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

Molenmaker, Welmer E.

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

Welmer E. Molenmaker

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

Welmer E. Molenmaker

This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO, Grant No. 404-10-026) awarded to Eric van Dijk and Erik W. de Kwaadsteniet.

ISBN: 978-94-6182-718-0

Layout and Printing: Off Page, Amsterdam

Photos: Tokyo, September–October 2014, by Welmer E. Molenmaker.

www.wemolenmaker.com

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

PROEFSCHRIFT

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden

op gezag van Rector Magnificus prof. mr. C. J. J. M. Stolker,

volgens besluit van het College voor Promoties

te verdedigen op donderdag 19 januari 2017

klokke 15:00 uur

door

Welmer Eduard Molenmaker

geboren op 10 december 1984

te Obdam

Promotor

Prof. dr. E. van Dijk

Co-promotor

Dr. E. W. de Kwaadsteniet

Promotiecommissie

Prof. dr. C. K. W. de Dreu	
Prof. dr. W. W. van Dijk	
Prof. dr. I. van Beest	Universiteit van Tilburg
Dr. L. B. Mulder	Rijksuniversiteit Groningen

Voor mijn ouders en Eliza

"I've seen things you people wouldn't believe. (…)
All those moments will be lost in time, like tears in rain."
Roy Batty – Blade Runner

Contents

Chapter 1

General introduction: Why study the willingness to sanction in social dilemmas?

■ Why study the willingness to sanction in social dilemmas?

The greatest challenge for all societies, regardless of how advanced they are, is to ensure and protect the collective welfare. This challenge arises from the fact that the interests of the collective do not necessarily coincide with the personal interests of the people belonging to that collective. Thus, on many occasions, people face a dilemma between furthering the collective interests or furthering their personal interests. Situations that revolve around this conflict of interests are called *social dilemmas* (Dawes, 1980). Consider, for instance, public goods and common resources to which people generally have unlimited access, such as medical care, public transport, national security, drinking water, energy, clean environments, etcetera. Despite the collective's interest to provide and/or preserve such entities, not all people may feel inclined to do so. After all, for an individual it is more profitable to consume from these entities without contributing to their provision or preservation (Olson, 1965; Samuelson, 1954). However, if too many people rather opt for their personal interests (i.e., non-cooperative choice behavior) than for the collective interests (i.e., cooperative choice behavior), public goods cannot be provided and common resources become depleted (Hardin, 1968). Thus, the pursuit of personal interests can have detrimental consequences for the collective welfare. How to prevent such collective failure? How to avoid the disastrous situation that unlimited access to public goods and common resources, as Garrett Hardin (1968) put it in his seminal paper *The Tragedy of the Commons*, eventually brings "ruin to all" (p. 1244)?

One of the most straightforward ways to protect the collective welfare is to make cooperative choice behavior more attractive and non-cooperative choice behavior less attractive (for overviews, see Kollock, 1998; Komorita & Parks, 1995; Messick & Brewer, 1983; Parks, Joireman, & Van Lange, 2013; Van Lange, Joireman, Parks, & Van Dijk, 2013; Weber, Kopelman, & Messick, 2004). To change the relative attractiveness of the options at hand, sanctions readily come to mind as effective solutions to accomplish this goal. Indeed, a recent meta-analysis including nearly 200 effect sizes demonstrated that both the use of positive sanctions (i.e., *rewards*) for cooperative choice behavior and negative sanctions (i.e., *punishments*) for non-cooperative choice behavior can effectively enhance cooperation (Balliet, Mulder, & Van Lange, 2011).[1] This issue of effectiveness has long dominated the scientific thinking on the use of sanctions in social dilemmas. Although it is an important insight that rewards and punishments are effective means to stimulate cooperative choice behavior, a critical question remained unanswered: Are people actually willing to sanction? This question is of critical importance, if only for the obvious reason that someone should first be willing to administer rewards and punishments before they can actually show their effects. For long, hardly any research dealt with this question (for an overview, see Van Dijk, Molenmaker, & De Kwaadsteniet, 2015). To shed more light on this neglected topic, the central theme of this dissertation is the (un)willingness to reward cooperation and punish non-cooperation.

---

[1] *Sanction* is the standard term to refer to both punishment (negative sanction) and reward (positive sanction) in various scientific domains (see e.g., Baldwin, 1971; De Kwaadsteniet, Rijkhoff, & Van Dijk, 2013; O'Reilly & Puffer, 1989; Van Lange, Rockenbach, & Yamagishi, 2014; Weiss & Sachs, 1991).

In the remainder of this first chapter I further introduce this central theme and argue why it is important to study the ingness to sanction in social dilemmas. I first elaborate on defining social dilemmas. This is followed by a brief review of the working of sanctions. Next, I address the ingness to sanction. That is, I explain why it is not self-evident that sanctions, even if they are considered effective, are administered in social dilemmas. Moreover, I address the value of having a better understanding of sanctioning. Finally, I outline the aims of this dissertation and, in doing so, discuss the methods used before I close with a short overview of the remaining chapters.

Defining social dilemmas

Why do social dilemmas pose such a key challenge for societies? Why is it often necessary to protect the collective welfare with the use of sanctions? Why does Hardin (1968) refer to social dilemmas as a 'tragedy'? At their core, social dilemmas are defined by two features: (1) for each individual it is more beneficial not to cooperate than to cooperate, but (2) all individuals are worse off if no one cooperates than if all cooperate (Dawes, 1980). Whereas the first feature prescribes that non-cooperation is the optimal strategy for the individual (i.e., it maximizes an individual's benefit at the lowest cost), the second feature prescribes that it is detrimental to the collective when everyone would follow this strategy. One can basically recognize these two features in almost any situation that involves interdependence among people (i.e., when people's actions affect one another). Social dilemmas take many forms. Two of the most important types are *public good dilemmas* and *common resource dilemmas* (Camerer, 2003; Dawes, 1980; Parks et al., 2013). I first elaborate on these two types of social dilemmas before I proceed to the question of why the conflict between personal interests and collective interests is considered a dilemma.

Public good dilemmas deal with situations in which goods and services can be realized for the public. Whereas people generally are free to use such public goods, the realization of public goods requires that people contribute to their provision (Olson, 1965; Samuelson, 1954). Consider, for instance, blood transfusions that are given to those in (medical) need, as long as there are enough blood donors to supply their provision. As the first feature of social dilemmas prescribes, the optimal strategy for the individual would be not to donate blood because contributing comes with a cost, while yielding no direct benefits. After all, people always get blood transfusions (if necessary) because it is not allowed to exclude anyone from blood transfusions, irrespective of whether or not they donated blood themselves. However, if too many people choose not to donate blood themselves, blood transfusions cannot be provided and all people will potentially be worse off, as is prescribed by the second feature of social dilemmas. Societies face many challenges that are essentially public good dilemmas, such as the provision of medical care, public transport, national security, education, and in a way one could even argue that world peace also is a public good dilemma.

In contrast, common resource dilemmas deal with situations in which common (scarce) resources can be consumed by a group of people. Whereas people generally are free to harvest from common resources, the preservation of common resources requires that people restrict their harvesting (Hardin, 1968; Ostrom, 1990). Take, for example, forests that provide wood,

as long as enough trees are left available to replenish the forest. Since excessive harvesting is more beneficial to the individual than restricted harvesting, the first feature of social dilemmas prescribes that the optimal strategy for the individual would be to cut down as many trees as possible. However, as the second feature of social dilemmas prescribes, if too many people choose to harvest excessively, forests deplete and all people will be worse off. After all, trees can only be harvested once and forests replenish very slowly. Many of the challenges that societies face are essentially common resource dilemmas, such as the preservation of drinking water, energy, clean environments, food, and on a global scale also our planet.

If non-cooperative choice behavior can have such disastrous consequences for the collective interests, why then is it so hard for people to constrain their pursuit of personal interests? Why is it not self-evident that people serve the collective interests and cooperate? Put differently, why are situations in which the collective interest collides with the personal interest considered *dilemmas*? To answer this question, it is important to note that it is generally assumed that people strive to maximize their own benefits at the lowest costs (the self-maximizing assumption). For example, in psychology the hedonic principle of seeking pleasure and avoiding pain is often considered basic to behavior (e.g., Kahneman, Diener, & Schwarz, 1999). All choices that people make can basically be traced back to the primary motive of maximizing pleasant feelings and minimizing painful feelings, which is perhaps best illustrated by the fact that both human and nonhuman species' choice behavior is modulated by brain regions associated with pleasure and pain (e.g., Leknes & Tracey, 2008). The self-maximizing assumption – which is besides psychology also leading in the other scientific disciplines that study people's choice behavior (i.e., anthropology, biology, economics, mathematics, political science, and sociology) – is essentially rooted in evolutionary theory.

