W02-6

INCORPORATING MOTIVATIONAL HETEROGENEITY INTO GAME THEORETIC MODELS OF 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

Paper to be presented at the 2002 meeting of the Public Choice Society
San Diego, California, March 22-24

Incorporating Motivational Heterogeneity into Game Theoretic Models of Collective Action

ABSTRACT

Understanding cooperation in the context of social dilemma games is fundamental to understanding how alternative institutional arrangements may foster collective action in such settings. An abundance of experimental evidence is inconsistent with predictions from game theoretic models based strictly on self-regarding utilities. In recent years, scholars have turned to alternative representations of utility in an attempt to capture motivational heterogeneity across individuals. In the research reported here, we examine two models of heterogeneous utility, linear-altruism and inequity-aversion, as complements to the standard model based on purely self-interested motivations. We examine these models in the contest of two-person social dilemma games. In addition, we examine data from experiments and survey instruments that provide evidence related to the empirical robustness of models based on different types of players characterized by heterogeneous utility functions.

Incorporating Motivational Heterogeneity into Game Theoretic Models of Collective Action

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

I. INTRODUCTION

Mancur Olson's *The Logic of Collective Action {The Logic*, hereafter) advanced powerful arguments 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 initiated 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 means by which collective action can be organized successfully: selective incentives and political entrepreneurs. A major contribution of *The Logic*, therefore, was not that it demonstrated the failure of collective action, but that it opened up a field of inquiry into the reasons for successful collective action.

In addition to the conditions posited by Olson, scholars have turned in recent years to the question of how 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 expanding the range of other individuals that may be included in a utility function.

In addition to the standard motivational model based on selfish interests, this paper theoretically and empirically examines two well-known models of motivational heterogeneity: linear-altruism and inequity-aversion. We examine these two models not as alternatives to the standard, self-regarding model found in *The Logic*, but as complements. The two models we examine by no means exhaust worthy candidates for analysis. We select them to examine the implications that can be derived in a formal comparison of these models and how these implications can be tested empirically. In addition, both models are quite intuitive and have been adopted by multiple scholars.

The most important difference between linear-altruism and inequity aversion models and the standard self-interest model is that these two models incorporate heterogeneity of preferences. The standard self-interest model assumes that *everyone* is selfish; homogeneity and selfishness are the key defining characteristic of the standard model. In contrast, both of the models we examine in this paper make a fundamental assumption of inter-individual heterogeneity regarding utility derived from benefits achieved by others.

The model of inequity aversion does not assert that *everyone* has the same degree of aversion towards inequity; it argues that some individuals are selfish and some are not, those who are not selfish differ among themselves in the extent to which they are not selfish. In this view, the best conceptual dimension to describe inter-individual heterogeneity is an individual's degree of aversion toward inequity. Nor, does the model of linear altruism argue that everyone takes all other individuals' interests into account. The linear altruism model posits the extent to which one takes others' interests into

account as the key conceptual dimension on which the inter-individual heterogeneity exists.

Some of the basic features and implications of these two models have previously been studied, especially by their original advocates. In this paper we expand those efforts in several ways. First, we attempt to develop full-blown game-theoretic models of 2x2 social dilemmas based on the two models. The key question in this regard is how to use game theory to study implications of motivational heterogeneity in collective action situations. We conduct this by developing a series of models based on: (1) the inequity aversion model, (2) the linear altruism model, (3) the presence of absence of the assumption of common knowledge, and (4) simultaneous and sequential decision-making. Second, we derive empirically testable hypotheses based on the 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. We regard these results as broadly consistent with evolutionary theories of reciprocity.

The remaining sections of this paper are organized as follows. In Section II, we introduce the alternative motivational models in social dilemma settings. In Section III, we analyze the models in terms of the implied preference ordering possibilities of individuals in the context of a 2x2 social dilemma. In Section IV, we analyze the implication of the models for outcomes. In Section V, the models' implications are examined using evidence from survey data and one-shot experiments. Section VI offers summary comments.

XL ALTRUISM AND EQUITY IN A SOCIAL DILEMMA SETTING

The Social Dilemma Setting

A social dilemma is an action situation (as explained in Ostrom, Gardner, and Walker, 1994) in which, if individuals act purely on self-regarding incentives, the resulting outcome is one in which each individual is worse off in comparison to the case in which group members acted otherwise.[1] For this study, we focus on 2x2 social dilemmas.

Consider the dilemma shown in Figure 1, where outcomes are presented as pecuniary payoffs. There are two individuals: YOU and OTHER. Each of the two individuals has two choice options: C (Cooperation) and D (Defection). YOU receive a higher monetary payoff when YOU choose D, regardless of what OTHER chooses. More specifically, if

[1] See Ahn (2001) for a discussion of the various ways to define a social dilemma and their implications for research.

OTHER chooses C, YOU receive $40 when YOU choose D and $30 when YOU choose C. If OTHER chooses D, YOU receive $20 when YOU choose B, and $10 when YOU choose C. The payoffs are symmetric. Acting independently, and maximizing own monetary payoff, implies both players choose D, resulting in a payoff of $20 each. This outcome is deficient relative to the case where both choose C and receive $30, hence the dilemma.

