

Tilburg University

The effect of feedback on support for a sanctioning system in a social dilemma

Mulder, L.B.; van Dijk, E.; De Cremer, D.; Wilke, H.

Published in:
Journal of Economic Psychology

Publication date:
2005

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Mulder, L. B., van Dijk, E., De Cremer, D., & Wilke, H. (2005). The effect of feedback on support for a sanctioning system in a social dilemma: The difference between installing and maintaining the sanction. *Journal of Economic Psychology*, 26(3), 443-458.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Economic Psychology 26 (2005) 443–458

JOURNAL OF
**Economic
Psychology**

www.elsevier.com/locate/joep

The effect of feedback on support for a sanctioning system in a social dilemma: The difference between installing and maintaining the sanction

Laetitia B. Mulder^{a,*}, Eric van Dijk^a, Henk A.M. Wilke^a,
David De Cremer^b

^a *Department of Social and Organizational Psychology, Leiden University, PO Box 9555,
Leiden 2300 RB, The Netherlands*

^b *Department of Economic and Social Psychology, University of Tilburg, PO Box 90153,
5000 LE Tilburg, The Netherlands*

Available online 22 January 2005

Abstract

When do people support sanctioning systems in social dilemmas? Sanctions increase collective efficiency, but have the disadvantage of restricting people's autonomy. This paper studies the effects of feedback about collective (in)efficiency and the influence of the presence or absence of a sanctioning system. The results show that, except after feedback about collective inefficiency, people were reluctant to support installation of a sanctioning system. When a sanctioning system was already present, however, sanction support was strong and not affected by feedback. Interestingly, the presence of a sanctioning system increased pessimism about attaining collective efficiency. This suggests that the mere presence of a sanctioning system creates the need to have that sanctioning system, and that installing one can have irreversible consequences.

© 2005 Elsevier B.V. All rights reserved.

* Corresponding author. Tel.: +31 0 71 527 3718; fax: +31 0 71 3619.
E-mail address: lbulder@fsw.leidenuniv.nl (L.B. Mulder).

JEL classification: H41; D8

PsycINFO classification: 3000

Keywords: Sanctions; Social dilemma; Feedback

1. Introduction

Many decisions consist of a choice between furthering our personal interests and furthering group interests. Such choice situations are studied in research on “social dilemmas”. In social dilemmas, acting on personal interests (defection) yields higher outcomes for individual group members than acting on group interests (cooperation). At the same time each individual attains a higher outcome when all group members act on group interests rather than when all group members act on personal interests (see for overviews: Dawes, 1980; Komorita & Parks, 1995; Kopelman, Weber, & Messick, 2002; Messick & Brewer, 1983). An example of a social dilemma is the decision whether or not to pay a television license fee to have public television provision. In this example, public television is more likely to be realized when everybody pays the fee than if nobody pays the fee. However, not paying the fee is more advantageous for an individual than paying the fee, as one individual’s contribution only makes a small difference in realizing public television. Many other situations in society (e.g. paying taxes, using public transport, environmental friendly waste treatment), can also be defined as social dilemmas.

One way to increase cooperation in social dilemmas is the introduction of a sanctioning system (Caldwell, 1976; Fehr & Gächter, 2002; Harvey, Bell, & Birjulin, 1993; Wit & Wilke, 1990; Yamagishi, 1986, 1988b, 1992). Such sanctioning systems sanction defection and therefore make it less attractive. For example, people who do not pay their TV license fee could be fined or removed from cable subscription. These kinds of punishments decrease the attractiveness of not paying and, therefore, by definition, increase the attractiveness of paying. The goal of a sanctioning system is to promote collective efficiency. In accordance with this, sanctioning systems have indeed been found to increase cooperation (Caldwell, 1976; Eek, Loukopoulos, Fujii, & Gärling, 2002; Fehr & Gächter, 2002; McCusker & Carnevale, 1995; Van Vugt & De Cremer, 1999; Wit & Wilke, 1990; Yamagishi, 1986, 1988b, 1992).

Sanctioning systems may increase cooperation in two possible ways. Firstly, it makes defection relatively less attractive for the individual group member. Secondly, because a sanction also punishes the other defecting people involved in the social dilemma, it can increase the trust that others will cooperate (Yamagishi, 1986, 1988a, 1988b, 1992). Hobbes in his *Leviathan* (1651/1909) already regarded the latter reason as the explanation of a sanctioning system’s success in increasing cooperation (in: Yamagishi, 1992).

