

# Unintended Consequences of Cooperation Inducing and Maintaining Mechanisms in Public Goods Dilemmas: Sanctions and Moral Appeals

**Xiao-Ping Chen**

*University of Washington*

**Madan M. Pillutla**

*London Business School*

**Xin Yao**

*Wichita State University*

Research suggests that contribution to public goods (i.e. cooperation) will increase when groups use sanctions. We argue that when groups use rewards and punishments to induce members to cooperate in a social dilemma, individuals' natural propensity to cooperate may be reduced. Results from two laboratory experiments provide consistent support for our hypothesis. Specifically, we found that cooperation (in groups that adopted sanctioning systems, including both reward and punishment) decreased significantly following the removal of sanctioning systems. We also found that a moral appeal to cooperate was as effective as sanctions in inducing cooperation. Moreover, cooperation induced through appeals was more likely to sustain than that induced through sanctions. We found that people's trust in others' cooperation mediated the relationship between the use of sanctions and appeals and the level of cooperation after these inducing practices were removed. Implications of these results are discussed in the group and organizational context.

**KEYWORDS** cooperation, moral appeals, public goods dilemma, sanctions

In his classic book *The Gift Relationship*, Titmuss (1970) contrasted the (then) primarily volunteer-driven blood donation scheme in England and Wales with the paid blood donor system in the United States and came up with the surprising finding that the British system was more efficient in terms of the quality and quantity of blood that

## **Author's note**

Address correspondence to Xiao-Ping Chen, Department of Management and Organization, Michael G. Foster School of Business, University of Washington, Box 353200, Seattle, WA 98195–3200, USA [email: [xpchen@u.washington.edu](mailto:xpchen@u.washington.edu)]

Copyright © The Author(s), 2009

Reprints and permissions: <http://www.sagepub.co.uk/journalsPermissions.nav>  
12:2; 241–255; DOI: 10.1177/1368430208098783

it could make available to needy recipients. The fact that blood quality and donor relationships suffer with the introduction of commercialization of blood as well as donor relationships is not counterintuitive; clearly individuals who are paid for their blood have more of an incentive to lie about their health and suitability to give blood than those who give it for free. However, the finding that commercialized blood markets were inefficient in terms of matching demand and supply is counterintuitive. His findings also indicated that the rise in the number of commercial blood banks that pay for blood coincides with a diminishing number of donors in the voluntary sector.

Based on these findings, and a survey that indicated that most volunteers gave moral reasons when asked why they donated blood, Titmuss (1970) concluded that the commercialization of blood donor system ‘represses the expression of altruism and erodes the sense of community’ (p. 314). Titmuss’ work is insightful in its analysis of the consequences of commercializing voluntary exchanges. His conclusion that reward (or punishment) for donating blood corrupts those who have given blood for free, turning them from altruists to narrow self-interest seekers, seems accurate for the most part. What is lacking in the analysis, however, is a clear model of the micro processes that explain why this could be the case.

The voluntary blood donor system in England and Wales that Titmuss (1970) described in his study is an example of what Olson (1965) referred to as a ‘public goods dilemma’. A public good is a commodity that can be provided only if group members contribute something towards its provision; however, all persons—contributors and non-contributors—may use it (Komorita & Parks, 1995). The problem of the provision of public goods has attracted the attention of a number of scholars because of an interesting phenomenon—individually reasonable behavior (i.e. that benefits the individual) by every person in a collective can lead everyone in the collective to be worse off than if each had adopted some other behavior. This is called a ‘social dilemma’ (Dawes, 1980), and in situations involving the provision of public goods, it is referred to as a public goods dilemma. For example, each one

of us would be better off if we could make use of a public resource, such as a blood bank, or a public library, without making a contribution toward the provision of the good. However, if everyone acts in this manner, no contribution can be made and we would have no blood bank or public libraries—a situation where everyone is worse off than if the public good were available.

For over 30 years, researchers from different disciplines have examined many factors (e.g. structural factors such as group size and payoff structure, and motivational factors such as group identity and reciprocity) that influence contributions in public goods dilemmas. An experimental paradigm that is often used to study social dilemmas can be described as follows. Participants are brought to a laboratory in groups and are given a fixed endowment at the beginning of the experiment. They are then asked to make one decision (individually and anonymously) to invest any amount of their endowment to one of two accounts: a joint (J) account or a personal (P) account. The investment in the J account receives a certain amount of interest but the total investment and interest are divided equally among all group members. The investment in the P account belongs entirely to the individual member but it receives no interest. A member’s payoff is the sum of the investment in the P account and an equal share from the J account.

Take for example a four-person group where each member is given 1000 points as his/her endowment and where the money invested in the J account receives an interest of 100%. Suppose individual ‘i’ decides to invest some points (call it  $j_i$ ) in the J account, and the remaining points in her individual P account (call it  $p_i = 1000 - j_i$ ), then the total points that member i earns can be calculated by the following formula:

$$\text{Payoff for member } i = p_i + \frac{1}{4}(\sum(j_i) * (1 + 100\%)), i = 1, 2, 3, 4.$$

Table 1 presents the payoff matrix of the above example. It can be seen that investing fewer points in the J account always leads to a higher

Table 1. Payoff matrix for a public goods dilemma

Contribution to J account	Mean contribution of others to J account					
	0	200	400	600	800	1000
0	1000	1300	1600	1900	2200	2500
200	900	1200	1500	1800	2100	2400
400	800	1100	1400	1700	2000	2300
600	700	1000	1300	1600	1900	2200
800	600	900	1200	1500	1800	2100
1000	500	800	1100	1400	1700	2000

Note: Each member of a four-person group is asked to invest 0 to 1000 points of an endowment to a joint (J) account or to a personal (P) account. The total investment in the J account earns an interest of 100% (doubles in value), and the investment in the P account earns no interest. Each individual's total payoff is an equal share from the J account and whatever they have invested in.

payoff for an individual than investing more points (numbers in the higher row are greater than those in the lower ones). When no points are invested in J, however, the payoff (1000 points) is smaller than when all are invested in J (2000 points). Thus, this payoff matrix meets the condition of a social dilemma (Dawes, 1980).

