ON GETTING PEOPLE TO COOPERATE WHEN FACING A SOCIAL DILEMMA: MORALIZING HELPS

by

Robyn M. Dawes, Harriet Shaklee

Oregon Research Institute

and

F. Talarowski

University of California at Santa Barbara

Decision Research
Report 76-8

ACKNOWLEDGMENT

Support for this research performed by Oregon Research Institute was provided by the Advanced Projects Agency of the Department of Defense and was monitored under Contract N00014-76-C-0074 with the Office of Naval Research, under subcontract from Decisions and Designs, Inc. The authors wish to express their thanks to Paul Slovic, Sarah Lichtenstein and Baruch Fischhoff for their comments.

On Getting People to Cooperate When Facing a Social Dilemma: Moralizing Helps

In a large battle, each individual soldier may reason that he is best off if he takes no risks; for that way he assures his own personal survival, while the effect of his decision on the outcome of the battle is minimal. Yet, if all the soldiers reason that way, the result will be a rout; and all would then be worse off than they would have been had they all been willing to take risks.

The above example illustrates social dilemmas. Briefly, a social dilemma occurs when (i) each individual reasons that he or she is best off defecting from the group effort no matter what other people in the group do, yet (ii) everyone is better off if all cooperate than if all defect (Dawes, 1975). Other examples of social dilemmas may be found in the decision to pollute (one might save considerable money while making a negligible contribution to the overall pollution problem), the decision in agrarian societies to have as many children as possible, the decision to cheat on income tax returns, and the decision to "rip-off" large organizations.

Recently, Dawes (1975) has devised an experimental paradigm for studying reactions to social dilemmas, and an extensive series of studies has been conducted by Dawes, McTavish and Shaklee (1976) to determine what factors affect the decision to cooperate or defect in a social dilemma situation. The paradigm involves offering subjects in a group the choice of either earning a small amount of

money with no fine to any other group member or receiving a large amount of money with an equivalent fine spread out equally among all members of the group. For example, in the studies reported by Dawes, McTavish and Shaklee, groups typically consisted of eight subjects. Each group member could make either a cooperative choice of earning $2.50 for participation with no fine to any group member--or a defecting choice of earning $12 with a $1.50 fine to each group member including themselves. Thus, each subject was $8 better off defecting ($12.00 - $1.50 - $2.50); yet, if all members chose the defecting choice, none would receive anything--because the eight fines of $1.50 exactly matched the $12 for defecting. Subjects viewed this task with extreme seriousness.

Several factors influencing choice were discovered. One was communication. When subjects were allowed to communicate with one another, the rate of cooperation was three times that of subjects who were not allowed to communicate (approximately 75% versus 25%). Dawes, McTavish and Shaklee had initially hypothesized that communication might result in subjects committing themselves to cooperate; yet a condition in which subjects were required to make a non-binding announcement of their intentions to the group elicited no more cooperation than did a condition in which subjects merely discussed the problem and were in fact prohibited from making such a commitment. Thus, it appeared that communication about the problem in and of itself was the important variable.

One informal observation made during these studies was that

subjects often brought up the moral and ethical implications of the two choices when discussing them. For example, statements such as "if you fink out on the rest of us you are going to have to live with this for the rest of your life" were not at all uncommon. Evidence that such moral exhortation may in fact work can be found in a field study by Schwartz and Orleans (1973). These authors interviewed taxpayers a month before the filing deadline; the questions in the interviews emphasized either the positive moral aspects of paying taxes--e.g., good citizenship--or the legal sanctions for nonpayment. Those taxpayers who had received an interview emphasizing morality paid an average of $243 more than they had the previous year while those receiving the sanction interview paid an average of only $11 more. (Two control groups paid an average of $40 and $57 less.)

Thus, if moralizing works to control behavior in social dilemma situations, both inside the laboratory and in a field setting, the experimenters themselves should be able to increase cooperative behavior by emphasizing moral and ethical concerns, even though the subjects themselves are unable to communicate with one another. The following two experiments are devoted to testing the hypothesis that moral exhortation might in fact be effective in controlling responses to a social dilemma.

Experiment 1

The same game was used as in Dawes, McTavish and Shaklee (1976). Subjects were not allowed to communicate with each other. In a control condition, the results of placing an 0 (the cooperative response) or an X (the defecting response) on the piece of paper were explained, while in the experimental ("morality") condition,

the experimenter read a rather extensive statement about the group versus individual welfare implications of choosing the 0 or the X. As in previous research, all decisions were made in private, and were never revealed to other group members. Of course, each group member could determine from his or her payoff how many people had cooperated or defected.

Method

Subjects, 73 females and 70 males, responded to newspaper ads indicating that they could receive anything from nothing to $10.50 for participation in a psychological experiment. Subjects were recruited in groups of 8, each group being randomly assigned to the control or the morality condition until there were 10 groups in each. When less than eight subjects arrived, the payoffs were modified appropriately. For example, if 6 subjects arrived, the payoff for defection (placing the X response) was $9 rather than $12, again with $1.50 fine to each participant. In all, there were six groups of 6 subjects, five of 6 and nine of 8.

