publication draft accepted for Journal of Conflict Resolution
Begun: 4 August 1993
This version: February 26, 1996

When Is Universal Contribution Best for the Group? Characterizing Optimality in the Prisoners' Dilemma

Norman Frohlich
Faculty of Management
University of Manitoba
Winnipeg, Manitoba R3T 2N2
(204) 474 6385
e-mail: Frohlic@ccm.umanitoba.ca

and

Joe Oppenheimer
Department of Government & Politics
University of Maryland
College Park, Maryland 20742
(301) 405 4136
e-mail: joppenhe@bss2.umd.edu

Abstract

Social scientists from a variety of disciplines have long been captivated by the simplicity and elegance of the two-person, binary choice, Prisoners' Dilemma (2x2 PD). Over the years, the domain of the research has been extended and applied to events that are neither two-person nor binary. We use a defining characteristic of the 2x2 PD to identify situations under which full levels of contribution are suboptimal. We propose, on the basis of that characteristic, an extended definition and categorization of Prisoners' Dilemmas to n-person and non-binary situation. The new distinction is shown to point to differing normative and strategic imperatives for the different categories of games.

Outline of Contents

Tables and Graphs:

Equations and Inequalities:

When Is Universal Contribution Best for the Group?
Characterizing Optimality in the Prisoners' Dilemma

Introduction

Many social scientists, captivated by the simplicity and elegance of the two person, binary choice, Prisoners' Dilemma (PD) Game (hereafter called a 2x2 PD), have extended analyses to classes of events which are neither two person, nor binary (consider for example, Iwakura and Saijo; Ostrom et. al.; Brams and Kilgour; Hardin; and Schelling). One consequence of these activities is a degree of ambiguity as to what exactly constitute Prisoners' Dilemmas. Schelling (1973, 1978) provided a graphical method for describing a variety of binary choice situations with externalities, among them a number of n-person PD's (n-PD's). He did not, however, attempt to identify which of the situations constituted direct extensions of the 2x2 PD. In his rich description and discussion of these situations Schelling noted that some situations with externalities had social optima which are obtainable with less than universal contribution. But he did not identify the analytic conditions which generate internal (v. corner) optima.

The discussion in this paper can best be thought of as a series of friendly clarifications and extensions of Schelling's analysis. We propose a generalized definition of the PD and identify the conditions which guarantee an optimum in an n-PD at less than universal contribution.[1] This implies a categorization of n-PD's into those with optima at universal contribution and those in which less than universal (we will refer to this as *partial*) contribution is optimal. This distinction will be shown to have both normative and practical implications.

Reclassifying the Prisoners' Dilemma

The Traditional 2x2 PD

The rich tapestry captured in the 2x2 PD is built upon a tension between two elementary principles: strategic dominance and Pareto optimality. In any game of normal form, a strategy, D, is said to be dominant if and only if for each other strategy C, D yields a better outcome for the player than does C under each contingent choice of the other player. Similarly, an outcome is said to dominate another strongly when all players prefer the dominating outcome to the dominated one. An outcome, P, P[2] is said to be Pareto suboptimal if there is another outcome, R, R which both players prefer to P, P.[3]

The traditional 2 person, binary, PD is represented in Table 1. The two players are referred to as **i**, and **j**. The traditional algebraic notation labels the outcomes: R (reward), T (temptation), S (sucker), and P (punishment). The dominant strategy, "don't contribute," is labelled (D) and the dominated strategy, "contribute," (C). The 2x2 PD is endowed with the following properties: 1) Both players have a dominant strategy: D; 2) the Nash equilibrium outcome P,P (the joint

---

1/ Contribution here refers to the allocation of some resource, in which there is an implicit maximum which corresponds to the individual's budget to this problem.

2/ We will not use subscripts for the payoffs to i, and j respectively unless they are necessary for clarity.

3/ Here we use the conventional notation in which the first entry in the ordered pair of payoffs refers to the payoff to the first (or Row) player and the second to the second, (or Column) player. The symmetry of the payoffs in the example is not required for a definition of dominance.

result reached via those strategies) is suboptimal in that it is smaller than R, R (the joint result reached via the cooperative strategy of mutual contribution C).

Table 1: The Two Person Binary PD		
	Player *j*'s Strategic Options	
Player *i*'s Strategic Options	**Contribute (C)**	**Don't Contribute (D)**
Contribute (C)	R, R	S, T
Don't Contribute (D)	T, S	P, P

Analytically, these two properties can be represented directly as relationships between the payoffs. For defect (D) to dominate contribute (C), both the inequalities in (1) must hold.

$$T > R \text{ and } P > S \qquad \text{(1) Dominance}$$

By further specifying, as in (2), that R is preferred to P we insure that the Nash equilibrium is dominated and the full ordering of the payoffs is established.

$$T > R > P > S \qquad \text{(2) Sub-optimality}$$

The Nash outcome, which results from both playing their dominant strategies therefore leads to the Pareto inferior outcome: P, P. The first two properties, have been considered to be adequate by some (Schelling, 1978, for example, p. 218), and is enough to generate the dominant strategies and sub-optimal equilibrium mentioned above.

