Volunteer's Dilemma

ANDREAS DIEKMANN
Institute for Sociology
University of Munich

A conflict game simulating social traps in which a collective good can be provided by a volunteer is discussed and some solution concepts are analyzed. There is a greater incentive for "free riding" than for the production of the collective good at the cost of the volunteer. However, if everybody defects, all players will lose. Such a result is frequently produced by "diffusion of responsibility" as described by Darley and Latané (1968). In contrast to other social traps, a dominant strategy does not exist. Also, the mixed-equilibrium strategy yields very low payoffs. The expected gain is not greater than the payoff achieved by the nonequilibrium maximin strategy. Superrationality might be a way out of the trap. However, this concept lacks the desirable equilibrium property. Only in the free communication version of the game can a definitive answer to the question of a rational strategy be given.

Social dilemmas as defined by Dawes (1975) have two characteristics. There is a dominant strategy, and the intersection of dominant strategies is a deficient Pareto-suboptimal equilibrium. As an example, Olson's problem of an insufficient supply of collective goods can be mentioned (Olson, 1965; Hardin, 1971). However, there are also situations possibly leading to social dilemmas in which a dominant strategy does not exist. Such a situation is formalized by the Volunteer's Dilemma game.

Imagine a somewhat different kind of a "prisoner's dilemma": N prisoners are sentenced for ten years if there is not at least one volunteer who confesses. In this case the volunteer(s) can expect one-year imprisonment, whereas the other prisoners will be released. So everybody has the choice either to confess his guilt with the sure gain of nine years' less

AUTHOR'S NOTE: This article was supported by the Deutsche Forschungsgemeinschaft (DFG). I would like to thank my colleagues and friends for the discussion of the game and hints to possible solution principles, particularly Arnold Hickelsberger, Eckchart Köhler and Peter Mitter. The equilibrium solution formula 3 and the asymmetric equilibrium point were first derived by Peter Mitter in a different way from that presented here.

JOURNAL OF CONFLICT RESOLUTION, Vol. 29 No. 4, December 1985 605-610

imprisonment or not to confess. In the latter case either he will face ten years or, if somebody else does the job, he will be released.

Clearly, a *volunteer* produces a collective good, but the production is costly. Nevertheless, he profits too (the sentence is reduced). If there is no volunteer in the group, all lose. Let us denote the utility of the collective good by U and the production costs by K. Then the N-person game is depicted as follows:

	0	1	2	. . .	N - 1	C-choices
C	U-K	U-K	U-K	. . .	U - K	$U - K > 0$[1]
D	0	U	U	. . .	U	$N \geq 2$

As usual, C denotes "cooperation" (volunteering) and D "defection" (somebody else should do the job).

The game can be played as a simple parlor game. For example, N persons are allowed to claim either 10¢ or $1 each. The amount shall be written on a covered piece of paper. Everybody will get his claim if there is at least one person who demands only 10¢.

The game is very similar to a variety of social conflict situations. Helping behavior is a prominent example. Darley and Latané's (1968) principle of "diffusion of responsibility" might be a case in point. If N persons witness an accident or crime, some of them feel their conscience is relieved if there is at least one person helping the victim. In this case bystanders achieve a psychic gain of U. But helping is costly (K). Therefore, each person might be inclined to defect hoping that somebody else will help the victim. An example from a different field of application is altruistic behavior in biology and genetics as described by Dawkins (1978: 6).

SOLUTION CONCEPTS

What is rational behavior in a volunteer's dilemma context? There is no definitive answer. Evidently D is not a dominant strategy. But if one

1. If $U - K < 0$, the game is structurally different. Schelling's (1971) "mattress problem" is a volunteer's dilemma with the property $K > U$. Cost sharing ($K = k/n$, with k the costs of the collective good, n the number of C-choices, and K the individual costs of a cooperative action) is an extension in another direction but will not be explored here. For $N = 2$ the game corresponds to a degenerate version of "chicken."

expects the other players to apply maximin, a better choice might be D. If everybody follows this reasoning, all players will be "trapped." Of course, maximin is not an equilibrium strategy.