For a more general analysis, we introduce the representation shown in Figure 2. Again, payoff entries are pecuniary. Given the ordinal relations among the payoff entries - T>R>P>S - the action situation contains the same decision-making problem as the one shown in Figure 1. Viewing the game as a one-time decision, each individual, acting independently, is always better off in terms of pecuniary payoff when choosing D.

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 one obtains by defecting, when the other cooperates. Fear is the magnitude of loss (P-S) that incurs to an individual when he cooperates while the other defects. Cooperators' Gain (R-P) is the magnitude of the gain each of the two players obtains when both choose to 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 payoff structures similar to those shown in Figures 1 and 2 have consistently found that a significant proportion of individuals choose the cooperative choice. In addition to the pecuniary payoffs of the game, a collection of variables has been shown to have systematic influences on the frequency of the cooperative choices. Schmidt et al. (2001) categorize these variables into four broad sets: the structure of pecuniary payoffs, player types based on motivational heterogeneity, information about player types, and the linkage between players.

A growing literature exists of models designed to incorporate alternative motivations; these include Andreoni, 1990; Rabin, 1993; Cain, 1998; Dufwenberg and Kirshsteiger, 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.

Linear Altruism and Inequity Aversion

The term altruism is often used to refer to any action that is not strictly self-regarding. In this study, we restrict the analysis to what is often called "linear-altruism" (Taylor, 1987; Cain, 1998; Dougherty and Cain, 1999). As shown in (1), a linear altruism model posits that individuals' utilities are 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 inter-individual 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 preference relationship proposed by Fehr and Schmidt (1999) and shown in (2) suggests 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

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

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

Figure 3 graphically 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. More specifically, the figures display indifference curves over own payoff and others. In the self-regarding model, shown in the first panel, own utility is not affected by the level of other's monetary 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. The region to the right of an indifference line is preferred to that to the left. An interesting aspect of the inequity-aversion model is that it predicts that some individuals with preference for equitable outcomes will reject some strictly Pareto-improving allocations, if the improvement is in the direction of increased inequity.

III. 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. In the context of the 2x2 social dilemma shown in Figure 2, this

implies that there can be preference-orderings other than the strictly self-regarding one shown in (3).

$$u(D,C)>u(C,C)>u(D,D)>u(C,D) \quad (3)$$

Types of preference orderings can be expressed in terms of the values of model parameters and the normalized payoff parameters G_n, C_n, and F_n. Combining Figures 3 and Figure 4 illustrates how alternative preference types can occur.

Figure 4 displays the four outcomes of a 2x2 social dilemma, with T=4, R=3, P=2, and S=l. Now, imagine this outcome space is superimposed on the panels of Figure 3. Note that, when Figure 4 is superimposed on the indifference mapping for a self-regarding individual (panel 1), the outcome in which own payoff is larger is always the most preferred outcome. We refer to this preference type as a PD type.

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 Figure 3) relative to the slope of the line that connects the outcomes (C,C) and (D,C) shown in Figure 4. 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). Note that, if the slope of the line connecting (C,C) and (D,C) is steeper than the indifference curve, Greed (T-R) is therefore smaller. In summary, if altruistic preferences are sufficiently strong and the temptation to defect is sufficiently small, (C,C) will be 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 the Assurance types; likewise, Greed>Fear is a necessary condition for the 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 5 presents the type space of the linear altruism model linearly and divides the space into three subspaces (the PD type, the Assurance type, and the Strong Altruism type) as a function of the parameter theta and the material payoff parameters of a social dilemma.

Types of Preferences Based on Inequity-Aversion

Now consider preference types based on the inequity-aversion model (see the indifference curves corresponding to panel 3 in Figure 3). 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 3) is smaller than that of the line connecting (C,C) and (D,C) in Figure 4. In this case, the individual prefers (C,C) to (D,C) and is an Assurance type. Note, as β_i approaches 1, the upper limit of the parameter restriction, 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 4. 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. Unlike the model of linear-altruism, preference types based on the inequity-aversion model are not dependent upon the relationships among the normalized payoff parameters. Substantively, this means that the preference types defined with this model are possible regardless of the relative magnitudes of the three material payoff parameters G_n, F_n, and C_n. 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.

IV. EQUILIBRIA IN GAMES WITH VARYING LEVELS OF INFORMATION

In this section, we discuss equilibria predictions for the 2×2 social dilemma game, based on alternative preference types and assuming different levels of information among players. We begin with models where each of the two players knows with certainty the preference type of the other player. It is reasonable to assume, however, that in many situations, a player does not have complete information regarding the type of other players. When information is incomplete, a player's belief, his probability assessment of another player's type and likely behavior, becomes a crucial factor in determining behavior. With this in mind, we discuss the decision-making problem from the viewpoint of a single player with incomplete information. Then, we proceed to analyze the 2×2 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.

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 games with complete information, when play is simultaneous and sequential. Recall, the preference types generated 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 2×2 box, embedded in the 4×4 box for the feasible equilibria derived from the linear-altruism model. We use the solution concept of a Nash equilibrium for the simultaneous games, and that of a subgame perfect equilibrium for the sequential games.