In addition to these two conditions, a third condition is often identified as characterizing the 2x2 PD: $(S+T)/2 < R$ (see Rapoport, p. 34). This condition is traditionally justified by its necessity to insure that R, R is in the Pareto set. Without this condition, the solution to the dilemma could focus on the probabilistic coordination over the pair of outcomes, S, T and T, S in repeated plays of the game.

To see this, consider the example in Table 2.

Table 2: A 2x2 Binary PD when Universal Contribution is not Pareto Optimal		
	Player *j*'s Strategic Options	
Player *i*'s Strategic Options	**Contribute (C)**	**Don't Contribute (D)**
Contribute (C)	2, 2	0, 5
Don't Contribute (D)	5, 0	1, 1

A cooperative strategy which coordinated the players' asymmetric choices (where first one chose D, while one chose C, and then the choices were reversed) would yield an expectation of payoffs dominating the reward payoff R.[4] In this example, a 50-50 taking of turns would yield each

4/ If the sum of the temptation and sucker payoffs is large enough to violate inequality (3), the outcome of both players avoiding their dominant strategy is no longer Paretian. The Pareto set would be made up of π,{T_i, S_j} and (1-π),{S_i, T_j}, (where π designates a probability of selection associated with the outcome). It is an outcome which, dominates $R_{i,j}$ as long as both $\pi S + (1-\pi)T > R$ and $\pi T + (1-S)\pi > R$. Of course, if the two players adopt mixed strategies in a non-coordinated game to take advantage of a violation of condition (3), P and R would enter as payoffs in a probabilistic fashion as well. There could still be a set of mixed strategies which yields an expected player an expectation of 2.5 (more than the 2 of both contributing). So we can see that the additional condition precludes a coordinated, asymmetric, cooperative, outcome of alternating contribution with non-contribution, as a means of generating optimal returns to the two players in repeated plays.

It is convenient to interpret this third requirement from a slightly different perspective and to use this to gain insight into how Prisoners' Dilemma games can be categorized. To say that $(S+T)/2 < R$ is to require that:

$$R-S > T-R \qquad \text{(3) Externality larger than internality}$$

In terms of the payoffs, we can interpret these differences as potential benefits and costs of defection from a choice to contribute. The internality of the decision (or the player's own benefit from her defection) is T - R and the externality (the cost borne by the other contributor) is R - S. Then (3) can be interpreted as meaning that the negative effect of a given player's defection from a universal contribution outcome on the other player (the *externality* of the decision) outweighs the benefits obtain by the defector via that defection (the *internality* of the decision).[5] But this characterization implies an interpersonal comparison of the payoffs. By taking advantage of the symmetry of the game, however, one can re-interpret the difference, and avoid interpersonal comparisons. Condition (3) can be interpreted as requiring that the externality of **j**'s choice of defection on **i** be greater than the internality of **i**'s choice to defect. This insures that the Pareto set contain a realizable outcome which dominates the suboptimal Nash equilibrium (and also dominates all coordinated mixed strategies). This will be the interpretation which we will use for condition (3). In the continuous strategy extension of the 2X2 PD, proposed below, those games which violate condition (3) will be shown to we have optima at less than universal levels of contribution.

Expansion #1: The Continuous 2-PD

An obvious analytic limitation of the 2x2 PD, as defined above, is that it is restricted to binary choices. Many non-laboratory situations involve a much richer set of choices. Some research has been done to deal with the more general two-person situations which are non binary (see for example, Brams and Kilgour, Chapter 3, 1988, Miller, 1977). To deal with such situations we can expand the strategy set to a continuous set.[6] Each player can be thought of as having a continuous resource, any portion of which can be contributed with the residual being withheld.

---

payoff that dominates R,R. But only by coordinating mixed strategies can they achieve a Paretian lottery of mixing T and S.

5/ This characterization of the conditions for defection in a 2x2 PD as involving an externality larger than the internality is a direct consequence of the work of Bernholz (1976), Miller (1977) and Aldrich (1977) which showed the standard 2x2 PD to be a special case of the Liberal Paradox (Sen 1970). But it is not a customary way of describing the properties of the PD.

6/ One could do this by insisting on either a finite set of strategic options, or by specifying a continuous decision interval. The latter presents us with some advantages of analysis, and is pursued here. The former can be derived as a special case of the continuous representation.

The graphic mode introduced by Schelling (1973, 1978) simplifies the presentation of the problem. It maps the payoffs of one player as a function of the possible choices of the other player. But it does so in terms of the two extreme strategies of the former; see Figure 1. The two parallel lines represent **i**'s extreme strategies. The bottom line represents "make a full contribution" the top "make no contribution." The *y* coordinate represents the payoff to player **i** as a function of player **j**'s contribution as measured along the *x* axis. Thus the height of the lines representing **i**'s payoff is dependent upon the level of **j**'s contribution.

