Heterogeneous Preferences and Collective Action

T. K. **Ahn**
Workshop in Political Theory and Policy Analysis
Indiana University

Elinor Ostrom
Department of Political Science
Workshop in Political Theory and Policy Analysis
Indiana University

James M. Walker
Department of Economics
Workshop in Political Theory and Policy Analysis
Indiana University

January 9, 2003

Incorporating motivational heterogeneity into game-theoretic models of collective action

Abstract. In recent years, scholars have turned to alternative representations of utility to capture motivational heterogeneity across individuals. In the research reported here, we examine two models of heterogeneous utility—linear-altruism and inequity-aversion—in the context of two-person, social dilemma games. Empirical tests are conducted drawing on data from experiments and surveys. We find that the model of inequity-aversion accounts for a substantial proportion of the preference types and behavior that are not explained by the standard model of self-interested preferences. In contrast, the altruism model does not provide a significant increase in explanatory power over the inequity-aversion model.

Key Words, collective action; social dilemma; game theory; motivation; heterogeneity; altruism; equity; cooperation

Incorporating motivational heterogeneity into game-theoretic models of collective action

T. K. Ahn, Elinor Ostrom, and James M. Walker

1. Introduction

Mancur Olson's *The Logic of Collective Action* (hereafter, *The Logic)* advanced a convincing argument for why groups of individuals with a common interest may not achieve that interest due to individual incentives. At the same time, *The Logic* also stimulated a large field of inquiry into why one observes behavior that appears inconsistent with the absence of collective action predicted by standard models. *The Logic* posited two ways that collective action can be successfully organized: selective incentives and political entrepreneurs. A major contribution of *The Logic* was not, therefore, that it demonstrated the failure of collective action, but rather that it opened a field of inquiry into the reasons for successful collective action.

To broaden the approach posited by Olson, scholars have turned in recent years to the possibility that motivational heterogeneity, across individuals, may be an additional factor in understanding collective action. Put simply, motivational heterogeneity implies that individuals differ in regard to values or social orientations.

In addition to the standard model based on purely self-interested motivations, this paper examines theoretically and empirically two well-known models of motivational heterogeneity: linear-altruism and inequity-aversion. Basic features of these two models have been studied previously, especially by their original advocates. In this paper we explore their implications in the game-theoretic setting of 2x2 social dilemmas. A

principal focus is how one can use a formal game-theoretic approach to study implications of motivational heterogeneity in collective action situations. We conduct the analysis for each of the preference assumptions by developing a series of models based on (1) the presence or absence of common knowledge and (2) simultaneous and sequential decision making. We derive testable hypotheses based on a game-theoretic analysis and use multiple data sets to evaluate the relative explanatory power of the alternative motivational models.

Theoretically, we show that the homogeneous, self-interest model is embedded in the inequity-aversion model, which in turn is embedded in the linear-altruism model. Empirically, we find that while the standard self-interest model is too narrow to provide explanations for the empirical observations, the linear-altruism model is too broad, under-restricted, and often redundant. The inequity-aversion model occupies the middle ground; it explains most of the anomalies generated by the standard self-regarding model without sacrificing parsimony.

The remaining sections of this paper are organized as follows. In Section 2, we introduce the alternative motivational models in the context of a 2x2 social dilemma setting. Section 3 examines implied preference-ordering possibilities and Section 4 examines implications for strategic outcomes. In Section 5, the models' implications are tested using evidence from survey data and one-shot experiments. Section 6 offers summary comments.

2. Altruism and equity in a social dilemma setting

If individuals act purely on self-regarding incentives, a social dilemma is an action situation that generates an outcome where each individual is worse off in comparison to

the case in which individuals act cooperatively. For this study, we focus on 2x2 social dilemmas.

2.1. The social dilemma setting

For a more general analysis, we introduce the representation shown in Figure 1 where the payoffs are denoted generically as T(temptation), R(reward), P(punishment), and S(sucker's payoff). The payoff entries are pecuniary. Viewing the game as a one-time decision, each individual, acting independently, is always better off in terms of pecuniary payoffs when choosing D.

(Figure 1 about here)

The concepts of Fear, Greed, and Cooperators' Gain (Rapoport and Chammah, 1965; also see Ahn et al., 2001) help simplify the analysis. Greed (T-R) is the magnitude of gain to an individual that defects when the other cooperates. Fear is the magnitude of loss (P-S) to an individual that cooperates when the other defects. Cooperators' Gain (R-P) is the magnitude of the gain to the two individuals when both cooperate. When these three measures are normalized by the range of possible payoffs, (T-S), they are called normalized Greed (G_n), normalized Fear (F_n), and normalized Cooperators' Gain (C_n). The relationship among these three quantities characterizes a 2x2 social dilemma quite well. The sum of the measures equals one.