As the evolutionary principle of adaptation stipulates, strategies that maximize an individual's benefit at the lowest cost in a particular environment (i.e., optimal strategies) are the most likely strategies to evolve and persist (Darwin, 1859/1962). After all, the individual's reproductive success – referred to as inclusive fitness (Hamilton, 1964) – would be lower with more costly strategies (i.e., suboptimal strategies). This process of natural selection thus causes optimal strategies to triumph over suboptimal strategies because they are more evolutionary adaptive. In evolutionary terms, optimal strategies are fitness maximizing strategies. To illustrate this principle of adaptation before applying it to social dilemmas, consider for instance the evolutionary advantage that great apes species (i.e., orangutans, gorillas, chimpanzees, bonobos, and humans) had over their ancestors more than 14 million years ago when they evolved the capacity to save and transport tools for gathering food (e.g., rocks to crack open nuts). While their ancestors had to depend on the tools available onsite, great apes species were able to bring the tools necessary to gather food (Mulcahy & Call, 2006). Gathering food was a less effortful endeavor (and thus more optimized), which gave great apes species an evolutionary advantage over their now extinct ancestors. This example demonstrates that optimal strategies are the most likely strategies to survive. As such, there is an evolutionary basis to assume that maximizing benefits at the lowest costs to the individual is hardwired into the genetic build of humans and any other species.

The idea that people choose optimal strategies over suboptimal strategies – often referred to as 'rational' choice behavior – is also one of the key assumptions in the mathematical analysis of interdependent situations, known as game theory (Von Neumann & Morgenstern, 1944). Game theory assumes that rational decision makers strive to maximize their own benefits at the lowest costs in interdependent situations, like chess players who play to win without losing any pieces. Based on this assumption, game theory provides a benchmark to predict the choice behavior of several rational decision makers in interdependent situations. This benchmark is known as a Nash equilibrium and reflects a rational decision maker's optimal strategy when all strategies of the other rational decision maker(s) are taken into account (Nash, 1950). The Nash equilibrium is called a dominant strategy because it is the optimal response to all strategies. This is often illustrated with a well-known dyadic game called the prisoner's dilemma (Poundstone, 1992), which also served as the blueprint for defining social dilemmas (Dawes, 1980). In the prisoner's dilemma, two individuals choose between furthering their personal interest (i.e., non-cooperative choice) or their joint interest (i.e., cooperative choice). As shown in Table 1.1, both individuals benefit more if they mutually cooperate (€3) than if they mutually not cooperate (€2). Yet, if only one individual cooperates, this cooperative individual benefits the least (€1), while the other non-cooperative individual benefits the most (€4). Since an individual always benefits more with the non-cooperative choice than with the cooperative choice – regardless of whether the other individual opts for the cooperate choice (€4 instead of €3) or the non-cooperative choice (€2 instead of €1) – the optimal and thus dominant strategy for each individual is not to cooperate.

From the above it becomes apparent that there are dominant strategies in interdependent situations like social dilemmas that optimize the personal interests. However, the prisoner's dilemma also nicely illustrates that Nash equilibria do not necessarily optimize the interests of the collective. Despite the fact that rational decision makers do not cooperate, collectively they are actually better off if they would cooperate. Behold the essence of social dilemmas: "individual rationality leads to collective irrationality" (Kollock, 1998, p. 183). It were these insights, although largely theoretical in nature, that made Hardin realize that collective disasters

Table 1.1. *Prisoner's dilemma*

		Individual 2	
		Cooperative choice	Non-cooperative choice
Individual 1	Cooperative choice	€ 3 / € 3	€ 4 / € 1
	Non-cooperative choice	€ 1 / € 4	€ 2 / € 2

would result from unrestricted freedom to pursue personal interests because this would eventually result in overpopulated societies that would overuse their scarce resources. In his seminal paper *The Tragedy of the Commons*, Hardin (1968) illustrates this point by describing a situation about herdsmen who are all allowed to let their herds graze on a common pasture. It is in each herdsman's personal interest (and thus individually rational) to let as many of their herds graze on the common pasture as possible. However, if all herdsmen do so, all grass will be eaten and destruction of the common pasture is inevitable (and thus collectively irrational). To protect the common pasture from destruction, herdsman should thus restrict their usage of the common pasture.

Since Hardin's seminal paper, a tremendous amount of empirical research has been done to examine the extent to which people indeed act as prototypical rational decision makers in social dilemmas (for comprehensive overviews, see e.g., Dawes, 1980; Gächter & Herrmann, 2009; Kollock, 1998; Komorita & Parks, 1995; Messick & Brewer, 1983; Parks et al., 2013; Pruitt & Kimmel, 1977; Van Lange, De Cremer, Van Dijk, & Van Vugt, 2007; Van Lange et al., 2013; Weber et al., 2004). Such empirical research has consistently shown that people are often not as 'rational' as game theory would predict, as they frequently cooperate with each other. The fact that cooperative choice behavior is quite ubiquitous in both human and nonhuman species thus suggests that cooperation has been evolutionarily adaptive in many situations. To give one example, when individuals are genetically related, cooperation could increase their reproductive success (for a clear overview of the evolutionary explanations why cooperation has evolved, see Barclay & Van Vugt, 2015). Whereas one can basically recognize a social dilemma in almost any situation that involves interdependence among people, there are circumstances that seem to disarm interdependent situations from its 'dilemma' (e.g., when people are close relatives). Yet, this cooperative course of action is very fragile and falls rapidly if people get the impression that others take advantage of their generosity (e.g., Fehr & Gächter, 2000; Hart, Bridgett, & Karau, 2001; Kameda, Tsukasaki, Hastie, & Berg, 2011; see Olson, 1965; Samuelson, 1954). Only a few non-cooperative 'bad apples' can already undermine the cooperative standards in a group (e.g., Kerr et al., 2009). So the challenges that social dilemmas pose are not so much rooted in the unwillingness of people to cooperate, it is the ease with which the balance may tilt towards non-cooperation that makes social dilemmas a real threat to the collective welfare of groups, organizations, and societies.

The working of rewards and punishments

Ever since political philosopher Thomas Hobbes published his influential book *Leviathan* (1651/1991), in which he argued that mutual cooperation can only be brought about by coercion, an often heard advice to 'solve' social dilemmas is to use punishments (e.g., Hardin, 1968; Olson, 1965; Ostrom, 1990). Punishments (but also rewards) are means that basically change the outcome structure of social dilemmas in such a way that cooperative choice behavior becomes more beneficial and non-cooperative choice behavior becomes less beneficial to the individual. However, because of its brutal Hobbesian appearance (Shinada & Yamagishi, 2007), punishment was often criticized and long ignored as proper solution to

social dilemmas (Crowe, 1969; Fox, 1985; Lynn & Oldenquist, 1986; Taylor, 1982). As a result, it took until the pioneering work of Toshio Yamagishi (1986) before punishment finally came into focus and sanctioning established itself as a major theme in social dilemma research (for an historical overview, see Shinada & Yamagishi, 2007). The purpose of the following paragraph is to provide a brief review of the working of sanctions.

Over the last decades, numerous experiments have consistently shown that punishments can enhance cooperation in social dilemmas (for overviews, see Balliet et al., 2011; Van Dijk et al., 2015; Van Lange et al., 2014). Yamagishi (1986, 1988) studied the effectiveness of punishment systems in repeated social dilemmas, and was the first to demonstrate that (after some time) the mere threat of punishment is enough to sustain high levels of cooperation. Whereas the studies by Yamagishi (1986, 1988) dealt with exogenous punishments (i.e., imposed by an external source), Ostrom, Walker, and Gardner (1992) revealed that people are also able to self-govern social dilemmas if they have the opportunity to punish each other (e.g., Fehr & Gächter, 2000, 2002; Gächter, Renner, & Sefton, 2008; Ostrom, Burger, Field, Norgaard, & Policansky, 1999). These kinds of endogenous punishments are often referred to as peer-to-peer punishment (Van Lange et al., 2014). Peer-to-peer punishment not only establishes high levels of cooperation in (small) groups, people also prefer groups in which they have the opportunity to punish each other over groups without any punishment opportunities (Gürerk, Irlenbusch, & Rockenbach, 2006). In their meta-analysis including 154 punishment effect sizes, Balliet et al. (2011) showed that punishment had a medium-sized (see also Cohen, 1988), positive effect on cooperation in social dilemmas (d = 0.70, 95% CI [0.60, 0.80]), especially if the punishment was costly to administer. According to Balliet et al., this indicates that the effectiveness of punishments seems to depend on whether people believe that these sanctions are administered with the intent to serve the collective interests.