The payoffs on the vertical axes refer to the values of the original 2x2 PD game. The payoffs associated with the strategies are represented as lines in the strategy/outcome space. In the simplest and characteristic case these payoffs are a linear increasing function of the amount contributed by the other player. Moreover, the two strategy/outcome lines are depicted as parallel to represent the assumptions that the marginal cost to the player of allocating additional resources to a contribution is constant as are the marginal increases in group benefits resulting from equal marginal contributions. Resources withheld are equally valuable at any level of contribution by the other and resources contributed by the other are equally productive at all contribution levels.[7] In this sort of diagram (see Figure 1) it should be obvious that the requirement of a dominant strategy (inequality (1)) demands that the line representing no contribution always has to be above the line representing a full contribution. The lines must not cross. The game represented here conforms to that property.

Although the values of the two strategies are shown explicitly as lines, given the two assumptions of linearity and constant cost of allocation, we can define the value of any strategy of partial contribution. The payoffs to **i**, associated with the strategy of allocating 50% of the budget for a contribution, for example by **i**, would be depicted by a line half way between the two lines shown and would be defined throughout the range of **j**'s possible choices, and so on.

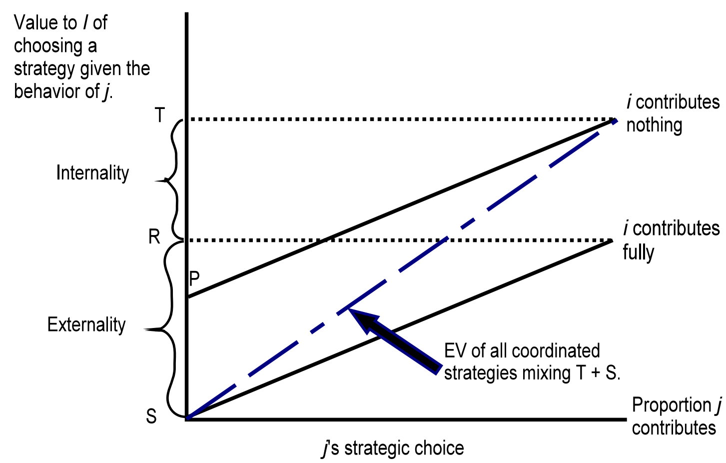

Figure 1: Displaying payoffs associated with extreme strategies in a continuous strategy space

The second property of the PD, that the Nash equilibria be suboptimal, is also satisfied in the game represented in the figure. The dominant strategy line, don't contribute, has, at its start, a value (P). This is the payoff at the left hand side when neither **j** nor **i** contribute. That point is lower than the right hand end of the contribution line (R), which is the payoff associated with both making full contributions. In other words, the value of mutual full contribution (R) is greater than the value of mutual non contribution (P).

7/ Nothing requires that the PD be linear, or that the cost of a contribution to a cooperative project be a constant. An early algebraic representation of the N-PD (Hardin 1971) made this assumption and it has generally been carried on in the literature.

But what of the third property of the PD? What is the analogue of the property in the representation in Figure 1? To get a feel for the extension of inequality (3) to the payoffs associated with strategy choices in a continuous game, recall that starting from mutual contribution, (3) insists that the effect of **j**'s decision not to make a contribution on **i**'s payoffs (R-S) must be larger than the effect of **i**'s defection on her own payoff (T-R). The effect of **j**'s defection from the full contribution of both parties on **i**'s payoff can be read directly from the graph. It is the vertical distance between R and S: (R - S). On the other hand, **i**'s choice to (fully) defect from a joint and full contribution strategy represents a gain of the distance between R and T: (T - R).

It is easy to show that the third traditional requirement of the 2x2 PD (that R-S > T-R) holds in any game which conforms to the assumptions of dominance, the suboptimality of joint defection and linear parallelism (i.e. constant and equal marginal returns).[8] To violate the inequality, the line representing defection must be steeper than the contribution payoff line.[9] But not all steeper lines will do. Only some lines are sufficiently steep to insure that 2R < T+S.

<u>The social optimum</u>: The introduction of a continuous set of strategies raises other substantive questions. Specifically, in the traditional PD, optimality is achieved in a corner solution: it pays for the group to have *everyone* contribute fully. This is insured by the third condition, as specified in (3). When the condition is violated, in the 2x2 PD the Pareto set can be made up exclusively of lotteries over T, S and S, T attainable only via the adoption of coordinated mixed strategies. The analogue of such a set of coordinated mixed strategies in the continuous case is the dashed line ST in Figure 1. There, the Paretian outcome which spreads the risk evenly between the 2 players would be the midpoint of ST.