Normalized Gain: $G_n = (T-R)/(T-S)$

Normalized Fear: $F_n = (P-S)/(T-S)$

Normalized Cooperators' Gain: $C_n = (R-P)/(T-S)$

$G_n + F_n + C_n = 1$

Experimental studies using pecuniary payoffs similar to those shown in Figure 1 have consistently found that a significant proportion of individuals choose to cooperate. A growing literature is developing that incorporates alternative motivations; these include Andreoni, 1990; Rabin, 1993; Cain, 1998; Dufwenberg and Kirchsteiger, 1998; Falk and Fischbacher, 1998; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000. In the analysis here, we apply two of these models, linear-altruism (Cain, 1998) and inequity-aversion (Fehr and Schmidt, 1999), to the 2x2 social dilemma decision setting.

2.2. *Linear altruism and inequity aversion*

A "linear-altruism" model (Taylor, 1987; Cain, 1998; Dougherty and Cain, 1999) posits that an individual's utility function is constructed as a linear-weighted sum of one's own pecuniary payoff (π_i) and the pecuniary payoff to others (π_j). The magnitude of the type parameter, (θ_i), varies across individuals within the interval [0,1], reflecting interindividual heterogeneity.

- Linear-altruism utility function

$$u_i(\pi_i, \pi_j; \theta_i) = \pi_i + \theta_i \pi_j \qquad (1)$$

where $0 \leq \theta_i \leq 1$.

The inequity-aversion model proposed by Fehr and Schmidt (1999) assumes that individuals have preferences that relate to income equality. An individual with an aversion to income inequality would sacrifice own income to increase or decrease others' income so as to achieve a more equal allocation. The strength of such preferences is increasing in the magnitudes of the type parameters α_i and β_i.

- Inequity-aversion utility function

$$u_i(\pi_i, \pi_j; \alpha_i, \beta_i) = \pi_i - \alpha_i \max(\pi_j - \pi_i, 0) - \beta_i \max(\pi_i - \pi_j, 0) \qquad (2)$$

where $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$.

Figure 2 presents the characteristics of the three motivational models: self-regarding, linear-altruism, and inequity-aversion. In each panel, an individual's utility is shown as a function of own income and other's income. In the self-interest model, shown in the upper-left panel, own utility is not affected by the level of other's payoff. In contrast, the indifference curves based on linear-altruism and inequity-aversion illustrate the tradeoff an individual is willing to make between own earnings and others, with the slope of the indifference curves varying with parameter specifications.

(Figures 2 and 3 about here)

3. Types based on preference-orderings

Models of linear-altruism and inequity-aversion allow for the possibility of preference orderings over outcomes that are not determined solely by the amount of one's own pecuniary payoff. Figure 3 displays the four outcomes of a 2x2 social dilemma with T=4, R=3, P=2, and S=1. Now, imagine this outcome space is superimposed on the panels of Figure 2. Note that, when Figure 3 is superimposed on the indifference mapping for a self-regarding individual (upper-left panel), the outcome in which own payoff is larger is always the most preferred outcome. We refer to this preference type as a PD type.

3.1. *Types of preferences based on linear-altruism*

Consider the case of linear-altruism. Preferences over the cooperative outcome (C,C) versus the outcome (D,C) depends on the slope of the indifference curves (shown in the upper-right panel of Figure 2) relative to the slope of the line that connects the outcomes (C,C) and (D,C) shown in Figure 3. If the line that connects (C,C) and (D,C) is steeper than the slope of the indifference curve, (C,C) is preferred to (D,C).

Further, an individual with preferences consistent with linear-altruism may prefer (C,D) to (D,D). In fact, if Greed (T-R) and Fear (P-S) are the same, such an individual prefers (C,D) to (D,D) whenever he prefers (C,C) to (D,C). This in turn implies that when Greed is larger than Fear, there can be individuals who prefer (C,D) to (D,D), but prefer (D,C) to (C,C). We refer to this preference type as Utilitarian. On the other hand, when Fear is larger than Greed, there can be individuals who prefer (C,C) to (D,C), but prefer (D,D) to (C,D). We refer to such preference types as Assurance. When both conditions hold, i.e., one prefers (C,C) to (D,C) and (C,D) to (D,D), one has a dominant strategy in Cooperation, which we refer to as Strong-altruism. Notice that Fear > Greed is a necessary condition for the existence of Assurance types; likewise, Greed > Fear is a necessary condition for Utilitarian types.

The necessary and sufficient conditions for each preference type can be expressed more precisely in terms of a player's type parameter and the normalized material payoff parameters as shown in Table 1. The top panel of Figure 4 presents the preference-type space of the linear-altruism model and divides the space into three subspaces (the PD type, the Assurance type, and the Strong-altruism type) as a function of the parameter 0i and the material payoff parameters of a social dilemma.

(Table 1 about here)

(Figure 4 about here)

3.2. *Types of preferences based on inequity-aversion*

Now consider preference types based on the inequity-aversion model shown in the bottom panel of Figure 4. When weak orderings are assumed away, the inequity-aversion model allows only two types of preferences: PD and Assurance.

First, consider the case where the slope of the indifference curve below the 45-degree line (in Figure 2) is smaller than that of the line connecting (C,C) and (D,C) in Figure 3. In this case, the individual prefers (C,C) to (D,C) and is an Assurance type. Note, as β_i approaches 1, the slope of the indifference curve below the 45-degree line becomes smaller. This implies that if the slope becomes sufficiently small, an individual could have an Assurance preference where the individual prefers (D,D) to (D,C).

