



Universiteit
Leiden
The Netherlands

The (un)willingness to reward cooperation and punish non-cooperation

Molenmaker, Welmer E.

Citation

Molenmaker, W. E. (2017, January 19). *The (un)willingness to reward cooperation and punish non-cooperation*. Kurt Lewin Institute Dissertation Series. Retrieved from <https://hdl.handle.net/1887/45536>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/45536>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/45536> holds various files of this Leiden University dissertation

Author: Molenmaker, Welmer E.

Title: The (un)willingness to reward cooperation and punish non-cooperation

Issue Date: 2017-01-19





Chapter 3

The impact of personal responsibility on the (un)willingness to punish non-cooperation and reward cooperation

This chapter is based on: Molenmaker, W. E., De Kwaadsteniet, E. W., & Van Dijk, E. (2016). The impact of personal responsibility on the (un)willingness to punish non-cooperation and reward cooperation. *Organizational Behavior and Human Decision Processes*, 134, 1-15. doi: 10.1016/j.obhdp.2016.02.004

■ Abstract

To promote cooperation, people often rely on the administration of sanctions. However, from previous research we know that those in control of sanctions are generally reluctant to punish non-cooperative choice behavior and prefer to reward cooperative choice behavior, which is consistent with the do-no-harm principle. We propose that people are reluctant to punish because they feel personally responsible for the harm done. As such, we argue and demonstrate that the relative preference for rewarding over punishing is more pronounced when people decide individually than jointly (Experiments 3.1 and 3.2). Moreover, we show that the effect of grouping individuals on the reluctance to punish is mediated by feelings of personal responsibility (Experiment 3.3). These findings corroborate our reasoning that the feeling of personal responsibility has a self-restraining impact on the willingness to punish those who impair others' interests, but not on the willingness to reward those who serve others' interests.

■ Introduction

Sanctions are ubiquitous within societies, organizations, and many other groups. Fines and subsidies are installed to steer the behavior of citizens in the desired direction, penalties and imprisonment are imposed on offenders to prevent future offenses, and employees are promised bonuses and promotions to stimulate productivity. While sanctioning often benefits the collective welfare, it is not self-evident that those in control of negative sanctions (i.e., punishments like fines, penalties, or restrictions) and positive sanctions (i.e., rewards like bonuses, prizes, or privileges) are always willing to incur the costs of administering them. Recent research has, for instance, shown that people punish non-cooperative choice behavior less often and to a lesser extent than they reward cooperative choice behavior (Chapter 2; Molenmaker, De Kwaadsteniet, & Van Dijk, 2014; see Sutter, Haighner, & Kocher, 2010; see also Molm, 1997; Wang, Galinsky, & Murnighan, 2009). In fact, when people have both sanction means available, they tend to refrain from punishing and opt for rewarding.

Thus, although punishments and rewards can both be effective in enhancing the level of cooperation (e.g., Fehr & Gächter, 2000; Komorita & Barth, 1985; Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009; Wit & Wilke, 1990; Yamagishi, 1986, 1988; for overviews, see Balliet, Mulder, & Van Lange, 2011; Van Dijk, Molenmaker, & De Kwaadsteniet, 2015; Van Lange, Rockenbach, & Yamagishi, 2014), people usually are not as willing to punish those who impair others' interests as they are willing to reward those who serve others' interests. This general preference for the use of rewards over punishments is consistent with the do-no-harm principle, which states that people are reluctant to inflict harm on others (Baron, 1993, 1995; Baron & Jurney, 1993; Baron & Ritov, 1994; Ritov & Baron, 1990; Spranca, Minsk, & Baron, 1991; see also Van Beest, Van Dijk, De Dreu, & Wilke, 2005). After all, someone only harms another person directly with the use of punishments and not with the use of rewards. The fact that people are reluctant to punish non-cooperative choice behavior and prefer to reward cooperative choice behavior thus seems to be rooted in the do-no-harm principle (Molenmaker et al., 2014).

An important question that remains, however, is *why* people adhere to the do-no-harm principle when making sanctioning decisions. What does it mean that people are reluctant to punish non-cooperative choice behavior? Does this mean that they generally feel that no harm should be done, even when it is directed at someone who has impaired the interests of others? Or does it perhaps mean that they could live with the infliction of harm, but that their reluctance to administer punishments results from the fact that they are the ones doing it? That is, could it be that people are not merely concerned about the moral 'wrongness' of inflicting harm, but also about their own part in it? It is our central premise that this indeed is the case. We argue and show that an important reason why people apply the do-no-harm principle to their use of sanctions is because they feel personally responsible for the harm done. That is, we propose that people are reluctant to punish to the extent that they feel personally responsible for the harm done. It is the aim of the present chapter to identify personal responsibility as a determinant of the relative preference for rewarding cooperative choice behavior over punishing non-cooperative choice behavior.

To investigate the impact of personal responsibility on the willingness to sanction, we draw attention to the fact that people not necessarily need to be solely responsible for the (negative and positive) sanctions they administer; this responsibility can also be shared when sanctions are administered by groups of people. Yet, individual decision making has been the primary focus in research on sanctioning decisions (for overviews, see Gächter & Herrmann, 2009; Van Dijk et al., 2015), thereby leaving the willingness to sanction jointly largely unaddressed (see Putterman, 2014). This lack of knowledge about sanctioning by groups is unfortunate since prior research revealed that people often act very differently as members of a group than as individual decision makers. In contrast to individuals, groups for instance are less likely to help others in emergencies (i.e., bystander apathy; Darley & Latané, 1968; Latané & Darley, 1968; Latané & Nida, 1981), take more risks (i.e., risky shifts; Kogan & Wallach, 1967; Wallach & Kogan, 1965; Wallach, Kogan, & Bem, 1962, 1964), are more competitive (i.e., discontinuity effect; Insko et al., 1987; McCallum et al., 1985; Schopler et al., 1995; Wildschut, Pinter, Vevea, Insko, & Schopler, 2003), and are more aggressive (e.g., Festinger, Pepitone, & Newcomb, 1952; Le Bon, 1903; Milgram & Toch, 1969; Sherif, Harvey, White, Hood, & Sherif, 1961; Zimbardo, 1969). An often proposed explanation for these group phenomena is the fact that feelings of responsibility are reduced by the presence of others with whom responsibility can be shared. This so-called diffusion of responsibility essentially entails that individuals in groups are less restrained by a sense of personal responsibility for their actions. As such, the comparison between individual versus group decision making can teach us more about the self-restraining impact of feelings of personal responsibility on the willingness to administer (negative and positive) sanctions.

Feelings of personal responsibility restrain the infliction of harm

Our proposition that the feeling of personal responsibility is an important reason why people adhere to the do-no-harm principle accords with earlier research on this principle. It has for instance been shown that the reluctance to harm is stronger when people are directly (as opposed to indirectly) responsible for the anticipated harm (Milgram, 1974; Royzman & Baron, 2002). In a similar vein, the reluctance to harm is stronger when harmful outcomes result from people's actions rather than their inactions (Cushman, Young, & Hauser, 2006; Ritov & Baron, 1990, 1992; Spranca et al., 1991). Thus, doing harm is considered worse than not preventing harm from happening. Given that those who consider doing harm as worse also feel more personally responsible for the harm done (see Baron & Ritov, 2009; Spranca et al., 1991), these early studies suggest that the experience of personal responsibility for the anticipated harm amplifies the reluctance to inflict it on others (Baron & Ritov, 2004). From this work it follows that the infliction of harm itself may not be the only reason why people adhere to the do-no-harm principle. It could very well be that people feel that those who impaired others' interests deserve some form of punishment, but that their personal responsibility for the sanction restrains the tendency to inflict harm. That is, when people feel personally responsible for the anticipated harm, they may be more concerned about the harm they are about to inflict on others. Thus, we argue that people's reluctance to punish non-cooperation, as opposed to their willingness to reward cooperation, is a self-restraining tendency that originates from their feeling of personal responsibility for the harm done.

Note that our reasoning so far is that people monitor their own actions, and if they anticipate that an action would cause harm to others, they restrain it to the extent that they feel personally responsible for the action (see Schlenker, Britt, Pennington, Murphy, & Doherty, 1994; Shafir, Simonson, & Tversky, 1993; Shaver, 1975). In a way, one could say that decision makers basically hold *themselves* accountable for the harm they may inflict. In contrast to such an internal type of accountability (Lerner & Tetlock, 1999; Schlenker et al., 1994), one could also argue that people may restrain their willingness to punish because they expect they might be called on to explain their actions to *others* (i.e., external accountability). After all, people make most of their decisions in social contexts and often have to explain their actions to others (Semin & Manstead, 1983). Accountability toward others has indeed also been identified as an important amplifier of self-restraining tendencies (e.g., Lerner & Tetlock, 1999; Scott & Lyman, 1968; Tetlock, 1992). People's reluctance to punish non-cooperation may therefore also be a self-restraining tendency that originates from the fact that they are externally accountable for the harm done.