Social scientists have proposed a number of solutions that collectives can use to get their members to contribute to the provision of the social good or more generally to cooperate in social dilemma situations. Some of the solutions are structural and others social or motivational (see Kollock, 1998; Yamagishi, 1986 for overviews). Rewarding cooperators, punishing non-cooperators, and monitoring group members are a few structural solutions that have been offered to affect cooperation in social dilemma situations. Past research has demonstrated that penalties can increase cooperation (Caldwell, 1976; Fehr & Gächter, 1999; Platt, 1973; Yamagishi, 1986, 1988). For example, experimental studies have shown that cooperation is more likely if a sanction system is provided (Andreoni, Hardbaugh, & Vesterlund, 2003; McCusker & Carnevale, 1995) or if individuals have the ability to punish defectors (Caldwell, 1976; Komorita, 1987). Yamagishi (1986, 1988) used a penalty-based sanctioning system whereby individuals had a choice to contribute to a 'penalty fund', which was used to set up a monitoring system to identify and punish the least cooperative individual in four-person groups. He found that group members did choose to contribute to the

penalty fund and that subsequent cooperation increased.

Olson (1965) proposed that individuals could be made to contribute to a public good when selective incentives were offered to contributors. Field research on conservation behavior has shown that selective incentives in the form of monetary rewards are effective in decreasing the consumption of water and electricity (Maki, Hoffman, & Berk, 1978; Winett, Kagel, Battalio, & Winkler, 1978). The above research indicates that selective rewards and punishments are effective means of ensuring cooperation in social dilemmas. In the extreme, the provision of rewards or punishment could completely alter the structure of the situation such that there is no dilemma at all, which makes the cooperation more 'instrumental' rather than 'elementary' (Yamagishi, 1986). In other words, the change in the payoff structure provides incentives other than intrinsic desires to cooperate for contribution to public goods. An example is the complete marketization of blood donors that Titmuss (1970) discussed.

Social dilemmas are replete within organizations where most complex work is organized around groups. Group work provides an opportunity for opportunistic individuals to not contribute their share and yet reap the rewards that accrue to the group. Organizational scholars who study free-riding in organizational groups have labeled the phenomenon social loafing (see for example George, 1992). Like their colleagues in the social sciences, these scholars

have theorized and empirically validated various solutions to free riding; and they have reached similar conclusions, i.e. that punishments (e.g. Miles & Greenberg, 1993) and rewards (Albanese & Van Fleet, 1985) will deter free riding.

Like Titmuss (1970), we argue that while the application of rewards and punishment may temporarily raise cooperation rates and induce individuals to contribute toward the establishment of a public good (or more generally to cooperate), they may actually undermine individuals' natural propensities to cooperate in the long run. A recent study by Mulder, van Dijk, De Cremer, and Wilke (2006) has shown that using punishment to induce cooperation led to lower trust of internal motivations of others' cooperation, as well as less cooperation once the punishment is removed. We thus propose:

*Hypothesis 1a:* Cooperation induced by sanctions (reward or punishment) will wither away in the absence of these sanctions.

Mulder et al. (2006) found that trust in others' motives dropped after the removal of punishment and the drop left the trust level lower than that of teams that experienced no punishment. This was not the case with cooperation: although the removal of sanctions resulted in a lower level of trust, the post-removal cooperation was not any lower than that by groups that had not faced sanctions to begin with (the no-sanction condition). Thus, sanctions undermined trust but not cooperation. This finding could be due to the fact that Mulder et al. (2006) did not provide any feedback regarding other people's cooperation level over the course of the experiment. In other words, participants could not see the actual outcomes of sanctions and make attributions about others' behaviors accordingly. Thus, Mulder et al.'s (2006) participants formed opinions about others' trustworthiness while making later contribution decisions without actual information of others' decisions. We suspect that, if feedback is provided in which people see some team members fail to cooperate even in the face of sanctions, people's beliefs about others' trustworthiness as well as their cooperation will be further undermined (see Pillutla & Chen, 1999 about the detrimental

effects of unexpected non-cooperative behavior in social dilemma situations). Granted that when sanction is first implemented, cooperation may be higher among people who undergo sanctions than those who do not; we anticipate the drop in cooperation after the removal of sanctions will be so dramatic that the cooperation level of those who have experienced sanction will be lower than those who have not.

*Hypothesis 1b:* When sanctions are removed, people will behave less cooperatively than those who experienced no sanction in the first place.

In the following sections we further propose that the mechanism through which sanctions, including both rewarding contributors and/or punishing non-contributors, damage individuals' decisions to cooperate is that of lowering their trust in other group members.