Note that the inequity-aversion model does not allow a preference ordering of (C,D) > (D,D). Imagine an indifference curve that passes through (D,D) in Figure 3. This line will always have a non-negative slope above (D,D). Therefore, the point (C,D) always falls on the left side of the indifference line passing (D,D).

The possible preference-ordering types based on the inequity-aversion model and their conditions are shown in Table 2. The conditions for preference types are expressed only in terms of β_i, the weight in one's utility function attached to the utility loss due to disadvantageous inequity. Parameter α_i plays a role in equilibrium analysis, but is not a determining factor of an individual's preference-ordering type for a 2×2 social dilemma game. The bottom panel of Figure 4 presents the preference type space of the inequity-aversion model. The bold solid line bounds the type space. Because α_i is unbounded above, the type space is not closed. For a given value of G_n, the dotted line separates the two preference types as a function of β_i.

(Table 2 about here)

4. Equilibria in games with varying levels of information

In this section, we discuss equilibria predictions for the 2x2 social dilemma game. We begin with models where each of the two players knows the preference type of the other

player with certainty. Then we proceed to analyze the 2x2 social dilemma game with incomplete information based on an assumption of common knowledge that players share a common and accurate prior about the probability distribution of types.

4.1. *Games with complete information*

A variety of pure strategy equilibria, other than mutual defection, are feasible when there are multiple player types. Based on player type, Tables 3 and 4 display the feasible equilibria for simultaneous and sequential games with complete information. Recall, the preference types allowed by the inequity-aversion model are a proper subset of those allowed by the linear-altruism model. Therefore, the feasible equilibria derived from the inequity-aversion model are shown in a shaded 2x2 box, embedded in the 4x4 box for the feasible equilibria derived from the linear-altruism model. We use the solution concept of a Nash equilibrium for the simultaneous games. For the sequential games, we refine the Nash equilibria using the solution concept of subgame perfection.

(Tables 3 and 4 about here)

Two points are worth noting. First, the sequence of play (whether the game is played simultaneously or sequentially) and the preference type of the first mover in sequential games, are both important. In general, when player's types are known, the mutually cooperative outcome is more likely achieved when the game is played sequentially. Second, while the inequity-aversion model predicts only mutual defection or mutual cooperation in equilibrium, the linear-altruism model allows outcomes in which one player defects and the other cooperates.

4.2. *Games with incomplete information*

The assumption we adopted in the analysis of complete information, that each of the two players knows the type of the other player, is not generally true in field settings or experiments. In this subsection, we analyze the decision situation assuming that players do not know the exact type of the other. We analyze the action situation as a game of incomplete information and present Bayesian equilibria of the game based on the two models.

A. Incomplete information: Game equilibria with linear-altruism

Below we discuss the equilibria of the 2x2 social dilemma game graphically and develop hypotheses regarding the relative frequencies of cooperative choice for games played simultaneously or sequentially. Four information situations are considered: (1) play in a simultaneous game, (2) first mover in a sequential game, (3) second mover in a sequential game following play of C by the first mover, and (4) second mover in a sequential game following play of D by the first mover. Formal conditions and proofs for equilibria are not presented here. Instead, Figures 5 and 6 graphically summarize behavior in equilibrium.[1]

(Figures 5 and 6 about here)

Figure 5 relates to behavior based on linear-altruism. In each panel of Figure 5, the bold segment of the line shows the parameter space of players who cooperate in an equilibrium. Players whose types do not belong to the bold segment defect in the equilibrium. We present the case in which $F_n > G_n$. This allows for three preference types: PD, Assurance, and Strong-altruism.

The first panel of Figure 5 shows for the simultaneous game, all Strong-altruism types and a subset of Assurance types will cooperate. The second panel shows that all Strong-altruism and a subset of Assurance types will cooperate as first movers in a sequential game. Defining the cooperative subsets precisely, for both a simultaneous game and for first movers in a sequential game, depends on the exact distribution of types. The third panel of Figure 5 shows that as a second mover of a sequential game, when the first mover cooperates, all Assurance types and Strong-altruism types cooperate. Conditional on the first mover defecting, the fourth panel shows that the second mover cooperates only in the case of Strong-altruism types.

B. Incomplete information: Game equilibria with inequity-aversion

Figure 6 relates to equilibrium behavior based on the inequity-aversion model. In each panel of Figure 6, the cross-hatched area of the type space shows the parameter space of players who cooperate in an equilibrium. Players whose types are not contained in the cross-hatched area defect in the equilibrium. In the simultaneous game, the strategy profile in which both players defect always exists as an equilibrium, regardless of the distribution of types.[2]

We will assume that whenever there exists an equilibrium of the simultaneous game in which a subset of Assurance types cooperate, the equilibrium will be played. The upper-left panel of Figure 6 shows the potential for this equilibrium. As noted earlier, this subset could be an empty set, the whole set, or anything in between the two. As the shaded area indicates, an Assurance type player is more likely to cooperate the larger the value of β_i and the smaller the value of α_i.