The conditions which permit coordinated mixed optima, internal to the strategy space, now defined in expected value terms, are easily identified for the continuous 2-PD. Preserving conditions 1 and 2 from above to insure dominance and the Pareto suboptimality of defection, it is specifically the absence of condition 3 that guarantees the existence of an internal optimum for the group. If R-S < T-R, or R < (T+S)/2, the social optimum occurs at less than full and universal levels of contribution. Graphically, this implies that the slope of the 'don't contribute' line is greater than the slope of the contribute line. The two strategy lines cannot be parallel. Further, the optimum would lie on ST. This is not the equivalent of any non-coordinated mix of the two players' strategies. The proofs of these claims is provided in Appendix A. To see, intuitively why condition 3 is needed to guarantee that a full universal contribution is optimal, recall that when it holds, full universal contribution is optimal precisely because the internality is smaller than the externality.

However, the traditional interest in the Prisoners' Dilemma is *not* when the game is played with players who can make binding agreements.[10] Rather, the game is of most interest when such

---

8/ R > P by Pareto Inferiority, and R-S = T-P by parallelism, imply R-S > T-R.

9/ It is easy to see that the defection line can be *less* steep than SR. If it is less steep then T-P < R-S and hence T+S < R+P < 2R.

10/ What an unfortunate use of words: such games are called **cooperative**. Traditional prisoner dilemma games have a direct solution as a cooperative game: the core is RR. When condition 3 is violated the core shifts to a range (as described in footnote 4) of mixes of S and T.

agreements are not achievable.[11] What precisely is required for the choice of such coordinated mixed strategies? In fact, without such binding agreements, the choice of the coordinated mix of outcomes becomes impossible, for by backward induction, consider the last stage: there the player who would be asked to contribute would have an incentive to defect and not contribute. By doing so she could insure a payoff of at least 1, rather than 0. The story from there unravels. So more than simple coordination is required: some sort of punishment is needed for deviation.[12] But without the possibility of enforced cooperation, the best that can be achieved are uncoordinated strategies with some proportions contributed, and the residual withheld.

Expansion #2: The Continuous N-PD

Are there parallel conditions which insure that full and universal contribution, rather than an *interior* solution will be the social optimum for a continuous n-person PD (N-PD)? And if there is an interior optimum, can we characterize some of its more general properties and their implications for empirical situations?

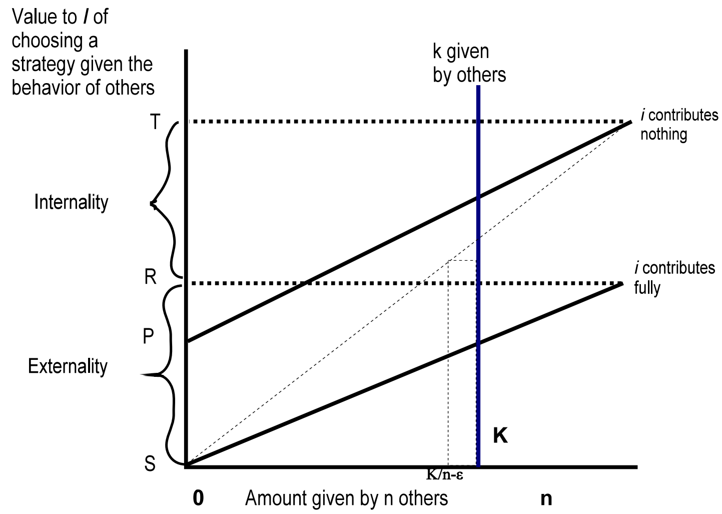

Figure 2 Payoffs to i given aggregate choices of others

To answer these questions, let us sketch the general linear N-PD using the notation used above. Suppose there are n+1 individuals involved in a linear N-PD and each player has 1 unit of resources - any proportion of which can be contributed or not. We represent the payoff to any individual as a function of the aggregate choices of the n others in Figure 2,. There we have labelled the payoffs associated with the four corners (contingencies under which the individual in question contributes all or none of the resources available) with the traditional symbols T, R, P and S. Although the letters T, R, P, and S in the linear N-PD occupy the corners as they do in the continuous 2-PD, they must be somewhat reinterpreted. They represent the payoff to one individual when she allocates all resources to either cooperation or defection and *all* others do one or the other. The first two properties of the 2x2 PD are reflected in the facts that the line representing the payoffs for defection is always above the line for cooperation and that R > P. Traditionally, social welfare, W, is a weighted function of each individual's welfare. To keep the discussion simple, and parallel to that of Schelling (1978, p. 218), we assume symmetry, in resources and tastes, as well as an equal weighting (i.e. utilitarian) of the individuals. Thus armed, we can calculate the group's maximum welfare for a generalized

---

11/ These games are referred to as non-cooperative. See any general text on game theory such as Luce and Raiffa (1957).

12/ Folk theorems covering the sort of incentives which would be able to yield subgame perfect choices which lead to optimality exist. The folk theorems all require considerably more apparatus to be carried through (e.g. a specification of the time horizon, discount rates, etc.). These are discussed well in Fudenberg and Tirole (1991). That such possibilities exist is a note of optimism.

linear N-PD. The social welfare of the group can be expressed as a function of one individual's strategic choice of contribution level along with the aggregated contributions of all other players.