Several 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 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 predicts outcomes in which one player defects and the other cooperates.

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 drop the assumption of complete information and analyze the decision situation assuming that players do not know the exact type of the other. First, we analyze the decision-making problem from the perspective of a single player. The analysis will show that an individual's belief about the other's type and behavior plays a crucial role, along with the individual's own preference type, in his decision-making. Secondly, by introducing the common knowledge assumption, we analyze the action situation as a game of incomplete information and present Bayesian equilibria of the game based on the two models.

Incomplete Information: The Individual Decision Problem

Figure 6 illustrates the decision-making problem a player faces when the preference type of the other player is unknown. The figure is based on preferences motivated by linear-altruism, with simultaneous play. The extent to which beliefs about others type matters is dependent on one's own type. Suppose that Player 1 is a PD-type. No matter the other player's choice, D exists as the dominant strategy because $T + \theta_1 S > R + \theta_1 R$ and $P + \theta_1 P > S + \theta_1 T$. Similarly, a Strong Altruism-type player has a dominant strategy, C.

Now, consider the case where Player 1 is an Assurance-type, to whom beliefs about the other player's type does matter. Belief is treated as a probability assessment on the likely decision of the other player. The belief has two relevant components: (1) the probability

that Player 2 is either an Assurance-type or a Strong-Altruism type, and (2) if Player 2 is an Assurance-type, the probability that Player 2 will play C.

Let μ_1^A and μ_1^L denote Player 1's probability assessments that Player 2 is an Assurance-type and a Strong-Altruism type, respectively. Let μ_1^{AC} denote Player 1's probability assessment that Player 2, if an Assurance- type, will play C. Then the probability assessment of Player 1 that Player 2 will play C, denoted p_1, is: $p_1 = \mu_1^A \times \mu_1^{AC} + \mu_1^L$.

Given p1, Player 1 calculates the expected utility of playing C versus D. The expected utility from C is $u_1(C) = p_1 (R + \theta_1 R) + (1-p_1)(S + \theta_1 T)$. The expected utility from D is $u_1(D) = p_1 (T + \theta_1 S) + (1-p_1)(P + \theta_1 P)$. Player 1 plays C if $u_1(C)$ is greater than $u_1(D)$, or

$$\theta_i \succ \frac{Fn - p(Fn - Gn)}{1 - Fn + p(Fn - Gn)}. \qquad (5)$$

Simply stated, (5) shows that the larger the altruism parameter, θ_i, and the smaller the normalized Fear, Fn, the larger the probability that Player 1 will play C.[2]

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 and 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 7 and 8 summarize behavior in equilibrium graphically. More specifically, for each situation, the shaded area signifies the parameter conditions under which cooperation would be observed in equilibrium.[3]

Figure 7 relates to behavior based on the model of linear-altruism. We present the case in which $F_n > G_n$. This allows for three preference types: PD, Assurance, and Strong-

[2] An important issue is the origin of player's beliefs about others. There are two bases for approaching this question. First, beliefs can be treated as purely personal traits, as in the early psychology literature on trust (Rosenberg, 1956; Deutsch, 1958). The other extreme is the standard game theoretic assumption of common knowledge in games of incomplete information. When belief is treated as a personal attribute, the outcome of a 2x2 social dilemma is simply a set of two independent decisions. On the other hand, treating the situation as an incomplete information game implies that each player knows the utility payoff function of his counterpart probabilistically; that the belief is accurate; and that each player knows that the probability distribution is also known to the counterpart, ad infinitum. One way to think about this kind of common knowledge is that there is a population from which the two players are randomly and independently drawn to play the game; the objective distribution of the type parameter within the population is known to the two players and each of the two players know that the other knows, etc. The assumption of common knowledge is rarely met in field settings. When a group of individuals shares common experience, however, and each group member has a sense of how cooperative others are which is not so different from the beliefs others have, then the common knowledge assumption may be a reasonable approximation to the shared experience and knowledge within the group.

[3] See Ahn (2001) for derivation of equilibria conditions.

Altruism. See Figure 6 for the exact expression of the two cut points in terms of the normalized payoff parameters. The upper left-hand panel of Figure 7 shows that in the information set for the simultaneous game, all Strong-Altruism types and a subset of Assurance-types will cooperate. The upper right-hand panel shows that all Strong Altruism and a subset of Assurance types will cooperate as first movers in a sequential game. Which of the two cooperative subsets, that for the information set of the simultaneous game or that for the first mover's information set in the sequential game, cannot be determined unless the exact distribution of types is known.

The lower left-hand panel of Figure 7 indicates that as a second mover of a sequential game, when the first mover cooperates, all Assurance-types and Strong-Altruism-types will cooperate. The lower right-hand panel shows that as a second mover of a sequential game and when the first mover defects, only Strong-Altruism types cooperate.

Incomplete Information: Game Equilibria with Inequity Aversion

Figure 8 relates to equilibrium behavior based on the inequity-aversion model. The strategy profile in which both players defect always exists as an equilibrium, regardless of the distribution of types.[4] However, we will assume that whenever there exists an equilibrium of the simultaneous game in which a subset of Assurance-types cooperates, the equilibrium will be played.