Even though avoiding blame by others is an important motive in social interactions (Shaver, 1985), and people can often get blamed for the punishments they administer (e.g., Atwater, Waldman, Carey, & Cartier, 2001; Eriksson, Andersson, & Strimling, 2015; Herrmann, Thöni, & Gächter, 2008; Kiyonari & Barclay, 2008; Nikiforakis, 2008; Strimling & Eriksson, 2014; Trevino, 1992), we propose that personal responsibility may have a self-restraining impact on the willingness to sanction, regardless of people's external accountability. That is, we argue that personal responsibility has an impact on the willingness to punish because people hold *themselves* internally accountable for the harm they might inflict. Consistent with this notion, prior research revealed that the relative preference for rewarding cooperators over punishing non-cooperators even emerged under conditions of complete anonymity without the possibility of getting blamed by others (Chapter 2, Molenmaker et al., 2014; see also Baron, 1995; Baron & Ritov, 2004; Royzman & Baron, 2002; Spranca et al., 1991). The fact that people feel personally responsible for the anticipated harm may thus already be enough to amplify their reluctance to harm, and increase the relative preference for the use of rewards over punishments.

Sanctioning individually versus jointly

If personal responsibility indeed plays a self-restraining role in the infliction of harm, any factor that decreases personal responsibility may in fact decrease the reluctance to punish non-cooperative choice behavior as well. As we mentioned above, we believe that a group of people with whom responsibility can be shared is such a key factor. But how do groups decrease feelings of personal responsibility? To answer this question, we turn to the Triangle Model of Responsibility (Pennington & Schlenker, 1999; Schlenker, 1986; Schlenker et al., 1994; Schlenker, Weigold, & Doherty, 1991). This model states that the experience of personal responsibility for an anticipated action in a given situation (e.g., the punishment of non-cooperative choice behavior) is determined by the extent to which one (1) knows what action should be performed, (2) is obligated to perform the anticipated action, and (3) has personal control over the anticipated action. As these determinants decrease in magnitude, so will the feeling of personal responsibility (Schlenker et al., 1994). Although the grouping

of individuals can affect all three determinants, group members who jointly make decisions will definitely have less personal control over the eventual sanction decision than each of them would have as individual decision maker. A decrease in personal control may therefore explain how joint decision making may decrease the feeling of personal responsibility (see also Skinner, 1996). As a result, one can expect that individuals feel less personally responsible for the (anticipated) actions they perform as a group. Indeed, studies consistently show that people attribute less responsibility to themselves for decisions they made jointly as compared to decisions they made individually, especially if this concerns decisions that had negative outcomes (e.g., Forsyth, Zyzanski, & Giammanco, 2002; Li et al., 2010; Mynatt & Sherman, 1975).

We propose that the same happens with sanctioning decisions. However, we argue that individuals in groups do not only *feel* less personally responsible for the (negative and positive) sanctions they administer, their reduced sense of personal responsibility may also attenuate their tendency to restrain the infliction of harm (see also Schlenker et al., 1994). As a result, we expect that individuals in groups are less reluctant to punish non-cooperative choice behavior. That is, we predict that the relative preference for rewarding over punishing is particularly dominant when individuals decide alone, and less so when they decide in groups. For the current purposes, it is important to note that our reasoning – which hinges on the attenuating effect of sharing responsibility – applies to the administration of punishments and not to the administration of rewards. The key issue is that feelings of personal responsibility for sanctions has a self-restraining impact in the case of harming others (i.e., punishment), but not in the case of favoring others (i.e., reward). If anything, one could even argue that it would be good to be solely responsible for rewarding cooperative choice behavior (see Kiyonari & Barclay, 2008). However, so does sharing this responsibility, as long as cooperative choice behavior is rewarded. So we reason that whether people decide as individual decision makers or as groups particularly affects their use of punishments and lesser their use of rewards.

The grouping of individuals may, however, not only affect the use of sanctions through feelings of personal responsibility. The fact that people make sanction decisions jointly can also reduce the concerns they may have about their entitlement to impose sanctions on others. Even though punishing non-cooperation and rewarding cooperation generally is beneficial to the collective, individual decision makers may have more doubts than groups of people about whether they are entitled to impose these sanctions on others (see also Miller & Effron, 2010; Miller, Effron, & Zak, 2009). After all, why can only they, and not others in their group or the group as a whole, determine whether cooperative choice behavior should be rewarded, and even more importantly, whether non-cooperative choice behavior should be punished? People who lack such a subjective sense of entitlement seem to be reluctant to take action (see e.g., Effron & Miller, 2015; Hornsey & Imani, 2004; Hornsey, Trembath, & Gunthorpe, 2004; Miller, 1999; Miller & Ratner, 1996, 1998; Ratner & Miller, 2001). As such, one could also argue that joint decision making may affect the willingness to use (negative and positive) sanctions in general because groups may be less concerned about their entitlement to sanction than individual decision makers. As argued above, however, we believe that grouping individuals particularly affects people's willingness to punish non-cooperation (and less so

their willingness to reward cooperation) because they are about to inflict harm on others (as opposed to favoring others).

Aggression committed by individuals versus groups

As we mentioned earlier, it is suggested that the experience of personal responsibility plays an important role in many group phenomena. Indirect support for our reasoning about doing harm in groups can particularly be found in the fact that individuals are more aggressive in groups (e.g., Festinger et al., 1952; Le Bon, 1903; Milgram & Toch, 1969; Sherif et al., 1961; Zimbardo, 1969). Aggression has similarities with punishment as both imply that harm is done. Although direct comparisons in aggression committed by individuals versus groups are scarce, the few experimental studies that were conducted consistently show that groups harm others more severely than individuals do. For instance, a study using the hot sauce paradigm (e.g., McGregor et al., 1998) demonstrated that groups in contrast to individuals allocate more hot sauce to others to consume (Meier & Hinsz, 2004; see also Van Beest, Carter-Sowell, Van Dijk, & Williams, 2012). In earlier studies on aggression, electric shocks were often used as a measure of aggression (e.g., Buss, 1961; Milgram, 1974). Research by Jaffe and colleagues showed that groups administer more severe electric shocks to a confederate who failed on a task than individuals did (Jaffe, Shapir, & Yinon, 1981; Jaffe & Yinon, 1979; see Bandura, Underwood, & Fromson, 1975). Finally, and most strongly related to the present research, groups use larger monetary fines to take revenge than individuals (Mathes & Kahn, 1975).

Although the above studies demonstrate that groups are more aggressive than individuals, and thereby show that sharing responsibility attenuates the reluctance to harm others, the question remains whether this also is the case for the willingness to punish non-cooperative choice behavior, which generally is – in contrast to aggression – beneficial to the collective. In addition, despite the fact that aggression and punishment both involve the infliction of harm, the harm done with the used measures of aggression is potentially much more extreme (i.e., physical pain) than with the sanction means that people generally have available to promote cooperation in real-life situations (i.e., loss of [access to] material resources). Mathes and Kahn (1975) were the only researchers who compared the use of monetary fines by individuals and groups, but in their study participants could only use these fines to respond to an individual who insulted them personally. As a result, it is yet unknown whether the grouping of individuals would have the same effect on people's willingness to administrate (monetary) punishments in situations in which others' choice behavior impairs the collective interests and not necessarily is a harmful act toward them personally. Do those who are solely responsible for sanctions indeed restrain their willingness to punish non-cooperation (as opposed to their willingness to reward cooperation), and does sharing this responsibility attenuates their reluctance to punish, or is the impact of sharing responsibility not that strong? The present research is specifically designed to answer this question.

Punishment and reward in social dilemmas

Although our reasoning is applicable to (negative and positive) sanctioning in general, social dilemmas (Camerer, 2003; Dawes, 1980) are an appropriate context to investigate the willingness to punish non-cooperative choice behavior and reward cooperative choice behavior. Social dilemmas revolve around a conflict between group members' personal interest and the collective interest (for overviews, see Parks, Joireman, & Van Lange, 2013; Van Lange, Joireman, Parks, & Van Dijk, 2013; Weber, Kopelman, & Messick, 2004). In real-life, people frequently face such conflicts of interests. Consider, for instance, an important type of social dilemma called the *common resource dilemma*. Common resource dilemmas deal with the problem of maintaining scarce common resources, such as energy, clean water, and food (see Hardin, 1968). For an individual it is profitable to consume from such common resources. However, if people consume excessively, these resources may deplete and the collective will be worse off than if people would restrain their harvesting. Due to this mixed-motives structure, the occurrence of mutual cooperation is not self-evident and cooperation generally needs to be enforced by promoting cooperative choice behavior and deterring non-cooperative choice behavior (Hardin, 1968; Olson, 1965). As a result, social dilemmas are ideal for investigating the general preference for administering rewards over punishments.