Let us now characterize the situation in which the group optimum is not a corner solution involving full and universal contribution. As a start, let us try to find the conditions under which an interior solution based on symmetric play can do better for the group than a corner solution.

First note that we can calculate the proportion of the group assets which should be contributed to attain a welfare maximum by differentiating the group welfare function, Equation 4, with respect to the proportion allocated to cooperation (k/n).

That maximum will appear the same whether it were calculated from an assumed *equal* contribution by each member of the group or as the *mean* level of contribution needed to support the maximum group welfare.[13] But note that we are not interpreting the results as based on symmetric donations but rather as based on the *mean* level of contribution needed to support the maximum group welfare.

Let us start with the assumption of symmetric levels of contributions and let the total amount contributed by the n others be k. Then the per capita allocation of others is k/n where $0 < k/n < 1$. Now consider the welfare of the group viewed as a function of the strategic choice of the remaining player. Using the symmetry assumption the remaining player's contribution would also be k/n. This yields the following equation for group welfare when each player contributes the same amount:

$$(n+1)\ \{\frac{k}{n}\ [S+\frac{k(R-S)}{n}\ ]\ +\ (1-\frac{k}{n})\ [P+(T-P)\frac{k}{n}]\ \} \qquad \textbf{(4)}\ \text{Group Welfare}$$

Differentiating the welfare function with regard to k/n and setting the derivative equal to zero yields a solution for the optimum (non coordinated) per-capita contribution level:

$$\frac{k}{n} = \frac{2P-(S+T)}{2[(R-S)-(T-P)]} \qquad \textbf{(5)}\ \text{Optimal allocation to cooperation in linear N-PD}$$

Expression (5) specifies the conditions for optimal *symmetric* contributions in the N-PD. To identify the conditions that guarantee an internal optimum simply add the requirement that $0 < k/n < 1$. Algebraically this reduces Expression (5) to the following inequality: R-S < T-R (or 2R < T + S). This exactly mirrors condition (3) in the linear 2-PD. The graphical interpretation is that the slope of the defection line is sufficiently greater than the slope of the cooperation line to support the inequality. Although algebraic expression of the conditions for a non-coordinated symmetric internal optimum in this case parallel those in the continuous 2-PD, the conditions require reinterpretation. We can, however, continue to interpret the result in terms of externalities and internalities. R-S can be seen to be the effect on (i.e. externality) one individual who contributes fully when the rest of the group switches from fully contributing to non contributing. It is the maximum externality the group can impose on a fully contributing individual. T-R, on the other hand, is the increase in benefits (i.e. internality) which the individual can obtain by defecting from a universal contribution situation while the rest of the

---

13/ This is possible because with our symmetry conditions all the individual payoff functions are identical and the group welfare function is utilitarian.

group continues to contributes fully. It is the maximum internality the individual can get by changing strategies.

In the continuous N-PD, R-S < T-R, implies that the maximum externality a fully contributing player can experience from the choices of all others not to contribute (R-S) is less than the maximum internality she can gain under the same circumstances by defecting (T-R). This maximum *externality* occurs when the individuals shift from universal full contribution (i.e. on the cooperation line) to no contribution. The maximum *internality* occurs at the right hand side of the Figure 4$: the contingency under which others cooperate fully and are taken advantage of.

Note again, that with the assumption of symmetry, the first part of the expression (R-S) can be reinterpreted. To see this, notice that the slope of the cooperation line is simply (R-S)/n. Hence an incremental contribution of 1 by some other **j** yields an incremental gain of (R-S)/n for **i**. And since there are n others the total externality of this cooperation on all other players save **j** is simply (R-S). Thus in the N-PD the condition is similar to that in the 2-PD except now the relevant comparison is between one player and all others. For a symmetric optimum to exist *other than that of full cooperation* it must be that for each player 1) the maximum externality she receives from the cooperation of all others be less than the maximum internality obtainable by her defecting; and similarly, 2) assuming an additive welfare measure -- the maximum total externality produced by her actions on others be less than the internality gained by her own choice.[14]

Perhaps surprisingly, however, the internal maxima based on symmetric play are not the best that the group can do! It is not as good as some outcomes available with *asymmetric* play.[15] To

---

14/ There is yet another interpretation of the condition which yields insight into its ethical, and hence behavioral, impact. The condition for a symmetric internal optimum R-S < T-R can be rewritten as:

$$(R-P) < (T-P) - (R-S)$$

This expression has an intuitive interpretation. R-P is the difference to one player of all contributing and no one contributing. T-P is the total externality on one non-contributing player of all others switching from fully contributing to not contributing. R-S is that same externality on a contributing player. Thus, when the difference in the two externalities is larger than the difference between the all-cooperate and all defect payoff to that player, there is an optimum without full cooperation. In other words, if the added reward to one individual of a steep defect (relative to the cooperation) line is enough to outweigh the punishment associated with moving from all cooperate to all defect, a less than full cooperation optimum exists. The limit of these symmetric optima is the situation under which 2R = S + T. In that special case the symmetric optimum is attained when everyone contributes their full endowment: k/n = 1.