The upper left-hand panel of Figure 8 shows that in the simultaneous game, a subset of Assurance-types will cooperate in equilibrium. The same caveat applies; the 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-hand panel of Figure 8 indicates that in the first mover's information set, equilibrium behavior is a function only of α_i; the smaller the value of α_i the more likely cooperation. This implies that in that in this situation, 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-hand 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-hand panel of Figure 8 shows that no player, regardless of type, will cooperate as the second mover of a sequential game when the first mover defects.

[4] This is not the case with the linear-altruism model. If there is a sufficiently large proportion of Strong-Altruism-types, a strategy profile in which all Assurance-types defect will not be an equilibrium.

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 implications exist

- The inequity-aversion model predicts no occurrences of cooperation by a second mover's following defection by the first mover.
- Both the linear-altruism model and inequity-aversion model predict weakly less cooperation in the simultaneous setting than for second mover's following cooperation by the first mover.
- Both the models predict weakly less cooperation by the second movers of sequential settings following first movers' defection than in simultaneous settings.

Discussion of Theoretical Results

The analyses in Section III and IV show how the relationships among the three motivational models are embedded within each other. The standard self-interest model is embedded in the inequity aversion model, which in turn is embedded in the linear altruism model. The nesting exists in terms both of the predicted preference ordering types and the equilibria in game theoretic models. The self-interest model predicts only 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 (D,D) outcome, the inequity aversion model adds (C,C) outcome, and the linear altruism model further adds (C,D).

In terms of the relative frequencies of cooperation in different information sets in games with incomplete information, all predictions derived from the inequity aversion model are also derived from the linear altruism model. While the inequity aversion model predicts zero cooperation between the second movers of sequential games when first movers defect, the linear altruism model is consistent with any behavior under that condition.

The embededness relations imply that the player-type space of the linear altruism model is the largest among the three. Thus, whenever the linear altruism model is refuted by empirical evidence regarding player types the other two are automatically refuted. However, this does not mean that the linear altruism model is necessarily the "better" model. When we evaluate the models based on empirical data in the next section, therefore, we will use the criterion of "marginal explanatory power." We will ask the question of what additional proportion of data that cannot be explained by a model is explained by another model with a wider explanatory space. For example, what marginal proportion of data on preference ordering do the additional two types added by the linear altruism model account? We will acknowledge a model's usefulness only when the marginal explanatory power is strong enough.

V. EMPIRICAL RESULTS

In this section, we provide evidence from recently collected survey data, and two existing experimental studies, to examine the "empirical validity" of the motivational models presented in this paper.[5] In particular 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.[6]

The U.S. survey data were collected during the spring semester of 1999, in three undergraduate courses (introductory microeconomics, honors introductory microeconomics, and a sophmore 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 represented in Figure 9. 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, I would choose B). The students then completed the questionnaire shown in Figure 10.[7] Similar questions were asked in the questionnaire administered by Walker and Ostrom related to the experiments reported in Hayashi et al., (2001) although the wording was slightly different due to a difference in the framing of the decision.

Preference Orderings from Surveys

Using the preference/type classifications summarized in Tables 1 and 2, in Table 5, we report data on subjects' types based on the subjects preference orderings over the four possible outcomes of the 2x2 social dilemma game. As one can see, PD preference type is the most frequently observed, consistent with 42% of the observations from the class survey and 19.8% from the Hayashi, et al. study.

Recall that the model of inequity aversion implies the existence of one added preference type, Assurance. As shown in Table 5,10.5% of the observations from the class survey and 18.6% of the data from Hayashi, et al. is consistent with an Assurance type. We also consider an additional preference type, referred to as a PD-Assurance-Indifference type. Subjects in this category revealed a preference ordering over outcomes of DC=CC>DD>CD. Interestingly, this type accounts for 19% of the classroom survey data and 15.9% of the questionnaire data reported in Hayashi, et al.

---

[5] An earlier study by Rachel Croson (1999) utilized a series of public good experiments to distinguish between theories of altruism and theories of reciprocity. Croson's results provide support for "reciprocity theories" — of which inequity aversion would be one example - over altruism theories.

[6] See Hayashi et al. and Cho and Choi for the details of the experimental procedure.

[7] In addition to these questions, the students were also asked several more questions related to the satisfactory level of each of the four possible outcomes, their belief about others' choice, and their level of general trust.

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 for Hayashi, et al. In addition, we include three additional "indifference" types which account for 3% of the in classroom survey and 6% in the Hayashi, et al. study. Finally, note that 25% of the classroom respondents (39% of the Hayashi, et al. respondents) reveal preference inconsistent with any of the three motivational models.

Behavior in One-shot Settings

In this section, we examine behavior in two decision settings, previously reported in Hayashi, et al. (2001), using subjects from the U.S. and Japan and Cho and Choi (2000), using subjects from Korea. **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. In the sequential move setting, we observe subjects making decisions in three information settings: (1) when they know that they are the first mover, (2) when they know they are the second mover and they know that the first mover cooperated, and (3) when they know they are the second mover and they know that the first mover defected.