To test our reasoning, we conducted three experiments in which participants took part in a common resource task with a sanction opportunity implemented (see De Kwaadsteniet, Rijkhoff, & Van Dijk, 2013; De Kwaadsteniet, Van Dijk, Wit, & De Cremer, 2010; Molenmaker et al., 2014). In Experiment 3.1, participants observed the harvest decision of another person and subsequently *voted* individually or jointly about whether or not they wanted to assign a fixed sanction (punishment versus reward). In Experiment 3.2, we used a similar setting as in the first experiment, but now participants individually or jointly determined the *size* of a variable sanction (punishment versus reward). Thus, in the first experiment the dependent variable was dichotomous (vote for or against sanctioning), whereas in the second experiment it was continuous (the size of the sanction). In Experiment 3.3, we focused on punishment decisions to test whether the effect of responsibility (Individual versus Joint) on the size of the punishment is independent of external accountability (Accountable versus Unaccountable). Moreover, we also examined whether the effect is mediated by feelings of personal responsibility, while we sought to rule out alternative explanations such as feelings of external accountability or entitlement to sanction.

■ Experiment 3.1

Our first experiment provided an initial test of the willingness to administer sanctions individually or jointly. Participants performed a common resource task in which they could harvest chips from a common pool. Harvested chips could be kept for oneself, while the chips left in the common pool would be doubled and divided equally among the group members. Thus, harvesting chips was best for one's self-interest, whereas leaving chips in the common resource was best for the collective interest. Within this social dilemma context, we confronted participants with a non-cooperative/cooperative group member and let them decide whether

or not they wanted to administer a fine/bonus to that person. This sanction decision was either made individually or jointly. Based on our reasoning, we predicted that the relative preference for rewarding over punishing is more pronounced when individuals decide alone than when they decide in groups. More specifically, we predicted that individuals would vote less often for punishment of a non-cooperative group member than groups, whereas the votes for reward of a cooperative group member would not necessarily differ between groups and individuals.

Method

Participants and design

We recruited 165 students at Leiden University (111 women and 54 men; $M_{\text{age}} = 21.22$ years, $SD_{\text{age}} = 4.62$) to participate in an experiment on “group decision making”.¹ For their participation, students received a monetary compensation (€3). This experiment employed a 2 (Responsibility: Individual versus Joint) $\times$ 2 (Sanction Type: Punishment versus Reward) between-participants factorial design.

Procedure

When participants arrived at the laboratory to take part in the experiment, they were seated in separate cubicles. Each cubicle contained a personal computer that was used to present the instructions and register the data. Participants were randomly assigned to one of the four conditions by a computer automated procedure.

The instructions informed participants about the joint task they had to perform together with three fellow participants whose identities were unknown. Participants learned that the first part of the joint task consisted of a common resource task in which they could earn extra money on top of their initial participation fee. In the common resource task, each person could harvest up to 10 chips (each worth €0.10) from a common pool containing 40 chips. The chips they harvested would accrue totally to themselves and the chips they left in the common pool would be doubled and divided equally among the four persons (see e.g., Molenmaker et al., 2014; Van Dijk & Wilke, 1995, 1997, 2000; Van Dijk, Wilke, & Wit, 2003). The instructions also explained to the participants that in the second part of the joint task there would be an opportunity to decrease (punishment conditions) or increase (reward conditions) the personal outcome of one person. However, this would be explained more thoroughly after they performed the common resource task. To ensure that participants understood these instructions correctly, we posed five practice questions. After answering each practice question, the correct answer was disclosed. Next, participants decided how many chips to harvest from the common pool.

Before any feedback was provided about others' harvest decisions and their personal outcome, participants received instructions about the second part of the joint task. Here

¹ For each experiment, we aimed to recruit as many participants as possible within the given time available in the lab (approximately two weeks per experiment).

we introduced our manipulations. Participants were informed that we had selected one person (person C) whose personal outcome could be decreased by giving that person a fine (punishment conditions) or increased by giving that person a bonus (reward conditions; for a comparable procedure, see Molenmaker et al., 2014). The fine was equal to the number of chips person C had harvested. The bonus, by contrast, was equal to the number of chips person C had left in the common pool. Thus, participants had no influence on the size of the fine/bonus. In the individual conditions, participants learned that only they were randomly selected to decide whether person C would receive the fine/bonus, whereas in the joint conditions the group (except person C) had to vote as to whether or not the fine/bonus should be administered. The majority rule (at least 2 out of 3 persons) would determine whether they as group would assign the fine/bonus (see e.g., Putterman, Tyran, & Kamei, 2011). After reading the instructions about the opportunity to assign a fine/bonus, we posed practice questions to ensure comprehension of the instructions. For example, we asked who would decide whether person C would receive the fine (*bonus*). The correct answers were disclosed after answering each question.

Next, we gave participants preprogrammed feedback about the harvest decision of the selected person (person C). In the punishment conditions, person C had harvested the maximum of 10 chips from the common pool, whereas in the reward conditions person C had harvested the minimum of 0 chips (i.e., had left 10 chips in the common pool). In other words, in the punishment conditions participants could punish a non-cooperator and in the reward conditions they could reward a cooperator. Subsequently, participants voted for or against assigning the fine/bonus to person C.

After participants made this decision, we posed some manipulation checks. First, we checked the experience of personal responsibility for the sanction. Participants indicated on a 9-point rating scale ranging from 1 (*not at all*) to 9 (*totally*) to what extent four statements applied to them (Cronbach's $\alpha = .88$). With the four statements, presented in random order, we measured how personally responsible and liable participants thought and felt they were (e.g., “*I am personally liable for the fine [bonus] that person C receives*”, “*I feel personally responsible for the fine [bonus] that person C receives*”).² Finally, we asked two questions to determine whether participants understood whether they could assign a fine or a bonus (i.e., sanction type manipulation) and whether they assigned this individually or jointly (i.e., responsibility manipulation). In addition, some additional questions measured the general comprehension

² In Experiment 3.1 we also measured how responsible and liable participants thought the *other persons* in the joint task were. Since the pattern of results was consistent with the results of the experienced personal responsibility, we did not pose this measure of others' responsibility in the following experiments. In addition, we posed questions – for exploratory purposes – to measure whether participants' made their sanction decision to avoid feeling guilty, feeling regretful, being held responsible, and being held liable. No significant effects of responsibility manipulation (Individual versus Joint) emerged, so we also did not pose these questions in the following experiments.

and the believability of the joint task.³ At the end of the experiment, we thanked, debriefed, and paid all the participants. Payment consisted of the participation fee plus an additional €1 from the common resource task.

Results

Manipulation checks

We asked participants whether they could assign a bonus or a fine to check the manipulation of sanction type, and we asked participants whether they individually or jointly decided to assign the fine/bonus to check the manipulation of responsibility. All participants (100%) answered these questions correctly. Furthermore, the 2 (Responsibility: Individual versus Joint) $\times$ 2 (Sanction Type: Punishment versus Reward) ANOVA on felt personal responsibility demonstrated that participants felt more responsible for the sanction decision in the individual conditions ($M = 7.19$, $SD = 1.84$) than in the joint conditions ($M = 5.34$, $SD = 2.16$), $F(1,161) = 34.85$, $p < .001$, $\eta^2 = .18$, 90% CI [.10, .26]. No other effects on felt personal responsibility were significant ($p_s > .10$, $\eta^2 < .01$). Altogether, these results indicate that our manipulations were successful. The data of all 165 participants were included in the analyses.

Sanction behavior

We performed a binary (Sanction Vote: 0 = no sanction, 1 = sanction) logistic regression with responsibility (Individual versus Joint) and sanction type (Punishment versus Reward) as independent variables. This analysis yielded a significant Sanction Type main effect ($B = 0.86$, $SE = 0.39$, Wald ($df=1$) = 4.975, $p = .026$, Odds Ratio = 2.36, 95% CI [1.11, 5.02]), which indicated that the proportion of participants choosing to punish (69.1%) was lower than the proportion of participants choosing to reward (84%). In line with our expectations, the Responsibility $\times$ Sanction Type interaction effect was (marginally) significant ($B = 1.50$, $SE = 0.79$, Wald ($df=1$) = 3.61, $p = .057$, Odds Ratio = 4.49, CI [0.95, 21.12]) and the Responsibility main effect was non-significant ($B = 0.33$, $SE = 0.38$, Wald ($df=1$) = 0.81, $p = .369$, Odds Ratio = 1.40, CI [0.67, 2.92]). See Table 3.1 for the frequencies.

As expected, the results indicated that in the individual condition the proportion of participants choosing to punish (59.5%) was lower than the proportion of participants choosing to reward (87.8%), $\chi^2(1) = 8.52$, $p = .004$, Odds Ratio = 4.90, 95% CI [1.60, 15.01], while the proportion of participants choosing to punish (78.6%) and the proportion of participants choosing to reward (80.0%) in the joint condition did not differ significantly, $\chi^2(1) = 0.03$, $p = .873$, Odds Ratio = 1.09, CI [0.37, 3.18]. This effect was particularly driven by punishment since the results also showed that the proportion of participants choosing to punish was (marginally)

³ Besides direct measures of the general comprehension and the believability of the joint task (e.g., “Could you also harvest chips from the common pool?”), we also posed more indirect measures (i.e., how important they thought it was that person C got a fine/bonus, how much influence they thought they had on the outcome of person C, and how they evaluated the choice by person C).