One may conclude from the above that the challenges that social dilemmas pose can simply be solved by implementing punishments. However, punishment can also have negative effects. For example, Tenbrunsel and Messick (1999) showed that people generally perceive the decision they make in social dilemmas as an ethical decision (i.e., they consider the ethical aspects of their decision). However, when punishments are installed, this perception changes into a 'business' decision (i.e., they consider the costs and benefits of their decision). Consequently, people cooperate even less with weak punishments than they would do without punishments because they are more focused on the personal benefits of non-cooperation over cooperation (see Gneezy & Rustichini, 2000). Besides these negative effects of punishments on the willingness to cooperate, Mulder and colleagues (2006a) revealed that punishments also have negative effects on people's trust in others. The installation of punishments may communicate that the group does not consist of cooperative group members and that extrinsic means are needed to establish cooperation (see also Mulder, Van Dijk, & De Cremer, 2009; Mulder, Van Dijk, De Cremer, & Wilke, 2006b; Mulder, Van Dijk, Wilke, & De Cremer, 2005). Moreover, the effectiveness of punishment opportunities is undermined when people start to abuse them (e.g., Herrmann, Thöni, & Gächter, 2008; Rand & Nowak, 2011; for a review, see Sylwester, Herrmann, & Bryson, 2013). Non-cooperators may retaliate against

those who punished them (called counter-punishment), which usually is even more detrimental to the collective welfare than the mere presence of non-cooperators (Cinyabuguma, Page, & Putterman, 2006; Denant-Boemont, Masclet, & Noussair, 2007; Nikiforakis, 2008). In addition, cooperative choice behavior may be punished by non-cooperators (called antisocial punishment) when this cooperative choice behavior makes those who do not cooperate look bad (Parks & Stone, 2010). Against this background, one could say that punishments can solve social dilemmas, but only if those in control of punishments use them 'wisely'.

Whereas previous research has mainly focused on the effectiveness of punishments, rewards also change the outcome structure of social dilemmas in such a way that cooperative choice behavior becomes more beneficial and non-cooperative choice behavior becomes less beneficial to the individual. Yet, studies on the effectiveness of rewards are relatively scarce (for overviews, see Balliet et al., 2011; Van Dijk et al., 2015; Van Lange et al., 2014). This lack of attention might stem from the idea that cooperators already display desired choice behavior and it is therefore not necessary to reward them. However, rewarding cooperative choice behavior may also be an indirect punishment for those who are not rewarded. Moreover, the reward of cooperation may also send a signal about the desired behavior, also to those who do not cooperate (Van Dijk et al., 2015). As such, rewards may not only be a stimulant for those who already cooperate, it may also stimulate non-cooperators to change their choice behavior. The meta-analysis on 33 reward effect sizes by Balliet et al. (2011) showed that reward had a medium-sized (see also Cohen, 1988), positive effect on cooperation in social dilemmas (d = 0.51, 95% CI [0.31, 0.70]), especially if the reward was costly to administer. For example, McCusker and Carnevale (1995) adapted Yamagishi's (1986, 1988) experimental design and extended it with a reward system. Their results showed that rewards also sustain high levels of cooperation in social dilemmas. More recently, Rand and colleagues (2009) revealed that rewards even outperform punishments when both sanction means are available. However, this only seems to be the case in repeated social dilemmas in which people can build (positive) reputations (see also Rapoport & Au, 2001; Sefton, Shupp, & Walker, 2007; Sutter, Haigner, & Kocher, 2010; Walker & Halloran, 2004), which indicates that rewards are particularly effective in repeated interactions.

Just as for the implementation of punishments, rewards too can have negative effects. Research in other domains than social dilemmas, for example, showed that rewards may undermine autonomy and the intrinsic motivation to cooperate (e.g., Deci & Ryan, 2000; Ryan & Deci, 2000; for an overview, see Deci, Koestner, & Ryan, 1999). Furthermore, Chen, Pillutla, and Yao (2009) demonstrated that not only punishments but also rewards can have a negative effect on people's trust in others. The installation of extrinsic means – both punishments and rewards – may communicate that they are apparently needed because the group does not consist of intrinsically cooperative members (see Mulder et al., 2005). However, whereas people tend to evaluate (harsh) punishment of non-cooperation negatively (e.g., Atwater, Waldman, Carey, & Cartier, 2001; Eriksson, Andersson, & Strimling, 2015; Trevino, 1992; for a review, see Strimling & Eriksson, 2014), reward of cooperation is evaluated rather positively. For example, a study by Sutter et al. (2010) showed that people are more supportive of groups

that administer rewards than of groups that administer punishments. Furthermore, Kiyonari and Barclay (2008) revealed that those who administer rewards (but not punishments) are often rewarded in return (Milinski, Semmann, & Krambeck, 2002; Rand et al., 2009). Thus, the review presented above indicates that both rewards and punishments can be, and often are, effective tools to stimulate cooperative choice behavior, even though their use can also have negative effects. In addition, whereas the use of punishments is often accompanied with (some) public resistance, this seems less of an issue with the use of rewards. An emerging theme in social dilemma research therefore is whether people consider sanctioning the appropriate course of action. Central to this theme, and the prerequisite for any effect of sanctions, is the question whether people are actually willing to impose sanctions on others. In the following paragraph, I explain the value of addressing this important question.

The willingness to sanction

As a founding father of the contemporary research on sanctions in social dilemmas, Toshio Yamagishi (1986) was one of the first to stress the importance of the willingness to sanction. In social dilemmas, the administration of sanctions is often costly in terms of time, effort, and money. For example, the surveillance and monitoring of the Amazon rainforest to detect illegal logging is an effortful and difficult endeavor for local governments. Since the costs of sanctioning may exceed the benefits for an individual (Edney & Harper, 1978), Yamagishi recognized that the sole reliance on sanctions to solve social dilemmas may give rise to a new *second-order social dilemma* (Oliver, 1980; Yamagishi, 1986). Although all people in a group benefit when high levels of cooperation are established through sanctioning, there should first be enough people willing to incur the costs of administering sanctions. As the two features of social dilemmas prescribe (see Dawes, 1980), the optimal strategy for the individual would be not to participate in costly sanctioning, but it is detrimental to the collective when everyone would follow this strategy because cooperative choice behavior would not be enforced. Given the assumption that people choose optimal strategies over suboptimal strategies, there is a theoretical reason to believe that sanctioning is as unlikely as cooperative choice behavior. However, Yamagishi's (1986, 1988) early work on sanctioning demonstrate that people are willing to install punishments at their own expense if they value others' cooperation but expect or fear that others will not cooperate. More recently, Fehr and Gächter (2002) revealed that some people are even willing to costly punish non-cooperation when any direct gain for themselves is absent. To conclude, despite the fact that costly sanctioning is not self-evident, there are often people willing to use sanction opportunities (for an overview, see Van Dijk et al., 2015).

However, an important question that is often overlooked but needs to be addressed as well, is *why*? Why would people be willing to incur the costs of sanctioning in social dilemmas? Surprisingly enough, this fundamental question about the determinants of sanctioning has long remained unaddressed. Whereas it was generally assumed that people sanction to promote cooperative choice behavior and deter non-cooperative choice behavior (e.g., Yamagishi, 1986), more and more empirical evidence suggests that people do not

necessarily have the collective interest in mind when they reward cooperation and punish non-cooperation. For example, research on the motives underlying punishment indicates that retribution and not deterrence is the primary motive for punishment because people often want to give norm-violators their just deserts (e.g., Carlsmith, 2006; Carlsmith, Darley, & Robinson, 2002; Crockett, Özdemir, & Fehr, 2014; see also Mooijman, Van Dijk, Ellemers, & Van Dijk, 2015). In addition, studies have shown that people are even willing to sanction in single interactions without any repetition (i.e., one-shot games), where deterrence cannot play a role (Bone & Raihani, 2015; Gächter & Herrmann, 2009). These findings fit with the notion that the emotions that people experience in response to others' choice behavior typically fuel their willingness to incur the costs of sanctioning (e.g., De Kwaadsteniet et al., 2013; De Quervain et al., 2004; Fehr & Fischbacher, 2004; Fehr & Gächter, 2002; Nelissen & Zeelenberg, 2009; Pillutla & Murnighan, 1996; Seip, Van Dijk, & Rotteveel, 2014; Wang, Galinsky, & Murnighan, 2009). Taken together, it is too simplistic to assume that people use sanctions just to enforce cooperation.

I argue that there are several reasons why it is important to broaden the focus in social dilemma research to the determinants of the willingness to reward cooperation and to punish non-cooperation. First and foremost, this is important for our theoretical understanding of the psychological processes involved in the use of sanctions in social dilemmas. As mentioned above, emotions are often identified as a proximate mechanism underlying sanctioning (e.g., De Kwaadsteniet et al., 2013; Nelissen & Zeelenberg, 2009; Pillutla & Murnighan, 1996; Seip et al., 2014). However, other psychological processes – which received far less attention in experimental research – may also play a role (for an overview, see Van Dijk et al., 2015). Most importantly, the underpinnings of the willingness to sanction are not necessarily determinants that *foster* sanctioning, but may also be determinants that *hamper* sanctioning. Research on the *do-no-harm principle* has, for instance, shown that people tend to be reluctant to inflict harm on others (e.g., Baron, 1995; Baron & Jurney, 1993; Spranca, Minsk, & Baron, 1991), which may also apply to the use of sanctions. That is, the reluctance to harm may tone down the willingness to punish non-cooperative choice behavior, but not the willingness to reward cooperative choice behavior, even if this would imply that non-cooperation remains unpunished. Examining not only the determinants of sanctioning, but also their boundary conditions, may thus provide a more comprehensive understanding of the psychological mechanisms underlying the (un)willingness to sanction.