In both simultaneous and sequential experiments, the subjects faced the following decision problem, 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 diverse 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, one observes that first movers cooperate at rates of 56%, 82%, 52% across the three subject pools, and that cooperation is reciprocated by second movers at very high levels, although not at a 100% level.

We view these results as broadly consistent with the survey data presented above and the motivational models presented earlier. While behavior in the simultaneous game and behavior by first movers does not directly indicate a player's type, it does suggest that many subjects are motivated by incentives beyond pure pecuniary payoffs to self. On the other hand, a choice of Cooperation by a second mover of the sequential game, following the first mover's Cooperation, is an indicator of either Assurance or Altruism type. Choice of Cooperation by a second mover in the sequential game following the first mover's Defection is an indicator of an Altruism type, but inconsistent with an Assurance type. Finally, note that the Japanese subjects' behavior reveals an additional behavioral

tendency. Three out of the Japanese 25 subjects chose Cooperation, knowing that the first movers' had already chosen Defection. The possibility of an unconditionally cooperating type, i.e. the Strong-Altruism, cannot be totally ignored, though their presence is relatively small in this sample.

VI. SUMMARY COMMENTS

This paper has conducted a series of theoretical and empirical investigations of two models that posit alternative motivations to selfishness: linear altruism and inequity aversion. When applied to 2x2 social dilemma games, each of the two models generates a series of possible preference-ordering types over the four outcomes of the game, while the traditional assumption of self-interest allows for only one preference type. The model of linear altruism classifies individuals into four preference-ordering types: PD, Assurance, Utilitarian, and Strong-Altruism. The motivational model based on linear altruism allows for the existence of strong altruists who unconditionally cooperate. Further, the existence of types in this model is dependent on the structure of material payoffs of the game. On the other hand, the model of inequity aversion generates only two preference-ordering types: PD and Assurance, and the existence of these types in not dependent upon the pecuniary payoffs of the game. The model of inequity aversion precludes unconditional cooperation and divides individuals into two broad subsets of conditional cooperators and unconditional defectors.

When the social dilemma is framed as an incomplete information game, both motivational models specify conditions for cooperative equilibrium in simultaneous and sequential 2x2 social dilemma games. The equilibrium analysis allows us to derive hypotheses regarding the relative frequency of cooperation in the qualitatively different information sets of the games. The hypotheses based on the altruism model include a wider class of behavior than those that are based on the model of inequity aversion.

Empirical tests are conducted drawing on two sets of experimental data. In terms of preference ordering, 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 an additional explanation for the behavior of subjects that is not accounted for by 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 and to recruit a larger proportion of Assurance types rather than only PD types. The existence of multiple types of players helps to explain a consistent empirical finding that subjects involved in social dilemmas who can communicate on a face-to-face basis without the capability to commit each other to keep promises - mere cheap talk - are able to increase the proportion of individuals cooperation substantially (Ostrom, Gardner, and Walker, 1994; Isaac and Walker, 1991).

References

Ahn, T. K. 2001. *Foundations for Cooperation in Social Dilemmas.* Ph.D. Dissertation. Indiana University. Bloomington, Indiana.

Ahn, T. K., Elinor. Ostrom, David. Schmidt, Robert Shupp, and James. Walker. 2001. "Cooperation in PD Games: Fear, Greed, and History of Play." *Public Choice* 106(l/2):137-155.

Andreoni, James. 1990. "Impure Altruism and Donation to Public Goods: A Theory of Warm Glow Giving." *Economic Journal* 100:464-477.

Bolton, Gary E. and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90:166-193.

Cain, Michael. 1998. "An Experimental Investigation of Motives and Information in the Prisoner's Dilemma Game." *Advances in Group Processes* 15: 133-160.

Cho, Kisuk, and Byoung-il Choi. 2000. "A Cross-Society Study of Trust and Reciprocity: Korea, Japan, and the U.S." *International Studies Review* 3(2):31-43.

Crosson, Rachel. 1999. "Theories of Altruism and Reciprocity: Evidence from Linear Public Goods Games" Wharton School of Economics, University of Pennsylvania: Department of Economics. Working Paper

Deutsch, M. 1958. "Trust and Suspicion." *Journal of Conflict Resolution* 2(4):265-279.

Falk, Armin, and Urs Fischbacher. 1998. "A Theory of Reciprocity." Working Paper. Zurich: University of Zürich, Institute for Empirical Research in Economics.

Dougherty, Keith L. and Michael J.G. Cain. 1999. "Linear Altruism and the 2x2 Prisoner's Dilemma." Working Paper. Miami: Florida International University.

Dufwenberg, Martin, and Georg Kirchsteiger. 1998. "A Theory of Sequential Reciprocity." Discussion Paper, Center, Tilburg University.

Falk, Armin, and Urs Fischbacher. 1998. "A Theory of Reciprocity." Working Paper. Zurich: University of Zürich, Institute for Empirical Research in Economics.

Fehr, Ernst and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114:817-868.

Hayashi, Nahoko, Elinor Ostrom, James Walker, and Toshio Yamagishi. 1999. "Reciprocity, Trust, and the Sense of Control: A Cross-Societal Study." *Rationality and Society* ll(l):27-46.