The upper-right panel of Figure 6 indicates that for first mover's equilibrium behavior is a function of only cti; the smaller the value of a; the more likely cooperation. This implies that in this information set, being an Assurance type does not necessarily mean that the player is more likely to cooperate than another player who is a PD type. The lower-left panel shows that as the second mover of a sequential game, when the first mover cooperates, all Assurance types will cooperate. On the other hand, the lower-right panel shows that no player, regardless of type, will cooperate as the second mover of a sequential game when the first mover defects.

C. Incomplete information: Hypotheses

The analyses above do not assume a specific distribution of types and thus are limited in the extent to which specific testable hypotheses can be made. However, even without making specific distributional assumptions, several broad empirical conclusions can be derived:

- Following defection by the first mover, the inequity-aversion model predicts no occurrences of cooperation by second movers.
- Both models predict weakly less cooperation in the simultaneous game than in the sequential game when the first mover cooperates.
- Both models predict weakly more cooperation in the simultaneous game than in the sequential game when the first mover defects.

5. Empirical results

The analyses in Sections 3 and 4 show how the three motivational models are embedded within each other. Summarizing, the standard self-interest model is embedded in the inequity-aversion model, which in turn is embedded in the linear-altruism model. The

nesting is in terms of both predicted preference orderings and equilibria. The self-interest model predicts only a PD preference-ordering type. The inequity-aversion model adds an Assurance type. The linear-altruism model adds two more types: the Utilitarian and the Strong-altruism types. In terms of the feasible equilibria under complete information, the self-interest model predicts only the (D,D) outcome, the inequity-aversion model predicts either (D,D) or (C,C), and the linear-altruism model predicts (D,D), (C, C), or (C,D).

In this section, we provide evidence from recently collected survey data and two existing experimental studies. We use data from a recent class survey conducted by the authors, and from Hayashi et al. (1999) and Cho and Choi (2000). Hayashi et al. provide both survey and behavioral data from subjects in Japan and the U.S., while Cho and Choi provide behavioral data from Korea.[3] The U.S. survey data were collected during the spring semester of 1999, in three undergraduate courses (introductory microeconomics, honors introductory microeconomics, and a sophomore political science course) at Indiana University. Participation was voluntary. Students were not required to answer the survey as a requirement for the course.

The students were presented with the decision situation shown in Figure 7. They were asked to assume they would make a decision whose monetary outcome was affected by a similar decision made by another student in the class. Students were first asked to check off one of two decision boxes: I would choose A or I would choose B. A choice of A corresponds to defection, B to cooperation in our theoretical analyses. The students then completed the questionnaire shown in Figure 8.[4] Similar questions were asked in a questionnaire administered by Walker and Ostrom related to the experiments reported in

Hayashi et al. (1999), although the wording was slightly different due to a difference in the framing of the decision.

(Figures 7 and 8 about here)

5.1. *Preference orderings from surveys*

Using the preference-type classifications summarized in Tables 1 and 2, we report data on subject types based on the subjects' preference orderings over the four possible outcomes of the 2x2 social dilemma game in Table 5. As one can see, PD preference types are most frequently observed, consistent with 42% of the observations from the class survey and 20% from the Hayashi et al. study.

(Table 5 about here)

Recall that the model of inequity-aversion implies the existence of an Assurance type as one added preference type compared to the self-interest model. As shown in Table 5, 10% of the observations from the class survey and 19% of the observations from Hayashi et al. are consistent with an Assurance type. In deriving preference types in Section 3, we considered only strong preference orderings, for simplicity. Here, we consider an additional preference type, referred to as a PD-Assurance-Indifference type. Subjects in this category revealed a preference ordering of DC = CC > DD > CD. In the inequity aversion model, this preference implies $\beta_i = G_n$; in linear altruism model, $\theta_i = G_n/(1-G_n)$. This type accounts for 19% of the classroom survey data and 16% of the questionnaire data reported in Hayashi et al. We should note, given that the survey questions regarding preference orderings were restricted to integer values, we suspect that many of the observations of indifference may be linked to this artifact of the survey.

The linear-altruism model expands the space of possible preference types to account for what we refer to as Utilitarian and Strong-altruism. As shown in Table 5, however, these types account for none of the data in the classroom survey and only a small percentage of the data from Hayashi et al. In addition, we include three additional "indifference" types that account for 3% of the observations from the in-classroom survey and 6% of the data in the data from Hayashi et al. Finally, note that 25% of the classroom respondents and 39% of the Hayashi et al. respondents reveal preference types inconsistent with any of the three motivational models.

5.2. *Behavior in one-shot settings*

In this section, we examine behavior in two decision settings, previously reported in Hayashi et al. (1999) and Cho and Choi (2000). All experiments were conducted as one-shot decision settings, using double-blind procedures. In one setting the players moved simultaneously and in the other they moved sequentially, where second movers were informed of first mover decisions prior to their own decision. Because of the nature of the sequential move setting, observations on subject's behavior are observed in three situations: (1) first mover, (2) second mover when the first mover cooperated, and (3) second mover when the first mover defected.

