AD/A-006 188

FORMAL MODELS OF DILEMMAS IN SOCIAL DECISION-MAKING

Robyn M. Dawes

Oregon Research Institute

Prepared for:

Office of Naval Research
Advanced Research Projects Agency

December 1974

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

Unclassified

Security Classification

AD/A-006 188

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Oregon Research Institute
Eugene, Oregon

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

Formal Models of Dilemmas in Social Decision-Making

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Robyn M. Dawes

6. REPORT DATE

December, 1974

7a. TOTAL NO. OF PAGES

31

7b. NO. OF REFS

17

8a. CONTRACT OR GRANT NO.

N00014-75-C-0093

b. PROJECT NO.

NR(170-770)

c. ARPA Order No. 2449

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

Oregon Research Institute Research Bulletin Vol. 14 No. 12

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale. Its distribution is unlimited.

11. SUPPLEMENTARY NOTES

PRICES SUBJECT TO CHANGE

12. SPONSORING MILITARY ACTIVITY

Office of Naval Research
Code 455
Arlington, Virginia 22217

13. ABSTRACT

Social dilemmas occur whenever two conditions are met: (1) each individual is better off personally choosing an anti-social course of action than choosing a pro-social one, yet (2) all individuals would be better off if all were to choose the pro-social course than if all were to choose the anti-social course. That is, a social dilemma is one in which: (1) the anti-social course is a dominating strategy, and (2) the result of its universal adoption is deficient in that everyone would prefer to have everyone choose the pro-social course. Interest in social dilemmas has increased radically in recent years because problems such as overpopulation, pollution, and selfish selfish use of energy appear to be the result of such dilemmas, as are many forms of "ripping off" groups or formal organizations.

Following Hardin's (1968) analysis of the "tragedy of the commons", Dawes (1973) has developed a simple game according to the principle that benefit for anti-social behavior in the game accrues directly to the individual whereas loss (which is greater than benefit) is spread out among all the players. Such a game results in a true social dilemma -- one which becomes worse the more players there are. The present paper proves formally that this game is equivalent to a multi-person "separable" prisoner's dilemma, which in turn is equivalent to a game with payoffs for anti-social and pro-social behavior that are linear functions with equal slopes of the number of people who choose the pro-social action (see Hamburger, 1973; Schelling, 1973). Finally, the "paradoxes" of joining unions or engaging in collective actions -- as discussed by Olsen (1965) and Messick (1972) -- are also shown to involve the same essential structure. Thus the dominating strategy with a deficient equilibrium description of of social dilemmas and the gain-for-self-loss-spread-out are essentially identical.

DD FORM 1473 1 NOV 65

S/N 0101-807-6801

Unclassified

Security Classification

Unclassified

Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Anti-Group Decision						
Commons Dilemma						
Competition						
Cooperation						
Social Decision-making						

DD FORM 1 NOV 65 1473 (BACK)

(PAGE 2)

Unclassified

Security Classification

Formal Models of Dilemmas in Social Decision-making[1]

Robyn M. Dawes

University of Oregon and Oregon Research Institute

Paper presented at

Human Judgment and Decision Processes Symposium
Northern Illinois University
October 16, 1974

To appear in
Human Judgment and Decision Processes:
Formal and Mathematical Approaches
[Kaplan & Schwartz, eds.]

Social dilemmas are easy to invent. Consider a game in which each of three participants must place either a blue poker chip or a red poker chip in an envelope in private. Each participant who places a blue chip in the envelope receives $1.00, and his choice has no effect on the other two participants. Each one who places a red chip in the envelope receives $2.00, and the other two are fined $1.00 each for this choice. (Equivalently, that individual receives $3.00 and then pays his share of a $3.00 fine assessed to the group as a whole.) Which chip should each participant choose? No matter what the other two people do, each is $1.00 better off choosing a red chip; moreover, the choice of the red chip is the only guarantee against losing money. But if all choose the red chip, no one gets anything; while if all had chosen the blue, each would have received $1.00.

Social dilemmas -- which are often described as involving a conflict between "individual rationality" and "group rationality" -- have become of increasing interest to both social scientists and laymen. Overpopulation and pollution are two dramatic examples of particular interest. Mathematically oriented psychologists and sociologists have developed formal models (usually algebraic or geometric) of social dilemmas. This chapter attempts a systematic review and integration of such models; it draws heavily on the work of Hamburger (1973) and Schelling (1973) -- attempting both to integrate their work and to delineate its relationship to a "commons dilemma game" devised by the author. In particular, three dilemma games discussed by other authors and the commons dilemma game are proved to be equivalent.

The simplest social dilemma is one involving two people, the well-known <u>prisoner's dilemma</u>. In the example from which it draws its name, the dilemma concerns two men who are known to have robbed a bank, who have been taken prisoner, but who cannot be convicted without a confession from one or both. The law enforcement people offer each an identical proposition: if you confess and your partner does not, you will go free and he will be sent to jail for ten years; if you both confess you will both be sent to jail for five years, while if neither confesses, I will send both of you to jail for a single year on a lesser charge. Each prisoner is now asked to consider his own best interests in light of what the other may do. If the other confesses, each is better off confessing, for then he will go to jail for five years rather than ten; if the other does not confess, each is still better off confessing, for then he will go free rather than go to jail for a year. Hence, the strategy of confessing is better is both circumstances; it is termed a dominating strategy. Both prisoners would be better off, however, if neither confessed; hence simultaneous choice of the dominating strategies (confession) leads to a deficient equilibrium, a result that is less preferred by both prisoners than is the result that would occur if neither chose his dominating strategy, i.e., if neither confessed. This result is termed in "equilibrium" because neither prisoner is motivated to change his choice given that the other has confessed.

In the game considered at the beginning of this chapter, the dominating strategy is choosing the red chip and the resulting deficient equilibrium is that no one gets anything -- while if all had chosen the blue chip, all would have received a dollar.

In general, a social dilemma may be defined as a situation in which each player has a dominating strategy and in which the choice of dominating strategies results in a deficient equilibrium. This definition may easily be stated formally when each player has a choice between two strategies (or choices of action) and all players have the same payoff structure, one that depends only on the number of people who choose the dominating strategy. [Condition (1) in Schelling's 1973 article]. Although the concept of social dilemma does not require that choice is limited to two alternatives or that all players have the same payoff structure, most formal theoretical work is within this framework.