15/ This should come as no surprise. After all, consider the discussion of the social optima of the 2 person case above, and their associated footnotes (4 and 10). As in the previous discussion, only within the framework of *cooperative* game theory, can the Paretian outcome be achieved. In this case, we will show that the solution will be for some to contribute fully, and others to withhold any contribution. The payoffs for the withholders will be calculated by the proportion giving (i.e. some k, as in Figure 2,) from the PT function.

see this one has only to delve a bit deeper into the interpretation of the condition permitting an internal solution to be better than the corner solution. Since the strategy lines diverge, the marginal net valuation of a single defection is greater the larger the number of others who have cooperated: with defection being the steeper, the further one moves to the right the greater the value of a defection. Hence, if an optimum level of cooperation is achieved via the fewest number of cooperators, the remaining defectors have the most to gain by defecting. Put another way, for any fixed level of aggregate contributions, the maximum gain to defections is obtained when as few individuals as possible contribute to achieve that level of cooperation. To illustrate the reasoning, note that if all individuals cooperated symmetrically with some proportion (k/n) of their budgets, they would all be at the same abscissa (see Figure 2,). Now if one individual (call her **j**) shifted to contribute ε more while the total contribution remained the same (i.e. others offset this by contributing less), **j** would be facing a total contribution level of others lower than previously. Her welfare would be calculated to the left, at a lower abscissa. But ε to the left the curves are closer together, and the area of welfare she loses is precisely the area between the curves of width ε as we move from k/n. On the other hand, let *i* be a single individual who makes the offsetting decreased contribution. Then *i* gains the area of the slice of width ε of the difference between the height of the curves. Since they are monotonically diverging, the optimum is always when the defection is concentrated in a minimum number of individuals. That is, the optimum calls for outcomes of maximum inequality.

To illustrate, consider a simple 5 person PD (see Figure 2, Figure 3) with an interior symmetric optimum solution of k/n = .8. For our illustration we calculate the value of T using equation (5), and assign the parameters the following values: S = 4, R = 20, P = 10. Solving for T, yields T = 42.67. The game can be characterized as having an optimal defection of one individual. When one individual defects in such a game, the sum of payoffs is about 6% higher than when all individuals contribute to yield the symmetric optimum.

Characteristics: 5 person symmetric PD with budget = 10. Contribution of 1 unit generates .4 units per capita.Optimal symmetric contribution is .8.

To show: Group payoffs are better when one gives 0, others all give 1 rather than when all give .8 symmetrically.

CASE 1: ASYMMETRIC COOPERATION			
Player	**Gives**	**Keeps**	**Payoff**
1	10	0	16
2	10	0	16
3	10	0	16
4	10	0	16
5	0	10	42.67
Group Welfare (sum of payoffs) =			106.67

CASE 2: SYMMETRIC COOPERATION			
Player	**Give**	**Keeps**	**Payoff**
1	8	2	20.026
2	8	2	20.026
3	8	2	20.026
4	8	2	20.026
5	8	2	20.026
Group Welfare (sum of payoffs) =			100.13

Figure 3: Illustration of Prisoner Dilemma with Internal Optima

Characterizing N-PD's: Symmetrically and Asymmetrically Optimal N-PD's

We would characterize all linear, continuous n-person games with dominant defect strategies which satisfy R-S > T-R as **Symmetrically Optimal N-PD's** and all those for which R-S ≤ T-R as **Asymmetrically Optimal N-PD's**. As we have defined them, an N-PD has a set of dominated strategies which, were they chosen, would yield a Pareto optimal outcome. An **Asymmetrically**

The cooperators will receive the corresponding value from the SR function.
Such an outcome is not available without binding agreements.

Optimal N-PD still has dominated strategies, but universal selection of those strategies *fail to yield the social optimum*. Although the symmetric maximum of universal partial contribution is better than full contribution, it is not the global optimum.

An implication of this is worth pointing out. The two classes of N-PD's respond differently to analysis under impartial reasoning. There are different ethical imperatives inherent in these two classes of games.

In a recent paper Frohlich (1992) argued that one can identify the normative imperative associated with a rational self-interested play of the 2x2 PD by doing a type of thought experiment: by playing the game from an impartial point of view. This involves imagining what strategy choice rational self-interested players would make in a stylized condition:

> Imagine each player to be faced with the task of choosing a strategy for one of the places in the game with the knowledge that after the choice, each would be randomly assigned the payoff associated with one of the two places in the game.

This situation requires that the players give equal weight in their calculations to the interests of both parties in the game. It induces impartial reasoning. Elsewhere (Frohlich and Oppenheimer 1992) we (and many others) have argued that decisions taken under conditions which induce impartial reasoning have a claim to ethical validity. In the 2x2 PD the incentive structure facing the individuals when they play the game from an impartial point of view has a dominant solution. From an impartial point of view rational self interested players have a dominant incentive to cooperate. Thus, it was argued, the ethical imperative coming from rationality and self interest in the traditional 2x2 PD is to cooperate.