In both the simultaneous and sequential settings, the decision problem was framed in the following way. Each was allocated $10. The subjects could choose to keep their $10 or give the $10 to the person with whom they were paired, in which case the $10 would be doubled. Subjects had common information that the subject with whom they were paired faced the same decision. Thus, the subjects faced a binary choice social dilemma. If each cooperated and gave away their $10, they each earned $20. If each

defected and kept their $10, they each earned $10. In the case where one cooperated and the other defected, the one cooperating received nothing and the one defecting earned $30.

Table 6 presents the relative frequency of cooperation in the four information settings. In the simultaneous setting, we observe cooperation rates of 36%, 46%, and 56% across the U.S., Japanese, and Korean subject pools. In the sequential setting, first movers cooperate at rates of 56%, 82%, and 52% across the three subject pools, and cooperation is reciprocated by second movers at very high levels, although not at a 100% level.

(Table 6 about here)

We view these results as broadly consistent with the survey data presented above and the motivational models presented earlier. Although observing subject behavior is often not sufficient to characterize the exact preference type of each subject, the levels of cooperation reported in Table 6 suggest that some subjects are motivated by factors beyond their own pecuniary payoffs. Based on the predictions drawn from the models presented above, a set of inferences on the preference type of a player can be made. Cooperation by a first mover in a sequential game is consistent with all of the three preference types of PD, Assurance, and Strong-altruism. On the other hand, cooperation by a player in the simultaneous game or by a second mover of the sequential game following cooperation by the first mover, can be seen as an indicator of either Assurance or Strong-altruism type. Further, cooperation by a second mover in the sequential game following defection by the first mover can be seen as an indicator of Strong-altruism, but inconsistent with an Assurance type. Three out of the twenty-five Japanese subjects

cooperated, knowing that the first movers had already defected. The possibility of an unconditionally cooperating type, i.e., Strong-altruism, cannot be totally ignored, though their presence is relatively small in this sample.

6. Summary comments

This paper presents a series of theoretical and empirical investigations of two models that posit alternative motivations to selfishness: linear-altruism and inequity-aversion. Empirical tests are conducted drawing on two sets of experimental data. In terms of preference orderings, the model of inequity-aversion accounts for a substantial proportion of the preference types not explained by the standard model of self-interested preferences. In contrast, the altruism model does not provide a significant increase in explanatory power over the inequity-aversion model.

Future theoretical and empirical work examining various types of collective action situations will be well advised to presume that there are multiple types of individuals likely to be involved in collective action. This leads the theorist to examine a wide variety of contextual factors—including those that Mancur Olson posited—that may operate so as to help participants identify the types of other players involved. The existence of multiple types of players helps to explain a consistent empirical finding that subjects involved in social dilemmas can achieve cooperation at substantially higher levels when they communicate face-to-face, even when the communication does not involve enforcement mechanisms.

Notes

1. See Ahn (2001) for derivation of equilibrium conditions.
2. This is not the case with the linear-altruism model, because Strong-altruism types always cooperate.
3. See Hayashi et al. and Cho and Choi for the details of the experimental procedure.
4. In addition to these questions, the students were also asked several more questions related to their belief about others' choice, and their level of general trust.

References

Ahn, T.K. (2001). Foundations for cooperation in social dilemmas. Unpublished doctoral dissertation. Indiana University, Bloomington.

Ahn, T.K., Ostrom, E., Schmidt, D., Shupp, R. and Walker, J. (2001). Cooperation in PD games: Fear, greed, and history of play. *Public Choice* 106(1/2): 137-155.

Andreoni, J. (1990). Impure altruism and donation to public goods: A theory of warm glow giving. *Economic Journal* 100: 464-477.

Bolton, G. E. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review* 90: 166-193.

Cain, M. (1998). An experimental investigation of motives and information in the prisoner's dilemma game. *Advances in Group Processes* 15: 133-160.

Cho, K. and Choi, B. (2000). A cross-society study of trust and reciprocity: Korea, Japan, and the U.S. *International Studies Review* 3(2): 31-43.

Dougherty, K. L. and Cain, M. (1999). Linear altruism and the 2x2 prisoner's dilemma. Working paper, Florida International University, Miami.

Dufwenberg, M. and Kirchsteiger, G. (1998). A theory of sequential reciprocity. Discussion Paper, CentER, Tilburg University.

Falk, A. and Fischbacher, U. (1998). A theory of reciprocity. Working paper, University of Zürich, Institute for Empirical Research in Economics.

Fehr, E. and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114: 817-868.

Hayashi, N., Ostrom, E., Walker, J. and Yamagishi, T. (1999). Reciprocity, trust, and the sense of control: A cross-societal study. *Rationality and Society* 11(1): 27-46.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of group.* Cambridge: Harvard University Press.

Rapoport, A. and Chammah, A.M. (1965). *Prisoner's dilemma.* Ann Arbor: University of Michigan Press.

Rabin, M. (1993). Incorporating fairness in game theory and economics. *American Economic Review* 83(5): 1281-1302.