Isaac, R. Mark and James Walker. (1991). "Costly Communication: An Experiment in a Nested Public Goods Problem," in *Laboratory Research in Political Economy*. Ann Arbor: University of Michigan Press.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Group*. Cambridge, MA: Harvard University Press.

Ostrom, Elinor, Roy Gardner, and James M. Walker. 1994. *Rules, Games, and Common-Pool Resources.* Ann Arbor: University of Michigan Press.

Rapoport, Anatol, and A. M. Chammah. 1965. *Prisoner's Dilemma.* Ann Arbor: University of Michigan Press.

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

Rosenberg, M. 1956. "Misanthropy and Political Ideology." *American Sociological Review* 21:690-695.

Schmidt, David, Robert Shupp, James M. Walker, T.K. Ahn, and Elinor. Ostrom. 2001. "Dilemma Games: Game Parameters and Matching Protocols." *Journal of Economic Behavior and Organization* 46(4):357-377.

Taylor, Michael. 1987. *The Possibility of Cooperation.* New York: Cambridge University Press.

FIGURE 1. A 2x2 Social Dilemma

		If OTHER Chooses	
		C	D
If YOU Choose	C	YOU get $30 OTHER gets $30	YOU get $10 OTHER gets $40
	D	YOU get $40 OTHER gets $10	YOU get $20 OTHER gets $20

FIGURE 2. General From Representation of a 2x2 Social Dilemma

		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 3. Indifference Mappings: Selfish, Linear Altruism, and Inequity Aversion.

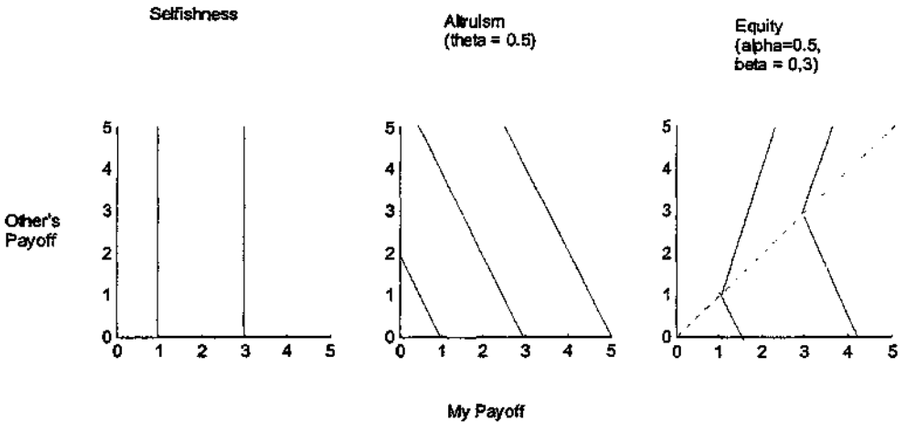
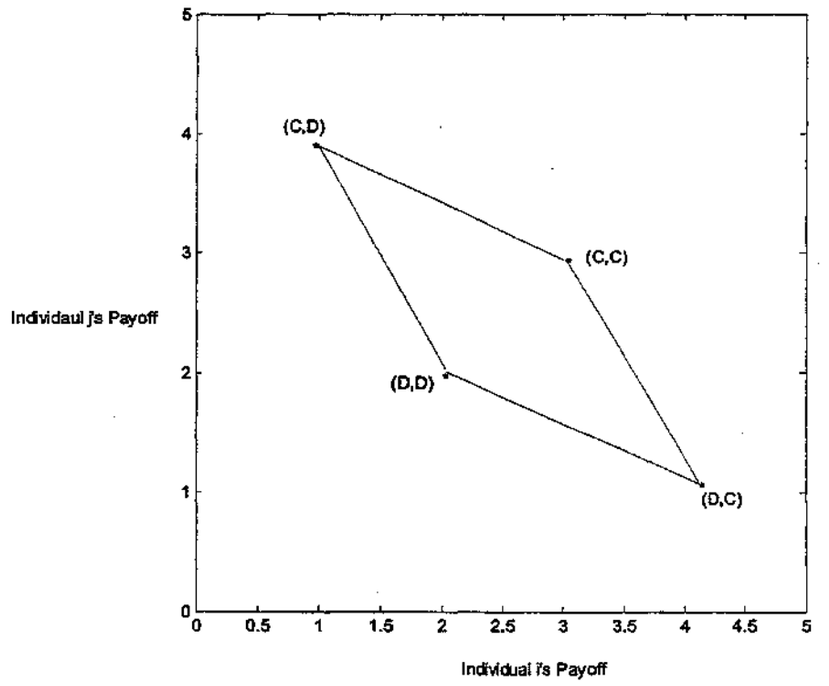
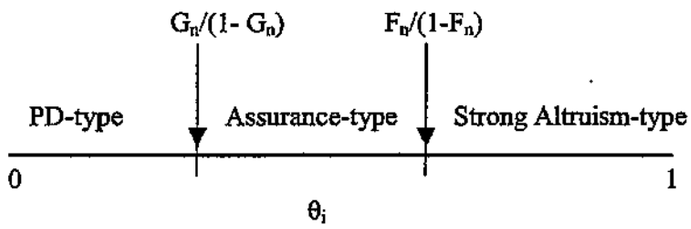