We might call the new game, obtained by playing the original game from an impartial point of view, the impartiality transform of the game. It was demonstrated in Frohlich 1992 that when condition (3) does not hold in otherwise 2x2 PD's the impartiality transform of the PD does not yield a dominant pure strategy of universal contribution. What we have shown here is that this reasoning can be extended to the continuous 2-PD and the continuous N-PD. Thus in the **Asymmetrically Optimal N-PD** an impartiality transform does not yield an imperative to cooperate fully. Rather, mathematically identifying the non-coordinated maximum yields only 1 solution: the imperative to contribute that amount which, if everyone else contributed the same amount, would yield a constrained maximum. But, as we have seen above, that is not the best the group can do. The Kantian imperative doesn't lead to a Paretian outcome. Rather, the asymmetric solution with a minimum number of individuals contributing to the point of the local maximum and the remainder free riding would yield the social optimum. That would have to be generated by a coordinated strategy of contribution and could not be achieved by individually independent applications of impartial reasoning.

There are obvious implications of these options. A solution such as (5) leaves one in a Pareto inferior position: one which can be bettered by a choice of a mixed coordinated strategy. On the other hand, an asymmetric coordinated strategy to contribute will, in the end, select some to free ride on others who are left "holding the bag." Putting it other ways, equality of actual burdens has its costs, as do Paretian goals when it is suboptimal for all to fully cooperate: the first in the sum of the payoffs, and the second in the lack of equality of payoffs.

We can further complicate the choice situation and follow Schelling (1973) in which each of the n players have only an all or nothing choice about contributing their resource. Such a game can be referred to as a binary N-PD, and we can assume properties (1) and (2) to hold. In this set of games the importance of condition (3) and the distinction between what we would call the

Binary Symmetrically Optimal N-PD and the **Binary Asymmetrically Optimal N-PD** becomes even more important. When choices are binary, the impartiality transform fails completely to yield a dominant solution for the transformed asymmetrically optimal game. It does, however, imply a dominant solution for the transformed symmetrically optimal game. This can be interpreted as meaning that rational self interested individuals facing a symmetrically optimal N-PD face an unambiguous moral tension. They know what is right (the solution of the impartiality transform) and they know what is in their immediate individual interest (the dominant strategy of the game they face). Individuals facing an asymmetrically optimal game do not experience the same clear tension. They still know what is in their immediate individual interest (the dominant strategy of the original game). But the impartiality transform of the game yields no unambiguous moral imperative since it does not indicate whether **they** should contribute - only what **proportion** of individuals should contribute. Indeed, the social optimum **requires** that some individuals **not** contribute!

Finally, to put some flesh on these theoretical bones it may be worthwhile pointing out some empirical examples of possible **Asymmetrically Optimal N-PD's**. If it is granted that the value of a barrel of oil produced by OPEC rises as more and more members withhold their oil from the market, and in addition the average cost of production declines for producers, then the OPEC problem is potentially a game of this sort. The slope of the defect line is steeper than the slope of the cooperate line - reflecting the higher price to be paid by cooperating as oil prices rise due to others' cooperation. A similar example might be provided by a shared fishery. Indeed any common pool problem where the price of the product rises as a function of scarcity and economies of scale obtain, would be subject to the same analysis. Viewed from that perspective, even the archetypical example of the medieval common, might, in some circumstances, be an **Asymmetrically Optimal N-PD**! It may require a far more sophisticated form of coordination to achieve the best outcome than the simple imperative which treats individuals equally.

Appendix A

Proof that if T+S > 2R, the mixed optimal coordinated strategy M is larger than the best mixed proportion strategy H

Consider Figure 4$. The intermediate line represents the payoff to **i** of contributing 50% of her potential contribution. It is the straight line which lies 1/2 of the distance between SR and PT. The dashed line ST, on the other hand, represents the payoffs associated with the mixed coordinated strategy.
To see that this follows from the notions at hand, consider the following conditions.

Suppose T + S > 2R. We can determine whether the midpoint on the line ST (M) lies above the midpoint (H) of the line which indicates the value to *i* of donating 1/2 when *j* does so also.
M can be expressed as follows: M = (S + T)/2