## Attributions and sanctions in social dilemmas

The Goal/Expectation Theory of cooperation in social dilemmas (Pruitt & Kimmel, 1977) posits that an individual's contribution towards the provision of a public good is a function of their expectations about how much others in the group would contribute; when individuals expect others to contribute more (less) towards the provision of a public good, they will in turn contribute more (less). These expectations are likely to be affected by the attributions that individuals make regarding others' contributing behavior in similar situations. If an individual believes that others are contributing because of certain contextual factors, they are likely to infer that individuals will not contribute in the absence of these factors. If, on the other hand, they believe that others are contributing because of intrinsic factors, they will have stable expectations of the others' contributions regardless of the presence or absence of the contextual factors. These stable expectations can be thought of as *group-level trust*, which is defined as a

... common belief among a group of individuals that another individual or group makes good faith efforts to behave in accordance with any commitments both explicit or implicit, is honest in whatever

negotiations preceded such commitments, and does not take excessive advantage of another even when the opportunity arises. (Cummings & Bromily, 1996, p. 303)

Theories of lay epistemology (e.g. Heider, 1958; Jones & Davis, 1965; Kelly, 1967) suggest that individuals infer the cause of another person's actions based on personal and environmental forces. These and other attribution models suggest that the importance of any potential cause of a person's action is reduced to the extent that there are other potential *facilitative* causes of the action, and increased to the extent that there are other potential *inhibitory* factors. Several studies have found that individuals make external attributions to actions that are caused by strong external forces influencing an actor and internal attributions to actions where the external forces are minimal (e.g. Jones, Davis, & Gergen, 1961; Thibaut & Riecken, 1955). These theories and associated empirical studies would suggest that, in situations where the extrinsic factors for cooperation are made salient, individuals will be more likely to view others' cooperation as extrinsically motivated and will therefore be less likely to expect them to continue to cooperate once such factors are absent. A punishment or reward system is a salient extrinsic factor and individuals can thus attribute others' cooperation to the presence of these factors. When these factors are removed, individuals are likely to expect others not to cooperate. Cooperation in the absence of these factors is likely to be attributed to intrinsic motives, which are expected to be stable. In other words, the presence of the reward or punishment affects internal or external attributions about others' cooperative behaviors, which will in turn affect expectations about others' future contributions. Therefore, we propose:

*Hypothesis 2:* Cooperation expectation of/trust in others will be lower in conditions where sanctions have been applied compared to conditions where they have not been present.

Malhotra and Murnighan (2002) and Mulder et al. (2006) demonstrated the detrimental effects of external pressure on cooperation; the former with contracts and the latter with

sanctions. Both studies found that these external factors that engendered interpersonal cooperation resulted in lowered levels of trust compared to conditions in which external factors were absent. The Malhotra and Murnighan (2002) study demonstrated that contracts undermined the development of interpersonal trust while the Mulder et al. (2006) paper showed that sanctions undermined group members' beliefs that others' contributions were due to internal motives.

We attempt to replicate the findings of the Mulder et al.'s (2006) paper and show that sanctions undermine group-level trust. In addition, we hypothesize that the lowered trust that follows the imposition and removal of sanctions mediates the effects of sanctions on cooperative behavior. Thus:

*Hypothesis 3:* Trust in others will mediate the effects of the removal of sanctions on group cooperation.

## Experiment overview

We report two experiments in the present article. The first experiment focused on testing Hypotheses 1a and 1b. In the second experiment, we test Hypotheses 2 and 3 in addition to replicating findings from Experiment 1. Furthermore, we examine the effect of providing intrinsic factors (i.e. a moral appeal) for cooperation as a solution to the proposed detrimental consequences of the sanctioning system. In Experiment 1, participants made contribution decisions first in the presence of a reward or punishment system, and then in the absence of such systems. We compared the contributions made following the removal of either sanctioning system with (a) the contributions made in the presence of such systems, and (b) the contributions made in the control condition where sanctions were never introduced.

## Experiment 1

### Method

**Participants** The participants were 160 undergraduate students enrolled in an introductory

course in organizational behavior in a major university in Hong Kong. Participation in the experiment partially fulfilled a requirement of the course; in addition they received monetary payment as a result of a lottery drawing embedded in the experimental task (see details later under 'Procedure').

**Experimental design and conditions** Two experimental conditions and one control condition were created to study the effects of reward/punishment on cooperative behavior. Each experimental condition involved two stages of decision-making: first when reward or punishment was introduced and second when such reward or punishment was removed. In the control condition, participants also made decisions in two stages, but without introduction of sanctions. The three conditions of the experiment are as follows: reward–no reward, punish–no punish, and control. We describe each of these conditions below.

*Reward–no reward condition* In this condition, group members were informed that the person who invested the highest amount to the J account would receive a bonus of six points for a given trial; and in the second stage, they were asked to make choices as before, but were told that there would be no bonus points given to the highest J investor.

*Punish–no punish condition* In this condition, group members were informed that a reduction of six points would be given to those who invested fewer than five points to the J account; and in the second stage, they were told that this rule was removed and they could freely choose how many points to invest in J without punishment.

*Control condition* In this condition, participants were asked to make their investment decisions in the same public goods dilemma as in the experimental conditions, but without reward or punishment in either stage of decision-making.

### **Procedure**

Upon arrival at the laboratory, participants were randomly assigned to one of the three

experimental conditions. In all conditions, four participants were seated in a large room with partitions so that they could not see each other. They were told that the four people in the room constituted a group and each participant was given a member ID (1, 2, 3, or 4). Written instructions, read aloud by the experimenter, described the experimental situation. Participants were informed that: (1) they were to make a series of decisions in the experiment; (2) they would be given an endowment of 10 points at the beginning of each decision from which they could choose any amount from 0 to 10 to invest into one of two accounts: a Personal account (P) or a Joint account (J); (3) the points invested in P would not gain any interest but would belong to oneself, whereas the points invested in J would gain 100% interest but would be divided equally among all four group members; (4) the payoff for each trial would be the sum of points invested in P and the points shared from J; and (5) the total number of points earned would be converted to lottery tickets at the end of the experiment (10 points = 1 lottery ticket) and four prizes (ranging from HK\$200 to HK\$1000) would be given to the winners of the lottery.