Second, focusing on the determinants of the willingness to sanction is also relevant for our theoretical understanding of the evolutionary functions of sanctioning in social dilemmas. In other words, analyzing the psychological determinants of sanctioning (i.e., the proximate level) can provide fruitful insights for analyzing the evolutionary functions of sanctioning (i.e., the ultimate level; Barclay & Kiyonari, 2014; Tinbergen, 1968). The emergence and maintenance of cooperation norms in groups has, for example, been suggested as an explanation of why sanctioning is evolutionary adaptive (Boyd, Gintis, Bowles, & Richerson, 2003; Fehr, Fischbacher, & Gächter, 2002; Fehr & Henrich, 2003; Gintis, 2000; Gintis, Bowles, Boyd, & Fehr, 2003; Henrich et al., 2010; Henrich et al., 2006).

This proposition about the evolutionary function of sanctioning has, however, been challenged by recent studies showing that the effectiveness of sanctioning is undermined when those who are being punished start to retaliate for the punishments they received (Cinyabuguma et al., 2006; Denant-Boemont et al., 2007; Nikiforakis, 2008; Rand, Armao, Nakamaru, & Ohtsuki, 2010). Consequently, alternative hypotheses about the evolutionary functions of sanctioning have been proposed that better fit these recent findings about the use of sanctions (e.g., Krasnow, Cosmides, Pedersen, & Tooby, 2012; Krasnow, Delton, Cosmides, & Tooby, 2016). The above illustrates how identifying determinants of sanctioning and their boundary conditions can steer the reasoning about the evolutionary forces that may or may not have selected for its design in new directions (see also Delton, Krasnow, Cosmides, & Tooby, 2011; Kenrick et al., 2009; Krasnow, Delton, Tooby, & Cosmides, 2013; Todd & Gigerenzer, 2007).

Third, besides the above theoretical considerations, there also is an important practical relevance to identify the determinants of the (un)willingness to sanction. Sanction opportunities may be implemented in real-life social dilemmas to make cooperation more attractive (by rewarding) and non-cooperation less attractive (by punishing). This requires, however, that the people in control of sanctions also use them for that purpose. For a clear understanding of how to implement sanction opportunities in real world social dilemmas, one should understand the conditions under which people actually use sanctions to promote cooperative choice behavior and deter non-cooperative choice behavior. Only then, one is able to organize sanction opportunities in such a way that they can be an effective solution to real-life social dilemmas. Thus, when it comes to solving social dilemmas in real-life, one first needs to understand the determinants of the (un)willingness to sanction before one can implement effective sanction opportunities. Research on the use of sanction may thus provide useful insights about the conditions that enable groups, organizations, and societies to ensure and protect the collective welfare.

■ The present research

The aim of the present dissertation is to identify determinants of the (un)willingness to reward cooperation and punish non-cooperation. Moreover, this research explores the psychological processes underlying sanctioning decisions. In doing so, I take the social psychological standpoint that an individual's choice behavior not only results from personal determinants, but also from situational determinants, as well as the interaction between both types of determinants (Lewin, 1951; Ross & Nisbett, 1991). Adopting this perspective is worthwhile since the use of sanctions in social dilemmas has been approached (almost) exclusively from an evolutionary perspective. As a consequence, prior research has mainly dealt with the question whether evolution selected for a disposition to promote cooperation – which would be a personal determinant of sanctioning – and its underlying mechanisms (Boyd et al., 2003; Fehr et al., 2002; Fehr & Henrich, 2003; Gächter & Herrmann, 2009; Gintis, 2000; Gintis et al., 2003; Henrich et al., 2010; Henrich et al., 2006). Yet, hardly any social dilemma research on the willingness to sanction dealt with the question of how the characteristics of

the situation affect the individual, which does in fact also teach us more about the motives that may underpin the (un)willingness to sanction. To fill this gap in the literature, the present work focuses on situational determinants of the (un)willingness to reward cooperation and punish non-cooperation.

Another aim of this dissertation is to examine whether people are as willing to punish non-cooperation as they are willing to reward cooperation. In the present dissertation, I argue and demonstrate that the willingness to punish differs markedly from the willingness to reward. In the context of social dilemmas, non-cooperative choice behavior tends to signal that the collective welfare may be jeopardized, while cooperative choice behavior tends to signal that things are going well. From this perspective, social dilemmas particularly call for deterrence of non-cooperation. Since non-cooperation is deterred more directly by punishing non-cooperation than by rewarding cooperation, one could argue that people should be more willing to punish non-cooperation than to reward cooperation. However, the use of punishments – in contrast with the use of rewards – implies that one directly inflicts harm on another person. Thus, punishment seems to come with the psychological 'cost' of harming others, whereas reward seems to come with the psychological 'benefit' of favoring others. As such, it may very well be that people actually prefer rewarding cooperation over punishing non-cooperation. The present work empirically tests this proposition. In doing so, I investigate whether the type of sanction people have at their disposal – either reward or punishment – is a primary determinant of their (un)willingness to sanction in social dilemmas.

Research approach

To set up the experiments presented in this dissertation, I used (1) economic games as experimental paradigm for social dilemmas and (2) financial sanctions as a measure of the willingness to sanction. With economic games (like the prisoner's dilemma described earlier), one creates rather straight-forward social decision making situations, which has a couple of key advantages. First, economic games capture the essence of social dilemmas (i.e., a conflict between personal and collective interests) without having to provide a context that has unintended connotations. Experimental research on economic games may thus generate fundamental insights about social decision making. Second, and related to the previous point, economic games are suited for an interdisciplinary approach, which is illustrated by the variety of scientific disciplines that use economic games to study social dilemmas, including anthropology, (evolutionary) biology, economics, mathematics, political science, psychology, and sociology. Third, economic games allow for direct comparisons between different types of social dilemmas because the games can be tailored in such a way that their payoff structures are identical. Fourth and finally, despite the simplicity in terms of structure, economic games are versatile in terms of the emotions, cognitions, and motives they may activate.

Furthermore, although sanctioning may manifest itself in many forms (from verbal compliments and reprimands to the infliction of physical pain), there are several reasons why financial sanctions – both bonuses and fines – provide a suitable approach to measure the willingness to sanction. First, financial sanctions are easy to implement in economic games

because both are numerical. Second, financial sanctions enable one to disentangle the choice to sanction from the size of sanctions. That is, the (un)willingness to sanction may not only be reflected by whether or not people engage in sanctioning, but also by the strength or size of the sanctions they administer. Third, and related to the previous points, financial sanctions allow for direct comparisons between the willingness to reward and the willingness to punish because the cost–consequence–ratio of both sanction means can be kept identical. Fourth and finally, financial sanctions can be presented without using the terms 'reward', 'bonus', 'punishment' or 'fine', which may have unintended connotations. To conclude, economic games and financial sanctions are the ideal methods to study the willingness to reward cooperation and punish non-cooperation in a laboratory setting.

Overview of the following chapters

The central theme of the present dissertation is the (un)willingness to reward cooperation and punish non-cooperation. In this first chapter, I introduced this central theme and argued why it is of critical importance to study the willingness to sanction in social dilemmas. The following three empirical chapters report the results of a series of experiments that I conducted to identify determinants of the (un)willingness to sanction. To identify sanction type (*Reward* versus *Punishment*) as a primary determinant, all empirical chapters (i.e., Chapters 2-4) revolve around the question how willing people are to reward cooperation and to punish non-cooperation. In the first two empirical chapters (i.e., Chapters 2 and 3), I argue that people generally prefer rewarding cooperation over punishing non-cooperation. In addition, each empirical chapter focuses on another situational factor, respectively *what* kinds of (non-)cooperative choice behavior people face (see Chapter 2), *how* they can sanction (see Chapter 3), and *when* they can sanction (see Chapter 4). In the present dissertation, I thus investigate whether the 'what', the 'how', and the 'when' are determinants of the (un) willingness to sanction. Note that the empirical chapters are based on separate articles that have either been published (Chapters 2 and 3) or are still under review for publication (Chapter 4). Although this makes it possible to read each chapter separately and in any order, it also creates some (minor) overlap between their opening paragraphs.

Chapter 2. This chapter – about what kinds of (non-)cooperative choice behavior people face – focuses on how the willingness to reward and punish depends on whether people face a public good dilemma or a common resource dilemma (Molenmaker, De Kwaadsteniet, & Van Dijk, 2014). I argue that people are less willing to punish non-cooperative choice behavior than to reward cooperative choice behavior. More importantly, however, I argue that the willingness to sanction is not only determined by the type of sanction people have at their disposal but is also moderated by the type of social dilemma they face. Specifically, I propose that people are less willing to punish and more willing to reward in public good dilemmas than in common resource dilemmas. The key findings of this chapter are that people punish less often and to a lesser extent than they reward, and that this general preference for rewarding cooperation over punishing non-cooperation is more pronounced in a public good dilemma than in a common resource dilemma (Experiments 2.1 and 2.2).