Table 3.1. *Number of participants who voted for and voted against sanctioning as a function of Sanction type and Responsibility (Experiment 3.1)*

	Punishment		Reward	
	Voted For	Voted Against	Voted For	Voted Against
Individual	25 _a	17	36 _a	5
Joint	33 _b	9	32 _a	8
Total	58	26	68	13

Note. Frequencies with differing subscripts within Voted For rows are (marginally) significantly different at the $p < .10$.

lower in the individual condition (59.5%) than in the joint condition (78.6%), $\chi^2(1) = 3.57$, $p = .059$, Odds Ratio = 2.49, CI [0.95, 6.52]. It should be noted, however, that the confidence intervals indicated that the precision of this estimation seemed low. In contrast, the proportion of participants choosing to reward in the individual condition (87.8%) and the joint condition (80%) did not differ significantly, $\chi^2(1) = 0.92$, $p = .339$, Odds Ratio = 1.8, CI [0.53, 6.06].

Controlling for harvesting decisions

In this experiment, participants did not only make a decision to sanction, they also made a harvesting decision themselves. Although participants made this decision before we introduced our manipulations, one may wonder whether the differences in sanction behavior could be explained by differences in their harvests (see also De Kwaadsteniet et al., 2010). For instance, participants who had harvested relatively large amounts themselves could have felt less entitled to punish non-cooperators (see also e.g., De Quervain et al., 2004; Miller et al., 2009; Ratner & Miller, 2001). On average, participants harvested 2.92 chips ($SD = 3.18$) from the common pool.

To obtain more insight on the influence of participants' own harvest, a 2 (Responsibility: Individual versus Joint) $\times$ 2 (Sanction Type: Punishment versus Reward) ANOVA on harvest decision was performed, which yielded no significant effects ($p_s > .10$, $\eta^2 < .01$). Next, we added harvest decision as covariate to the binary (Sanction Vote: 0 = no sanction, 1 = sanction) logistic regression on sanction behavior. This analysis showed that harvest decision was a non-significant predictor of sanction behavior ($p > .10$, Odds Ratio ≈ 1) and none of the initial effects became non-significant when the harvests were included as covariate.⁴ We can therefore conclude that participants' own harvest decision did not explain the differences in sanctioning behavior we found.

⁴ In our experiments, we also recoded the harvesting decisions to distinguish between participants who either had harvested chips from the common pool (i.e., took 1–10 chips) or had not harvested from the common pool (i.e., took 0 chips). Including this factor in our analyses did not reveal any significant effect of harvesting decisions on participants' willingness to sanction and this did not influence the significance of other effects in our experiments.

Discussion

The results of Experiment 3.1 support our reasoning. First of all, we replicated previous research by showing that non-cooperative choice behavior was punished less often than cooperative choice behavior was rewarded (Chapter 2; Molenmaker et al., 2014). In accordance with the do-no-harm principle, which states that people are reluctant to harm others (e.g., Baron, 1993, 1995; Baron & Jurney, 1993), we thus revealed a reluctance to administer punishments. Note that the participants generally were high cooperators themselves since they harvested on average only 2.92 out of 10 chips. Thus, while participants' highly cooperative choice behavior was exploited by a non-cooperative group member, they were reluctant to assign a punishment (in comparison to the assigned rewards for cooperation), even though participants could administer the (negative and positive) sanction without any cost to themselves.

More importantly, however, we also demonstrated that this relative preference for rewarding over punishing was more pronounced when participants decided alone than when they decided in groups. When participants decided individually, they were less willing to assign a punishment to a non-cooperator than when they decided jointly. No such difference was found for the assignment of a reward to a cooperator. These results corroborate our reasoning that, when it comes to punishment, people are not merely concerned about the moral 'wrongness' of the harm done, they are also very much affected by their own part in it. That is, people apply the do-no-harm principle to their use of sanctions because they are personally responsible for the harm done. The fact that people have personal responsibility for the administered sanctions thus seems to trigger the tendency to restrain their infliction of harm. After all, the results of Experiment 3.1 demonstrated that sharing this responsibility with a group of people attenuated the reluctance to punish those who impaired the collective interests. Thus, whether the sanction decision is made by individual decision makers or groups particularly affects the willingness to punish and not necessarily the willingness to reward.

Although in Experiment 3.1 we focused on situations in which one could administer sanctions with predetermined sizes, one could also think of settings in which sanction sizes are not fixed or predetermined. For instance, when employees violate certain company rules, their managers often have to decide whether, and for how many days they should be suspended (e.g., for one day, a week, until further notice, permanently, etc.). As sanction decisions often involve determining the *size* of sanctions, we focused on such situations in a second experiment. By doing so, we are able to test whether personal responsibility not merely restrains the willingness to use punishments, but also the size of punishments people are willing to administer. That is, predetermined punishments – like the one in Experiment 3.1 – may not necessarily be what people consider the appropriate punishment to administer, either because they find it too large or too small. As such, people may opt for administering no punishment at all, while in fact they might have wanted to administer a punishment of a different size. After all, people sometimes feel that those who impaired others' interests deserve some form of punishment. In Experiment 3.2, we therefore examined whether feelings of personal responsibility also have an impact on what size of punishment people consider appropriate for non-cooperative choice behavior.

At first glance, the fact that people may sometimes want that non-cooperation is punished might seem to contradict the do-no-harm principle, as this principle states that people are

reluctant to inflict harm on others, even if the overall benefit outweighs the harm done (e.g., Baron, 1995). We believe, however, that it actually demonstrates that people do not necessarily consider it morally ‘wrong’ that harm is inflicted, as long as the inflicted harm gives non-cooperators their just deserts or at least signals disapproval about their non-cooperative choice behavior (see e.g., Carlsmith, 2006; Carlsmith, Darley, & Robinson, 2002; Crockett, Özdemir, & Fehr, 2014; De Quervain et al., 2004). Thus, despite the fact that people may deem it desirable that non-cooperation is punished, feelings of personal responsibility for the anticipated harm may have an impact on the size of punishment they consider appropriate. That is, we argue that people are reluctant to punish to the extent that they feel personally responsible for the harm done. As a result, people may opt for punishing to ensure that non-cooperative choice behavior is punished, but punish to a lesser extent as an individual decision maker than as a group of people because of their feelings of personal responsibility for the harm done. Although the results of Experiment 3.1 are in line with our reasoning, we thus conducted a second experiment to replicate and extend the previous experiment using a different way of measuring sanction behavior.

In Experiment 3.1, participants voted as to whether or not they wanted to administer a sanction and in the joint conditions this would be determined by the majority rule (see e.g., Putterman et al., 2011). With this majority rule, it strongly depended on others’ voting behavior whether one’s own vote would affect the group decision. An individual could vote for punishment, but the punishment would only be administered if the others voted for punishment as well. As a result, expectations about others’ voting behavior could have had a strong effect on participants’ own voting behavior. In Experiment 3.2, we gave them the opportunity to decide on the size of the sanction. To operationalize the joint conditions, the sanction size was determined by averaging the sanctioning decisions of the individual group members (see e.g., Bandura et al., 1975). Due to this procedure, the second experiment differed markedly from the first because, although in both experiments there is a decrease of control in the joint conditions, expectations about others’ sanction behavior may be less influential in Experiment 3.2 than in Experiment 3.1. Furthermore, an additional advantage of using a continuous dependent variable, as was the case in Experiment 3.2, is that effect size estimations are generally more precise than when using a dichotomous dependent variable (e.g., Greenland, Schwartzbaum, & Finkle, 2000; Jewell, 1984; Nemes, Jonasson, Genell, & Steineck, 2009). Therefore, we provided another test of the impact that the grouping of individuals has on sanctioning with sanction size as indicator of the willingness to administer (negative and positive) sanctions.

■ Experiment 3.2

Similar to the first experiment, we presented participants in Experiment 3.2 with a non-cooperative/cooperative group member within the context of a social dilemma. However, this time we let them decide on the size of the punishment/reward they wanted to administer to that person. This sanction decision was either made individually or jointly. Based on our reasoning, we predicted that individuals would punish a non-cooperative group member to

a lesser extent than groups, whereas the size of the rewards for a cooperative group member would not necessarily differ between groups and individuals.

Method

Participants and design

We recruited 156 students at Leiden University (125 women and 31 men; $M_{\text{age}} = 20.42$ years, $SD_{\text{age}} = 2.65$) to participate in the experiment for a monetary compensation (€3) (see Footnote 1). Similar to the first experiment, the second experiment also employed a 2 (Responsibility: Individual versus Joint) $\times$ 2 (Sanction Type: Punishment versus Reward) between-participants factorial design.

Procedure

The procedure of this second experiment was identical to the procedure of the first experiment, although now the joint task was performed together with four fellow participants and the common pool contained 50 chips (each worth €0.10). As in Experiment 3.1, the instructions explained that we had selected one person (person C) whose personal outcome could be decreased (punishment conditions) or increased (reward conditions). In Experiment 3.2, however, participants could do this by assigning decrement points (punishment conditions) or by assigning increment points (reward conditions) to person C. The number of points participants could assign ranged from 0 to 100 points. Each assigned point would decrease/increase the personal outcome of person C with €0.01.