Taylor, M. (1987). *The possibility of cooperation.* New York: Cambridge University Press.

		Individual 2	
		Cooperation	Defection
Individual 1	Cooperation	R, R	S, T
	Defection	T, S	P, P

T, R, P, and S are pecuniary payoffs: $T > R > P > S$; $2R > T + S$.

Figure 1. Generic representation of a 2×2 social dilemma

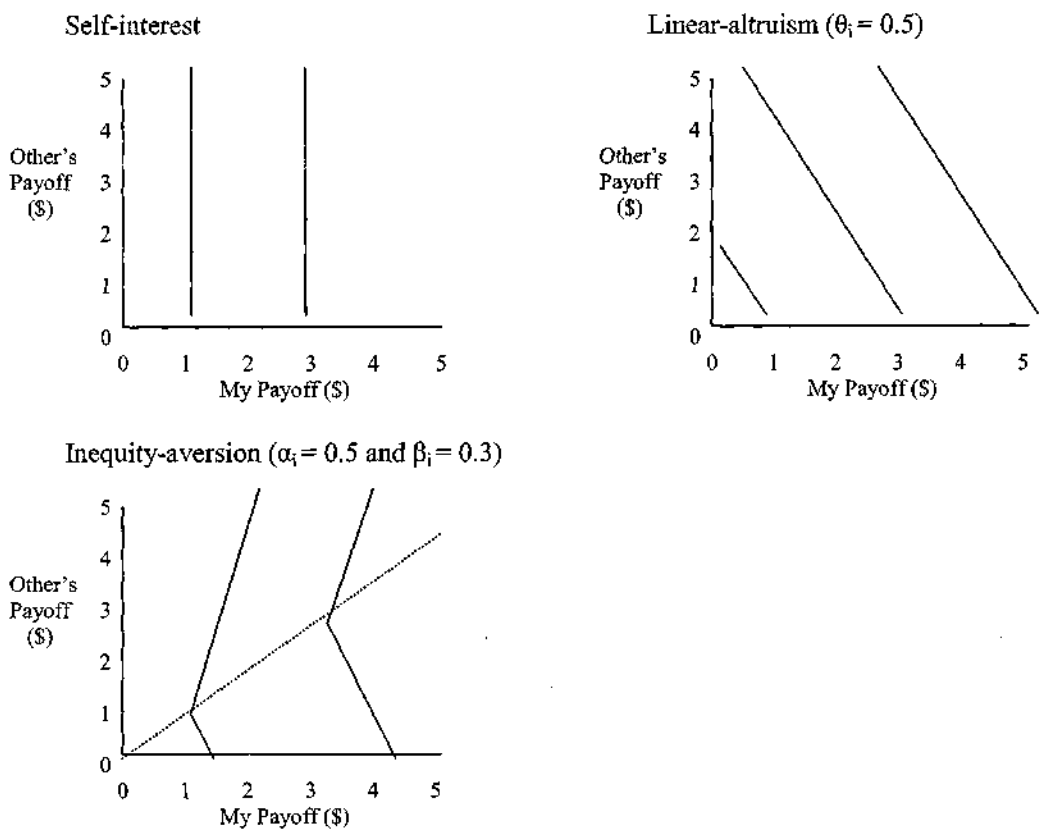
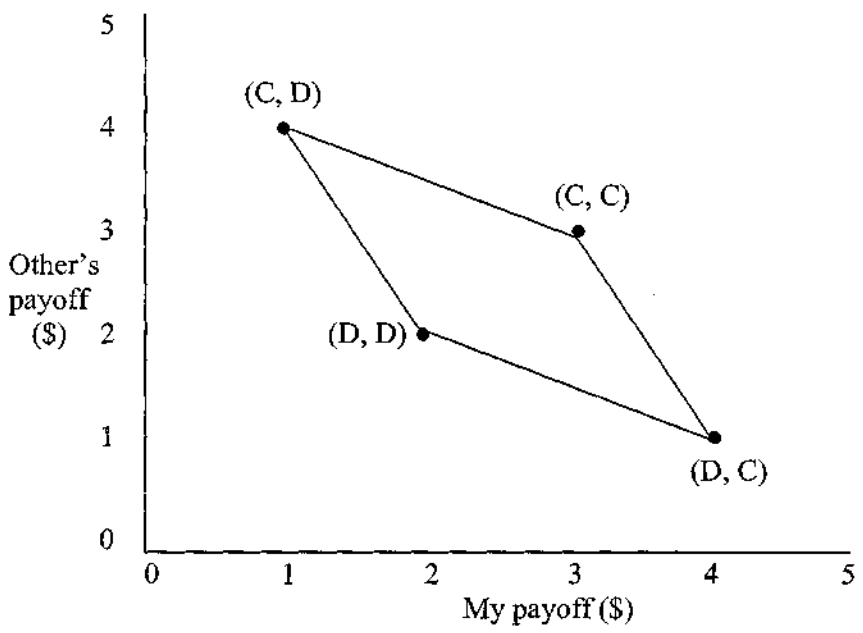