In a *vocal game* with free communication the players can agree to determine a volunteer by lot. The *asymmetric strong equilibrium* point is (N – 1) D-choices and a C-choice by the unlucky volunteer with an expected utility of U – K / N. This is also an equilibrium strategy in an iterated supergame. However, some form of *coordination* or social norms will be required. But in many social conflict situations such as those investigated by Darley and Latané, coordination by communication is not possible. Those situations correspond to a *tacit game*.

There is also another *weak equilibrium point that is symmetric* if mixed strategies are introduced. Letting q_i be the probability of player i's D-choice the expected utility is:

$$E_i = q_i U(1 - \prod_{j,\, i \neq j}^{N} q_j) + (1 - q_i)(U - K) \qquad [1]$$

Taking the derivative with respect to q_i:

$$\frac{dE_i}{dq_i} = K - U \prod_{j,\, i \neq j}^{N} q_j \qquad [2]$$

and letting $dE_i/dq_i = 0 (i = 1,2 \ldots N)$, the resulting system of simultaneous equations yields the equilibrium strategy with the symmetric solution:

$$q_o = \left[\frac{K}{U}\right]^{1/(N-1)} \qquad [3]$$

Surprisingly, q_0 substituted into equation 1 yields an expected value of (U – K)—not more than can be assured by maximin. In Harsanyi's (1977: 106) terms, for all players the symmetric equilibrium point is "unprofitable." Also, the mixed equilibrium strategy is weak—that is, if other players arrive at the equilibrium strategy, player i can switch in this game to any possible strategy without penalty.

SUPERRATIONAL STRATEGY

Payoffs are higher if players are *superrational* in Hofstadter's sense (1983). This is to say that all participants in the game maximize their expected utility under the restriction of Kant's categorical imperative. In a symmetric game this requires that all q_i's be identical. Equation 1 takes the following form:

$$E = qU(1 - q^{N-1}) + (1 - q)(U - K) \qquad [4]$$

Maximization of this expression yields the superrational strategy:

$$q^* = \left[\frac{1}{N}\,\frac{K}{U}\right]^{1/(N-1)} \qquad [5]$$

which prescribes a higher probability of cooperative action ($1 - q^*$) than in the symmetric equilibrium case. In the tacit game version without communication, superrational players achieve the maximal symmetric and Pareto-optimal payoff vector. Substitution of q^* in equation 4 yields, after some algebraic manipulations,

$$E^* = U - mK, \quad 1 > m > \frac{1}{N}, \quad m = 1 - q^*(1 - \frac{1}{N}) \qquad [6]$$

The expected value E^* is greater than the maximin value and the symmetric equilibrium value but less than the asymmetric equilibrium value. If coordination by communication is impossible, superrationality seems to be an optimal principle.[2] In Hofstadter's terms, coordination is achieved by the "bond of logic." However, there is always an incentive to defect, due to the fact that superrationality is not an equilibrium strategy.

2. Because the expected payoff of the symmetric equilibrium strategy is identical to the maximin value, the weak equilibrium point is *nonprofitable for all players* (Harsanyi, 1977: 106). On the other hand, the profitable strong equilibrium point cannot be attained without communication. The superrational strategy might be considered a solution of the coordination problem in the tacit game, although this is *not a best-reply strategy* as required by Harsanyi's (1977: 116) rationality postulate A2. However, there is no other pure or mixed best-reply strategy solving the coordination problem because a third equilibrium point does not exist. In Harsanyi's terms no rational player can expect more than the maximin payoff, and therefore the game is "unprofitable." By the postulate A1 the Harsanyi rational strategy is *maximin*. Note that both strategies, maximin and superrationality, are not best-reply strategies. However, superrationality yields higher payoffs for all players.