In the individual conditions, participants learned that only they were randomly selected to decide how many points person C would receive. Participants in the joint conditions learned that they as group (except person C) decided how many points person C would receive. Each person first had to indicate privately how many points they wanted to assign. Next, the decisions of each person would be combined by taking the average number of points (see e.g., Bandura et al., 1975). This average number of points determined how many points they as group would assign to person C.

Again, we posed practice questions to ensure comprehension of the instructions. For example, we asked who would decide how many decrement/increment points person C would receive. The feedback about the harvest decision of person C was the same as in the first experiment, namely that (s)he had harvested the maximum/minimum of 10 chips from the common pool. After participants decided how many decrement points/increment points they wanted to assign we again posed some manipulation checks, such as the personal responsibility participants experienced for the sanctions (Cronbach's $\alpha = .84$).

Results

Manipulation checks

To check the manipulation of sanction type, we asked participants whether they could assign decrement points or increment points. All participants except one (i.e., 99.4%) answered this question correctly. To check the manipulation of responsibility, we asked participants whether

they decided individually or jointly to assign points. All participants except four (i.e., 97.4%) answered this question correctly. Furthermore, we analyzed felt personal responsibility with a 2 (Responsibility: Individual versus Joint) $\times$ 2 (Sanction Type: Punishment versus Reward) ANOVA. As expected, this analysis revealed that participants felt more responsible for the sanction decision in the individual conditions ($M = 6.63$, $SD = 1.70$) than in the joint conditions ($M = 5.24$, $SD = 1.94$), $F(1,152) = 22.54$, $p < .001$, $\eta^2 = .13$, 90% CI [.06, .21]. No other effects on felt personal responsibility were significant ($p > .10$, $\eta^2 < .01$). These results suggest that our manipulations were successful. The data of all 156 participants were included in the analyses.

Sanction behavior

A 2 (Responsibility: Individual versus Joint) $\times$ 2 (Sanction Type: Punishment versus Reward) ANOVA yielded a significant Sanction Type main effect ($F(1,152) = 21.01$, $p < .001$, $\eta^2 = .12$, 90% CI [.05, .20]), which indicated that the size of the punishments ($M = 50.92$, $SD = 36.82$) was smaller than the size of the rewards ($M = 76.45$, $SD = 33.38$). In line with our expectations, the Responsibility $\times$ Sanction Type interaction effect was also significant ($F(1,152) = 4.25$, $p = .041$, $\eta^2 = .02$, CI [.00, .08]) and the Responsibility main effect was non-significant, $F(1,152) = 1.01$, $p = .316$, $\eta^2 < .01$, CI [.00, .04]. See Table 3.2 for the mean number of points and standard deviations per condition.

Table 3.2. *Number of points assigned as a function of Sanction type and Responsibility (Experiment 3.2)*

	Punishment		Reward		Overall	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Individual	42.38 _a	36.67	79.38 _a	30.08	60.88 _a	38.17
Joint	59.46 _b	35.40	73.51 _a	36.53	66.49 _a	36.43
Overall	50.92	36.82	76.45	33.88	63.69	37.30

Note. Means with differing subscripts within rows are significantly different at the $p < .05$.

To interpret the interaction, we performed simple-effect analyses. As expected, in the individual condition the size of the punishments was significantly smaller than the size of the rewards, $F(1,152) = 22.08$, $p < .001$, $\eta^2 = .12$, 90% CI [.05, .21], but in the joint condition this difference was only marginally significant, $F(1,152) = 3.18$, $p = .076$, $\eta^2 = .02$, CI [.00, .07]. Furthermore, this effect was particularly driven by punishment since the simple-effect analyses also showed that the size of the punishments was significantly smaller in the individual condition than in the joint condition ($F(1,152) = 4.70$, $p = .032$, $\eta^2 = .03$, CI [.00, .08]), while the size of the rewards did not differ significantly between the individual condition and the joint condition ($F(1,152) = 0.56$, $p = .457$, $\eta^2 < .01$, CI [.00, .03]).

Controlling for harvesting decisions

To ensure that the observed differences in sanction behavior were not explained by differences in harvest decisions, a 2 (Responsibility: Individual versus Joint) $\times$ 2 (Sanction Type: Punishment versus Reward) ANOVA on harvest decision was performed, which yielded no significant effects ($p > .10$, $\eta^2 < .02$). On average, participants harvested 2.60 chips ($SD = 2.91$) from the common pool. Moreover, when we added harvest decision as covariate to the 2 (Responsibility: Individual versus Joint) $\times$ 2 (Sanction Type: Punishment versus Reward) ANOVA on sanction behavior, this covariate was a non-significant predictor of sanction behavior ($p > .10$, $\eta^2 < .01$) and none of the initial effects became non-significant when harvests were included as covariate (see Footnote 4). Thus, participants' own harvest decision did not explain the differences in sanctioning behavior we found.

Discussion

The results of Experiment 3.2 are again in line with our reasoning. Participants – most of whom showed high levels of cooperation – were generally less willing to punish non-cooperative choice behavior than to reward cooperative choice behavior (Chapter 2; Molenmaker et al., 2014). As predicted, this relative preference for reward over punishment was more pronounced when participants decided individually than when they decided jointly. In accordance with early research on the do-no-harm principle (e.g., Ritov & Baron, 1990; Royzman & Baron, 2002; see Milgram, 1974), participants individually punished a non-cooperator to a lesser extent than jointly, whereas no such a difference was found on the rewarding of a cooperator. The results of Experiment 3.2 thus further corroborate our idea that personal responsibility is an important determinant of the willingness to punish non-cooperative choice behavior, but not of the willingness to reward cooperative choice behavior.

■ Experiment 3.3

Sanction decisions take place in social contexts and in such contexts people often are accountable toward others for the (negative and positive) sanctions they administer (see Semin & Manstead, 1983). In the first place, one generally is accountable toward the person who is punished, but there are also many occasions in which people are called on to explain their sanction decisions toward other people as well. Since the administration of punishments may often be disapproved of by others (e.g., Atwater et al., 2001; Eriksson et al., 2015; Herrmann et al., 2008; Kiyonari & Barclay, 2008; Nikiforakis, 2008; Strimling & Eriksson, 2014; Trevino, 1992), it may be that it is external accountability why people restrain their willingness to punish. Consistent with this notion, there is prior research suggesting that those who are accountable toward others seem less willing to punish than those who are unaccountable (Lerner, Goldberg, & Tetlock, 1998; Piazza & Bering, 2008; but see Kurzban, DeScioli, & O'Brien, 2007). Moreover, people sometimes hide their punishments, especially severe ones, even when this comes at a cost to themselves (Rockenbach & Milinski, 2011).

Based on this, one may wonder whether the reluctance to punish is reduced among individuals who jointly decide about administering punishments because their group provides

a shield of anonymity (e.g., Insko et al., 2001; Schopler et al., 1995). After all, the person who is punished often remains uninformed about what each individual group member has decided about their punishment; they often only learn about what punishment they received by the group as a whole. For example, juries in UK criminal courts are by law not allowed to discuss their deliberations outside the jury, even long after a verdict has been reached. Thus, could it be that sharing responsibility for the harm done may only attenuate the reluctance to punish when one also feels less accountable toward others (because one's input in the group decision remains unknown to others), or may a reduced sense of personal responsibility for the harm done already be enough to attenuate the reluctance to punish non-cooperative choice behavior? To answer these questions, we decided to conduct a third experiment in which we used the same procedure as in Experiment 3.2, the only difference being that we now focused on punishment only and manipulated responsibility and external accountability independently from each other. That is, in Experiment 3.3 we tested the willingness to punish individually versus jointly, when the punishment decision was public and needed to be explained to the group members (i.e., accountable) versus private and did not need to be explained to the group members (i.e., unaccountable).

Moreover, in Experiment 3.3 we also want to address the possibility that grouping individuals may not only reduce feelings of personal responsibility, but also affect the entitlement one may feel to impose sanctions on others. In the previous experiments, participants were randomly selected to make the sanction decision individually (as opposed to making this decision jointly) and may therefore have felt less entitled to administer sanctions. Even though people may feel that they are entitled to punish and reward when their own outcome is affected by others' choice behavior (see Miller et al., 2009; Ratner & Miller, 2001), this feeling may be less strong when they administer these sanctions individually than jointly. Experiments 3.1 and 3.2 showed, however, that grouping individuals affected the willingness to punish and not the willingness to reward. Thus, if people would – besides their personally responsibility – also be concerned about their entitlement to administer (negative and positive) sanctions in general, this only seemed to have affected their willingness to punish and not their willingness to reward. As such, it would be interesting to also measure feelings of entitlement and test its role in the effects of grouping individuals on the willingness to punish non-cooperative choice behavior.