Consider that each of N players has a choice between two strategies D and C (D for "defecting" and C for "cooperating"). Let D(m) be the player's payoff for a D choice when m players choose C and let C(m) be the payoff for a C choice when m choose C[2]. A <u>social dilemma game</u> is one in which:

(1) $D(m) > C(m+1)$ [Hamburger's Condition P3, Schelling's Condition (2)]

That is, whenever any number m of other people choose C each player is better off choosing D than choosing C and becoming the m + 1st cooperator, and

(2) $C(N) > D(0)$ [Hamburger's Condition P7]

(1) and (2) guarantee that D is a dominating strategy that results in a deficient equilibrium.

Hamburger (1973) has discussed these conditions at length, in relation to other conditions.

Two other aspects of most social dilemmas are that both the individuals in the society and the society as a whole are better off the more people who cooperate. In the present context of two choice games with identical outcome structure across players, these conditions may be expressed as:

(3) $D(m+1) > D(m)$ [Schelling's Condition (3)]

$C(m+1) > C(m)$, and

(4) $(m+1)C(m+1) + (N-m-1)D(m+1) > mC(m) + (N-m)D(m)$ [Hamburger's Condition P12]

Conditions (1) and (2) guarantee only that D is a dominating choice for everyone and that the <u>end result</u> of everyone's choosing D is deficient.

They do not in and of themselves imply conditions (3) and (4). In fact, as will be demonstrated shortly, games can satisfy (1), (2), and (3) but not (4) or (1), (2), and (4) but not (3).

Two person prisoner's dilemmas do necessarily satisfy condition (3) because by conditions (1) and (2), $D(0) > C(1)$, $D(1) > C(2)$, and $C(2) > D(0)$; it follows that $C(2) > C(1)$ and $D(1) > D(0)$. They do not, however, necessarily satisfy condition (4).[3]

As shown by Schelling (1973), two choice games can be simply and neatly characterized by graphing $D(m)$ and $C(m)$ as a function of m, an empirical demonstration appearing in Kelly and Grzelak (1972). Condition (1) is then that the curve for D at point m must always lie above that for C at point m+1. (There is occasionally some confusion here; it is not enough that the curve for D simply dominate that for C; rather it is at the point at which a player may choose to become the m + 1st cooperator that $D(m)$ must dominate.) Condition (2) is that the end point on the C curve must be higher than the 0 point on the D curve. Condition (3) stipulates that both curves must be monotone, and condition (4) involves a rather complex averaging property. An example of C and D curves satisfying conditions (1) through (4) is given in Figure 1. (Note that it is necessary to specify some metric on the absyssa in order to insure that condition (1) is satisfied.)

Insert Figure 1 about here

Figure 2 represents a game in which conditions (1), (2) and (3) are met but (4) is not.

Insert Figure 2 about here

4N

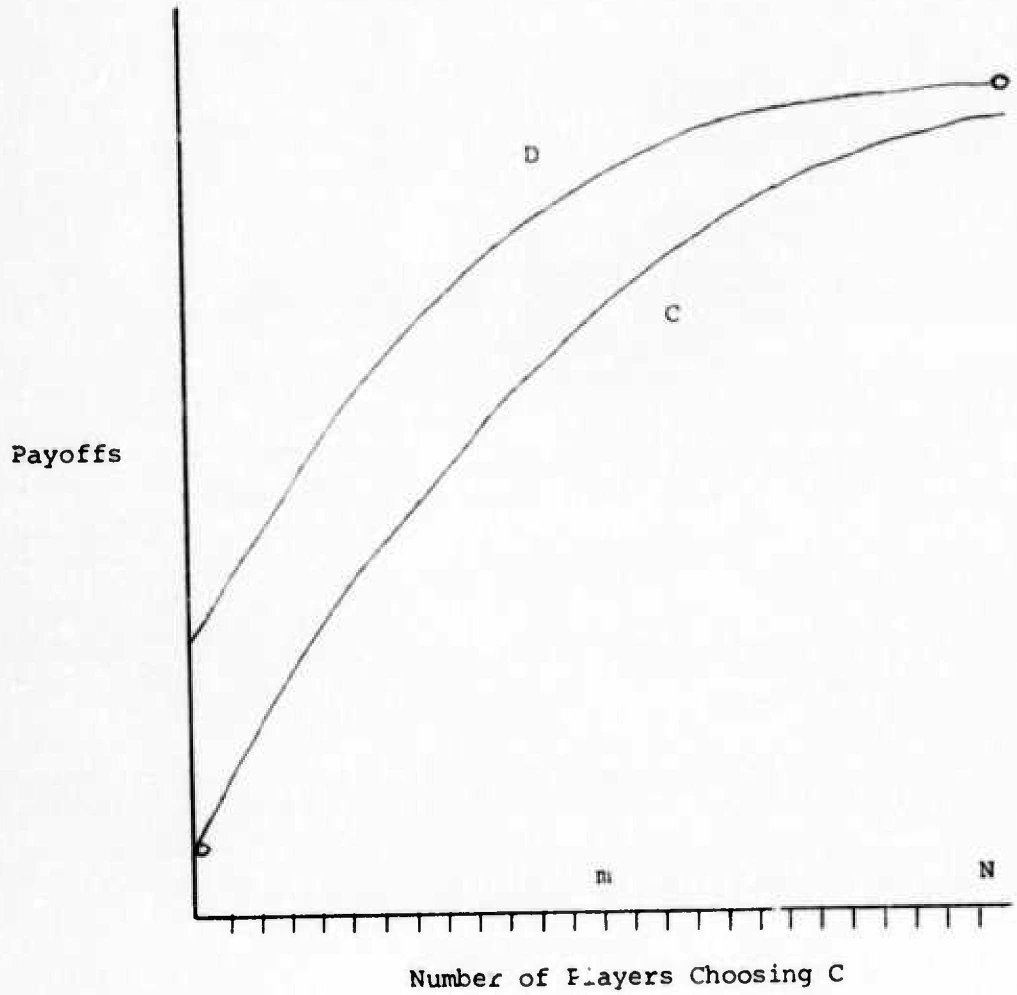


Figure 1. A Social Dilemma Game Satisfying Conditions (1) - (4)