Chapter 3. This chapter – about how people can sanction – focuses on how personal responsibility has an impact on the (un)willingness to punish and reward (Molenmaker, De Kwaadsteniet, & Van Dijk, 2016). I argue that the general preference for the use of rewards over punishments is particularly pronounced when people feel personally responsible for administering the sanction. That is, I propose that people are reluctant to punish to the extent that they feel personally responsible for the harm done. Feelings of personal responsibility are manipulated by distinguishing between individual decision makers and joint decision makers (i.e., groups of people). The key findings of this chapter are that personal responsibility is an important determinant of the willingness to punish non-cooperation, but not of the willingness to reward cooperation (Experiments 3.1 and 3.2). In addition, this chapter reveals that feelings of personal responsibility have a self-restraining impact on the willingness to punish, regardless of people's external accountability (Experiment 3.3).

Chapter 4. This chapter – about when people can sanction – focuses on how the willingness to reward and punish is influenced by the timing of sanctioning decisions (Molenmaker, De Kwaadsteniet, & Van Dijk, 2016). I argue that people are less willing to sanction before than after the occurrence of others' choice behavior. In the decision environment beforehand others' actual choice behavior still has to take place in the future, whereas in the decision environment afterwards the choice behavior did actually take place in the past. I propose that people are less willing to sanction if the choice behavior has not occurred yet. The key findings of this chapter are that people are less willing to sanction choice behavior that may possibly occur in the future than choice behavior that did actually occur in the past. More specifically, people are less willing to reward and punish when they decide beforehand than when they decide afterwards (Experiments 4.1 and 4.2), regardless of whether they decide directly afterwards or after a time delay (Experiment 4.2).

Chapter 5. The final chapter provides an integration of all the previous chapters. Specifically, Chapter 5 contains a summary and discussion of the findings presented in this dissertation. In doing so, I report the results of two meta-analyses, one on the choice to sanction and one on the sanction size, that I performed on data reported in the empirical chapters of this dissertation and from experiments not included in these chapters (see Appendices A and B). Moreover, general implications of these findings are discussed and directions for future research are presented.

Chapter 2

On the willingness to costly reward cooperation and punish non-cooperation: The moderating role of type of social dilemma

This chapter is based on: Molenmaker, W. E., De Kwaadsteniet, E. W., & Van Dijk, E. (2014). On the willingness to costly reward cooperation and punish non-cooperation: The moderating role of type of social dilemma. *Organizational Behavior and Human Decision Processes, 125*(2), 175-183. doi: 10.1016/j.obhdp.2014.09.005

■ Abstract

Sanction opportunities are often introduced to promote cooperative choice behavior. Experimental studies have repeatedly demonstrated that the use of both rewards and punishments can indeed effectively increase cooperation. However, research has only recently begun to identify the determinants of the willingness to sanction. We investigate the use of costly sanctions to promote cooperation in the context of social dilemmas. We argue and demonstrate that people's willingness to costly reward and punish is not only determined by the type of sanction (reward versus punishment) but is also moderated by the type of social dilemma people face (public good dilemma versus common resource dilemma). In two experiments, we demonstrate that people punish less often and to a lesser extent than they reward, especially in a public good dilemma compared to a common resource dilemma.

■ Introduction

As a member of groups, organizations, and societies, we frequently encounter situations that require us to cooperate with others. This may involve cooperation with relatives, colleagues, and neighbors, but also with strangers. In many of these situations, we may be confronted with others who do not feel inclined to cooperate. The fact that groups often include members who do not cooperate can be detrimental to the collective. For example, group performance may suffer from group members who expect that others will compensate for their lack of effort (i.e., free-riders), organizations may be less efficient when employees work independently of each other, and the natural environment is jeopardized by the many environmental-unfriendly choices people make. Thus, the welfare of the collective is often influenced by the individual choices people make, either positively (in case of cooperative choice behavior) or negatively (in case of non-cooperative choice behavior).

From a collective point of view, it comes as no surprise that authorities often employ sanctions to promote cooperative choice behavior. Sanctions can either be positive means to increase cooperation (i.e., rewards, such as a bonus, price, or privilege) or negative means to decrease non-cooperation (i.e., punishment, such as a fine, penalty, or restriction). Research from a variety of disciplines, such as social psychology (e.g., Blau, 1964; Eisenberger, Lynch, Aselage, & Rohdieck, 2004; Gouldner, 1960; Komorita & Barth, 1985; Thibaut & Kelley, 1959; Wit & Wilke, 1990; Yamagishi, 1986, 1988), organizational behavior (Cropanzano & Mitchell, 2005), and economics (e.g., Abbink, Bolton, Sadrieh, & Tang, 2001; Brosig, Weimann, & Yang, 2004; Fehr & Gächter, 2000, 2002; Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009) have repeatedly shown that both means can effectively promote cooperation (for an overview, see Balliet, Mulder, & Van Lange, 2011). However, to effectively promote cooperative choice behavior, decision makers in control of rewards and punishments should of course first be willing to provide and impose them. After all, sanction opportunities can only show their effect if they are actually administered.

In this chapter, we address this important aspect of implementing sanction opportunities by investigating people's willingness to costly reward cooperation and costly punish non-cooperation. Specifically, we focus on two factors that may determine whether people consider sanctioning the appropriate course of action (see March, 1994; Messick, 1999): the type of sanction they can administer and what kind of (non-)cooperative choice behavior they face.

The need for sanctions

To investigate the willingness to sanction, it is first important to understand why it is often necessary for authorities to promote cooperative choice behavior. Although cooperation is socially beneficial, the occurrence of mutual cooperation is not self-evident. After all, the collective interest does not necessarily coincide with the personal interest (Hardin, 1968; Olson, 1965; Samuelson, 1954). As a consequence, people often face the dilemma whether to further the collective interest or their personal self-interest. Situations that revolve around such a conflict are often referred to as social dilemmas (for overviews, see Parks, Joireman, & Van Lange, 2013; Van Lange, Joireman, Parks, & Van Dijk, 2013; Weber, Kopelman, &

Messick, 2004). Social dilemmas constitute the context in which we investigate the willingness to sanction.

Two important types of social dilemmas are the *public good dilemma* and the *common resource dilemma* (Camerer, 2003; Dawes, 1980). Public good dilemmas model the problem of realizing public goods from which all people may benefit, irrespective of whether or not they individually contributed to their provision. Blood transfusions, public broadcasting, and medical care are all real-world examples of public good dilemmas. For an individual it is more profitable not to contribute because contributing is costly, and eventually everybody can make use of public goods. However, if too many people choose not to contribute, public goods cannot be provided and the collective will be worse off than if people would decide to contribute. In common resource dilemmas, by contrast, people have to decide whether or not to restrict harvesting from scarce common resources. For example, energy conservation, overfishing, and water scarcity are all problems arising from excessive consumption. While it is in the individual's interest to consume from such common resources, these resources will deplete if people do not constrain their harvesting.

The use of sanctions is usually proposed as a means to promote cooperation (Hardin, 1968; Olson, 1965), and early social dilemma research on the willingness to sanction showed that there are also people willing to incur costs for punishments if they expect or fear that others will defect (Yamagishi, 1986, 1988). In fact, people prefer societies with sanctioning institutions over sanction-free societies (Gürerk, Irlenbusch, & Rockenbach, 2006). Furthermore, the level of cooperation increases when there are people present who are prepared to sanction at their own expense (e.g., Fehr & Fischbacher, 2004; Fehr & Gächter, 2000, 2002; Milinski, Semmann, & Krambeck, 2002; Ostrom, Walker, & Gardner, 1992; Rand et al., 2009; Sefton, Shupp, & Walker, 2007; Walker & Halloran, 2004). Consequently, the willingness to costly reward cooperators and punish non-cooperators is considered to be a prerequisite for cooperation (e.g., Boyd & Richerson, 1992; Fehr & Rockenbach, 2004; Gintis, 2000; Gintis, Bowles, Boyd, & Fehr, 2003; Gintis, Henrich, Bowles, Boyd, & Fehr, 2008). Altogether, the general picture emerging from these earlier studies is that there are indeed people who are willing to provide and impose sanctions to promote cooperation, even if it is costly to do so.

Whereas people may use costly sanctions, very little research focused on the distinction between the willingness to use rewards for cooperation versus punishments for non-cooperation (for an exception, see Sutter, Haigner, & Kocher, 2010; see also Molm, 1997; Wang, Galinsky, & Murnighan, 2009). Are people equally willing to costly reward cooperation as they are willing to costly punish non-cooperation, or do they have a preference for one over the other? This question needs to be addressed to identify the determinants of people's willingness to administer sanctions in social dilemmas. In the present chapter, we propose that people's willingness to use punishments may differ markedly from their willingness to use rewards. More importantly, we argue and show that people's willingness to reward and punish depends on whether they face a public good dilemma or a common resource dilemma.