In sum, we thus measured feelings of personal responsibility, external accountability and entitlement. To test for mediation, these feelings were measured before (instead of after; MacKinnon, Fairchild, & Fritz, 2007) participants decided about administering the punishment. As such, Experiment 3.3 provided another test to examine our central premise that people are reluctant to punish non-cooperative choice behavior to the extent that they feel personally responsible for the harm done. Based on our reasoning we predicted that individuals would punish a non-cooperative group member to a lesser extent than groups, irrespective of whether they were accountable or unaccountable toward others. Furthermore, we predicted that this effect of grouping individuals would be mediated by (reduced) feelings of personal responsibility.

Method

Participants and design

199 students at Leiden University (151 women and 48 men; $M_{\text{age}} = 20.38$ years, $SD_{\text{age}} = 2.43$) were recruited to participate in the experiment for a monetary compensation (€3.50) (see Footnote 1). This experiment employed a 2 (Responsibility: Individual versus Joint) $\times$ 2 (Accountability: Accountable versus Unaccountable) between-participants factorial design.

Procedure

In Experiment 3.3, we used almost the same procedure as the punishment conditions in Experiment 3.2. Thus, the instructions explained that we had selected one person (person C) whose personal outcome could be decreased by assigning decrement points. The number of points participants could assign ranged from 0 to 100 points. Each assigned point would decrease the personal outcome of person C with €0.01. Participants in the individual conditions learned that only they were randomly selected to decide how many points person C would receive, whereas participants in the joint conditions learned that the decisions of each person in the group (except person C) would be combined by taking the average number of points and that would determine how many points person C would receive (see Bandura et al., 1975).

In contrast to Experiment 3.2, the instructions now informed participants whether or not they were accountable toward others for the decision they would make. In the accountable conditions, participants learned that it would afterwards be made public how many points they assigned and that they would have to explain their decision to all other persons in the joint task (for a similar induction, see De Kwaadsteniet, Van Dijk, Wit, De Cremer, & De Rooij, 2007; see also Lerner & Tetlock, 1999). In the unaccountable conditions, participants learned that person C would only be informed about how many points (s)he would receive and that it would remain private how this decision was reached. Thus, the other persons in the joint task (including person C) would not know that they made the sanction decision (individual condition) or what the individual sanction decision of each person was from which the average number of points was taken (joint condition). After the instructions, we posed practice questions to ensure comprehension of the joint task. For example, we asked who would learn about how many decrement points they want to assign to person C. Next, we gave participants the same feedback about the harvest decision of person C as in the previous experiments, namely that (s)he had harvested the maximum of 10 chips from the common pool.

Before participants decided how many decrement points they wanted to assign and answered some manipulation checks, we first posed the measures of felt personal responsibility (Cronbach's $\alpha = .88$), felt external accountability (Cronbach's $\alpha = .88$) and felt entitlement (Cronbach's $\alpha = .90$).⁵ For each measure, participants indicated on a 9-point rating scale

⁵ We also measured how confident participants felt about their ability to make the punishment decision with four items (e.g., "I am confident that I am able to determine the number of decrement points that person C is going to receive"; Cronbach's $\alpha = .86$). No significant effects on this measure emerged.

ranging from 1 (*not at all*) to 9 (*totally*) to what extent four statements applied to them. For the measure of felt personal responsibility for the sanction, we adapted the four statements from the previous experiments (e.g., “*I feel personally responsible for the decrement points that person C will receive*”). In addition, four statements measured how accountable toward others participants felt for the sanction (e.g., “*I have the feeling that the others can hold me accountable for my decision about the decrement points I want to give to person C*”; adapted from De Kwaadsteniet et al., 2007) and four statements measured how entitled participants felt to administer a sanction (e.g., “*I feel legitimated to determine how many decrement points person C will receive*”; adapted from De Cremer & Van Dijk, 2005). To prevent that the order of measuring these constructs would influence the results, the four statements within a measure and the three measurements themselves were presented in random order.⁶

Results

Manipulation checks

To check the manipulation of responsibility, we asked participants whether they decided individually or jointly to assign decrement points. All participants except one (i.e., 99.5%) answered this question correctly. To check the manipulation of accountability, we asked participants whether the others or none of the others would know how many decrement points they decided to assign. All participants except two (i.e., 99%) answered this question correctly. Thus, these results indicate that our manipulations of responsibility and accountability were successful. The data of all 199 participants were included in the analyses.

Punishment behavior

We started with analyzing the influence of responsibility and accountability on the size of the punishments. A 2 (Responsibility: Individual versus Joint) $\times$ 2 (Accountability: Accountable versus Unaccountable) ANOVA only yielded a significant Responsibility main effect ($F(1,195) = 5.26, p = .023, \eta^2 = .03, 90\% \text{ CI } [.002, .07]$), which indicated that the size of the punishments was significantly smaller in the individual conditions ($M = 45.61, SD = 34.90$) than in the joint conditions ($M = 56.81, SD = 33.98$). The Accountability main effect ($F(1,195) = 0.20, p = .656, \eta^2 < .01, \text{ CI } [.00, .02]$) and the Responsibility $\times$ Accountability interaction effect ($F(1,195) = 1.14, p = .287, \eta^2 < .01, \text{ CI } [.00, .04]$) were both not significant. See Table 3.3 for the mean number of decrement points and standard deviations per condition.

⁶ An exploratory factor analysis using direct oblimin rotation on all items indicated that the felt personal responsibility items uniquely loaded on a first factor (Rotated factor loadings between $-.75$ and $-.85$, Eigen value = 1.35, Explained variance = 8.41%), the felt external accountability items uniquely loaded on a second factor (Rotated factor loadings between $.73$ and $.83$, Eigen value = 4.03, Explained variance = 25.18%), the felt entitlement items uniquely loaded on a third factor (Rotated factor loadings between $.76$ and $.89$, Eigen value = 3.78, Explained variance = 23.61%), and the felt confidence items (see Footnote 5) uniquely loaded on a fourth factor (Rotated factor loadings between $.40$ and $.95$, Eigen value = 1.80, Explained variance = 11.25%). Thus, the items successfully measured the four unique constructs we intended to measure.

Table 3.3. *Number of points assigned as a function of Accountability and Responsibility (Experiment 3.3)*

	Accountable		Unaccountable		Overall	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Individual	41.90 _a	33.77	49.31 _a	35.93	45.61 _a	34.90
Joint	58.35 _b	34.58	55.31 _b	33.68	56.81 _b	33.98
Overall	49.88	34.98	52.25	34.80	51.07	34.82

Note. Means with differing subscripts within rows are significantly different at the $p < .05$.

Controlling for harvesting decisions

A 2 (Responsibility: Individual versus Joint) $\times$ 2 (Accountability: Accountable versus Unaccountable) ANOVA on harvest decision yielded no significant effects ($p > .10$, $\eta^2 < .01$). On average, participants harvested 2.51 chips ($SD = 2.76$) from the common pool. Moreover, when we added harvest decision as covariate to the 2 (Responsibility: Individual versus Joint) $\times$ 2 (Accountability: Accountable versus Unaccountable) ANOVA on punishment behavior, this covariate was a non-significant predictor of punishment behavior ($p > .10$, $\eta^2 < .02$) and none of the initial effects became non-significant when harvests were included as covariate (see Footnote 4). Thus, the observed differences in punishment behavior were not explained by differences in harvest decisions.

Felt personal responsibility

We tested the influence of responsibility and accountability on felt personal responsibility with a 2 (Responsibility: Individual versus Joint) $\times$ 2 (Accountability: Accountable versus Unaccountable) ANOVA. This analysis showed that both the Responsibility main effect ($F(1,195) = 28.72$, $p < .001$, $\eta^2 = .12$, 90% CI [.06, .19]) and the Accountability main effect ($F(1,195) = 11.77$, $p = .001$, $\eta^2 = .05$, CI [.01, .11]) were significant, whereas the Responsibility $\times$ Accountability interaction effect was non-significant ($F(1,195) = 0.80$, $p = .373$, $\eta^2 < .01$, CI [.00, .03]). As expected, the Responsibility main effect indicated that participants felt more responsible for the sanction decision in the individual conditions ($M = 6.33$, $SD = 2.11$) than in the joint conditions ($M = 4.83$, $SD = 1.92$). In addition, the Accountability main effect indicated that participants felt more responsible in the accountable conditions ($M = 6.09$, $SD = 2.07$) than in the unaccountable conditions ($M = 5.12$, $SD = 2.13$).

Felt external accountability

A 2 (Responsibility: Individual versus Joint) $\times$ 2 (Accountability: Accountable versus Unaccountable) ANOVA was conducted to test how accountable toward others participants felt. This analysis only yielded a significant Accountability main effect ($F(1,195) = 47.91$, $p < .001$, $\eta^2 = .19$, CI [.12, .28]) and a (marginally) significant Responsibility main effect ($F(1,195) = 3.49$, $p = .063$, $\eta^2 = .01$, CI [.00, .05]). As expected, the Accountability main effect indicated that participants felt more accountable toward others for the sanction decision in the accountable conditions ($M = 5.99$, $SD = 1.91$) than in the unaccountable conditions ($M = 3.93$, $SD = 2.28$).