1.1. Two-edged sword

Although findings from experimental research suggest that a sanctioning system increases collective efficiency, it may also be a two-edged sword. On the one hand

a sanctioning system promotes collective efficiency, but on the other hand it could make people feel that they are losing the right to make their own behavioral choices. Indeed, a system in which people are being punished for a certain behavior constitutes a threat to people's freedom of choice (Brehm, 1966) and could therefore lead to reactance. In line with this, research on leadership has shown that installing a leader is perceived as a threat to freedom of choice (Van Dijk, Wilke, & Wit, 2003). We expect that the same may hold for supporting the installation of a sanctioning system. Thus, when deciding whether or not to install a sanctioning system, the threat to freedom of choice could conflict with the wish to reduce collective inefficiency. Put differently, the wish to reduce collective inefficiency may induce people to prefer a sanctioning system, but the wish to maintain freedom of choice may induce people to oppose a sanctioning system.

When considering when people would prefer a sanctioning system and when they would not, one may be inclined to concentrate on situations in which a sanctioning system is not present yet. However, in many situations in society, a sanctioning system is already present. In spite of this, there has been no investigation of support for a sanctioning system under circumstances where a sanctioning system is part of the status quo. We argue that support for a sanctioning system will depend on whether in the current situation there is a sanctioning system or not. Considering having a sanctioning system in a situation without a sanctioning system implies either *installing* a sanctioning system (gaining collective efficiency) or *not installing* (maintaining decisional freedom), whereas in a situation with a pre-existing sanctioning system this would imply either *maintaining* a sanction (maintaining collective efficiency) or *abolishing* it (gaining decisional freedom).

How does the presence of a sanctioning system influence people's preference for the sanctioning system? An answer to this question can be found in the concept of "loss aversion" (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991). Loss aversion refers to the finding that "losses loom larger than gains", which means that losses or disadvantages have a greater impact on people's preferences than equal gains or advantages. When the status quo situation functions as reference point for evaluating different options, the disadvantages of leaving the status quo situation are weighted more heavily than its advantages, resulting in a bias in favor of the status quo (Kahneman, 1992; Samuelson & Zeckhauser, 1988). These insights on the "status quo bias" suggest that support for a sanctioning system depends on whether the sanctioning system is currently present or not. The disadvantages of having a sanctioning system (i.e., having little decisional freedom) will be weighted more heavily in a status quo situation without sanctioning system than in a status quo situation with a sanctioning system. Similarly, the disadvantages of not having a sanctioning system (i.e., possible collective inefficiency) will be weighted more heavily in a status quo situation with a sanctioning system than in a status quo situation without a sanctioning system. This would mean that support for a sanctioning system is lower when there is no sanctioning system present than when there is a sanctioning system present.

Thus, on the basis of the status quo bias we expect that in the absence of a sanctioning system, people will more strongly support a sanctioning system than in the

presence of a sanctioning system. Baron and Jurney (1993) used similar reasoning within the context of supporting laws or regulations. In their scenario experiment people were reluctant to vote for certain proposed laws, because people felt that they harmed the individual's right to choose. However, people did not want to repeal the same laws when they already had been installed.

1.2. *The influence of feedback*

As we noted earlier, the goal of sanctioning systems is to increase collective efficiency. It therefore seems reasonable to assume that information about collective (in)efficiency plays an important role in the decision to support a sanctioning system or not. This reasoning relates to the notion proposed by Samuelson (1993) that the status quo bias is due to insecurity about unknown situations: people often know little about what to expect in the future. Prior information about how (inefficient) the collective has been in furthering the group interest, may reduce such insecurity. Thus, in a social dilemma situation, information about collective efficiency or collective inefficiency may reduce the status quo bias. Bettenhausen and Murnighan (1985) stated that feedback activates information that serves as a basis for future expectations and interactions. In a similar vein, we reason that providing feedback about collective (in)efficiency will decrease the extent to which people show a status quo bias when deciding whether or not to support a sanction. In other words, when feedback is available, people would be less subject to the status quo bias (cf. Inman & Zeelenberg, 2002).

This basic hypothesis, however, also raises two new questions. What is the influence of feedback on collective efficiency or inefficiency on sanction support in a situation in which a sanctioning system is absent? And what is the influence of feedback in a situation in which a sanctioning system is present?

1.3. *Sanction support when "things go wrong"*

The effect of collective inefficiency information on support for a sanctioning system, has, as far as we know, never been investigated. Earlier research on other structural solutions has shown that people are willing to appoint a leader (Rutte & Wilke, 1984; Samuelson, 1993; Van Dijk et al., 2003; Van Vugt & De Cremer, 1999), to privatize common supplies (Samuelson, 1993), or to give up free access to common supplies in general (Samuelson & Messick, 1995), particularly when they are informed that their group failed to use the collective good efficiently. So, there is support for the notion that people want to change something about the initial dilemma situation and give up some autonomy when there is an indication that "things go wrong" (i.e., in case of collective inefficiency). Therefore, in addition to our previous hypothesis, we argue that people will be in favor of installing a sanctioning system when there is collective inefficiency.

