Tatsuyoshi . 1988. "Strategy Space Reduction in Maskin's Theorem: Sufficient Conditions for Nash Implementation." *Econometrica*, 56(3): 693-700.

Saijo, Tatsuyoshi, and Yoshitaka Okano. 2009. "Six Approval Rules Whose Outcomes are Exactly the Same as the Mate Choice Rule." mimeo.

Saijo,Tatsuyoshi, Yoshikatsu Tatamitani, and Takehiko Yamato. 1996. "Toward Natural Implementation." *International Economic Review*, 37(4): 949-80.

Saijo, Tatsuyoshi, Tomas Sjöström, and Takehiko Yamato. 2007. "Secure Implementation." *Theoretical Economics*, 2(3): 203-229.

Sekiguchi, Sho. 2012. " Relation between Cooperative Behavior and Approval in the Public Good Provision Games." (in Japanese) Master Thesis presented to Tokyo Institute of Technology.

Selten, R. 1975. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory*, 4(1): 25-55.

Shinagawa, Hideo, Masao Nagatsuka, Yoshitaka Okano and Tatsuyoshi Saijo. 2012. "Cerebral Blood Flow Changes in the Frontal Region of Approval Mechanism and Prisoner's Dilemma Game: A Study with Event-Related Near-Infrared Spectroscopy." In Preparation.

Shubik, Martin. July 2011. "The Present and Future of Game Theory," Cowles Foundation Discussion Paper 1808, Yale University.

Smith, Adam. 1759. *The Theory of Moral Sentiments*. Glasgow, Scotland.

Sugden, Robert. 1984. "The Supply of Public Goods Through Voluntary Contributions." *Economic Journal*, 94(376): 772-787.

Takaoka, Masanori, Yoshitaka Okano, and Tatsuyoshi Saijo. 2011. "Institutional Stress between Approval and Costly Punishment Mechanisms: A Salivary Alpha-Amylase Approach." In Preparation.

Varian, Hal R. 1994. "A Solution to the Problem of Externalities When Agents Are Well-Informed." *American Economic Review*, 84(5): 1278-93.

Yamagishi, Toshio. 1986. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology*, 51(1): 110-116.