In addition, the Responsibility main effect indicated that participants felt more accountable toward others for the sanction decision in the individual conditions ($M = 5.23$, $SD = 2.36$) than in the joint conditions ($M = 4.66$, $SD = 2.29$).

Felt entitlement

Furthermore, we tested the influence of responsibility and accountability on felt entitlement with a 2 (Responsibility: Individual versus Joint) $\times$ 2 (Accountability: Accountable versus Unaccountable) ANOVA, which only showed a significant Responsibility main effect ($F(1,195) = 7.81$, $p = .006$, $\eta^2 = .14$, CI [.07, .22]). This indicated that participants felt significantly less entitled to decide about sanctioning in the individual conditions ($M = 4.23$, $SD = 2.01$) than in the joint conditions ($M = 5.03$, $SD = 2.07$).

Mediation analysis

Next, we analyzed whether felt personal responsibility, felt external accountability, felt entitlement or a combination of those mechanisms mediated the effect of responsibility (Individual versus Joint) on punishment behavior. To do so, we conducted a bootstrapping analysis for multiple mediator models (with 10,000 re-samples and bias corrected and accelerated confidence intervals; Preacher & Hayes, 2008) using the PROCESS Macro (Hayes, 2013). This method allowed us to test the mediating role of the potential mechanisms against each other because this method not only generates separate indirect effects, but also the differences between those indirect effects.

The direct effect of responsibility (Individual versus Joint) on punishment behavior (total effect = 11.47, $p = .015$) became non-significant by including felt personal responsibility, felt external accountability, and felt entitlement in the model (direct effect = 4.82, $p = .34$). While felt external accountability did not have a significant indirect effect ($b = 1.10$, $p = .36$, indirect effect = -0.62 , 95% Bootstrapping CI [-2.83 , 0.40]), both felt personal responsibility ($b = -2.87$, $p = .026$, indirect effect = 4.29, CI [0.71 , 9.70]) and felt entitlement ($b = 3.65$, $p = .003$, indirect effect = 2.98, CI [0.75 , 6.81]) did.⁷ Furthermore, the indirect effects of felt responsibility and felt entitlement were not significantly different from each other (indirect effect contrast = -1.32 , CI [-6.63 , 3.25]).⁸ Thus, we can conclude with 95% confidence that both felt personal

⁷We also statistically controlled for the manipulation of accountability (Accountable versus Unaccountable), harvesting decisions, and felt confidence, which did not influence the significance of the other effects. Thus, significant effects remained significant and non-significant effects remained non-significant when our manipulation of accountability, harvesting decisions, and felt confidence were added as covariates in the analysis.

⁸We used a parallel multiple mediation analysis (Hayes, 2013; Preacher & Hayes, 2008) – which implied that we a-priori modeled the mediators to be uncorrelated – because we consider them as separate mechanisms. However, there seemed to be a partial correlation between felt personal responsibility and felt entitlement ($r = .159$, $p = .027$), even when we statistically controlled for our manipulation of responsibility (Individual versus Joint) and accountability (Accountable versus Unaccountable), harvesting

responsibility and felt entitlement (but not felt external accountability) independently mediated the effect of responsibility (Individual versus Joint) on punishment behavior (see Figure 3.1).

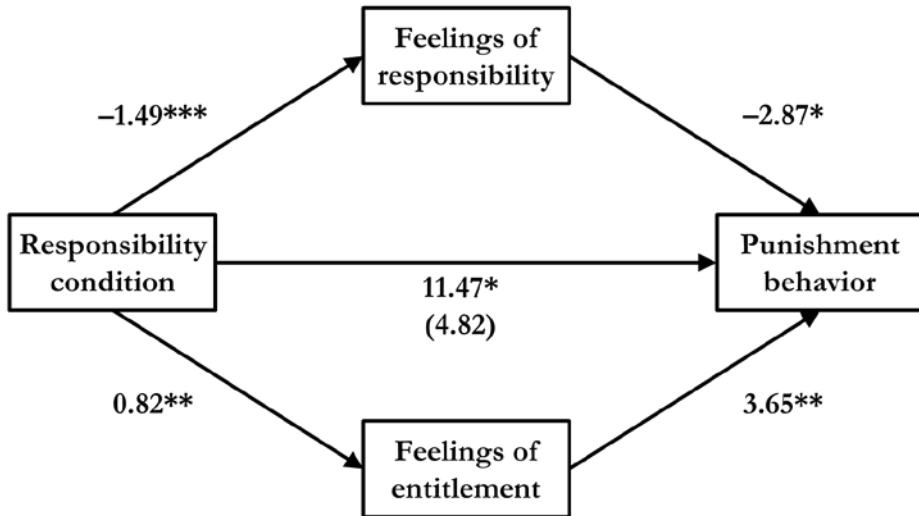


Figure 3.1. Mediation analysis for Experiment 3.3. Values are unstandardized regression coefficients. The parenthetical value is the direct effect of responsibility condition (Individual versus Joint) on punishment behavior, controlling for feelings of responsibility and feelings of entitlement (see Footnote 7). * $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

Experiment 3.3 once again corroborated our reasoning that the personal responsibility one feels for the sanctions is an important determinant of the willingness to punish non-cooperative choice behavior. In line with our predictions, participants individually punished

decisions, and felt confidence. Therefore, we also conducted two serial multiple mediation analyses (Hayes, 2013; Preacher & Hayes, 2008) to test whether the effect of responsibility (Individual versus Joint) on punishment behavior was mediated through felt personal responsibility and felt entitlement in a specific causal order. While (and without) statistically controlling for our manipulation of accountability, harvesting decisions, and felt confidence, these analyses showed support for both (1) an indirect effect through felt personal responsibility followed by felt entitlement ($b = -0.86$, 95% CI $[-2.37, -0.21]$) and (2) an indirect effect through felt entitlement followed by felt personal responsibility ($b = -0.38$, CI $[-1.32, -0.05]$). This indicated that the causal relation went in both directions and additional research is required to determine the exact causal relation between both mediators. However, including a causal order between felt personal responsibility and felt entitlement in the multiple mediator model did not influence the significance of the other effects. Thus, significant effects remained significant and non-significant effects remained non-significant when we tested for (1) a causal relation from felt personal responsibility to felt entitlement and (2) a causal relation from felt entitlement to felt personal responsibility.

a non-cooperator to a lesser extent than jointly. More importantly, the results of Experiment 3.3 extended our previous findings by revealing that this attenuating effect of sharing responsibility on the willingness to punish was mediated by both felt personal responsibility and felt entitlement. Thus, these findings are in line with our reasoning that the willingness to punish those who impair the interests of others is restrained by one's sense of personal responsibility for the harm done. Note that we found these results while controlling for external accountability by both manipulating and measuring it. We further reflect on these findings in the general discussion of this chapter.

■ General discussion

Recent research on the willingness to sanction revealed that people tend to punish non-cooperative choice behavior less often and to a lesser extent than they reward cooperative choice behavior (Chapter 2; Molenmaker et al., 2014; see Sutter et al., 2010; see also Molm, 1997; Wang et al., 2009). Based on these findings one may be tempted to conclude that people are generally reluctant to administer punishments to those who impair the collective interests and prefer to administer rewards to those who serve the collective interests. Despite the fact that the present research replicated this earlier work, we also demonstrated that this is not the complete picture and that such a conclusion would be premature. Going beyond prior research, we argued and showed that the relative preference for rewarding cooperation over punishing non-cooperation is particularly strong when sanctioning decisions are made by individuals, but less so when such decisions are made by groups. By doing so, we revealed that people are not merely concerned about the moral 'wrongness' of inflicting harm, they are also very much affected by their own part in it. That is, people are reluctant to punish those who impaired others' interests because they feel personally responsible for the harm done.

In accordance with early studies on the do-no-harm principle, which suggested that people's reluctance to harm is amplified by their sense of personal responsibility (see e.g., Baron & Ritov, 2009; Cushman et al., 2006; Milgram, 1974; Ritov & Baron, 1990; Royzman & Baron, 2002; Spranca et al., 1991), we demonstrated that the willingness to punish non-cooperation, as opposed to the willingness to reward cooperation, is restrained by feelings of personal responsibility for the harm done. After all, sharing responsibility with a group of people attenuated this tendency to self-restrain the infliction of harm. Non-cooperative choice behavior was punished more often (Experiment 3.1) and to a larger extent (Experiments 3.2 and 3.3) when people decided jointly than when they decided individually, while no such differences were found for reward of cooperative choice behavior. Thus, the relative preference for reward over punishment was less dominant when individuals decided as groups. More importantly, however, we revealed that it was indeed a reduced sense of personal responsibility (together with an enhanced feeling of entitlement) that caused this attenuating effect of sharing responsibility on the reluctance to administer punishments (Experiment 3.3). Altogether, these findings corroborate our reasoning that personal responsibility is an important determinant of the willingness to punish non-cooperative choice behavior, but not necessarily of the willingness to reward cooperative choice behavior.