The willingness to costly reward and punish

The majority of research on costly sanctioning focused on punishment of non-cooperation, thereby neglecting the possibility to reward cooperation. This is surprising since rewarding cooperation also proved to be an effective means to promote cooperation (Balliet et al., 2011). We argue that people may have a general preference for administering rewards over punishments as a means to promote cooperation. Why do we think this is the case? Research on the *do-no-harm principle* showed that, even if the overall benefit outweighs the harm done, 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). The same reasoning may apply to the use of rewards and punishments since both are beneficial in the sense that they can enhance cooperation. However, only the use of punishments – in contrast with the use of rewards – implies that one directly inflicts harm to another person. Based on this reasoning, we thus propose that people may be more reluctant to punish than to reward (cf. Abbink, Irlenbusch, & Renner, 2000; Offerman, 2002).

The do-no-harm principle has never been related to costly sanctioning in social dilemmas. Some earlier studies, however, provide indirect evidence for the above reasoning. For instance a study by Sutter et al. (2010) showed that people are more supportive of sanctioning institutions that administer rewards than sanctioning institutions that administer punishments. Furthermore, research on the use of secondary sanctions demonstrated that people who punished non-cooperators were punished themselves, whereas people who rewarded cooperators were rewarded themselves (Cinyabuguma, Page, & Putterman, 2006; Denant-Boemont, Masclet, & Noussair, 2007; Kiyonari & Barclay, 2008; Milinski et al., 2002; Nikiforakis, 2008; Rand et al., 2009). Such secondary sanctioning suggests that people evaluate punishments negatively and rewards positively. We believe that such differences may also be observed for first-order sanctioning. In fact, we argue that people may be less willing to costly punish non-cooperative choice behavior than to costly reward cooperative choice behavior. More importantly, however, the willingness to sanction may not only be determined by the type of sanction (reward versus punishment) but may also be moderated by the type of social dilemma people face (public good dilemma versus common resource dilemma).

Sanctioning in public good dilemmas versus common resource dilemmas

Both public good dilemmas and common resource dilemmas refer to the same conflict of interests (i.e., self-interest versus collective interest), and can be structured as each other's equivalents in terms of payoffs (Camerer, 2003; Dawes, 1980). When it concerns the willingness to costly sanction, however, we argue that public good dilemmas and common resource dilemmas should certainly not be treated similarly because they differ in the way in which the initial property is distributed (e.g., Camerer, 2003; Dawes, 1980; Van Dijk & Wilke, 1997, 2000). In public good dilemmas, people initially possess property themselves (private property), and decide whether or not they contribute this property to a public good. In common resource dilemmas, the property is initially located in a common resource (collective property), and

people decide whether or not they consume from this common resource. As a consequence, the property rights in public good dilemmas are considered private, whereas the property rights in common resource dilemmas are considered collective (Van Dijk & Wilke, 1997; see also Van Dijk, Wilke, & Wit, 2003).

Prior research showed that property rights (e.g., people's perception that money they decide on is their own) may lower people's willingness to allocate parts of their property to others (e.g., Cherry, 2001; Hoffman, McCabe, Shachat, & Smith, 1994; Muehlbacher & Kirchler, 2009; Oxoby & Spraggon, 2008). This indicates that people who consider themselves to have rights over a property feel more entitled to retain it. As such, the initial distribution of property may be an important factor. In a study on ultimatum bargaining, Leliveld, Van Dijk, and Van Beest (2012) showed that bargainers felt less entitled, and were less willing to allocate money to themselves, if they felt that the money was initially owned by their opponent than if the property was initially owned by themselves. We argue that the same reasoning may not only apply to the own versus other's property distinction, but also to the distinction between private and collective property. In other words, people may feel they are more entitled to private property than to collective property (i.e., the starting point in public good dilemmas versus common resource dilemmas, respectively). Moreover, we reason that this perception is not egocentric in the sense that it only affects how people perceive their own rights, but that it is more general, and that people also assign these rights to others. Thus, people may also feel that others are more entitled to their own private property than to the collective property.

Importantly, we argue that the dissimilarity in property rights across public good dilemmas and common resource dilemmas may lead to a difference in how people respond when they observe cooperation or non-cooperation. More specifically, since cooperation in a public good dilemma requires one to give up private property, people may consider this highly commendable (and thus more rewardable). After all, in the public good dilemma, people freely give up their property rights to further the collective interest. By contrast, cooperation in the context of a resource dilemma primarily means that one keeps collective what was collective. Although this may be seen as commendable, it may be less in need of an explicit reward because one had no (personal) rights to the collective resource to start with. However, when one does infringe on collective property in a common resource dilemma, people may consider such non-cooperation more objectionable (and thus more punishable) than not giving up private property in a public good dilemma. Thus, we argue that people may be less willing to costly punish non-cooperative choice behavior and more willing to costly reward cooperative choice behavior in public good dilemmas than in common resource dilemmas.

This proposition connects to research by Janoff-Bulman, Sheikh, and Hepp (2009) on the attribution of credit and blame across proscriptive and prescriptive morality (see also Goodwin & Darley, 2012; Janoff-Bulman & Carnes, 2013). In their research, they distinguish between morality that prescribes *what to do* (prescriptive morality) and morality that proscribes *what not to do* (proscriptive morality). Although both forms of morality can serve collective interests to the same extent, Janoff-Bulman et al. (2009) argue that they derive from distinct underlying systems. Prescriptive morality is activation-based and focusses on *doing what is good*

(for others). Proscriptive morality, on the other hand, is inhibition-based and focusses on *not doing what is bad* (for others). As they argued, prescriptive morality has a less mandatory and less strict nature than proscriptive morality. In agreement with this, they noted that participants indicated greater disapproval (moral blame) for people who hurt rather than did not help others (Janoff-Bulman et al., 2009, Study 5). Moreover, participants indicated greater approval (moral credit) for people who helped rather than did not hurt others (Studies 6 and 7).

Since cooperation requires people *to do what is good* for the collective in public good dilemmas (i.e., giving up private property) and *not to do what is bad* for the collective in common resource dilemmas (i.e., not infringing on collective property), we can draw a parallel between these types of social dilemmas and the two forms of morality that Janoff-Bulman et al. (2009) distinguish (prescriptive and proscriptive morality). The difference in moral (dis)approval between prescriptive morality and proscriptive morality is in line with our reasoning that (non-) cooperative choice behavior may be punished less and rewarded more in public good dilemmas than in common resource dilemmas.

■ Experiment 2.1

As a first test of our ideas, we used a *third party sanction paradigm* (see Fehr & Fischbacher, 2004) in which participants observed the choice behavior of two persons in a one-shot social dilemma task. This social dilemma context was either presented as a public good dilemma or a common resource dilemma. Subsequently, participants could either costly punish or costly reward one of the persons. Thus, participants themselves were not involved in the social dilemma (third party perspective), they only had the opportunity to sanction others in either a public good dilemma or a common resource dilemma. Based on our reasoning, we expected that participants (1) would punish a low cooperator less than that they would reward a high cooperator and (2) that they would punish a low cooperator less and reward a high cooperator more in a public good dilemma than in a common resource dilemma.

Method

Participants and design

Participants were 122 students at Leiden University (92 women and 30 men; M_{age} = 19.67 years, SD_{age} = 3.48).[1] A 2 (Social Dilemma Type: Public Good versus Common Resource) × 2 (Sanction Type: Reward versus Punishment) between-participants factorial design was used. Afterwards, two participants indicated on an open question about the purpose of the experiment that they did not understand their task correctly and that they made a mistake

[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).

on our main dependent variable. Therefore, we excluded the data of these participants from our analyses.[2]

Procedure

We invited participants to the laboratory to participate in an experiment on "group decision making". Upon arrival we seated them in separate cubicles, each containing a personal computer that was used to give instructions and to register their responses. A computer automated procedure assigned the participants randomly to one of the four conditions.

In the instructions we informed participants they had to perform a joint task with two fellow participants whose identities were unknown. In the joint task, one person (named person C) would observe the other two persons (named person A and B) performing a task. Subsequently, this person would have the opportunity to assign decrement coins (*punishment* conditions) or increment coins (*reward* conditions) to one of the two persons. Random selection would ostensibly determine which role each participant got. All participants then learned that they were assigned to the role of observer (person C). The instructions also explained to the participants that their choices made in the joint task determined how much extra money they could earn on top of their initial participation fee. After the study was conducted, a lottery would randomly select participants that would actually receive this extra money.

After this, the task that person A and B would perform (i.e., the social dilemma context) was explained. The type of social dilemma was manipulated in such a way that the payoff structures of the two social dilemmas were identical (see e.g., Dawes, 1980; Van Dijk & Wilke, 1995, 1997; Van Dijk et al., 2003). In the *public good* conditions person A and B would be endowed with 100 coins worth €0.10 each (in total €10, which is approximately US $13). They could contribute these coins to a common pool or keep the coins for themselves. The contributed coins in the common pool would be multiplied by 1.5 and divided equally among the two persons. The kept coins would accrue totally to oneself. In the *common resource* conditions person A and B could harvest maximally 100 coins each from a common pool of 200 coins (also worth €0.10 each). The harvested coins would accrue totally to oneself and the coins left in the common pool would be multiplied by 1.5 and divided equally among the two persons. After participants read the rules of the social dilemma task, we posed five practice questions to ensure comprehension of the task. For example, we asked participants how many coins the persons possessed (*public good* conditions) or how many the persons could maximally harvest from the common pool (*common resource* conditions). After answering each question, the correct answer was disclosed.