Participants were given three examples of some payoff distributions: (a) payoffs if all invest 10 points to J; (b) payoffs if all keep 10 points in P; and (c) the payoff when varying amounts in J are invested by various members. They were then given a short quiz to test their understanding of the task and were presented with correct answers afterwards.

For all conditions, during each stage of the two-stage decision-making task, participants were asked to make decisions for five consecutive trials, with real feedback given after every trial. In the feedback, they were informed about each member's amount invested in J and the associated payoff for each member (member was identified by ID only). The procedure was identical for the second stage except that the structural changes (be it reward or punishment) were removed for the two experimental conditions. Participants made choices for another five trials, again with real feedback provided after every trial. After all the trials were completed,



participants were debriefed and dismissed individually. The lottery was administered and the winners got paid after all experiment sessions were finished.

### Measurement

**Cooperation** Cooperation is operationalized as the amount of points participants invested in the J account. Because there were five trials in each stage, the average amount of points over the first five trials is used as the cooperation index of stage 1, and the average amount of points over the second five trials is used as the cooperation index of stage 2.

### Results and discussion

**Mean cooperation after the sanction was removed** Because we do not differentiate between the effects of reward versus punishment on cooperation, we grouped the two experimental conditions together as one sanction condition. Group-level ANOVA with repeated measure on trial blocks was conducted to test H1a. We examined whether the mean cooperation in stage 2 was significantly lower than in stage 1 in the sanction condition. The analysis revealed a significant difference in cooperation in stage 2 (mean = 3.86) than in stage 1 (mean = 5.55,  $F_{1,27} = 16.95$ ,  $p < .01$ , Effect Size = .39). These results support H1a that the cooperation induced by sanctions withers after such sanctions are removed.

To test H1b that cooperation after the removal of sanctions will be lower than where such sanctions are not introduced in the first place, we conducted group-level ANOVA to examine whether the stage 2 cooperation in the sanction condition was lower than in the control condition. Since stage 2 cooperation was contingent upon stage 1 cooperation, we used the mean cooperation at stage 1 as the covariate. This analysis yielded significant difference between the sanction (mean = 3.86) and the control conditions (mean = 5.34,  $F_{1,37} = 11.45$ ,  $p < .01$ , Effect Size = .23), providing support for H1b.

These results provide strong support for our hypothesis that using monetary rewards or

punishments to induce cooperation will result in lower cooperation than if such mechanisms had never been introduced. Our results stand in contrast to the Mulder et al.'s (2006) findings in that sanctions also undermined cooperation in our study. The results, however, provide no direct empirical evidence for why sanction systems have detrimental effects on cooperation. It would be useful to measure the trust which we posit as the psychological mechanism through which sanctions undermine cooperative behaviors. Moreover, we should examine interventions that solicit and sustain cooperation.

Experiment 2 was thus conducted (a) to test the underlying psychological mechanisms for the phenomenon we observed in Experiment 1; and (b) to examine an intervention (i.e. a moral appeal) that successfully sustains increased cooperation.

### Experiment 2

While sanctions highlight the extrinsic reasons for cooperation in a social dilemma, moral appeals may direct one's attention to the greater purpose of the group and the non-instrumental value of cooperation. This could lead to enhanced levels of trust or a greater tolerance for others' non-cooperation. Moral appeals are less likely to be seen as external pressure to cooperate and more as an appeal to one's better nature. We reasoned that when the importance of cooperation is made salient, people would be more likely to come up with intrinsic justifications for their cooperative behavior. For this reason, their co-operation will be more likely to sustain. Therefore, we propose

*Hypothesis 4:* Moral appeals will be effective in generating and sustaining cooperation in social dilemmas.

### Method

**Participants** The participants were 128 undergraduate and graduate students at a major university in the United States. The majority came from an introductory course in organizational behavior and their participation in the experiment partially fulfilled a requirement of the

course. The rest of them were recruited by advertisements across campus. All participants were randomly assigned to one of the experimental conditions and received monetary payment at the end of the experiment for participation.

**Experimental task and design** In this experiment, we created a scenario in which members were told that they worked for a company named 'Talking Fish' that specialized in multilingual web page design, Internet security, and graphics design. Each person was grouped with three other participants s/he has not met before. They were assumed to be working alone from home on a short-term project in which they translated web pages from English into Hungarian for a US shipping company called 'Explore'. Each person needed to decide how many hours – out of 10 hours per day – to put into the project in addition to other projects they each had. Because each person was paid based on the total logged in hours, yet there was no way to tell who logged in for the Explore project, hours for the Explore project was divided among team members, which created a public goods dilemma situation. As in Experiment 1, participants were told that they could put hours into their other projects (equivalent to an individual account), which would be compensated at a certain hourly rate; while their hours devoted to the Explore project would be compensated at double the hourly rate of time put in the personal projects, but would get divided among four members (equivalent to a joint account).

We created three experimental conditions and one control condition. In every condition participants played a three-stage public goods dilemma game where the stage was presented as a 'week' with five working days participants made decisions about (one decision per day). In the first stage (week 1), participants were asked to decide the number of hours they put into the group project (J account) without any experimental treatment. In the second stage (week 2), one of the three experimental interventions was introduced to a group: punishment, reward, and appeal. In the third stage (week 3), participants in the sanction conditions were

told that the management decided to remove the interventions. Participants in the control condition did not experience any of the interventions, they just made decisions through three stages (weeks). Thus each participant needs to allocate a total of 150 hours (10 hours/day for 5 days/week over three weeks) in this experiment. Below we describe in detail each of the three experimental conditions and the control condition.

*Punishment condition* Participants were informed that the group member who contributed the lowest amount to the J account (group project) on a randomly selected trial (work day) would receive a fine of 25 hours for that stage (week). If there was a tie among group members in terms of investment, everyone in the tie received the fine (procedure adopted from McKusker & Carnevale, 1995).