Other related research on delegation of decision rights also supports our line of reasoning. For instance, a recent study revealed that people often are willing to give others control over allocating unfair offers (e.g., Bartling & Fischbacher, 2012; see also Fershtman & Gneezy, 2001; Hamman, Loewenstein, & Weber, 2010). In a similar vein, control over punishments is frequently outsourced (Andreoni & Gee, 2012; see also Kamei, Putterman, & Tyran, 2014; Markussen, Putterman, & Tyran, 2014; Putterman et al., 2011). What these findings suggest is that people may want to avoid being personally responsible for inflicting harm, even if this means that they give up their own decision rights. Put differently, the fact that non-cooperators are punished is not necessarily what people deem undesirable, it is being personally responsible for administering those punishments what they may want to avoid. Therefore, it would be interesting to examine what factors may determine whether people want to take the responsibility of deterring non-cooperative choice behavior.

As we argued in our introduction, personal responsibility has a self-restraining impact on the willingness to punish non-cooperation, irrespective of whether people are accountable toward others. We proposed that feelings of personal responsibility have an impact on sanctioning because people hold *themselves* internally accountable for the harm done. Accordingly, our findings revealed that, even though participants expected that punishment decisions had to be explained publicly and the group could thus not serve as a shield of anonymity, the grouping of individuals still attenuated their reluctance to punish non-cooperative choice behavior (Experiment 3.3). While there is prior research suggesting that external accountability may have a self-restraining impact on the willingness to punish (Lerner et al., 1998; Piazza & Bering, 2008), our results, as well as other research (Kurzban et al., 2007), do not support this proposition. As such, the findings about the role of external accountability in sanctioning are inconclusive and future research should investigate what moderators seem to determine whether external accountability has an impact on the willingness to sanction. Nevertheless, our findings clearly indicate that this potential impact of external accountability is a different process than the self-restraining impact that personal responsibility has on sanctioning.

Although our findings indicate that accountability toward others does not necessarily affect the willingness to punish, our reasoning concerning the impact of personal responsibility does align with related research concerning others' negative reactions to the administration of punishments. Research has shown that people may frequently be blamed by others for the punishments they administer (e.g., Atwater et al., 2001; Eriksson et al., 2015; Strimling & Eriksson, 2014; Trevino, 1992). Along similar lines, punishers of non-cooperation often are punished in return, which has detrimental effects on their willingness to cooperate (Cinyabuguma, Page, & Putterman, 2006; Denant-Boemont, Masclet, & Noussair, 2007; Herrmann et al., 2008; Nikiforakis, 2008). Administered punishments can thus backfire because non-cooperators may retaliate for the punishments they receive. In contrast, this is not an issue for administering rewards because rewarders of cooperation often are rewarded in return (Kiyonari & Barclay, 2008; Milinski, Semmann, & Krambeck, 2002; Rand et al., 2009). Against this background, it is understandable that people are reluctant to be solely responsible for punishments (but not for rewards). From an evolutionary perspective, these findings may

even suggest that retaliation has been the selective force that explains why feelings of personal responsibility have a restraining impact on punishment decisions but not on reward decisions. Even though retaliation was not an issue in our research, as there was no possibility to respond to the sanctions, we showed that the reluctance to punish is less strong when people decided jointly. A suggestion for future research would therefore be to further investigate whether people would also prefer administering punishments jointly instead of administering them individually when both sanction opportunities are available (see Chapter 5), and whether this especially is the case when retaliation is possible.

Although one should always be cautious when generalizing experimental results to practice, we do want to discuss two practical implications of our findings. First of all, policymakers should realize that people do not use punishments in a similar manner as they use rewards, especially when they feel personally responsible for the sanctions. Whereas both sanction means can effectively promote cooperative choice behavior (Balliet et al., 2011), people are generally reluctant to punish non-cooperative choice behavior and prefer to reward cooperative choice behavior. Consequently, the possibility exists that non-cooperation is punished too little and cooperation is rewarded too much. The type of sanction (i.e., a punishment or reward) that policymakers may implement in real-life situations can thus be decisive for how willing those in control of sanctions are to actually enforce cooperation.

Another point worth mentioning is that policymakers who introduce sanction opportunities should be aware of the fact that people are less reluctant to punish when responsibility for the infliction of harm is shared with others. In real-life, situations in which groups decide jointly about sanctioning occur quite frequently. For instance, in US criminal courts guilty verdicts are given by twelve-person juries, government policies are often determined by task forces that consist of several members, and activities of chief executives in organizations are jointly reviewed and evaluated by the boards of directors. In contrast to our experimental paradigms, people in these real-life groups often discuss what course of action is considered appropriate before they as group actually administer the sanctions. It was beyond the scope of the present research, but it would be an interesting direction for future research to examine the consequences of group discussions on joint sanctioning.

Before closing, we also want to address evidence for future research that can build on the experimental paradigm we used. First, we compared group versus individual decision making, as this method has proven to be effective in attenuating both the feelings of personal responsibility and the infliction of harm (Bandura et al., 1975; Jaffe et al., 1981; Jaffe & Yinon, 1979; Mathes & Kahn, 1975; Meier & Hinsz, 2004). However, whereas this prior research primarily focused on aggressive acts committed by groups and individuals, we are the first to demonstrate that the grouping of individuals amplifies the willingness to punish those who impair the collective interests, which generally is – in contrast to aggression – beneficial to the collective (e.g., Balliet et al., 2011; Fehr & Gächter, 2002; Yamagishi, 1986). It would therefore be a good idea for future research to explore how effective joint sanction opportunities are in promoting cooperation (see Putterman, 2014). By doing so, it may be particularly relevant to address the long-term effects of joint sanctions

(see e.g., Chen, Dang, & Keng-Highberger, 2014; Chen, Pillutla, & Yao, 2009; Mulder, Van Dijk, De Cremer, & Wilke, 2006). It has, for example, been suggested that individually administered sanctions are less effective in sustaining cooperation in the long-run than non-monetary sanctions, such as moral appeals (Chen et al., 2014; Chen et al., 2009). Future research should also investigate whether sanctions administered by groups – which thus are supported by the collective – are more effective in sustaining long-term cooperation than sanctions administered by individuals.

Second, our research also revealed that joint decision making has the potential to reduce the concerns that people may have about their entitlement to impose punishments on others. In our experiments, participants in the joint conditions learned that they as a group decided about sanctioning, while participants in the individual conditions learned that they, and not the other group members, were randomly selected to individually decide about sanctioning. As a result, groups did not only feel more entitled to punish non-cooperative choice behavior than individuals, this also amplified their willingness to administer punishments. Thus, we demonstrated that the feeling of entitlement (besides the feeling of personal responsibility) also has an impact on the (un)willingness to punish. In addition, the grouping of individuals did not attenuate the willingness to administer rewards (Experiments 3.1 and 3.2), which seems to suggest that people are not concerned about a lack of entitlement to (negatively and positively) sanction in general, but about their entitlement to inflict harm on others (Experiment 3.3). Future research should therefore investigate whether entitlement is another reason why people generally prefer rewarding cooperation over punishing non-cooperation.

Finally, it is worth mentioning that another characteristic of our experimental paradigm was that participants took part in the social dilemma themselves (i.e., a second party perspective). Consequently, their personal outcomes were affected by the behavioral feedback we confronted them with. Whereas non-cooperation in such situations seems to ignite revenge-like tendencies (see e.g., Crockett et al., 2014; De Quervain et al., 2004), we showed that participants – who generally were high cooperators themselves – punished less often and to a lesser extent than they rewarded. The fact that we observed a reluctance to punish under these conditions emphasizes again how unwilling people are to do harm (see e.g., Baron, 1993, 1995; Baron & Jurney, 1993). Thus, we showed that the reluctance to punish non-cooperators not only occurs in situations in which people are an impartial third party (see Chapters 2 and 4; Molenmaker et al., 2014), but also in situations in which people are personally involved in the social dilemma at hand. In future research it would be interesting to experimentally manipulate whether a second or third party perspective actually affects the general preference for rewarding cooperation over punishing non-cooperation (see Appendix A).

Conclusions

The present chapter contributes to a more comprehensive understanding of the willingness to punish those who impair the collective interest and reward those who serve the collective interest. By distinguishing between individual decision makers and groups of people, we

reveal that personal responsibility is an important determinant of the willingness to punish non-cooperative choice behavior, but not of the willingness to reward cooperative choice behavior. Although people are generally reluctant to punish non-cooperation and prefer to reward cooperation, our research shows that this relative preference is particularly pronounced when sanctioning decisions are made by individuals, but less so when such decisions are made by groups. That is, we demonstrate that people are reluctant to punish non-cooperation to the extent that they feel personal responsible for the harm done, whereas they are very willing to reward cooperation, regardless of their feelings of personal responsibility. As such, the present chapter sheds new light on people's willingness to enforce cooperation with the use of punishments and rewards.