4-1

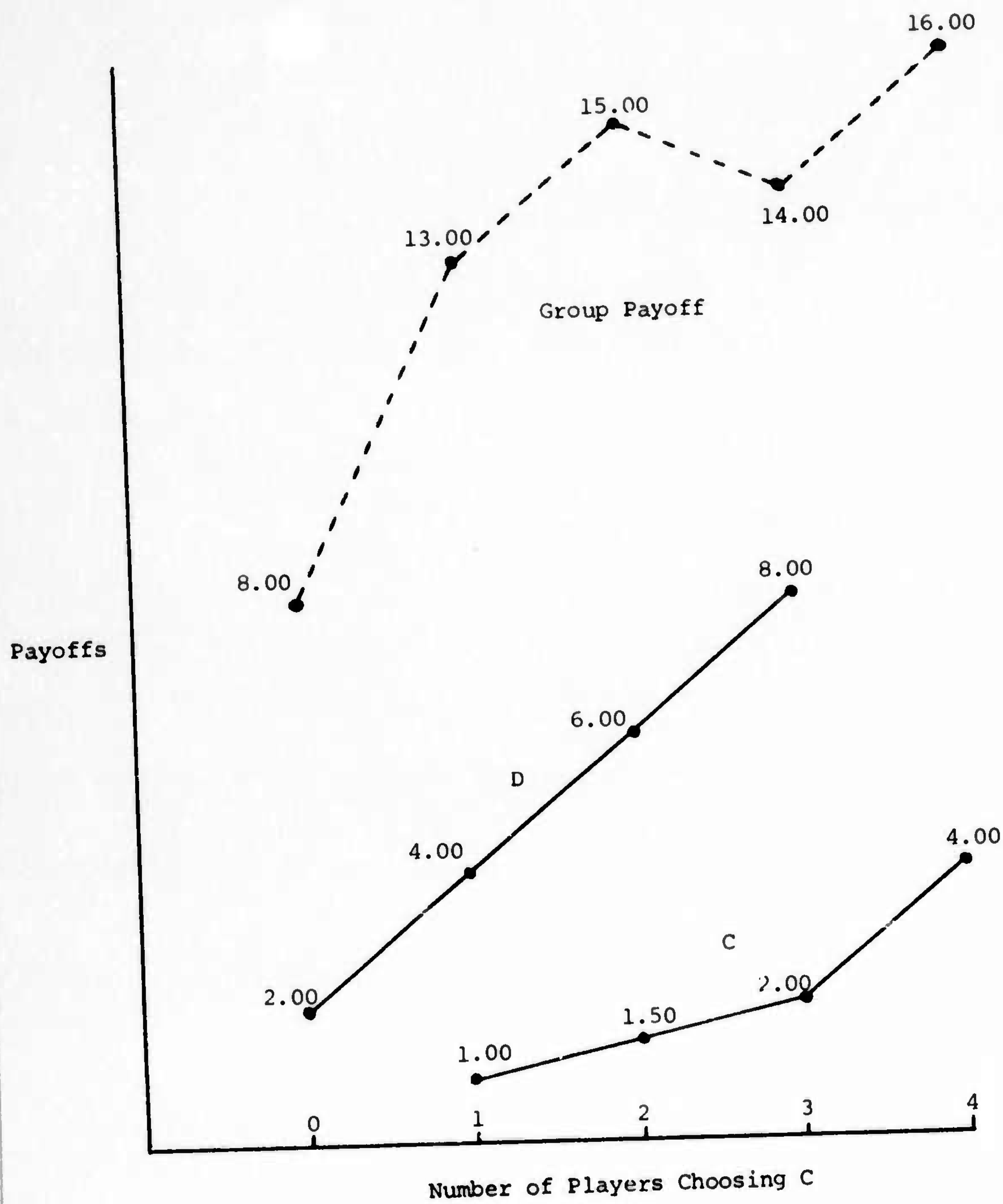


Figure 2. A Four Person Social Dilemma Game Satisfying Conditions (1), (2), and (3) But Not (4)

In contrast, condition (4) implies condition (2).

proof: by condition (4), society as a whole is better off if one player chooses C than if none do. That is, $C(1) + (N-1)D(1) > ND(0)$. Again by condition (4), society is better off if two players choose C than if one does. That is, $2C(2) + (n-2)D(2) > C(1) + (N-1)D(1)$. Iterating and combining inequalities yields $NC(N) > ND(0)$, which reduces to (2) by dividing by N.

Figure 3 represents a game in which conditions (1), (2), and (4) are met but (3) is not.

Insert Figure 3 about here

One type of social dilemma game of particular interest is that which generalizes a two person separable prisoner's dilemma. A prisoner's dilemma is defined as separable if and only if:

(5) $D(1) - C(2) = D(0) - C(1)$ [This is a restriction of Hamburger's Condition P9 to a situation of identical payoff structure for both players]

That is, the increment for defection is constant whether the other player cooperates (in which case the player receives D(1) for defecting and C(2) for cooperating) or defects (in which case the player receives D(0) or C(1)).

The origin of the term "separable" comes from Evans and Crumbaugh (1966), Pruitt (1967) and Messick and McClintock (1968), who independently noted that when condition (5) is satisfied, each player's choice of C or D may be conceptualized as choosing between the two options:

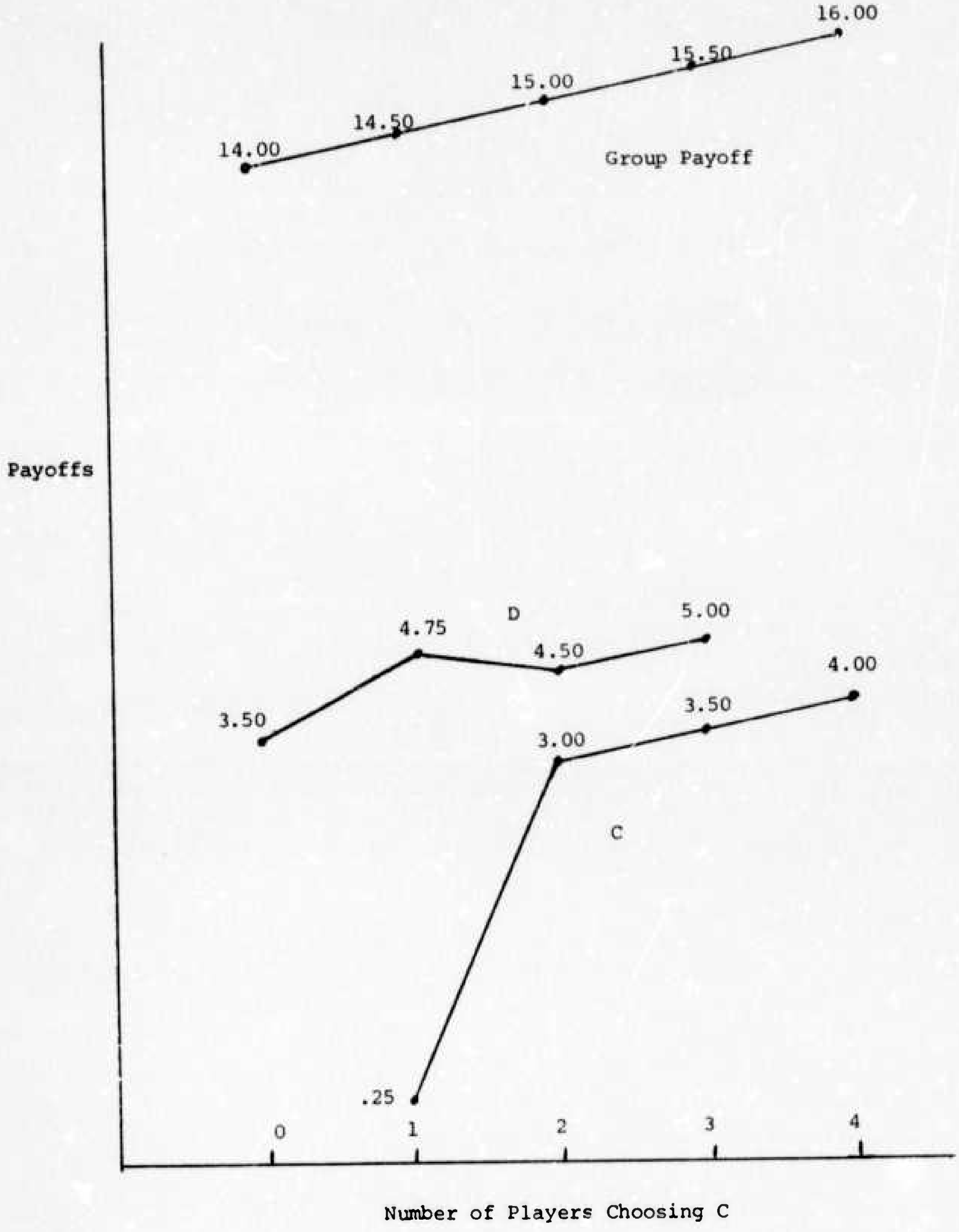


Figure 3. A Four Person Social Dilemma Game Satisfying Conditions (1), (2), and (4) But Not (3)

C: give the other player C(2), give me nothing

D: give the other player C(1), give me D(0) - C(1)

The outcome is the result of these two separable options if and only if:

$$(5') \quad D(1) = D(0) - C(1) + C(2),$$

which is just a restatement of condition (5), i.e., if and only if the payoff to a single defecting player can be expressed as the sum of D(0) - C(1) from his or her own choice and C(2) from that of the other player. Clearly it is also true that is both players choose C both get C(2) (as a result of the other's choice); if both choose D both receive D(0) [D(0) - C(1) as a result of their own choice and C(1) as a result of the other's], and a single cooperating player gets only C(1) (from the defector's choice).

Consider, for example, the separable two person prisoner's dilemma game in which D(1) = 9, C(2) = 6, D(0) = 3, and C(1) = 0. A C choice may be conceptualized as having the experimenter give 6 to the other player, a D choice as having the experimenter give the chooser 3 and the other player nothing. If both choose D, both get 3; if one chooses D and the other chooses C, the D chooser gets 9 and the other player 0; if both choose C both get 6. The term separable refers to the fact that each choice may be conceptualized as yielding one payoff for the chooser and another for the other player in such a way that the final payoffs are simply the sum of these payoffs. If, for example, D(1) $\neq$ 9 but C(2), D(0) and C(1) were still 6, 3, and 0 respectively, the game could not be separated in the above manner.

Condition (5) can also be restated as:

$$(5'') \quad D(1) - D(0) = C(2) - C(1),$$

which implies that the graph of the game consists of two straight lines of equal slope, as illustrated in figure 4(a). Figure 4(b) is a graph of a generalization to an N person game in which C(m) and D(m) are linear functions of m with equal slopes.

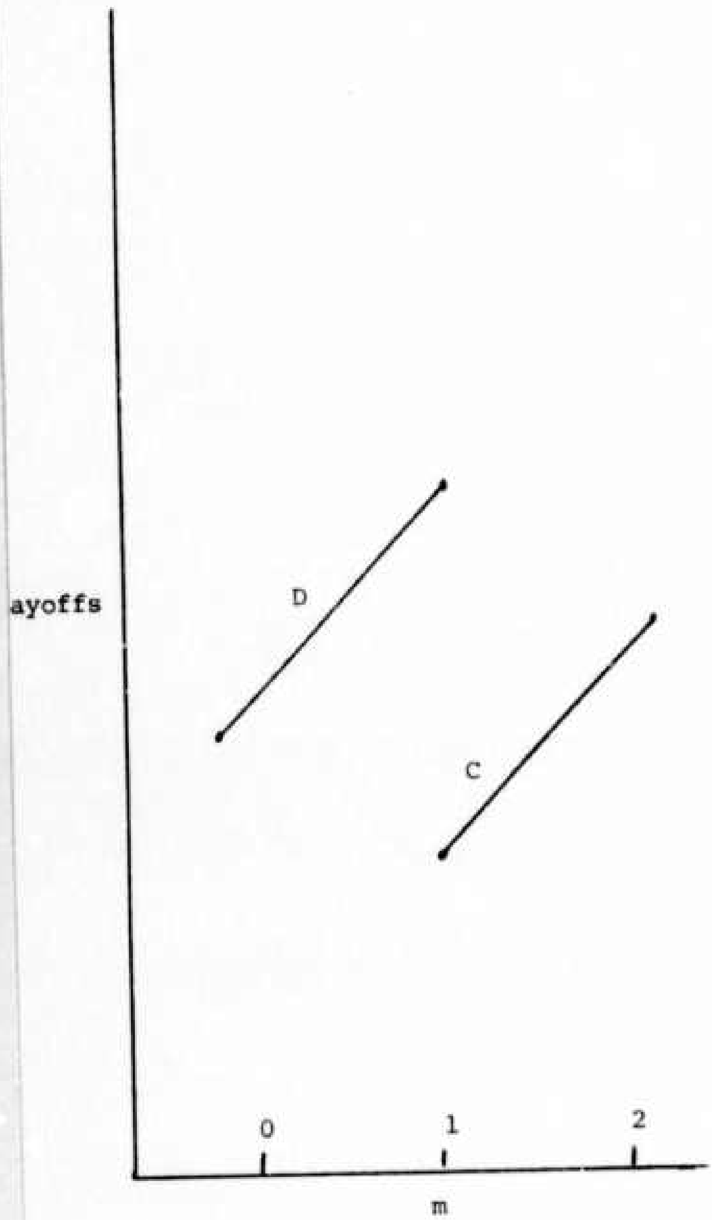
Insert Figure 4 about here

Hamburger (1973, p. 38) has proved that games characterized by a graph in which C(m) and D(m) are linear functions correspond to simultaneous prisoner's dilemma games in which each of the N players plays against each of the N-1 others. The payoffs for each of these pairwise prisoner's dilemmas are D(0), D(1), C(1) and C(2) (subject to the usual constraints that $D(0) > C(1)$, $D(1) > C(2)$ and $C(2) > D(0)$), and the equations for C(m) and D(m) are given by:

$$C(m) = [C(2) - C(1)]m + C(1)N - C(2) \tag{6}$$

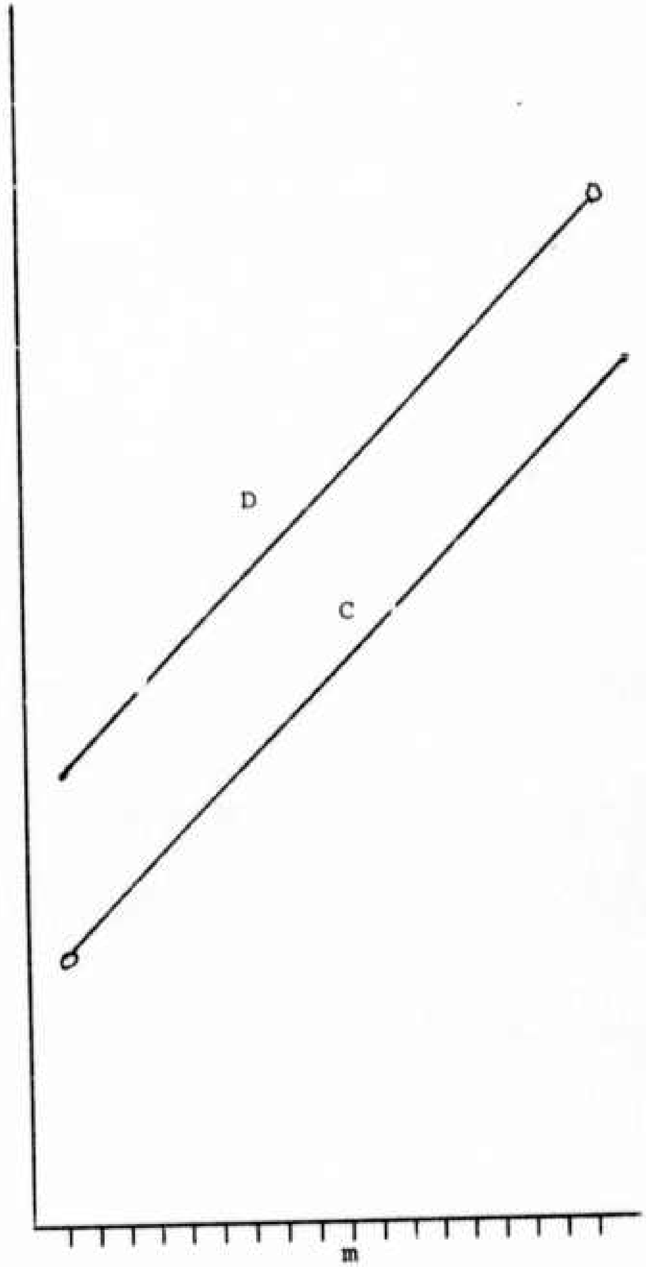
$$D(m) = [D(1) - D(0)]m + D(0)[N-1]$$

The proof is straightforward. First, each individual who cooperates when m-1 others also cooperate receives C(2) for those games and C(1) for the remaining $(N-1) - (m-1) = N-m$ games. Hence, that individual's payoff is $(m-1)C(2) + (N-m)C(1) = [C(2) - C(1)]m + C(1)N - C(2)$. Similarly, each individual who defects when m others cooperate receives D(1) for those m games and D(0) for the remaining N-m-1; i.e., he or she receives $(m)D(1) + (N-m-1)D(0) = [D(1) - D(0)]m + D(0)[N-1]$. Conversely, if α is the intercept of the C(m) function and β the slope, it is possible to solve for C(1) and C(2) in the first part of (6). Specifically, $C(1) = (\alpha+\beta)/(N-1)$ and $C(2) = (\alpha+N\beta)/(N-1)$. Similarly, if γ is the intercept of D(m) and δ its slope, $D(0) = \gamma/(N-1)$ and $D(1) = [\gamma + (N-1)\delta]/(N-1)$. Q.E.D.



Number of Players Choosing C

Figure 4 (a)



Number of Players Choosing C

Figure 4 (b)

Note that the relationship between games as defined by the graphs and as defined by the pairwise prisoner's dilemmas is not independent of N; that is for any pairwise structure there is a different graph depending on N and for graphs with different values of N there are different values of D(0), D(1), C(1), and C(2) in the pairwise games. Note also that such games satisfy condition (3) (trivially, since linear functions are monotone). They need not, however, satisfy condition (4).

The linear functions C(m) and D(m) will have the same slope if and only if C(2) - C(1) = D(1) - D(0), i.e., if and only if the pairwise games are separable. It will be proved later that games in which C(m) and D(m) are linear functions with the same slope satisfy condition (4), (a result implicit in Theorem 2, p. 34 of Hamburger).

An essential equivalence has now been established. Games described by graphs in which C(m) and D(m) are linear functions with equal slopes are identical to games in which each player simultaneously plays separable prisoner's dilemmas with each of the remaining N-1 players. (This equivalence has previously been proved by Hamburger, but it is reiterated here with slightly different proof and terminology because of its importance in what follows.)

Another approach to N person social dilemmas has been taken by Dawes (1973), who proposed a simple algebraic structure for the commons dilemma as expounded by Hardin (1968). (This dilemma is based on a somewhat minor point made by Lloyd in 1833 in an essay on population; its exposition and development are due mainly to Hardin.) In the example from which it draws its name, each of 10 people owns one 1,000 lb. bull and all 10 bulls graze upon a common pasture that is capable of sustaining them all. The introduction of an additional bull would result in the weight of each bull decreasing to 900 lbs.; that is, with the introduction of an additional bull the pasture could support only 9,900 lbs. of cattle rather than 10,000. Any individual who introduces an additional bull has increased his wealth by 800 lbs., because he now has two 900-lb. bulls rather than only 1,000-lb. bull. But the total wealth has been reduced by 100 lbs., as has the wealth of each of the other individuals.