How will information about collective inefficiency affect support for a sanctioning system when there is already a sanctioning system present? When things go wrong, it is likely that people may want to change something about the situation. In the pres-

ence of a sanctioning system, however, it is less likely that the preferred course of action would be to abolish that sanctioning system. After all, by abolishing a sanctioning system there is danger of drifting even further away from the goal of collective efficiency. Consequently, abolishing a sanctioning system is not likely to be regarded as a helpful solution, and will probably even be regarded as a counterproductive solution to the problem of collective inefficiency. We therefore anticipate that when a sanctioning system is present, and there is information on collective inefficiency, people will still want to maintain the sanctioning system.

Summarizing, we expect that knowing about collective inefficiency will result in support for a sanctioning system, irrespective of whether the sanctioning system was absent or already present.

1.4. Sanction support when “things go right”

How will information about collective *efficiency* affect support for a sanctioning system? In a situation without a sanctioning system, we expect that acting on information about collective efficiency is not likely to lead to support for installation of a sanctioning system. If “things go right” anyway, there is not much reason for people to feel that they would gain a lot by installing a sanctioning system.

It may be a different story, however, if a sanctioning system is already present. Using all the information at hand, people would take into account both collective efficiency and the presence of the sanctioning system. Considering these two facts, a highly plausible conclusion is that people will reason that collective efficiency is due to the presence of the sanctioning system. In other words, people might hold the sanctioning system, at least partly, responsible for the collective efficiency and attribute other people’s cooperation to the sanctioning system (Mulder, Van Dijk, De Cremer, & Wilke, submitted for publication; Weiner, 1985, 1986). Consequently, it may be regarded as risky to abolish the sanctioning system because that could bring about a decline in cooperation and thus a reduction in collective efficiency.

In sum, when there is information about collective efficiency, we expect the presence of the sanctioning system to lead to stronger sanction support than the absence of a sanctioning system.

1.5. This study

Summarizing, in the present study we expect that without any information about collective (in)efficiency, people show the status quo bias, so the presence or absence of a sanctioning system will determine sanction support. Hence, in a situation with a sanctioning system, sanction support will be stronger than in a situation without a sanctioning system. When there is information about collective (in)efficiency, however, people will make use of that information instead of showing a status quo bias. In the case of collective inefficiency this will lead to high sanction support irrespective of the presence or absence of a sanctioning system. In the case of collective efficiency, stronger sanction support will be given in the presence of a sanctioning system than in the absence of a sanctioning system.

These predictions may also be presented differently by comparing the effect of feedback in a situation without a sanctioning system with the effect of feedback in a situation with a sanctioning system. The predictions we formulated above imply that when there is no sanctioning system present, people will only support a sanctioning system when there is collective inefficiency. When there is already a sanctioning system present, however, people will support the sanctioning system irrespective of the information people have about collective (in)efficiency.

To test these expectations we conducted a public good experiment in which we varied the presence of a sanctioning system and feedback about collective (in)efficiency. Participants' reactions were measured by means of their sanction support.

1.6. Success estimates

Our main interest in this study is sanction support. In addition, we expect support for a sanctioning system to be related to people's expectations about achieving collective efficiency. After all, if people estimate that the chance of attaining collective efficiency is low, this may be a reason for them to support the installation of a sanctioning system. Moreover, people might estimate the chances of collective efficiency as higher with a sanctioning system than without a sanctioning system. To explore this, we included a measurement of how high people estimated the probability of succeeding to attain collective efficiency.

2. Method

2.1. Design and participants

The participants were 124 (39 male and 85 female) students from Leiden University, with a mean age of 21. They were each paid DFL 8.50 (at the time of the study was conducted one Dutch guilder equaled USD 0.40).

Each participant was randomly assigned to one of the 6 experimental conditions of a 2 (sanction: sanction or no sanction) \times 3 (feedback: no feedback, success feedback or failure feedback) between-participants design.

2.2. Procedure

Upon arrival in the laboratory, participants were guided into separate cubicles and were seated behind a computer. They did not have any contact with one another. It was suggested to participants that they were linked to three other participants via the computer and that in this way they formed a four-person group. Then, the properties of the decision-making game they would play were explained. Participants were told that each member of the group owned 10 chips. They could contribute chips to a "common fund" or keep chips for themselves. If the four group members contributed a total of at least 30 chips, the group was given a bonus of DFL 60 that

would be divided equally amongst the four group members, regardless of how much each member contributed. The chips that they did not contribute would directly yield money for themselves. In the *no-sanction conditions* they were told that the chips they kept for themselves would be worth DFL 1 each. In the *sanction conditions* they were told that these chips would be worth DFL 1.25 each. For each chip that they kept for themselves, the experimenter would take a DFL 0.25 (USD 0.10) fine. In this way, the pay-off structure was held constant across both no-sanction and sanction conditions (cf. Wit & Wilke, 1990).