The morality condition read as follows:

> Many decisions we make in society today involve a choice between individual gain and group welfare. It is, for example, to each individual's self interest to exploit the environment, since any gain accrues directly to the exploiter and the loss is shared by all members of the society. But, if everyone were to behave similarly, it would lead to rapid destruction of the world's resources. Thus, everyone is better off if everyone cooperates in protecting the environment than if each were to choose for his or her own gain. Consider, for example, the dynamics of the whaling industry. Each whaler has an individual

incentive to kill whales since he gets all the profit from the sale of meat, oil, and other whale byproducts. But at the same time, the whale population is decreased, threatening the survival of the species. Thus, that whaler's actions may cause the group to lose a valuable and fascinating species. Only if everyone cooperates in protecting whales will such a result be averted. In many other social decisions, the welfare of the group similarly depends on our laying aside our individual interests and cooperating with others for the group's welfare.

This is a study of ethical behavior in a similar social dilemma. In this experiment, we're interested in your choice in a decision between own gain and group welfare. We'd like you to consider the ethical implications of your behavior before deciding on your course of action today.

This table (Table 1) indicates the possible consequences of the decision each of you will be asked to make. You must decide whether to choose an **X** or an 0. As you will see, the X choice exploits the group, while the 0 choice is cooperative. You will have to mark an X or an 0 on a card in private. If you cooperate by choosing 0, you will earn $2.50 minus a $1.50 fine for every person who exploits the group by choosing an X. If you decide to exploit the group by choosing an X. you will earn $12.00 minus $1.50 for each exploiter, including yourself, who chooses an X. However, as you can see in the table, your payoffs do not go below zero. By looking at the second row of payoffs in the table, for example, you can see that

Payoff Matrix

Payoff to X	Number Choosing X	Number Choosing 0	Payoff to 0
	0	8	2.50
10.50	1	7	1.00
9.00	2	6	-0-
7.50	3	5	-0-
6.00	4	4	-0-
4.50	5	3	-0-
3.00	6	2	-0-
1.50	7	1	-0-
0.00	8	0	-0-

if 7 of you cooperate by choosing 0 and 1 of you exploits the group by choosing X, then those choosing 0 will earn $1.00 and the one choosing X will earn $10.50. Or, on the third row, if 6 of you cooperate and 2 of you exploit, then those choosing 0 earn $0.00 and those choosing X get $9.00.

If everyone chooses 0, everyone gets $2.50, and you can all leave with a fair share of the payment and the good feeling that you all cooperated for the group's welfare. Each time a person chooses X, that person decreases the assets of the group, since every group member is fined $1.50 for every X chosen. While the exploiter may enjoy a high payoff by choosing X, he or she does so by taking money away from the other group members, possibly causing the other members to get no payment at all for their participation today. Moreover, the person who chooses X gets a high payoff only if the other group members cooperate by choosing 0. We can see on the last row of the table that if everyone tries to exploit the group, no one gets any money.

In sum, there are both risks and benefits to each choice. For the 0 choice, the benefits are that the group as a whole is best served, everyone gets a fair share of the payment. The risk of cooperating is that others might choose X and benefit themselves at your expense.

The person who chooses X may get a higher payoff, but only if the others in the group cooperate.

If everyone exploits the group, nobody gets any payment

for their time invested today. Furthermore, if you decide to exploit the group you'll have to deal with the knowledge that your gain was arrived at by taking money away from others.

Your decision today will be totally private and none of the other participants in this group will know what you decided. You will each be paid and dismissed separately. Since both decisions and outcomes are private in this experiment, you cannot agree to meet after the experiment to split up the payoff. Once you have made your decision, you will write your code number and decision on the card in your envelope. You will also identify each other person in your group by code number, and indicate what decision you believe that person to be making.

You will be paid according to your accuracy in predicting the decisions of the other group members. After all experimental groups have been run, we'll identify the most accurate subjects and send them a check by mail. Thus, while we'll pay you for your decision immediately after this session is over, we'll have to delay your accuracy payment until we have evaluated your accuracy in relation to that of other participants in the experiment.

You will make your decision and predictions without communicating with each other in any way. I can answer any questions that you have.

As I mentioned before, we're interested in people's

decisions in a social dilemma of own versus group welfare. I once again urge you to consider the ethical implications of the choice you are about to make. I'd like to thank you in advance for your participation in this study of experimental ethics.

I can answer any questions you have at this time.

The instructions on the control conditions were identical except that there was no discussion of individual versus group benefit. These instructions are presented in Appendix A.

Results

The average percentage of defection in the morality conditions was 37.2. The average in the control condition was 75.8. Using the group as the unit of analysis, the t-value for testing these different percentages was 4.98 ($p < .01$).

Experiment 2

A parallel experiment was conducted at the University of California at Santa Barbara. There, college students in psychology who received experimental credit for participation played a game equivalent to a five-person game in which the payoff for the cooperative response was $2 and for the defecting response $10 with a $2 fine to each group member including the defector. This research was conducted by Talarowski (1976), who presented this game in many different formats. For purposes of this report, we will average across all formats.