This commons dilemma, gain to self with loss shared by everyone, is ubiquitous -- especially in large societies. In its most dramatic form, it may cause each single soldier to flee from a battle, because each reasons that his own participation makes little difference in the final outcome, yet it makes a great difference to him personally, and he thereby ensures rout and disaster for all the soldiers -- including himself (unless the soldiers on the other side are equally rational!). In a milder form, it may result in an academician's securing a job offer from another institution solely to achieve a better salary at his or her own institution. If he or she is successful, colleagues will, of course, suffer through restrictions of funds available to grant them raises, although the adverse effect on each individually will be quite small in a large institution. An intermediate form of the dilemma may be found in people's decisions to obtain unrealistically high payoffs from insurance companies because "after all, the company can afford it" (with the result that everyone's premiums skyrocket). Even the decision to have children may be regarded as involving a commons dilemma (Dawes, Delay and Chaplin, 1974, p. 3). "With the world as our commons, each of us may believe he stands to gain (fulfillment, 'eternal life', companionship and perhaps wealth) by having children, while the loss of each 'consumatory and polluting agent' to the commons is clearly distributed among all the living creatures in it, and particularly the other people. That this one type of pollution may underlie most other pollution problems makes the study and resolution of the class of such problems particularly timely."

These commons dilemmas all clearly involve two principles:

(A) gain for defection accrues directly to self

(B) loss, which is greater than gain, is spread out among all the members of the group (e.g., commons, or society, or world).[4]

Again within the context that each player has a choice between two actions and each has the same payoff structure dependent only on the number of cooperators and defectors, Dawes (1973) has defined the <u>commons dilemma game</u> as follows.

(i) each player who chooses D rather than C has his payoff incremented by an amount $d > 0$ above the payoff $C(N)$ for total cooperation.

(ii) players are collectively fined $d + \lambda$ ($\lambda > 0$) for each choice of D, each player's share of the fine being $(d + \lambda)/N$.

(iii) $d > \frac{\lambda}{N-1}$

Condition (iii) simply guarantees that the individual's increment for defection is not so small that it is offset by his or her share of the fine.

<u>Theorem 1: The commons dilemma game as defined by conditions (i) - (iii) satisfies conditions (1) - (4).</u>

Condition (1): $D(m) = C(N) + d - \frac{(N-m)(d+\lambda)}{N}$, while $C(m+1) = C(N) - \frac{(N-m-1)(d+\lambda)}{N}$. Hence, $D(m) - C(m+1) = d - \frac{d+\lambda}{N} = \frac{(N-1)d-\lambda}{N}$, which is greater than 0 by condition (iii). Note that $D(m) - C(m+1)$ is independent of m.

Condition (2): $D(0) = C(N) + d - \frac{N(d+\lambda)}{N} = C(N) - \lambda < C(N)$

Condition (3): $C(m+1) = C(N) - \frac{(N-m-1)(d+\lambda)}{N} > C(N) - \frac{(N-m)(d+\lambda)}{N} = C(m)$

$$D(m+1) = C(N) + d - \frac{(N-m-1)(d+\lambda)}{N} > C(N) + d - \frac{(N-m)(d+\lambda)}{N} = D(m)$$

Condition (4): Each choice of D decreases the outcome for the players as a whole by an amount λ. Q.E.D.

For example, the game proposed at the beginning of this chapter is a commons dilemma game in which $C(N) = \$1$, $d = \$2$, and $\lambda = \$1$.

The following theorem establishes that the commons dilemma game is identical to the two equivalent ones described earlier.

Theorem 2: Commons dilemma games, games described by graphs in which C(m) and D(m) are linear functions with equal slopes, and games in which each player simultaneously plays separable prisoner's dilemma games with each of the N-1 remaining players are all identical.

proof: Given the previous equivalence it is necessary only to establish the identity of commons dilemma games and those described by graphs in which C(m) and D(m) are linear functions with equal slopes.

$$C(m) = C(N) - \frac{(N-m)(d+\lambda)}{N} = \left(\frac{d+\lambda}{N}\right) m + [C(N) - (d+\lambda)]$$

$$D(m) = C(N) + d - \frac{(N-m)(d+\lambda)}{N} = \left(\frac{d+\lambda}{N}\right) m + [C(N) - \lambda]$$

which shows that C(m) and D(m) are linear functions with equal slopes. Conversely, if β is the slope of C(m) and D(m), α is the intercept of C(m) and γ is the intercept of D(m), it is possible to solve for d, λ, and C(N). Specifically, $d = \gamma - \alpha$, $\lambda = N\beta + \alpha - \gamma$, and $C(N) = N\beta + \alpha$. Q.E.D.

Corollary 2.1. Since the commons dilemma game satisfies condition (4), the other two do as well.

The relationships between the parameters of the three equivalent social dilemma games are outlined in Table 1.

Insert Table 1 about here

The commons dilemma game has a property not found in the other two. Even though it is strictly equivalent for any value of N, variation of N defines a whole additional dimension. Thus, while each commons dilemma game with a given N may be conceptualized as a game whose graph consists of linear functions with equal slopes, the entire class of commons dilemma games with d, λ, and C(N) fixed but N allowed to vary may be conceptualized as a graph consisting of planes in 3-space -- the dimensions being m, N and the resulting values of C(m) and D(m).

Moreover, the class of commons dilemma games formed by fixing d, λ, and C(N) and letting N vary has the property that the degree to which D(m) dominates C(m+1) increases as a function of N. That is,

(7) $[D(m) - C(m+1)] \uparrow N$

proof: As pointed out in the first part of Theorem 1,

$D(m) - C(m+1) = d - \frac{(d+\lambda)}{N}$. Q.E.D.

How can the commons dilemma game have property (7) given that the difference in intercepts of D(m) and C(m) is always d?

The answer is that the slope of both functions, $\frac{d+\lambda}{N}$, decreases with increasing N. (This reason sounds a bit "paradoxical" at first, but a few moments thought will reveal that for any given intercept difference, the smaller the slope, the larger the difference between D(m) and C(m+1).)

Table 1

Graph Parameters	Pairwise Prisoners' Dilemma Parameters	Commons Dilemma Parameters
α	$C(1)N - C(2)$	$C(N) - (d+\lambda)$
β	$C(2) - C(1) = D(1) - D(0)$	$(d+\lambda)/N$
γ	$D(0)\ [N - 1]$	$C(N) - \lambda$
δ	$C(2) - C(1) = D(1) - D(0)$	$(d+\lambda)/N$
$(\alpha+N\beta)/(N-1)$	$C(2)$	$C(N)/(N-1)$
$(\alpha+\beta)/(N-1)$	$C(1)$	$C(N)/(N-1) - (d+\lambda)/N$
$[\gamma + (N-1)\delta]/(N-1)$	$D(1)$	$(d+\lambda)/N + [C(N)-\lambda]/(N-1)$
$\gamma/(N-1)$	$D(0)$	$(C(N)-\lambda)/(N-1)$
$N\beta + \alpha$	$(N-1)C(2)$	$C(N)$
$\gamma - \alpha$	$N[D(0) - C(1)] - C(2) - D(0)$	d
$N\beta + \alpha - \gamma$	$(N-1)\ [C(2)\quad D(0)]$	λ

*Note that throughout $\beta = \delta$. Further, given that $D(1) - D(0) = D(2) - C(1)$ there are only three free parameters in each game.