After explaining the game, four questions were asked to check whether participants had understood the dilemma (example: “What will happen when you contribute all your chips to the common fund and you and the other group members together give at least 30 chips to the common fund?”). After answering each of these four questions, participants received feedback about the correct answer.

Next, participants were asked how many chips they contributed to the common fund (ranging from 0 to 10 chips). After having made their decision, participants in the success-feedback condition were informed that the group had contributed at least 30 chips and that they had thus succeeded in attaining the bonus. Participants in the failure-feedback condition were informed that the group contributed less than 30 chips and therefore failed to attain the bonus. Participants in the no-feedback condition were given no information about the success or failure of achieving the bonus.

2.3. Measurement of sanction support

Next, participants were asked to imagine that they were again in a similar decision-making situation. Sanction support was measured by presenting participants with a new decision-making situation. More specifically, they were asked whether they would be in favor or against the presence of a sanctioning system that would deduct a DFL 0.25 fine from each chip that group members kept for themselves (7-point scale, ranging from “1” [strongly against] and “7” [strongly in favor]). Note that if we were to take the same value for the chips, maintaining the sanctioning system in the *sanction condition* would imply that DFL 0.25 would be subtracted from DFL 1.25. In the *no-sanction condition*, however, installing a sanctioning system would imply that DFL 0.25 would be subtracted from DFL 1. So, in the *sanction condition* the sanction would result in the chips yielding DFL 1 and in the *no-sanction condition* the sanction would result in the chips yielding DFL 0.75. To deal with this issue the question of sanction support was asked twice: Firstly, sanction support was measured when the chips were worth DFL 1 and, secondly, when the chips were worth DFL 1.25. The situation that the participant had experienced before (i.e. chips worth DFL 1 in the no-sanction condition and chips worth DFL 1.25 in the sanction condition), was presented first and the other situation was presented last. In the analysis we used the average answer to these two questions as the measure for sanction support. Because the two questions were highly correlated ($r = 0.78$, $p < 0.001$) we can consider this measure to be a reliable one.

2.4. Measurement of success estimate without and with sanctioning system

To measure the perceived probability of attaining collective efficiency without a sanctioning system, we asked participants to estimate the probability of attaining the bonus in a future situation, in which there would be no sanction. We will refer to this as “the success estimate without sanctioning system (SE-without)”. The SE-without was measured by again presenting the two cases of chips worth DFL 1 or chips worth DFL 1.25. Participants were asked to estimate the probability of achieving the bonus without a sanctioning system (on a 7-point scale ranging from “1” [very low] to “7” [very high]). Like the measure of sanction support, the average answer to these two questions (that correlated highly, $r = 0.86$, $p < 0.001$) was used to analyze SE-without.

In the same way, we asked participants to estimate the probability of attaining the bonus when a sanction was present. We will refer to this as “the success estimate with sanctioning system (SE-with)”. Participants were asked to estimate the probability of attaining the bonus in the two cases of chips worth DFL 1 or DFL 1.25 if a sanctioning system were present (with a DFL 0.25 fine on every chip kept for them selves). These two questions correlated highly ($r = 0.79$, $p < 0.001$) and the average answer to these two questions was used to analyze SE-with.

2.5. Attitude towards a sanctioning system

After this, five additional questions were asked concerning participants’ opinions about the sanctioning system. An example of one of these questions is “I think a fine for each chip kept for oneself is necessary”. Participants could indicate the extent to which they agreed on a 7-point answering scale (ranging from “1” [totally disagree] to “7” [totally agree]). The five questions formed a reliable scale ($\alpha = 0.85$), which we will refer to as “Attitude towards a sanctioning system”. After this, the experiment was finished. Participants were debriefed, all the details about the experimental set-up were explained, and they were thanked and paid for their participation.

3. Results

3.1. Contribution to common fund

The average number of chips individual participants contributed was 7.74 (SD = 1.71). A 2 (sanction) \times 3 (feedback) ANOVA showed no effects of sanction or feedback.

3.2. Sanction support

We predicted that in absence of a sanctioning system, people would show little support for a sanctioning system unless there is collective inefficiency, and that in presence of a sanctioning system, people would support the sanctioning system, irre-

Table 1
Support for a sanctioning system as a function of presence of a sanctioning system and feedback

Feedback	<i>M</i> (SD)	
	No sanction	Sanction
No	3.17 (1.55)	4.80 (1.28)
Success	3.57 (1.58)	4.33 (1.60)
Failure	4.33 (1.28)	4.53 (1.11)

spective of the information people have concerning collective (in)efficiency. A 2 (sanction) \times 3 (feedback) ANOVA showed a main effect of sanction on sanction support, $F(1, 124) = 11.49$, $p = 0.001$. Participants in the sanction condition more strongly supported a sanctioning system ($M = 4.55$) than participants in the no-sanction condition ($M = 3.69$). Also, a marginally significant sanction \times feedback interaction was found, $F(2, 124) = 2.69$, $p = 0.07$. Because the pattern was in agreement with our expectations (see Table 1), further LSD post-hoc tests were performed to determine whether support for a sanction was stronger in the sanction condition relative to the no-sanction condition when no feedback or success feedback was given. In line with our prediction, these tests showed an effect of sanction in the no-feedback condition ($p < 0.001$) and in the success feedback condition ($p = 0.04$, one-tailed). As expected, no sanction effect was found in the failure feedback condition.