Next, participants read instructions about their role in the joint task. In the *punishment* conditions participants learned that they would be endowed with 200 coins (also worth €0.10 each) which they could assign as decrement coins to one of the persons in the task (we never used the word 'punishment'). The assigned decrement coins would be multiplied by

[2] Inclusion of the data of the excluded participants did not alter the pattern of results.

3 and subtracted from the individual outcome of the person concerned. The kept coins would accrue totally to oneself. The instructions in the *reward* conditions were identical, the only difference being that the participants learned that they could assign their coins as increment coins and that these assigned coins would be multiplied by 3 and added to the individual outcome of the person concerned (we also never used the word 'reward'). After participants read the instructions, we posed seven practice questions to ensure comprehension of their role in the joint task. For example, we asked participants how many coins would be subtracted from (*added to*) the individual outcome of the person concerned when they would assign 1 decrement coin (*increment coin*). After answering each question, the correct answer was disclosed.

After the practice questions, the joint task started and participants had to wait until person A and B finished reading and made their decision. This took about 2 minutes. Next, participants received (bogus) feedback about the individual decisions of the two persons. In the *public good* conditions participants learned that person A contributed 80 coins to the common pool (i.e., high cooperator) and person B 20 coins (i.e., low cooperator). In the *common resource* conditions participants learned that person A harvested 20 coins from the common pool (i.e., high cooperator) and person B 80 coins (i.e., low cooperator). After receiving the feedback about the individual decisions of the two persons, participants first had to decide if and to whom they wanted to assign coins as decrement coins in the *punishment* conditions or as increment coins in the *reward* conditions. When participants decided to assign coins to person A or person B, they had to indicate how many coins they assigned. The number of coins participants could assign to a person ranged from 0 to 200 coins.

Finally, we checked the type of social dilemma and type of sanction manipulations. Besides, participants had to fill in questions to measure their comprehension and the believability of the joint task. At the end of the experiment, we thoroughly debriefed, paid (1 course credit or €3 monetary compensation) and thanked the participants. When the experiment was performed by all the participants, a lottery randomly selected five participants who received their actual earnings from the joint task.

Results

Manipulation checks

To check the manipulation of type of social dilemma, we asked participants whether the two persons (A and B) could harvest or contribute coins in the joint task. All participants except one (i.e., 99.2%) answered this question correctly.[3] To check the manipulation of type of sanction, we asked participants whether they could assign decrement coins or increment coins. All participants (i.e., 100%) answered this question correctly. Altogether, these results indicated that our manipulations were successful and we included the data of all 120 participants in the analyses.

[3] We included the data of these participants in our experiments because they did not indicate that they made a mistake on our main dependent variables and exclusion of the data did not alter the pattern of results.

Choice to sanction

As a first test we analyzed the influence of the Social Dilemma Type and the Sanction Type on the proportion of participants who chose to sanction (N = 120). A binary (Sanction Choice: 0 = not sanctioned, 1 = sanctioned) logistic regression yielded a significant Sanction Type main effect (B = 1.22, SE = 0.47, Wald (df=1) = 6.82, p = .009, Odds Ratio = 3.40, 95% CI [1.36, 8.53]) and a significant Social Dilemma Type × Sanction Type interaction effect (B = 3.42, SE = 1.25, Wald (df=1) = 7.49, p = .006, Odds Ratio = 30.63, CI [2.64, 355.22]). The Social Dilemma Type main effect was non-significant (B = 0.15, SE = 0.45, Wald (df=1) = 0.12, p = .733, Odds Ratio = 1.16, CI [0.49, 2.79]).

In line with our expectations, we can conclude that the proportion of participants choosing to punish (66.1%) was lower than the proportion of participants choosing to reward (86.8%). In addition, the results indicated that the proportion of participants choosing to punish was lower in the public good dilemma (53.3%) than in the common resource dilemma (79.3%), $\chi^2(1) = 4.44$, p = .035, Odds Ratio = 3.35, CI [1.06, 10.59]. In contrast, the proportion of participants choosing to reward was higher in the public good dilemma (96.8%) than in the common resource dilemma (76.7%), $\chi^2(1) = 5.41$, p = .02, Odds Ratio = 9.13, 95% CI [1.05, 79.54]. See Table 2.1 for the frequencies.

Table 2.1. *Number of participants who sanctioned and not sanctioned, as a function of Sanction type and Social dilemma type (Experiment 2.1)*

	Punishment		Reward	
	Public Good	Common Resource	Public Good	Common Resource
Sanctioned	16	23	30	23
Not sanctioned	14	6	1	7
Total	30	29	31	30

The choice whether or not participants wanted to sanction also involved deciding *who* they wanted to sanction. Not surprisingly, a Chi-Square test (N = 92) showed that the large majority of participants chose to sanction the high cooperator (92.5%) in the reward conditions and the low cooperator (94.9%) in the punishment conditions, $\chi^2(1) = 69.35$, $p < .001$, Odds Ratio = 226.63, 95% CI [39.37, 1304.48].[4]

Sanction size

A 2 (Public Good versus Common Resource) × 2 (Reward versus Punishment) ANOVA on the number of coins (N = 120) yielded a significant Sanction Type main effect

[4] This indicates that the majority of participants did not sanction anti-socially (Herrmann, Thöni, & Gächter, 2008; Parks & Stone, 2010).

($F(1,116) = 30.29$, $p < .001$, $\eta^2 = .19$, 90% CI [.10, .29]) and a significant Social Dilemma Type × Sanction Type interaction effect ($F(1,116) = 8.60$, $p = .004$, $\eta^2 = .06$, CI [.003, .15]). The Social Dilemma Type main effect was non-significant ($F(1,116) = 1.763$, $p = .187$, $\eta^2 = .01$, CI [.00, .06]). The Sanction Type main effect indicated that the size of the punishments ($M = 11.47$, $SD = 11.94$) was smaller than the size of the rewards ($M = 30.31$, $SD = 24.29$), which explained 19% of the variance.

The Social Dilemma Type × Sanction Type interaction effect explained 6% of the variance. To break down this interaction effect we performed simple-effect analyses. This revealed that the size of the rewards was significantly larger in the public good condition than in the common resource condition ($F(1,116) = 9.23$, $p = .003$, $\eta^2 = .06$, 90% CI [.01, .14]), while the size of the punishments in the public good condition and the common resource condition did not differ significantly ($F(1,116) = 1.27$, $p = .263$, $\eta^2 < .01$, CI [.00, .54]). The simple-effect analyses also showed that in the public good condition the size of the punishments was significantly smaller than the size of the rewards ($F(1,116) = 36.18$, $p < .001$, $\eta^2 = .23$, CI [.13, .33]). In the common resource condition the size of the punishments was smaller than the size of the rewards ($F(1,116) = 3.25$, $p = .074$, $\eta^2 = .02$, CI [.00, .08]), although this difference was only marginally significant and the confidence intervals indicated that the precision of this estimation seemed low. See Table 2.2 for the mean number of coins and standard deviation per condition.

Table 2.2. *Number of coins assigned, as a function of Sanction type and Social dilemma type (Experiment 2.1)*

	Punishment		Reward		Overall	
	M	*SD*	*M*	*SD*	*M*	*SD*
Public Good	8.80	12.25	37.42	25.72	23.34	24.73
Common Resource	14.24	11.15	22.97	20.64	18.68	17.10
Overall	11.47	11.94	30.31	24.29	21.05	21.36

Discussion

In line with our reasoning, Experiment 2.1 demonstrated that participants were less willing to costly punish non-cooperative choice behavior than that they were willing to costly reward cooperative choice behavior. Furthermore, they were less willing to costly punish non-cooperative choice behavior and more willing to costly reward cooperative choice behavior in the public good dilemma than in the common resource dilemma. Thus, the results of Experiment 2.1 provide first evidence for our idea that people's willingness to costly reward and punish depends on the social dilemma context. This shows that people are less willing to punish those who do not cooperate by not giving up private property in public good dilemmas than those who do not cooperate by infringing on collective property in common resource dilemmas. Furthermore, people are more willing to reward those who cooperate by giving

up private property in public good dilemmas than those who cooperate by not infringing on collective property in common resource dilemmas.

Although the findings of Experiment 2.1 are in line with our reasoning, it should be noted that the difference in willingness to sanction between the public good dilemma and the common resource dilemma seemed to emerge particularly in the choice whether or not to sanction and the size of the rewards, and less so in the size of the punishments. Put differently, although fewer participants used punishments in the public good condition than in the common resource condition, the size of the punishments did not differ significantly between both conditions. This result is in line with earlier findings by Cubitt, Drouvelis, and Gächter (2011), who also found no difference in punishment *size* between both types of social dilemmas. Since people are reluctant to harm others (e.g., Baron, 1993, 1995; Baron & Jurney, 1993) and the participants' choice options were limited to punishment only in the punishment conditions (such as in the experiment by Cubitt et al., 2011), a possible explanation might be that the participants' willingness to punish a non-cooperator was so low, that a potential difference in the punishment size between both types of social dilemmas could not occur (i.e., a floor effect).