Both principles, symmetric equilibrium and superrationality, imply that the D-choice probability approaches 1 with increasing group size N, although this convergence is faster for the symmetric equilibrium strategy. For given group size N, both predict that the probability of defection is a monotonically increasing function of the cost-benefit ratio. The number of D choices in a group follows a binomial distribution with mean Nq. In an iterative game the emergence of a norm is predicted. Utility can be maximized by compliance with a rotation rule—that is, participants take turns in assuming the role of the volunteer. These and other hypotheses can be tested in an experimental framework.

PRODUCTION OF COLLECTIVE GOODS

There is no doubt that the collective good is produced under the maximin rule, or in the case of the asymmetric equilibrium, or in the trivial case N = 1. For the two remaining strategies the probability P that at least one player engages in action C takes the form

$$P = 1 - q^N \qquad [7]$$

This becomes for the symmetric equilibrium condition

$$P_O = 1 - \left[\frac{K}{U}\right]^{(N/(N-1))} \qquad [8]$$

and for superrational players

$$P^* = 1 - \left[\frac{1}{N}\ \frac{K}{U}\right]^{(N/(N-1))} \qquad [9]$$

Under the equilibrium strategy, the probability of supplying a collective good is a *decreasing* function of N, but the reverse is true for superrational players. Thus, on the group level the two principles lead to opposite predictions.[3] In the superrationality case, P^* approaches 1 in the limit,

3. Darley and Latané's (1968) experimental results confirm hypothesis 9 and disconfirm hypothesis 8. There is a monotonically decreasing tendency of individual helping be-

whereas the asymptotic value is 1 – K/U for the symmetric equilibrium strategy.

To summarize, an interesting counterintuitive feature of the game is that the symmetric equilibrium strategy yields very bad results, no better than the maximin principle. On the other hand, achieving the asymmetric equilibrium point requires coordination. Superrationality might be the way out of the trap in a volunteer's dilemma without communication. The conflicting predictions of the different solution concepts can be regarded as rival hypotheses that should be investigated in cross-experiments. Moreover, the game provides a good test of the predictive power of equilibrium concepts with respect to actual behavior. If the observed behavior is found to comply more with the symmetric equilibrium solution, despite the low expected payoff, this would be a strong confirmation of the equilibrium hypothesis.

REFERENCES

DARLEY, J. M. and B. LATANÉ (1968) "Bystander intervention in emergencies: diffusion of responsibility." J. of Personality and Social Psychology 8: 377-383.

DAWES, R. M. (1975) "Formal models of dilemmas in social decision making," in M. Kaplan and S. Schwartz (eds.) Human Judgment and Decision Processes: Formal and Mathematical Approaches. New York: Academic Press.

DAWKINS, R. (1976) The Selfish Gene. New York: Oxford Univ. Press.

HARDIN, R. (1971) "Collective action as an agreeable N-Prisoner's dilemma." Behavioral Sci. 16: 472-481.

HARSANYI, J. C. (1977) "Rational behavior and bargaining equilibrium," in J. C. Harsanyi (ed.) Games and Social Situations. Cambridge, MA: Cambridge Univ. Press.

HOFSTADTER, D. R. (1983) "The calculus of cooperation is tested through a lottery." Scientific American 6: 14-18.

OLSON, M. (1965) The Logic of Collective Action. Cambridge, MA: Harvard Univ. Press.

SCHELLING, T. (1971) "The ecology of micromotives." Public Interest 25: 61-98.

---

havior with group size but a slightly increasing chance of supplying a collective good. The probabilities for "ever helping" are (Fig. 1 in Darley and Latané 1968: 380) $1 - q_i = 0.82$ for a two-person group, and 0.62 for a five-person group. With only one bystander (N = 1) the probability of a C-choice is 1 as expected. By formula 7 we have a probability P = .97 for the two-person group and P = .99 for a five-person group.