*Reward condition* Participants were informed that the person who contributed the highest amount to the J account (group project) on a randomly selected trial (work day) would receive a bonus of 25 hours for that stage (week). If there was a tie among group members in terms of investment, everyone in the tie received the bonus.

*Appeal condition* In this condition, subjects were only given an email from the project supervisor at the beginning of the second stage. The appeal addressed the nature of the dilemma situation they were in, the importance of group work, and management's confidence in non-egotistic group members (see Appendix I).

Since we do not differentiate between the effects of reward versus punishment in our hypotheses, we grouped the punishment and reward conditions into one sanction condition. This experimental design can thus be viewed as 3 (sanction, appeal, or control) × 3 (trial blocks) factorial, with repeated measures on trial blocks.

### **Experimental procedure**

Participants were randomly assigned to one of the four conditions. Thirty-two four-person



groups were formed in the experiment, with eight groups in each condition. In all conditions, four people were seated apart from one another at four corners of a room so that they could not see each other's writing. Written instructions with simultaneous playback of audio recording described the experimental situation. Participants were not allowed to communicate with one another, nor was each person's randomly assigned ID (A, B, C, or D) revealed during the entire experiment. Participants were given one example of payoff distributions in which varying numbers of hours were contributed by various members. They were then given a short quiz to test their understanding of the task and were presented with correct answers afterwards. While reading/listening to the instructions, participants were shown the dollar bills that would be used to pay them after the experiment to ensure the understanding that their decisions had actual and significant consequences.

For all four conditions, participants reported their decisions to the experimenter at each trial by writing down their assigned IDs and the number of hours they decided to contribute to the group project on paper slips. The experimenter would then collect and record the decisions. Feedback was given at the end of each stage (week). In the feedback, they were informed about each member's contribution of hours to the group project, the corresponding payoff for each person for each trial, and sanction results at the end of the second stage (only in the reward and punishment conditions), all by assigned IDs. Payoffs were shown in hours (i.e. not the actual monetary amount they could receive). For conditions with sanction intervention, one day of the week was randomly chosen as the day for determination of reward and punishment at the end of the second week before feedback was given.

Following the feedback of week 2, participants were given a short questionnaire to complete. The questionnaire measured expectation of/trust in others' decisions to cooperate. Each experimental session lasted for about one hour. After completing all three stages of decision-making tasks, they were debriefed, paid (average of \$6), and dismissed individually.

### Measurement

All items were measured using a 7-point Likert scale running from 1 = 'Strongly Disagree' to 7 = 'Strongly Agree'.

### Expectation of/trust in others' cooperation

Five items were used to measure participants' expectations and trust in other group members' cooperation. This scale was an adaptation of the work of Pearce and associates (see Pearce, Bigley, & Branyiczki, 1998; Pearce, Branyiczki, & Bigley, 2000). Sample items included 'I can trust the other group members to put in their share of hours in the group project', and 'I felt that other group members were taking advantage of my contribution' (reverse coded). The alpha reliability for this scale was .79. The average value of within-group agreement ( $r_{wg}$ ) was .74, providing evidence for aggregating the trust measure at the group level.

**Cooperation** Cooperation was measured using the number of hours put into the group project. The average number of hours per day contributed to the group project of a certain week was used as an index of cooperation for that week. Three trial blocks (for three weeks) were created.

### Results

The mean and standard deviation for the mean cooperation in stage 1, stage 2, stage 3, and the mean level of trust in others are presented in Table 2. Before investigating whether cooperation induced through the various means was sustained over trials, we needed to first establish the fact that cooperation indeed increased significantly from stage 1 to stage 2 in our experimental conditions. As in study 1, we combined the reward and punishment conditions into one sanction condition as we were not interested in the differences between rewards and punishments. A group-level ANOVA with repeated measures on the first two trial blocks revealed a significant effect of Trial Block ( $F_{1, 29} = 11.94, p < .01, \text{Effect Size} = .29$ ), and a significant effect of Trial Block  $\times$  Condition interaction ( $F_{2, 29} = 7.34, p < .01, \text{Effect Size} = .34$ ). The means reported in Table 2 indicate that the

Table 2. Experiment 2: Mean cooperation (*SD*) in stages 1, 2, 3, and mean (*SD*) trust in others in stage 2 of the three experimental conditions

Conditions	Contribution (stage 1)	Contribution (stage 2)	Contribution (stage 3)	Trust in others (stage 2)
Punish	3.89 (2.02)	5.96 (2.44)	2.63 (2.56)	2.88 (1.21)
Reward	5.00 (2.95)	6.12 (2.86)	3.50 (2.57)	3.09 (1.40)
Sanction (punish & reward)	4.45 (2.48)	6.04 (2.65)	3.07 (2.56)	2.98 (1.30)
Appeal	5.34 (2.49)	6.28 (2.79)	5.01 (3.25)	3.72 (1.66)
Control	4.89 (2.71)	4.58 (2.90)	3.03 (2.89)	3.34 (1.35)

difference between Trial Blocks 1 (mean = 4.89) and 2 (mean = 4.58) was not significant in the control condition. In the two experimental conditions, mean cooperation rates increased from trial block 1 to trial block 2: it increased from 4.45 to 6.04 in the sanctions condition ( $t_{15} = 5.27, p < .01$ ) and from 5.34 to 6.28 in the appeal condition ( $t_7 = 2.40, p < .05$ ).