Property (7) is considered crucial to many people analyzing real-world commons dilemmas -- particularly Hardin (1972). The more people among whom the bad consequences of defecting behavior is spread out, the less each individual "suffers the consequences" of his or her own defection.

Another specific game of some interest is Messick's union game (Messick, 1973). This game is defined by the following three conditions:

(a) Each member of a potential union of size N must pay a fixed cost c to join.

(b) If the union succeeds in its goal each member of the potential union (not just each member who pays the cost c to join) receives a prize P, otherwise nothing.

(c) The probability that the union succeeds in its goal is equal to the number of members of the potential union who join (and pay c) divided by N.

Suppose, Messick reasons, m other people have joined the union. The expected value of joining the union when m others have joined is equal to:

$$\left(\frac{m+1}{N}\right) P - c$$

The expected value of not joining is equal to:

$$\left(\frac{m}{N}\right) P$$

An expected value maximizer will then join if and only if

$\left(\frac{m+1}{N}\right) P - c - \left(\frac{m}{N}\right) P > 0$, that is if and only if

$\frac{P}{N} - c > 0$ [equivalently $P/N > c$ or $P/c > N$]

Note that this result does not depend on m.

Now, let us reformulate the problem. Let c be regarded as the amount saved by not joining the union (i.e., a defecting payoff) and let $\left(\frac{m+1}{N}\right) NP - \left(\frac{m}{N}\right) NP = P$ be the expected loss to all the potential union members together from each defection. Thus, c is identified with d in the commons dilemma game, and P with $d + \lambda$. Hence, the expected value maximizer will join if and only if:

$\frac{d + \lambda}{N} - d > 0$, i.e., the player will <u>refuse</u> to join if and only if

$d - \frac{d + \lambda}{N} > 0$, i.e., if and only if

$$d > \frac{\lambda}{N - 1} \quad ,$$

which is condition (iii) of the commons dilemma game. That is, condition (iii) guarantees that the result of joining or not on the basis of maximizing expected value results in no joining -- which establishes a dilemma, because if all joined all would receive $P - c = d + \lambda - d = \lambda$, whereas if none joined none would receive anything.

The following theorem has been established.

<u>Theorem 3: The Messick union game results in a social dilemma for expected value maximizers if and only if it is equivalent to a commons dilemma game (hence equivalent to a game whose graph consists of linear functions C(m) and D(m) with equal slopes, hence equivalent to simultaneous separable prisoner's dilemmas in which each player plays against the N-1 remaining ones)</u>.

Corollary 3.1. <u>If the Messick union game results in a social dilemma for expected value maximizers it satisfies conditions (3), (4), and (7).</u>

Actually, conditions (3) and (4) are immediate, and condition (7) can be derived easily from Messick's formulation. Messick himself, who is concerned with when it is his game results in a dilemma for expected value maximizers, points out that when P and c are held constant some N is reached at which a dilemma occurs (1973, pg. 148).

Olsen (1965) has made a similar argument both with respect to the difficulty of getting laborers to join a union in an "open shop" situation and with respect to the difficulty of getting people to contribute to a public good or venture when a large number of contributions is necessary for success. The logic of Olsen's argument is essentially the same as that of Messick's. The main difference in mathematical development is that Olsen proceeds from differential equations (1965 pgs. 24-28) and hence considers a larger class of possible functions for determining whether or not an individual should join a union or contribute to a public effort. (Toward the end of his paper, Messick also broadens his scope -- by considering probabilities of union success that are monotonic in m but not necessarily linear.) Moreover, Olsen supports his argument with examples from the history of the labor union movement. The relative importance and influence of Olsen's work far outweigh its relative space in this chapter.

Messick and Olsen reach the same conclusions -- especially with respect to the importance of N. Frolich and Oppenheimer (1970) have challenged the idea that the type of social dilemma discussed by Olsen and others (and outlined in this chapter) necessarily becomes more acute as N increases. They argue that the probability of failing by exactly k units of effort (e.g., contributions) should be unrelated to N -- unless certain assumptions are made about "how subjective probabilities vary from situation to situation [1970, pg. 113]." Such an assumption is explicit in Messick's union model and is certainly reasonable in the contexts discussed by Olsen. The probability of failing by exactly k units should decrease with N. Isn't it reasonable to assume, for example, that a candidate for city council has a higher probability of failing by three votes than does a candidate for mayor of the city, who in turn has a higher probability of failing by three votes than does the candidate for governor of the state? Voters are clearly reasonable in assuming the contribution of their vote has less effect on the probability of victory for their favorite gubinatorial candidate than on the probability of victory for their favorite city council candidate. Hence, as Messick and Olsen argue, granted a certain amount of negative value involved in bothering to go to the poll, the expected value of voting for a city council candidate should be greater than that of voting for a gubinatorial candidate if the success of each candidate is equally valued.

The four equivalent games (hence single game) described above are (is) rather restricted. The way in which various constraints can be relaxed (hence the game generalized) can best be seen by considering the graph of the functions C(m) and D(m). First, these functions can remain linear but not have equal slopes; if so, the game is equivalent to one in which each player is engaged in a <u>nonseparable</u> prisoner's dilemma game with each of the N-1 remaining players. Monotone but nonlinear functions can describe social dilemmas which cannot correspond to pairwise prisoner's dilemmas. And then, of course, it is possible to consider the functions that do not satisfy one or both of the social dilemma conditions [(1) and (2)], functions which described games that lie beyond the scope of this chapter. Schelling (1973) has described a wide variety of such functions.

Also, it is possible to relax the assumption that the payoff structure is the same for all players. If such a relaxation is made, it is necessary to examine the game in some detail to see whether in fact it constitutes a social dilemma. For example, some players may profit so much by engaging in a defecting strategy and pay so little of the penalty that the resulting equilibrium is not deficient -- i.e., it benefits them although it hurts others severely. Such player may be analagous, for example to industries that share the dirty air they create with the rest of us but profit much more greatly per unit of pollution they create than we could by creating the same unit. Or perhaps their unit profit is the same but they pay the same fraction of the price as do the other "players," despite owning more units.