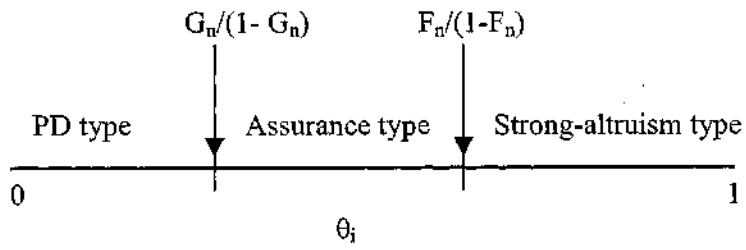
Figure 2. Indifference mappings: self-interest, linear-altruism, and inequity-aversion



The four outcomes of a 2×2 social dilemma game are marked with the strategy profiles (*s*, *s'*) such that *s* is my strategy and *s'* is other's strategy.

Figure 3. Mapping of four outcomes of a 2×2 social dilemma: T = 4, R = 3, P = 2, and S = 1

Linear-altruism Model ($F_n > G_n$)



Inequity-aversion Model

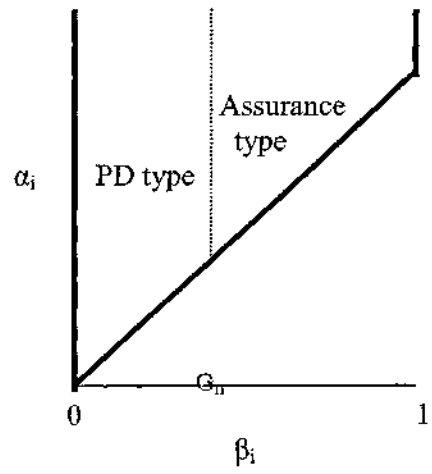
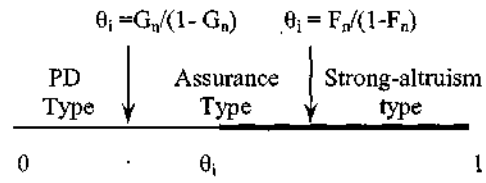
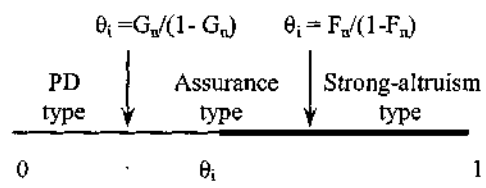


Figure 4. Type space and types

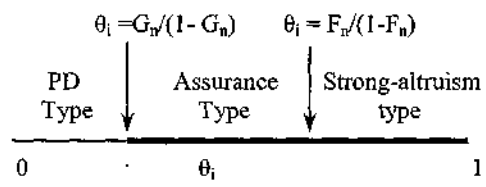
Simultaneous game



Sequential game 1st mover



Sequential game 2nd mover if 1st mover cooperates



Sequential game 2nd mover if 1st mover defects

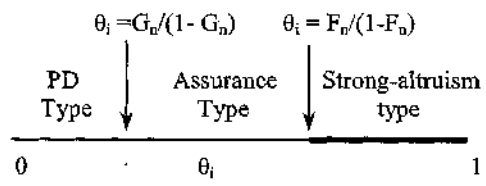
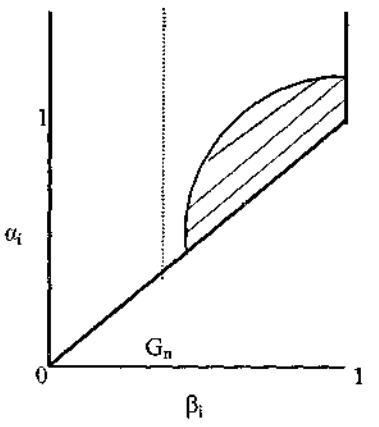
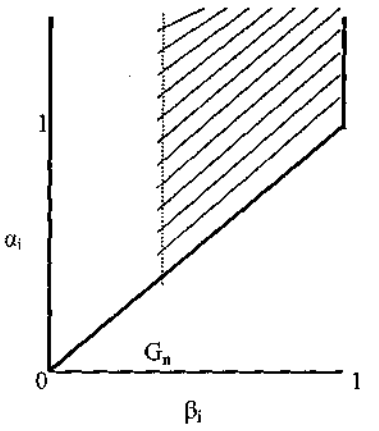


Figure 5. Type space of players who cooperate in equilibrium: linear-altruism with $F_n > G_n$