To test Hypothesis 1a (that cooperation induced by sanctions will wither away when the sanctions are removed) and Hypothesis 4 (that cooperation induced by moral appeal will be sustained even in the absence of appeals), we compared stage 3 cooperation (after sanctions and appeals were removed) with stage 2 in a group-level repeated measures ANOVA with trial block (stage 2 versus stage 3) and condition (sanctions, appeal, and control) as the two factors. This analysis revealed a significant effect for trial block ( $F_{1, 29} = 35.78, p < .01$ , Effect Size = .55) and for the interaction between trial block and condition ( $F_{2, 29} = 3.35, p < .05$ , Effect Size = .19). Follow up *t*-tests show that the mean cooperation in stage 3 (mean = 3.07) was significantly lower than mean cooperation in stage 2 (mean = 6.04) in the sanction condition ( $t_{15} = 9.09, p < .01$ ), but the decrease in cooperation rates was not significant at the .05 level in the moral appeal condition (mean = 5.01 at Stage 3 versus mean = 6.28 at Stage 2;  $t_7 = 2.10, ns$ ) or in the control condition (mean = 3.03 at Stage 3 versus mean = 4.58 at Stage 2;  $t_7 = 1.81, ns$ ). These results provide support for Hypotheses 1a and 4.

To test Hypothesis 1b (that cooperation following the removal of sanctions will be even lower than cooperation rates when sanctions were not

applied in the first place) we performed two comparisons. First, we compared cooperation rates in stage 1 (when sanctions were not applied) to cooperation in stage 3 (when sanctions were removed). Second, we compared cooperation rates in stage 3 between the sanction and control conditions, and between the appeal and control conditions.

A group-level ANOVA with repeated measures on trial block (block 1 and block 3) and condition (sanctions, appeal, and control) revealed a significant effect of trial block ( $F_{1, 29} = 9.55, p < .01$ , Effect Size = .25), and a marginally significant effect of condition ( $F_{2, 29} = 2.69, p = .085$ , Effect Size = .16). Follow up *t*-tests yielded significant differences in mean cooperation between trial blocks 1 and 3 in the sanction condition (mean = 4.45 at Stage 1 and mean = 3.07 at Stage 3,  $t_{15} = 4.02, p < .01$ ), whereas this difference was not significant at the .05 level in the appeal condition (mean = 5.34 at Stage 1 versus mean = 5.01 at Stage 3,  $t_7 = .36, ns$ ) or in the control condition (mean = 4.89 at Stage 1 versus mean = 3.03 at Stage 3,  $t_7 = 1.97, ns$ ). The results in the sanction condition supports Hypothesis 1b.

Another way to test Hypothesis 1b is to compare sanction with control conditions in stage 3. A group-level one-way ANOVA for trial block 3 revealed a marginally significant effect for condition ( $F_{2, 29} = 3.10, p < .06$ ; Effect Size = .18). Follow up *t*-tests suggest that sanctions did not differ from the control condition (mean = 3.07 for the sanction condition versus mean = 3.03 for the control condition,  $t_{22} = .05, ns$ ) but the appeal condition did marginally (mean = 5.01 for the appeal condition versus mean = 3.03 for

the control condition,  $t_{14} = 1.77, p < .10$ ). Thus we conclude that Hypothesis 1b is only partially supported in Experiment 2.

To test Hypothesis 2 (that trust in other group members following the application of sanctions will be lower than trust when sanctions were not applied) we compared trust in the sanction condition with trust levels in the combined appeal and control conditions. A one way ANOVA with trust as the dependent measure revealed a significant effect for condition ( $F_{1, 30} = 7.92, p < .01$ ). Follow up *t*-tests suggest that trust among group members was significantly lower in the sanction condition compared to the non-sanction conditions (mean = 2.99 for the sanction condition vs. mean = 3.53 for the non-sanction conditions,  $t_{30} = 2.81, p < .01$ ). This supports Hypothesis 2, and suggests that the mechanism through which sanctions decrease cooperation may be through reduction in trust in others.

Hypothesis 3 suggests that trust mediates the effect of the removal of sanctions on cooperation. We conducted the 3-step analysis (Baron & Kenny, 1986) to test the mediation effect comparing sanction, appeals, and control conditions. Step 1 and step 2 analyses show that there was a marginally significant effect of condition on stage 3 cooperation rates ( $F_{2, 29} = 3.10, p < .06$ ), a significant effect of condition ( $F_{1, 30} = 7.92, p < .01$ ) on trust, and a significant effect of trust on stage 3 cooperation rates ( $\beta = .52, p < .01$ ). Adding trust as a covariate in the step 3 analysis results in trust being the only significant effect ( $F_{1, 28} = 6.36, p < .05$ ) whereas the effect of condition was no longer significant ( $F_{2, 28} = 1.19, ns$ ). These results support the conclusion that trust mediates the impact of moral appeal on sustained increase in cooperation.

### Discussion

The results replicate and extend the findings in Experiment 1 in important ways. Similar to Experiment 1, we found that sanctions resulted in higher rates of cooperation than when they were not applied. Also, similar to Experiment 1, we found that the increase was not sustained. When sanctions were removed, cooperation rates

fell to a level lower than that achieved when they were not applied. This supports our view that sanctions undermine natural inclinations to cooperate. However, unlike Experiment 1, we did not find that the removal of sanctions resulted in lower cooperation than the control condition. This is an anomalous result that needs further examination.

We found that trust in others was undermined by the introduction of sanctions. This is similar to Mulder *et al.*'s (2006) results and allows us to confidently infer that sanctions do undermine important group outcomes such as group members' trust in one another.