FIGURE 4. Mapping of Four Outcomes of a 2x2 Social Dilemma
T=4, R=3, P=2, S=1

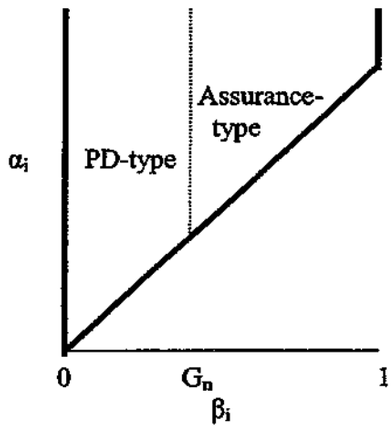


* The four outcomes of a 2x2 social dilemma game are marked with the strategy combinations (S,S') such that S is the strategy of individual i and S' is the strategy of individual j.

FIGURE 5. Type Space and Types



Linear Altruism Model ($F_n>G_n$)



Inequity Aversion Model

FIGURE 6. Decision-Making Problem When Partner's Type is Unknown:
Linear Altruism Model

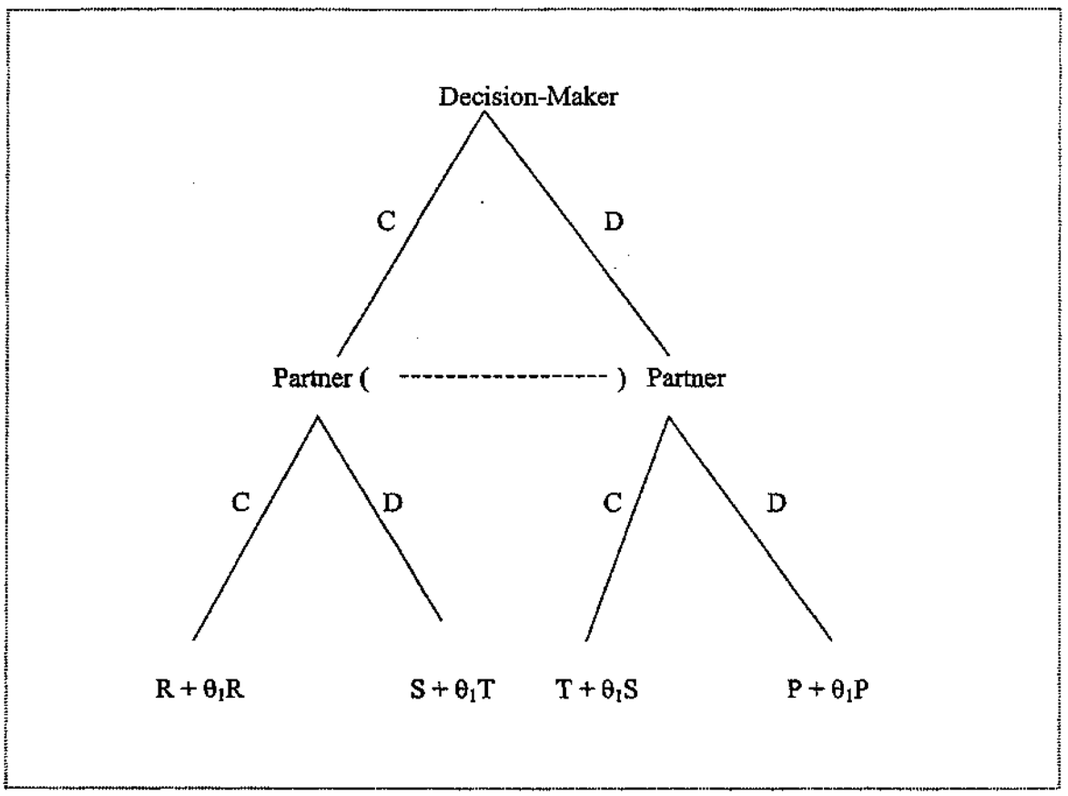


FIGURE 7. Behavior in Equilibrium: Linear Altruism (Fn>Gn)