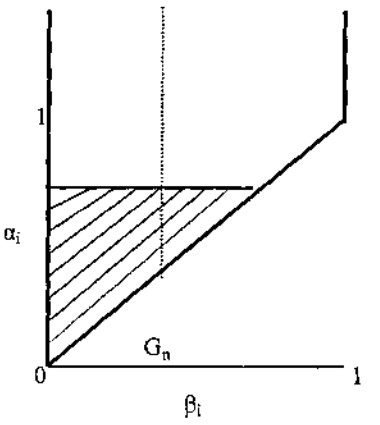
Simultaneous game



Sequential game 2nd mover if 1st mover cooperates



Sequential game 1st mover



Sequential game 2nd mover if 1st mover defects

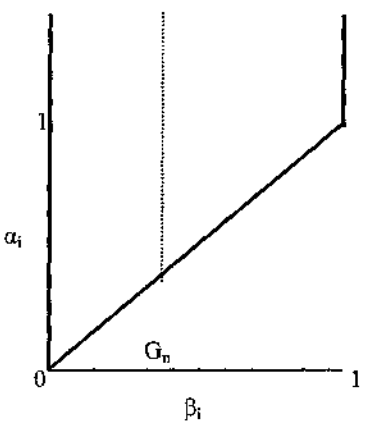


Figure 6. Type space of players who cooperate in equilibrium: inequity-aversion model

		OTHER	
		A	B
YOU	A	YOU: $10 OTHER: $10	YOU: $25 OTHER $5
	B	YOU $5 OTHER $25	YOU $20 OTHER $20

Figure 7. Decision problem in class survey

1. How satisfactory would it be to you if both you and the other player chose D?

1___2___3___4___5___6___7

Very unsatisfactory Very satisfactory

2. How satisfactory would it be to you if both you and the other player chose C?

1___2___3___4___5___6___7

Very unsatisfactory Very satisfactory

3. How satisfactory would it be to you if you chose D and the other player chose C?

1___2___3___4___5___6___7

Very unsatisfactory Very satisfactory

4. How satisfactory would it be to you if you chose C and the other player chose D?

1___2___3___4___5___6___7

Very unsatisfactory Very satisfactory

Figure 8. Class survey questionnaire

Table 1. Conditions for preference types: linear-altruism model

Type	Preference ordering	Condition	Necessary condition
PD	(D,C) > (C,C) > (D,D)> (C,D)	$\theta_i < \min[F_n/(1-F_n), G_n/(1-G_n)]$	None
Assurance	(C,C) > (D,C) > (D,D) > (C,D)	$G_n/(1-G_n) < \theta_i < F_n/(1-F_n)$	$F_n > G_n$
Utilitarian	(D,C) > (C,C) > (C,D) > (D,D)	$F_n/(1-F_n) < \theta_i < G_n/(1-G_n)$	$G_n > F_n$
Strong-altruism	(C,C) > (D,C) > (C,D) > (D,D)	$\max[F_n/(1-F_n), G_n/(1-G_n)] < \theta_i$	None

Table 2. Conditions for preference types: inequity-aversion model

Type	Preference ordering	Condition	Necessary condition
PD	(D,C) > (C,C) > (D,D) > (C,D)	$0 \le \beta_i < G_n$	None
Assurance	(C,C) > (D,C) > (D,D) > (C,D)	$G_n < \beta_i < G_n + C_n$	
	(C,C) > (D,D) > (D,C) > (C,D)	$G_n + C_n < \beta_i < 1$	

Table 3. Nash equilibria of complete-information, simultaneous games*

		Column player's type			
		PD	Assurance	Utilitarian	Strong-altruism
Row player's type	PD	(D,D)	(D,D)	(D,C)	(D,C)
	Assurance	(D,D)	(C,C); (D,D)	**	(C,C)
	Utilitarian	(D,C)	**	(D,C); (C,D)	(D,C)
	Strong-altruism	(C,D)	(C,C)	(C,D)	(C,C)

* First entry is for row player.
** No pure strategy equilibrium.

Table 4. Subgame perfect equilibrium outcomes of complete-information, sequential games*

		Second mover's type			
		PD	Assurance	Utilitarian	Strong-altruism
First mover's type	PD	(D,D)	(C,C)	(D,C)	(D,C)
	Assurance	(D,D)	(C,C)	(D,D)	(C,C)
	Utilitarian	(C,D)	(C,C)	(D,C)	(D,C)
	Strong-altruism	(C,D)	(C,C)	(C,D)	(C,C)

* First entry is for the first mover.

Table 5. Frequency results based on questionnaire*

Model			Preference-type	Class survey	Hayashi et al. U.S.
Linear-altruism model	Inequity-aversion model	Self-interest model	PD	42%	20%
			Assurance	10%	19%
			Indifference (DC=CC>DD>CD)	19%	16%
			Utilitarian	0%	1%
			Strong-altruism	0%	2%
			Indifference other*	3%	6%
Not explained:				25%	39%
			Anomaly 1: U(C,D) > U(D,C)	12%	6%
			Anomaly 2: U(C,C) = U(D,D)	6%	14%
			Other	7%	19%
Total subjects				162	198

* Other examples of indifference: DC > CC > DD = CD, CC = DC > DD = DC, CC = DC > CD > DD.

Table 6. Frequency results based on one-shot, double-blind experiments in three countries

	U.S.	Japan	Korea
Simultaneous game	36%	56%	46%
Sequential game 1st mover	56%	83%	52%
Sequential game 2nd mover if 1st mover cooperates	61%	75%	73%
Sequential game 2nd mover if 1st mover defects	0%	12%	0%

Sources: The U.S. and Japan data are from Hayashi et al. (1999) and the Korean data is from Cho and Choi (2000).