Also, LSD post-hoc tests were performed to investigate the effect of feedback within the no-sanction and sanction conditions. Within the no-sanction condition, sanction support was higher in the failure-feedback condition than in the no-feedback ($p = 0.004$) and success feedback condition ($p = 0.04$). Within the sanction condition, sanction support in the no-feedback condition did not differ from sanction support in the success- and failure-feedback condition. These results support our expectations.

3.3. Attitude towards a sanctioning system

A 2 (sanction) \times 3 (feedback) ANOVA was performed on the “Attitude towards a sanctioning system” scale. This analysis revealed a sanction \times feedback interaction, $F(2, 118) = 3.71$, $p = 0.03$. This interaction showed a similar pattern of results as the sanction support measure above (see Table 2). Again, LSD

Table 2
Attitude towards a sanctioning system as a function of presence of a sanctioning system and feedback

Feedback	<i>M</i> (SD)	
	No sanction	Sanction
No	3.27 (1.06)	4.21 (1.09)
Success	3.87 (1.04)	4.24 (1.15)
Failure	4.37 (0.78)	4.01 (1.29)

post-hoc tests indicated that when no feedback was given, sanction support was stronger under a sanction than under no sanction ($p = 0.006$). In the success feedback condition this was also the case, although the effect did not reach significance ($p = 0.13$, one-tailed). In the failure feedback condition there was no effect of sanction.

In the no-sanction condition, failure feedback revealed a more positive attitude towards a sanctioning system than no feedback ($p = 0.006$). Failure feedback also revealed a more positive attitude towards a sanctioning system than success feedback, although it was only marginally significant ($p = 0.07$, one-tailed). In the sanction condition, attitudes toward the sanctioning system did not differ between the different feedback conditions.

3.4. Success estimates

To investigate the effect of the presence of a sanctioning system and feedback on the perceived chances of collective efficiency, both without sanctioning system (SE-without) and with a sanctioning system (SE-with), we performed a 2 (sanction) \times 3 (feedback) ANOVA \times 2 (situation: with or without sanctioning system) ANOVA with the factor “situation” as a within participants variable.

Main effects were found of situation ($F [1, 122] = 87.60$, $p < 0.001$), sanction ($F [1, 118] = 16.43$, $p < 0.001$), and of feedback ($F [2, 118] = 7.06$, $p = 0.001$). The main effect of situation indicated that the success estimates were higher for the situation with a sanctioning system ($M = 4.87$, $SD = 1.25$) than for the situation without a sanctioning system ($M = 3.95$, $SD = 1.58$). The sanction main effect indicated that success estimates were higher in the no-sanction conditions ($M_{SE\text{-without}} = 4.28$, $SD = 1.52$; $M_{SE\text{-with}} = 5.39$, $SD = 0.98$) than in the sanction conditions ($M_{SE\text{-without}} = 3.61$, $SD = 1.57$; $M_{SE\text{-with}} = 4.34$, $SD = 1.28$). The main effect of feedback showed that success estimates were lower in the failure feedback conditions ($M_{SE\text{-without}} = 3.37$, $SD = 1.39$; $M_{SE\text{-with}} = 4.38$, $SD = 1.37$) compared to the success feedback conditions ($M_{SE\text{-without}} = 4.31$, $SD = 1.51$, $p = 0.005$; $M_{SE\text{-with}} = 5.33$, $SD = 0.91$, $p < 0.001$) and compared to the no-feedback conditions ($M_{SE\text{-without}} = 4.17$, $SD = 1.68$, $p = 0.02$; $M_{SE\text{-with}} = 4.89$, $SD = 1.27$, $p = 0.03$). Also, a situation \times sanction interaction was found, $F (1, 122) = 4.05$, $p < 0.05$. This interaction suggests that the prospect of a sanctioning system in a future situation, adds to high success estimates to a greater extent in the no-sanction condition ($M_{SE\text{-without}} = 4.28$ and $M_{SE\text{-with}} = 5.39$) than in the sanction condition ($M_{SE\text{-without}} = 3.61$ and $M_{SE\text{-with}} = 4.34$). No other significant effects were found.

Thus, participants estimated their chances of success as lower after failure feedback than after success feedback or no feedback. They estimated their chances of success to be higher with a sanctioning system than without a sanctioning system. However, participants who had experienced a sanctioning system estimated their chances of success (both in a situation with a sanctioning system and a situation without a sanctioning system) as lower than participants who had not experienced a sanctioning system.