Method

Fifty-six subjects were presented with 10 decompositions of commons dilemma games--five of a game in which the cooperative payoff was $2 and five of a game in which the cooperative

payoff was $2.40. In both games, a single defecting subject would cause the payoffs of the remaining four cooperating subjects to be zero; in both games universal defection resulted in a net payoff of zero to the group.

Half (28) of the subjects were randomly assigned to a "morality" condition similar to Experiment 1, and the other half randomly assigned to a control condition. The instructions were read aloud and they were also printed on the face sheet of the subjects' questionnaire. The morality instructions were highly similar to those in Experiment 1. The control instructions were the same as the morality instructions with the exception that no mention was made of the group and individual implications of the choices.

Subjects were paid off by being randomly grouped with four other subjects and having a single decomposition picked at random.

Results

In the morality condition, 32.5% of the responses were defections, while in the control condition, 46.4% were defections ($p < .05$). Note that here each subject is the unit of analysis.

Discussion

The effect replicated at two different locations, using two different games, using two different formats. While the results at Santa Barbara were much weaker than those at Eugene, the differences may be accounted for by differences in format or by (resulting?) differences in proportion of defection. The overall base rate of defection in Eugene was 52.6% while that in Santa Barbara was 39.5%. Of course, it is always possible the differences are a function of

geographic location, but we find it highly implausible that southern Californians are more cooperative than are Oregonians.

The important finding is: it worked. Of course, the manipulation is a form of "experimental demand," but that's just the point. Just as Milgram's (1963) subjects gave in to the demand to be obedient despite strong psychological resistances, ours gave in to the demand to be moral despite an $8 motive not to be.

The exact mechanism by which the moralizing works cannot be determined from these experiments, but must be the subject of future investigations. For example, moralizing may work by appealing to subjects' consciences. Or, alternatively, the moralization may work by simply making subjects aware of the consequences of their own behavior. The latter possibility is one that deserves much further consideration.

Previous attempts at experimental manipulation in which morality was introduced in a fairly subtle manner (in one or two sentences) did not work. Here we made the moral and ethical implications of the behavior as explicit as possible, and we did so in a highly redundant manner. We do not know the degree to which we could get the effect without preaching at quite such length.

Finally, it should be pointed out that the amount of cooperation in the morality condition in Experiment 1 (63%) is somewhat less than that obtained in the identical situation when the subjects were free to communicate with each other (see Dawes, McTavish and Shaklee, 1976). Since the conditions were so much different, it is not really appropriate to test for "significance"

of this difference. It is, however, suggestive; perhaps when people bring up moral and ethical considerations themselves, these considerations have more impact than when they are introduced in what is essentially a sermon.

References

Dawes, R. M. Formal models of dilemmas in social decision-making. In S. Schwartz & M. F. Kaplan (Eds.), Human Judgment and Decision Processes: Formal and Mathematical Approaches. New York: Academic Press, 1975.

Dawes, R. M., McTavish, J. & Shaklee, H. Behavior, communication, and assumptions about other peoples' behavior in a commons dilemma situation. Journal of Personality and Social Psychology, 1976, in press.

Milgram, S. Behavioral study of obedience. Journal of Abnormal and Social Psychology, 1963, 67, 371-378.

Schwartz, R. D. & Orleans, S. On legal sanctions. In M. Barker (Ed.) Law and the Social System. New York. Lieber-Atherton, 1973.

Talarowski, F. Effects of morality and decomposition on the commons dilemma. Unpublished masters thesis, University of California at Santa Barbara, 1976.

APPENDIX A

This table (Table 1) indicates the possible consequences of the decision each of you will be asked to make. You must decide whether to choose and X or an 0. You will have to mark an X or an 0 on a card in private. If you choose an 0, you will earn $2.50 minus a $1.50 fine for every person who chooses X. If you choose X, you will earn $12.00 minus $1.50 fine for each person, including yourself, who chooses X. However, as you can see, your payoffs do not go below zero.

By looking at the second row, for example, you can see that if 7 of you choose 0 and 1 of you chooses X, then those choosing 0 will earn $1.00 and the person choosing X will earn $10.50. Or, in the third row, if 6 of you decide to choose 0 and 2 of you decide to choose X, then those of you choosing 0 will earn $0 and those choosing X will earn $9.00.

Your decision will be totally private and none of the other participants in this group will know what you decided. You will each be paid and dismissed separately. Since both decisions and outcomes are private in this experiment, you cannot agree to meet after the experiment to split up the profit.

Once you have made your decision, you will write your code number and decision on the card in your envelope. You will also identify each other person in the group by code number, and indicate what decision you believe that person to be making.

You will be paid according to your accuracy in predicting

the decisions of the other group members. After all the experimental groups have been run, we'll identify the most accurate subjects and send them a check by mail. Thus, while we'll pay you for your decision immediately after this session is over, we'll have to delay your accuracy payment until we have evaluated your accuracy in relation to that of the other participants in the experiment.

You will make your decision and predictions without communciating with each other in any way. I can answer any questions that you have. I'd like to thank you in advance at this time for your participation in this study.