Our results indicate that moral appeals increase cooperation and they do so in a sustainable fashion. Preliminary evidence seems to indicate that appeals increase cooperation by increasing trust in fellow members of the group. One might speculate that the effects of appeals in a public goods dilemma on cooperation were due to the 'demand characteristics'—participants behave the way they thought the experimenter wanted them to behave. If this were the case, then punishing the 'least contributor' or rewarding the 'most contributor' would have sent a stronger message to participants than simply giving appeals. However, our data did not show a greater level of cooperation in the conditions that involved sanctions than that in the appeal condition. On sustaining cooperation, one may argue that participants in the appeal condition maintained a higher level of cooperation because one could not simply 'remove' an appeal from a situation; once the appeal was made, it stayed in their mind. This reasoning is plausible and suggests that means that induce intrinsic motivation and increase trust in coworkers tend to have a lasting effect. On the other hand, we did not introduce any 'counter appeals' that could threaten the buy-in of the appeal. Future research should examine further the strengths of benefits from appeals.

### General discussion and conclusions

The results of the two experiments echo earlier research findings of the over-justification effect

and Titmuss' 'surprising finding' of the negative effect associated with the commercialization of the blood donor system. They are also consistent with Mulder et al.'s (2006) finding on the relationship between punishment, trust in others' internal motivation for cooperation, and cooperation after punishment is removed. However, our study differs from Mulder et al.'s paper in several important ways. First, we showed that the removal of sanctions results in lowered levels of cooperation compared to when sanctions were not applied in one of the two experiments. Second, we demonstrated that appeal can be a viable alternative to sanctions without negative side effects. Third, we examined both reward and punishment as formats of sanctions and found that the negative effects existed. Thus, our findings make important contributions in deepening our understanding about *why* the effect of sanction would not sustain over time and more importantly, *how* it happened and *what* might be done to circumvent such problems.

The articulation of the psychological processes behind the phenomenon enables us to explain and predict behavior in situations beyond the ones investigated in this study. For example, in groups with strong cooperative norms, individuals are likely to cooperate; however, our results indicate that if the individual infers others' behavior as caused by mainly extrinsic reasons (i.e. cooperate because of the group norms), this behavior may not be sustained in situations where such pressure is lifted. Yet it is possible that if the individual believes that other members identify with and internalize the group norm, he/she might continue contributing to the group. The finding that cooperation increases with the use of sanctions and decreases once they are removed may be viewed as evidence of the need for the continuous use of such systems to promote cooperation. However, the adoption of the sanction systems could exacerbate the conditions that are supposed to make it necessary—they may become self-fulfilling prophecies in that they will induce the selfish motives that they are supposed to control. To avoid the 'addictive' effect (i.e. the more of it we have, the more we need it and the more we come to depend on it) of such systems,

we contend that extreme caution is needed in designing and implementing any sanction systems in inducing cooperation or other desirable behaviors.

While the conventional wisdom in capitalist societies advocates self-interest and material gain, we found that using appeals could actually achieve similar effects in cooperation induction as using monetary incentives. More importantly, cooperation induced through this approach was more likely to sustain than that induced through means involving sanctions only.

Our results also suggest that including the attribution process in the study of inducing and sustaining cooperation in social dilemmas will change our views on the real power of the motivational solutions that have been proposed to resolve social dilemmas. Unlike the social orientations that are part of a person's stable characteristics (Messick & McClintock, 1968), the extent to which a person makes an internal attribution of others' behavior is often a function of the social environments. Theories of and empirical research on social cognition (see Fiske & Taylor, 1984) may shed light on what to change in the social cues to bring out intrinsic motivation more easily and make internal attribution of others more likely to occur.

With numerous situations in organizations sharing the characteristics of a social dilemma, and cooperation among team members, between departments and the like so vital to the success of the organization, the managerial implication of our findings is evident. Organizations should not overly rely on external forces (e.g. reward, punishment, authority) to influence employee's behavior, for extensive use of such forces is likely to reduce their natural propensity to perform the behavior and to alter their view of why they are doing it. Meanwhile, organizations should be more deliberate in delivering moral appeals in promoting cooperation. This can be done at all levels of management. For example, organizational leaders can give regular motivational talks to employees to emphasize the value of cooperation in general. Recognition of cooperation can also be included in the organization's mission/value statement. Information about the nature of the dilemma and

how cooperation may benefit everyone can be disseminated at meetings and in conversation with individual group members. The organization and management should also set up examples to facilitate the social learning of cooperation (Bandura, 1977) both inside and outside the organization such as contributing to public goods such as public broadcasting or public urban constructions. These practices would help to build a culture of cooperation, which in turn serves as appeals to induce cooperation.

Dawes (1980) pointed out that knowledge, morals, and trust are three important factors that enhance cooperation in social dilemma situations. The appeal used in this study was related to the moral part, yet we believe that the more immediate factor for sustainable cooperation is one's belief about why they and others contribute. Even if one may be under pressure of moral standards, without internalizing the superiority of cooperation, one's cooperation is not likely to be sustained.

## Acknowledgment

This article is dedicated to Samuel Komorita, to remember his great contribution to the social dilemma research field and his great advice to the first author of this paper, Xiao-Ping Chen.