In general, it is possible to create a wide variety of N-person social dilemma games; all that must be guaranteed is that conditions (1) and (2) are met. The game discussed in the bulk of this chapter hopefully captures many characteristics of the real-world social dilemmas that motivate the study of experimental dilemmas; for example, conditions (3) and (4) seem ubiquitous in these real-world dilemmas -- as does condition (7) when size varies. Moreover, the game may be presented in a variety of manners: in terms of the graph of the payoff function for C(m) and D(m), in terms of the prisoner's dilemma, or in terms of the gain-for-self-loss-spread-out principle. (Whether different presentations result in different behaviors is an empirical question which may be of interest at least to propagandists.) As noted in a recent Western Psychological Association paper by Goehring, "a parsimonious representation of the N-player prisoner dilemma game matrix is possible if restrictions are imposed upon payoffs such that the incentive for defection and the payoff decrement incurred by individual players per player choosing his defection strategy are constant.

values independent of player identifications and of the distribution of player choices." These characteristics are precisely those of identity of payoff structure and of the independence of D(m) - C(m+1) of m, which of course guarantees that if C(m) and D(m) are linear functions then their slopes are equal.

A final note. As Amnon Rapoport (1967) has so persuasively argued in the context of prisoner's dilemma games, a social dilemma game that is repeated (iterated) may not constitute a dilemma at all. If there is the possibility of "tacit collusion" (pg. 140) or "that each player believes that his decision at Time t-u can partly effect what will happen at Time t" (pg. 141), then it is no longer clear that defection is a dominating strategy. In fact, the situation can become horribly complicated -- even more complicated than envisioned in Rapoport's "optimal strategies." We have a situation in which people are attempting to control the future behavior of others by dispensing rewards and punishments which simultaneously determine -- in a complex interactive way -- their own present rewards and punishments. It should not be surprising that few if any simple generalizations about "cooperative" or "competitive" behavior have arisen from studying people faced with such a complicated task, despite literally thousands of attempts to do so. In contrast, the social dilemma games discussed in this chapter do not involve iteration. They face the subject with a rather simple though compelling dilemma.[5] Perhaps subjects' behavior in these game situations -- and the effect of variables such as communication and humanization -- can shed some light on behavior in the real-world dilemmas the games were constructed to represent.

References

Dawes, R.M. The commons dilemma game: An n-person mixed-motive game with a dominating strategy for defection. Oregon Research Institute Research Bulletin, 1973, Vol. 13, No. 2.

Dawes, R.M., Delay, J., & Chaplin, W. The decision to pollute. Journal of Environment and Planning, 1974, 6, 3-10.

Evans, G.W. & Crumbaugh, C.M. Effect of prisoner's dilemma format on cooperative behavior. Journal of Personality and Social Behavior, 1966, 3, 486-488.

Frolich, N. & Oppenheimer, J.A. I get by with a little help from my friends. World Politics, 1970, 23, 104-120.

Goehring, D.J. Number of players and matrix competitiveness in prisoner's dilemma. Paper presented at the meeting of the Western Psychological Association, San Francisco, April 1974.

Hamburger, H. N-person prisoner's dilemmas. Journal of Mathematical Sociology, 1973, 3, 27-48.

Hardin, G. The trage of the commons. Science, 1968, 162, 1243-1248.

Hardin, G. Exploring new ethics for survival: The voyage of the spaceship beagle. Baltimore, Md.: Viking Press, 1972.

Kelley, H.H., & Grzelak, J. Conflict between individual and common interest in an n-person relationship. Journal of Personality and Social Psychology, 1972,21, 190-197.

Lloyd, W.F. Two lectures on the checks to population. Oxford University Press, Oxford, 1833.

Messick, D. M. & McClintock. Motivational bases of choice in experimental games. Journal of Experimental Social Psychology, 1968, 4, 1-25.

Messick, D. M. To join or not to join: An approach to the unionization decision. Organizational Behavior and Human Performance, 1973, 10, 145-156.

Olsen, M. The logic of collective action. Cambridge, Mass: Harvard University Press, 1965.

Pruitt, D. G. Reward structure and cooperation: The decomposed prisoner's dilemma game. Journal of Personality and Social Psychology, 1967, 7, 21-27.

Rapoport, Amnon. Optimal policies for the Prisoner's Dilemma. Psychological Review, 1967, 74, 136-148.

Rapoport, Anatol & Chammah, A. M. Prisoner's dilemma: A study in conflict and cooperation. Ann Arbor: University of Michigan Press, 1965.

Schelling, T. S. Hockey helmets, concealed weapons, and daylight savings time: A study of binary choices with externalities. Journal of Conflict Resolution, 1973, 17, 381-428.

Footnotes

1. This research was supported by the Advanced Research Projects Agency of the Department of Defense (ARPA Order No. 2449), and was monitored by ONR under Contract No. N00014-75-C-0093. I have received many valuable criticisms of earlier drafts of this paper. I would like to thank particularly: Baruch Fischhoff, Lita Furby, Sundra Gregory, Paul Hoffman, Len Rorer, Mick Rothbart, and Harriet Shaklee.

2. The m refers to the <u>number</u> of players who choose C, not to a particular set of m players. When m players choose C (i.e. cooperate), $N-m$ choose D (i.e. defect). Payoffs could be expressed in terms of the number of defectors rather than the number of cooperators -- and such a choice has seemed more "natural" to many readers of earlier versions of this paper -- but number of cooperators has been chosen in order to be consistent with past authors.

3. Some theorists, for example Rapoport and Chammah in their classic book on prisoners' dilemmas, require that $2C(2) > C(1) + D(1)$, -- in which case condition (4) is satisfied. The reason for this requirement is that the outcome yielding $C(1)$ and $D(1)$ may be preferable to that yielding $C(2)$ to each player if: (i) the subjects are permitted to redistribute the payoffs after the game, or (ii) the subjects may play the game many times and alternate who gets the $C(1)$ payoff and who gets the $D(1)$ payoff. Neither possibility is considered in this chapter; hence, this inequality is not used in the definition of a prisoners' dilemma.

4. If the loss to society as a whole did not outweigh the benefits to the defector, the result would be merely a redistribution of wealth -- perhaps with a net increase. Such a situation would scarcely constitute a dilemma.

5. My experience has been that moderate sized groups of students run for moderate amounts of money (e.g., $N = 8$, $C(N) = \$2.50$, $d = \$5.50$, $\lambda = \$2.50$) take the commons dilemma very seriously indeed