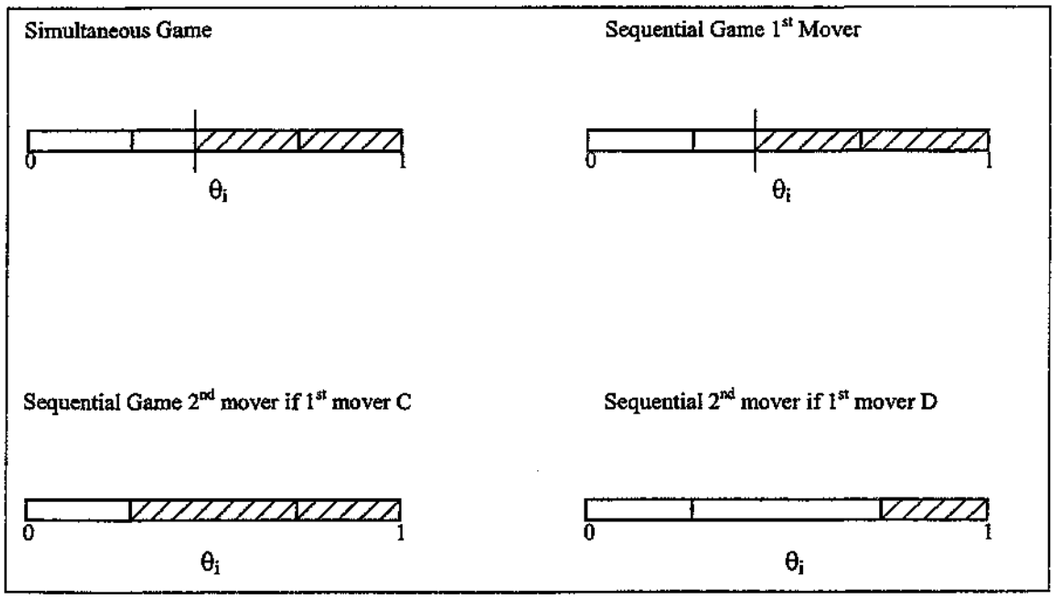


FIGURE 8. Behavior in Equilibrium: Inequity Aversion Model

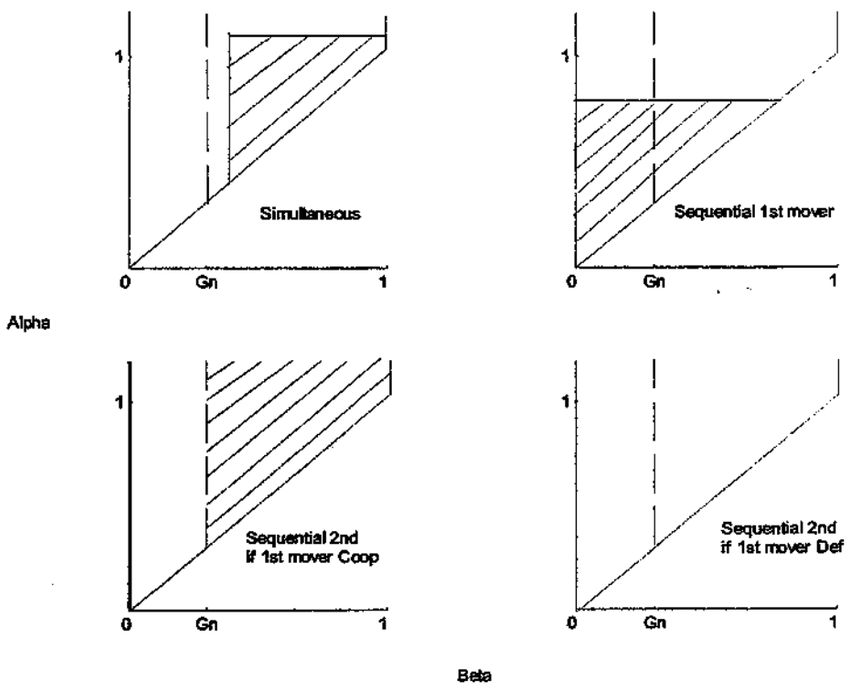


FIGURE 9. Decision Problem in CLASS SURVEY

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

FIGURE 10: Class Survey Questionnaire

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

Table 1. Conditions for Preference Types: Linear Altruism Model

Type	Preference Ordering	Condition	Necessary Condition
PD	(D,C) > (C,C)> (D,D)> (C,D)	$\theta_i < \min[Fn/(1-Fn), Gn/(1-Gn)]$	None
Assurance	(C,C)> (D,C) > (D,D) > (C,D)	$Gn/(1-Gn) < \theta_i < Fn/(1-Fn)$	Fn > Gn
Utilitarian	(D,C) > (C,C) > (C,D) > (D,D)	$Fn/(1-Fn) < \theta_i < Gn/(1-Gn)$	Gn > Fn
Strong Altruism	(C,C) > (D,C) > (C,D) > (D,D)	$\max[Fn/(1-Fn), Gn/(1-Gn)] < \theta_i$	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)	$0 \leq \beta_i < Gn$	None
Assurance	(C,C) > (D,C) > (D,D) > (C,D)	$Gn < \beta_i < Gn+Cn$	
	(C,C) > (D,D) > (D,C) > (C,D)	$Gn+Cn < \beta_i < 1$	

Table 3. Nash Equilibria of Complete-Information, Simultaneous Games

		Column Player's Type			
		PD	Assurance	Utilitarian	S-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)
	S-Altruism	(C,D)	(C,C)	(C,D)	(C,C)

First entry is for Row player. * No pure strategy equilibria

Table 4. Subgame Perfect Equilibrium Outcomes of Complete-Information, Sequential Games

		Second Mover's Type			
		PD	Assurance	Utilitarian	S-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)
	S-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%
			Chicken	0%	1%
			Angel	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					
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	36%	56%	46%
Seq 1st	56%	83%	52%
Seq 2nd if 1st C	61%	75%	73%
Seq 2nd if 1st D	0%	12%	0%

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