## References

- Albanese, R., & Van Fleet, D. D. (1985). Rational behavior in groups: The free-riding tendency. *Academy of Management Review*, *10*, 244–255.
- Andreoni, J., Hardbaugh, W., & Vesterlund, L. (2003). The carrot or the stick: Rewards, punishment, and cooperation. *American Economic Review*, *93*, 893–902.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.
- Caldwell, M. D. (1976). Communication and sex effects in five-person prisoner's dilemma game. *Journal of Personality and Social Psychology*, *33*, 273–280.
- Cummings, L. L., & Bromily, P. (1996). The organizational trust inventory (OTI): Development and validation. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in Organizations* (pp. 302–330). Thousand Oaks, CA: SAGE.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*, 169–193.
- Fehr, E., & Gächter, S. (1999). *Cooperation and punishment in public goods experiments*. Mimeo: Institute for Empirical Research in Economics, University of Zurich.
- Fiske, S. T., & Taylor, S. E. (1984). *Social Cognition*. Reading, MA: Addison-Wesley.
- George, J. M. (1992). Extrinsic and intrinsic origins of perceived social loafing in organizations. *Academy of Management Journal*, *35*, 191–202.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 220–226). New York: Academic Press.
- Jones, E. E., Davis, K. E., & Gergen, K. E. (1961). Role playing variations and their informational value for person perception. *Journal of Abnormal and Social Psychology*, *63*, 302–310.
- Kelly, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–238). Lincoln, NE: University of Nebraska Press.
- Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, *24*, 183–214.
- Komorita, S. S. (1987). Cooperative choice in decomposed social dilemmas. *Personality and Social Psychology Bulletin*, *13*, 53–63.
- Komorita, S. S., & Parks, C. D. (1995). Interpersonal relations: Mixed-motive interaction. *Annual Review of Psychology*, *46*, 183–207.
- Maki, J. E., Hoffman, D. M., & Berk, R. A. (1978). A time series analysis of the impact of a water conservation campaign. *Evaluation Quarterly*, *2*, 107–118.
- Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly*, *47*, 534–559.
- McCusker, C., & Carnevale, P. J. (1995). Framing in resource dilemmas: Loss aversion and the moderating effects of sanctions. *Organizational Behavior and Human Decision Processes*, *61*, 190–201.
- Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental

- games. *Journal of Experimental Social Psychology*, 4, 1–25.
- Miles, J. A., & Greenberg, J. (1993). Using punishment threats to attenuate social loafing effects among swimmers. *Organizational Behavior and Human Decision Processes*, 56, 246–265.
- Mulder, L. B., van Dijk, E., De Cremer, D., & Wilke, H. A. M. (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology*, 42, 147–162.
- Olson, M. (1965). *The logic of collective action*. Cambridge, MA: Harvard University Press.
- Pearce, J. L., Bigley, G. A., & Branyiczki, I. (1998). Procedural justice as modernism: Placing industrial/organisational psychology in context. *Applied Psychology: An International Review*, 47, 371–396.
- Pearce, J. L., Branyiczki, I., & Bigley, G. A. (2000). Insufficient bureaucracy: Trust and commitment in particularistic organizations. *Organization Science*, 11, 148–162.
- Pillutla, M., & Chen, X. P. (1999). Social norms and cooperation in social dilemmas: The effects of context and feedback. *Organizational Behavior and Human Decision Processes*, 78, 81–103.
- Platt, J. (1973). Social traps. *American Psychologist*, 28, 641–651.
- Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology*, 28, 363–392.
- Thibaut, J. W., & Riecken, H. W. (1955). Some determinants and consequences of the perception of social causality. *Journal of Personality*, 24, 113–133.
- Titmuss, R. M. (1970). *The gift relationship*. London: Allen and Unwin.
- Winett, R. A., Kagel, J. H., Battalio, R. C., & Winkler, R. C. (1978). The effects of monetary rebates, feedback, and information on residential energy conservation. *Journal of Applied Psychology*, 63, 73–80.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110–116.
- Yamagishi, T. (1988). Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly*, 51, 32–34.

Paper received 22 February 2007; revised version accepted 20 August 2008.

## Biographical notes

XIAO-PING CHEN is professor of management in the Michael G. Foster School of Business at the University of Washington. She is also associate editor for *Organizational Behavior and Human Decision Processes*. She received her PhD from the University of Illinois. Her research interests include cross-cultural studies of cooperation and competition in social dilemmas, leadership, organizational citizenship behavior, passion and creativity, and Chinese guanxi.

MADAN M. PILLUTLA is professor and chair of organizational behavior at the London Business School. His research interests include the investigation of fairness, trust, and competition in interpersonal and work relationships. Among his current research projects he is investigating the origin of fair behavior in exchange situations. Other research projects investigate how people can rebuild trust following transgressions in interpersonal relationships, and whether people like to associate with more competent or less competent people in their work-groups.

XIN YAO is an assistant professor of Management at the W. Frank Barton School of Business, Wichita State University. She received her PhD from the University of Washington. Her current research interests include entrepreneur motivation, early stage venture financing, social dilemmas, and emotion labor.

## Appendix 1

Dear All,

I am writing to you to help you better understand our current work situation and to encourage everyone's greater input in this group project.

Explore is one of the largest shipping companies in continental America. They constructed a comprehensive plan for fully entering the developing markets around the globe with improving online service being an essential part of the effort. It is our honor to play a critical role in implementing the strategy. Correspondingly, our dedication shown in the work is critical for building a strong long-term relationship with Explore. If we are able to demonstrate to our client that we value the opportunity to work with them and to be part of their mission, we can instill greater



trust in our relationship. This eventually leads to a good reputation and higher status in our field, and subsequently, business opportunities for our company.

We are a company among the few with a new form of employer–employee and between-employee relationships. The fact that we do not meet face-to-face for work presents both great benefits as well as challenges. This group project is a good example. The reality is that every group member can choose to withdraw his/her effort for the group project while still reaping the benefit as long as the other members contribute to this common effort. But the predicament exists because if no one puts in any time or everyone puts in only very few hours into the group project, the whole group will be worse off (in terms of each individual's payoff) than if everyone goes full speed for the project.

Since each group member does not really know the other members, your decision of contribution will be based on the faith that others also understand the quandary and that everyone sees the fact that common effort is a preferable solution to outguessing each other's contribution in order to benefit from others' hard work. We believe that once our employees understand the situation, they will be willing to put forth their individual resources for the benefit of all. Thus we hope to see each of our employees devote a fair amount of their time to common goals in similar situations.

Sincerely Yours,  
J. P.