4. Discussion

The current study shows that, without information about collective (in)efficiency, people wish to stick to the status quo situation: they support a sanctioning system to a greater extent in the presence of a sanctioning system than in the absence of a sanctioning system. When there is information about collective inefficiency, sanction support is high, irrespective of the presence of a sanctioning system: when there is no sanctioning system present, there is a desire to install a sanctioning system, and when it is there already, there is a wish to maintain it. When there is information about collective efficiency, people do not want to install a sanctioning system in the absence of one, and wish to maintain it in presence of one. Put differently, when there is no pre-existing sanctioning system, people do not support a sanctioning system in a public good dilemma, unless “things went wrong”. When a sanctioning system is already present, they support it irrespective of whether “things went wrong” or “things went right”.

4.1. *Installing a sanctioning system (or not)*

We reasoned that the loss of individual autonomy or the right to choose would inhibit the wish to install a sanctioning system when it does not exist. Yet, earlier research indicated that when the group has failed to use the collective good efficiently, there is an increased preference for structural solutions that reduce an individual's freedom of choice, such as leadership and giving up free access to the common good (Rutte & Wilke, 1984; Samuelson, 1993; Samuelson & Messick, 1995; Samuelson, Messick, Rutte, & Wilke, 1984; Van Dijk et al., 2003; Van Vugt & De Cremer, 1999). The leadership study undertaken by Van Vugt and De Cremer (1999) demonstrated this effect by focusing on leaders with sanctioning power. Our study shows that even without reference to leadership, a preference for a sanctioning system arises under conditions of collective inefficiency. This suggests that the wish for a sanctioning system “when things go wrong” extends beyond the domain of leadership.

On a more general level, the wish to install a sanctioning system when “things go wrong” and not to install a sanctioning system when “things go right”, may be understood by describing our findings in terms of what has been described as a strategy of “win–stay, lose–change” (Liebrand & Messick, 1996; Macy, 1995). When things go right, people do not want to change the situation (“win–stay”) and when something goes wrong, they want to change something about the situation (“lose–change”). This simple heuristic thus seems to help people decide that a situation of collective inefficiency needs to be changed, and one way of achieving this is the adoption of a sanctioning system.

4.2. *Holding on to a sanctioning system*

The second important contribution of this study is that we have also looked at situations in which a sanctioning system was already present. Sanctioning systems

are present in many real life situations and, as such, have often been taken for granted. Therefore, it is interesting and necessary to examine sanction support in a situation in which a sanctioning system already exists. The results of our study show that the presence of a sanctioning system indeed raises support for a sanctioning system and, importantly, that the feedback people receive does not alter sanction support.

This wish to support a pre-existing sanctioning system after having received *success* feedback can also be described as a “win–stay” strategy: because things have appeared to “go right”, people adhere to the existing situation and do not want to get rid of the sanctioning system. This wish to support a pre-existing sanctioning system after having received *failure* feedback, however, does not seem to be in line with the “lose–change” strategy. After all, failure feedback should lead to the wish to change the existing situation, something our present results did not show. However, as we already reasoned in our introduction, abolishing the sanctioning system might not be regarded as a helpful solution to the problem of collective inefficiency. Nevertheless, one could imagine that people might wish to change the situation, but all things being equal, they will probably prefer a change that would actually improve the situation, rather than make it worse by removing the sanction. In other words, people do not merely want “a change”, they want “a change for the better”.

4.3. Negative side effects of a sanction

In addition to the results with regard to sanction support, we found that people were generally more pessimistic about attaining the collective goal when they had experienced a sanctioning system than when they had not. So, the experience with a sanctioning system actually made people generally pessimistic instead of optimistic about collective efficiency. This new insight represents a third contribution to the knowledge of sanction systems, and in fact can be regarded as quite controversial because it places Hobbes’ notion – that a sanction creates positive expectations about fellow group members’ cooperation – in a different perspective. It is true that, in line with Hobbes’ ideas, participants in our experiment estimated the probability of attaining the bonus higher with a sanction than without a sanction. More interestingly, however, participants were more pessimistic about attaining the bonus when having experienced a sanctioning system than when not having experienced a sanctioning system. This specific observation suggests that a sanctioning system may create expectations that group members are non-cooperative; a finding that requires further research.

We can relate these findings to the ideas put forward by Cialdini (1996). He argued that surveillance systems communicate the message that group members are not to be trusted, which suggests that undesirable behavior can be expected. In our study, because of the presence of a sanctioning system, it may have occurred to people that group members would defect if it were not for the sanctioning system. In this way, people may conclude that their group actually needed a sanctioning system to attain collective efficiency (“Why else is there a sanctioning system?”) giving rise to a sense of inability to perform well without a sanctioning system (c.f. Taylor,

1976; Yamagishi, 1988a). Success in attaining collective efficiency may then be ascribed to the presence of the sanctioning system. Even when it is known that the group has failed to achieve collective efficiency, people may still want to stick to the sanctioning system because abolishing it could drive them even further away from collective efficiency. In this way, a sanctioning system may be regarded as indispensable as its mere presence creates the need to maintain that sanctioning system.