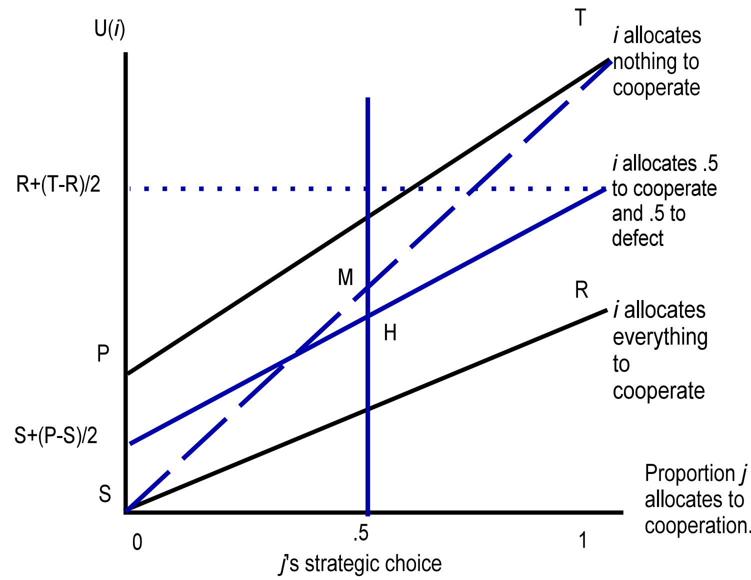

Figure 4 Comparing the values of the mixed coordinated and the mixed-proportion strategies

while H can be expressed as H = [S + P + (R - S + T - P)/2]/2.

Suppose H to be larger than M or:

$$S + P + (R - S + T - P)/2 > S + T.$$

Simplifying yields R + P > S + T. But given that R + P < 2R this contradicts the initial assumption: 2R < S + T, Hence M must be greater than H.

Derivation of optimal mixed proportion strategy

Without loss of generality let each player's resources in a continuous 2-PD be set at 1. Let k be interpreted as the proportion of each player's resources allocated to cooperation. The graph, in Figure 2 can be used to represent the game in which conditions 1 and 2 are satisfied. The equation representing the social welfare W(k) of the two players, (assuming symmetry of utility functions) can be written as:

$$W(k) = \{[S + (R-S)k]k + [P + (T-P)k](1-k)\}*2 \quad \text{(6) social welfare}$$

The quantity in the curlicue brackets is simply the payoff to one individual resulting from a strategy set of both players contributing k of their resources.

The socially optimum amount of cooperation is identified by setting the first derivative of W, in k, equal to 0. The first derivative of (6) is derivable as:

$$dW/dk = S + 2k(R-S) + (T-P) - 2k(T-P) - P \quad \text{(7) dw/dk}$$

Setting (7) equal to 0 and solving for k yields (8):

$$k = \frac{2P-(S+T)}{2[(R-S)-(T-P)]}$$

(8) optimal allocation to cooperation in a linear, 2

The conditions required for an internal social optimum are that the solution for k lie within the range $0 < k < 1$.[16] Imposing that condition and simplifying the expression yields the following condition:

$$T + S > 2R \quad \text{(9) Condition for interior solution}$$

16/ One immediate observation is that no such solution is possible when the lines are parallel (since the denominator vanishes under that condition).

References

Aldrich, J. 1977. Liberal Games, Further Comments on Social Choice and Game Theory. Public Choice 30: 29-34.

Bernholz, P. 1976. Liberalism, Logrolling and Cyclical Group Preferences. Kyklos 29: 26-37.

Brams, S. J. and D. M. Kilgour. 1988. Game Theory and National Security. New York, NY: Basil Blackwell.

Frohlich, N. 1992. An Impartial Reasoning Solution to the Prisoner's Dilemma. Public Choice 74: 447-460.

Frohlich, N., and J. Oppenheimer. 1992. Choosing Justice: An Experimental Approach to Ethical Theory, Berkeley, CA.: University of California Press.

Fudenberg, D. and J. Tirole. 1991. Game Theory. Cambridge, MA: MIT Press.

Hamburger, Henry 1990. Games as Models of Social Phenomena. New York, NY: W. H. Freeman.

Hardin, R. 1971. Collective Action as an Agreeable N-Prisoners' Dilemma. Behavioral Science 16: 472-479.

Iwakura, N. and T. Saijo. 1992. Payoff Information Effects of Public Good Provision in the Voluntary Contribution Mechanism: An Experimental Approach. Paper presented at the annual meeting of the Public Choice Society, New Orleans, LA, March.

Luce, D. and H. Raiffa. 1957. Games and Decisions. New York, NY: Wiley.

Marwell, G. and P. Oliver. 1993. The Critical Mass in Collective Action: A Micro - Social Theory. New York, NY: Cambridge Univ. Press.

Miller, G. 1977. Bureaucratic Compliance as a Game on the Unit Square. Public Choice 29: 39-52.

Miller, N. R. 1977. Social Preference and Game theory: A comment on the Dilemmas of a Paretian Liberal. Public Choice 30: 23-28.

Ostrom, E., J. Walker and R. Gardner. 1992. Covenants With And Without The Sword: Self Governance is Possible. American Political Science Review 86: 404-417.

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

Schelling, T. C. 1978. Micromotives and Macrobehavior. New York: W.W. Norton.

Schelling, T. C. 1973. Hockey Helmets, Concealed Weapons, and Daylight Savings: A Study of Binary Choices with Externalities. Journal of Conflict Resolution 17: 381-428.

Sen, A. K. 1970. The Impossibility of a Paretian Liberal. Journal of Political Economy 78: 152-7.