The current study was not designed to study behavioral consequences of a sanctioning system. It may be noted that we did not observe a difference in cooperation levels between the no-sanction and sanction conditions. This is in line with previous results on “step-level good” dilemmas (social dilemmas in which the collective goal is realized once the group members reach a certain level of contribution). Research has shown that in such dilemmas, people strongly anchor their decision on the “equal division rule” (e.g. Allison & Messick, 1990; Samuelson & Allison, 1994; Van Dijk, Wilke, Wilke, & Metman, 1999; Van Dijk & Wilke, 1996), which is the rule that group members should contribute an equal part of the total amount required to realize the collective goal. In the current study this came down to an average contribution of 7.5 chips by each group member. Indeed, participants appeared to have based their behavior on this equal division rule as the average number of chips contributed to the group in the current study was 7.74 and varied little.

Even though there were no behavioral effects of the sanctioning system in the social dilemma in the current study, we do anticipate that a sanctioning system may have behavioral implications that should not be disregarded. It should be noted that expectations about future outcomes for the group were negatively affected by the mere presence of the sanctioning system. Thus, it is conceivable that experiencing a sanctioning system may have a negative influence on behavior in new encounters. Experiencing a sanctioning system in one situation may lead to non-cooperative expectations and lack of cooperation in the other. Research by Mulder et al. (submitted for publication) provides some support for this. They showed that a sanctioning system could undermine interpersonal trust, which can result in decreased cooperation in a similar social dilemma situation without a sanctioning system.

The possible implications of sanctioning systems for expectations about fellow group members complement research that has demonstrated other negative effects of sanctions. Work by Deci and colleagues (e.g. Deci, 1971; Deci, Koestner, & Ryan, 1999) demonstrated that introducing incentives for behavior that people show spontaneously can undermine intrinsic motivation, that is, the motivation to choose behavior out of free will, with the behavior itself as the only incentive. Similarly, a sanction could “externalize” cooperative behavior in the sense that it takes away the intrinsic motivation to cooperate. Also, in the context of social dilemmas, Tenbrunsel and Messick (1999) and Gneezy and Rustichini (2000) showed that a (weak) sanctioning system can decrease cooperation, because it makes people define the social dilemma situation in business terms rather than in ethical terms.

Thus, there are indications that a sanctioning system may have adverse effects in at least two ways. It may not only undermine internal motivations to cooperate (Deci, 1971; Deci et al., 1999; Gneezy & Rustichini, 2000; Tenbrunsel & Messick,

1999), but it may also (as our study shows) make people pessimistic, possibly because it decreases expectations of non-cooperation.

4.4. Future research and concluding remarks

Future research is needed to investigate the relations between sanctions and behavioral expectations more systematically. Of course, there is good reason to believe that sanctions increase expectations of other people's cooperation (Yamagishi, 1986, 1988a, 1988b, 1992). However, if we combine our study with the ideas put forward by Cialdini (1996) and Taylor (1976), we think there is reason to suspect that a sanction may also decrease expectations of other people's cooperative intentions. Further research in which an existing sanctioning system is removed could indicate whether (having experienced) a sanctioning system reduces actual expectations of other people's cooperative intentions and consequently causes a decrease in cooperation.

All in all, the present findings provide sufficient evidence to conclude that the mere presence of a sanctioning system creates the need to have that system. Therefore, one should think twice about installing a sanctioning system in social dilemma settings, for it may have irreversible consequences.

Acknowledgement

The authors would like to thank Kate Hudson for her comments.

References

- Allison, S. T., & Messick, D. M. (1990). Social decision heuristics in the use of shared resources. *Journal of Behavioral Decision Making*, 3, 195–204.
- Baron, J., & Jurney, J. (1993). Norms against voting for coerced reform. *Journal of Personality and Social Psychology*, 64, 347–355.
- Bettenhausen, K., & Murnighan, J. K. (1985). The emergence of norms in competitive decision-making groups. *Administrative Science Quarterly*, 30, 350–372.
- Brehm, J. W. (1966). *A theory of psychological reactance*. Oxford, England: Academic Press.
- Caldwell, M. D. (1976). Communication and sex effects in a five-person Prisoner's Dilemma Game. *Journal of Personality and Social Psychology*, 33, 273–280.
- Cialdini, R. B. (1996). Social influence and the triple tumor structure of organizational dishonesty. In D. M. Messick & A. E. Tenbrunsel (Eds.), *Codes of conduct: Behavioral research into business ethics*. New York: Russel Sage Foundation.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, 31, 169–193.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105–115.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627–668.
- Eek, D., Loukopoulos, P., Fujii, S., & Gärling, T. (2002). Spill-over effects of intermittent costs for defection in social dilemmas. *European Journal of Social Psychology*, 32, 801–813.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.

- Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies*, 29, 1–17.
- Harvey, M. L., Bell, P. A., & Birjulin, A. A. (1993). Punishment and type of feedback in a simulated commons dilemma. *Psychological Reports*, 73, 447–450.
- Hobbes, T. (1651/1909). *Leviathan*. Oxford: Clarendon.
- Inman, J. J., & Zeelenberg, M. (2002). Regret in repeat purchase versus switching decisions: The attenuating role of decision justifiability. *Journal of Consumer Research*, 29, 116–128.
- Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, 51, 296–312.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Komorita, S. S., & Parks, C. D. (1995). Interpersonal relations: Mixed-motive interaction. *Annual Review of Psychology*, 46, 183–207.
- Kopelman, S., Weber, J. M., & Messick, D. M. (2002). Factors influencing cooperation in commons dilemmas: A review of experimental psychological research. In E. Ostrom, T. Dietz, P. C. Dolsak, P. C. Stern, S. Stonich, & E. U. Weber (Eds.), *The drama of the commons* (pp. 113–156). Washington, DC: National Academy Press.
- Liebrand, W. B. G., & Messick, D. M. (1996). *Frontiers in social dilemmas research*. Heidelberg and New York: Springer.
- Macy, M. W. (1995). PAVLOV and the evolution of cooperation: An experimental test. *Social Psychology Quarterly*, 58, 74–87.
- 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., & Brewer, M. B. (1983). Solving social dilemmas: A review. In L. Wheeler & P. Shaver (Eds.), *Review of personality and social psychology* (pp. 11–44). Beverly Hills, Ca: Sage.
- Mulder, L. B., Van Dijk, E., De Cremer, D., & Wilke, H. A. M. (submitted for publication). *Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas*.
- Rutte, C. G., & Wilke, H. A. M. (1984). Social dilemmas and leadership. *European Journal of Social Psychology*, 14, 105–121.
- Samuelson, C. D. (1993). A multiattribute evaluation approach to structural change in resource dilemmas. *Organizational Behavior and Human Decision Processes*, 55, 298–324.
- Samuelson, C. D., & Allison, S. T. (1994). Cognitive factors affecting the use of social decision heuristics in resource-sharing tasks. *Organizational Behavior and Human Decision Processes*, 58, 1–27.
- Samuelson, C. D., & Messick, D. M. (1995). When do people want to change the rules for allocating shared resources? In D. Schroeder (Ed.), *Social dilemmas; perspectives on individuals and groups*. (pp. 143–162). Westport, CT: Praeger.
- Samuelson, C. D., Messick, D. M., Rutte, C., & Wilke, H. (1984). Individual and structural solutions to resource dilemmas in two cultures. *Journal of Personality and Social Psychology*, 47, 94–104.
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1, 7–59.
- Taylor, M. (1976). *Anarchy and cooperation*. New York: Wiley.
- Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly*, 44, 684–707.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106, 1039–1061.
- Van Dijk, E., Wilke, H., Wilke, M., & Metman, L. (1999). What information do we use in social dilemmas? Environmental uncertainty and the employment of coordination rules. *Journal of Experimental Social Psychology*, 35, 109–135.
- Van Dijk, E., Wilke, H., & Wit, A. (2003). Preferences for leadership in social dilemmas: Public good dilemmas versus common resource dilemmas. *Journal of Experimental Social Psychology*, 39, 170–176.
- Van Dijk, E., & Wilke, H. A. (1996). Tacit coordination and fairness judgments in social dilemmas. In W. B. Liebrand & D. M. Messick (Eds.), *Frontiers in social dilemmas research* (pp. 117–134). Berlin: Springer Verlag.

- Van Vugt, M., & De Cremer, D. (1999). Leadership in social dilemmas: The effects of group identification on collective actions to provide public goods. *Journal of Personality and Social Psychology*, 76, 587–599.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92, 548–573.
- Weiner, B. (1986). Attribution, emotion, and action. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 281–312). New York, NY: The Guilford Press.
- Wit, A., & Wilke, H. A. (1990). The presentation of rewards and punishments in a simulated social dilemma. *Social Behaviour*, 5, 231–245.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110–116.
- Yamagishi, T. (1988a). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly*, 51, 265–271.
- Yamagishi, T. (1988b). Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly*, 51, 32–42.
- Yamagishi, T. (1992). Group size and the provision of a sanctioning system in a social dilemma. In W. B. G. Liebrand, D. M. Messick & H. Wilke (Eds.), *Social dilemmas: Theoretical issues and research findings. International series in experimental social psychology* (pp. 267–287).