Nevertheless, our results – as well as many other studies (e.g., Fehr & Gächter, 2002; Yamagishi, 1986, 1988) – showed that there are indeed people who are willing to costly punish non-cooperation, even though the punishments they administered were small in size. At first glance this might seem to contradict the do-no-harm principle (e.g., Baron, 1995), which states that people are reluctant to inflict harm on others. However, the fact that participants' choice options were limited to one type of sanction only, might provide an explanation for these results. Note that under these conditions, a decision not to punish in Experiment 2.1 means that one does not sanction at all. Thus, choosing not to punish would mean that one would not respond to the observed inequality. It may very well be that – despite their reluctance to inflict harm on others – participants felt it would be undesirable if they would not react at all (for other research indicating that people are willing to reduce inequality in outcomes, see e.g., Bazerman, White, & Loewenstein, 1995; Fehr & Schmidt, 1999; Schroeder, Steel, Woodell, & Bembenek, 2003; Tricomi, Rangel, Camerer, & O'Doherty, 2010). The solution would then be to punish, but punish mildly.

In our second experiment, we did present participants the opportunity to choose between rewarding and punishing. We anticipated that within an opportunity to address the observed inequality without having to punish, we might observe a greater reluctance to use the punishment option. Furthermore, this procedure increased the validity of our research since in real-life people are often able to choose between the use of punishments and rewards. Thus, in Experiment 2.2 we tested whether the type of social dilemma has an effect on the size of the sanctions when people can choose their own preferred type of sanction (reward or punishment). As such, we are able to measure people's default response to others' cooperative and non-cooperative choice behavior.

■ Experiment 2.2

In Experiment 2.2 we used the same third party sanction paradigm as in Experiment 2.1, with the only exception that participants now had the opportunity to choose between the use of costly rewards or costly punishments. Based on our reasoning, we again expected that participants (1) would punish a low cooperator less than that they would reward a high cooperator and (2) that they would punish a low cooperator less and reward a high cooperator more in a public good dilemma than in a common resource dilemma.

Method

Participants and design

Participants were 114 students at Leiden University (94 women and 20 men; M_{age} = 18.99 years, SD_{age} = 1.73) (see Footnote 1). We used a one-factor between-participants design with 2 levels (Social Dilemma Type: Public Good versus Common Resource).

Procedure

The procedure of this second experiment was almost identical to the procedure of Experiment 2.1. However, now participants could choose between assigning their coins as decrement coins (i.e., punishment) or as increment coins (i.e., reward). First participants indicated if and what kind of coins they wanted to assign.[5] Only when participants decided to assign decrement/increment coins, they indicated to whom and how many of these coins they wanted to assign. Thus, after they chose the type of coin, they could only assign their coins as that type of coin to only one of the two persons and not, for example, as decrement coins to one person and as increment coins to the other person. As in Experiment 2.1, the number of coins participants could assign to a person ranged from 0 to 200 coins (also worth €0.10 each).

Results

Manipulation checks

To check the manipulation of type of social dilemma, we asked participants whether the two persons (A and B) could harvest or contribute coins in the joint task. All participants except two participants (i.e., 98.3%) answered this question correctly (see Footnote 3). This result indicated that our manipulation was successful and we included the data of all 114 participants in the analyses.

---

[5] In the instructions and questions, the type of coins (decrement coins versus increment coins) were mentioned in a counterbalanced order. Including instruction order as variable in our analyses did not reveal any effect of the order of instructions on people's willingness to sanction and this did not influence the significance of other effects. Thus, significant effects remained significant and non-significant effects remained non-significant when instruction order was added as variable in the analyses.

Choice to sanction

First we analyzed the influence of the Social Dilemma Type on the proportion of participants who chose to sanction (N = 114). A binary (Sanction Choice: 0 = not sanctioned, 1 = sanctioned) logistic regression yielded a non-significant Social Dilemma Type main effect (B = 0.27, SE = 0.79, Wald (df=1) = 0.12, p = .733, Odds Ratio = 1.31, 95% CI [0.28, 6.13]). The proportion of participants who sanctioned (i.e. decided to either punish or reward) in the public good condition (i.e., 94.6%) was equal to the proportion of participants who sanctioned in the common resource condition (i.e., 93.1%). See Table 2.3 for the frequencies.

The choice whether or not participants wanted to sanction also involved deciding what type of sanction they wanted to use. We therefore compared the number of participants who punished and rewarded (N = 107), which showed that the proportion of participants choosing to punish (22.4%) was lower than the proportion of participants choosing to reward (77.6%), Z = 13.70, p < .001, Odds Ratio = 3.46, 95% CI [2.30, 7.70]. This lower proportion of participants choosing to punish compared to the proportion of participants choosing to reward was similar across the public good condition (punish = 19.9%; reward = 81.1%) and the common resource condition (punish = 26%; reward = 74%), $\chi^2(1) = 0.12$, p = .73, Odds Ratio = 1.51, CI [0.60, 3.77].

Furthermore, when participants chose to sanction, they subsequently indicated who they wanted to sanction (N = 107). As in Experiment 2.1, a Chi-Square test showed that the majority of participants chose to reward the high cooperator (92.5%) or punish the low cooperator (94.9%), $\chi^2(1) = 62.09$, p < .001, Odds Ratio = 103.12, 95% CI [20.39, 521.49] (see Footnote 4).

Table 2.3. *Number of participants who sanctioned (punished or rewarded) and not sanctioned, as a function of Social dilemma type (Experiment 2.2)*

	Public Good	Common Resource
Sanctioned	53	54
Punished	10	14
Rewarded	43	40
Not Sanctioned	3	4
Total	56	58

Sanction size

Since participants could choose what kind of coins they wanted to assign, the assigned coins could either be decrement coins or increment coins (or neither when participants choose not to assign any coins). Thus, participants chose a type of sanction in Experiment 2.2, whereas type of sanction was a between-participants factor in Experiment 2.1. To be able to compare

the sanction size of both types of sanction in one analysis, we coded decrement coins (i.e., punishment) as negative values (multiplied by -1), increment coins (i.e., reward) as positive values (multiplied by 1) and no coins assigned as zero. This enabled us to test our reasoning in one analysis since a negative mean implied that participants punished more than they rewarded, a positive mean implied that participants rewarded more than they punished, and a mean of zero implied that participants punished and rewarded equally.

A One-Way ANOVA on the coded number of coins (N = 114) yielded a significant Intercept ($F(1,112) = 47.58$, $p < .001$, $\eta^2 = .29$, 90% CI [.18, .39]). This indicated that the size of the rewards was larger than the size of the punishments since the value of mean coded number of coins differed positively from zero ($M = 24.98$, $SD = 39.42$), which explained 29% of the variance. In line with our expectations, a significant Social Dilemma Type main effect ($F(1,112) = 4.22$, $p = .042$, $\eta^2 = .03$, CI [.00, .09]) was found. Thus, the results indicated that the preference to reward over punishment was stronger in the public good condition (M = 32.59, SD = 46.98) than in the common resource condition (M = 17.64, SD = 28.97). This explained 3% of the variance, although the confidence intervals indicated that the precision of this estimation is low. See Table 2.4 for the mean coded number of coins and standard deviation of the participants who chose to reward or punish per condition.

Table 2.4. *Number of coins assigned by participants who chose to sanction, as a function of Sanction type and Social dilemma type (Experiment 2.2)*

	Punishment			Reward		
	N	*M*	*SD*	*N*	*M*	*SD*
Public Good	10	13.30	2.61	43	45.53	7.03
Common Resource	14	17.50	1.36	40	31.70	3.67
Overall	24	15.75	1.38	83	38.87	4.10

Discussion

The results of Experiment 2.2 were again in line with our reasoning. Thus, participants were less willing to costly punish non-cooperative choice behavior than that they were willing to costly reward cooperative choice behavior. Furthermore, they were less willing to costly punish non-cooperation than they were willing to costly reward cooperation in the public good dilemma than in the common resource dilemma. Experiment 2.2 thus further corroborated our idea that people's willingness to costly reward and punish depends on whether they face a public good dilemma or a common resource dilemma.

Experiment 2.2 showed strong support for our proposition that people have a general preference for administering rewards over punishments as a means to promote cooperation. When given the opportunity to choose between reward and punishment, the large majority of participants in both the public good dilemma and the common resource dilemma chose to

reward cooperation. As we suggested in the discussion of Experiment 2.1, having an option to choose reward rather than punishment as a way to address the observed inequality, can reduce the reliance on punishment because it allows one to address the observed inequality without inflicting harm. The fact that the proportion of participants who chose to punish in Experiment 2.2 (22.4%) was much lower than what we observed in the punishment condition of Experiment 2.1 (66.1%) accords with this notion. Note, however, that both patterns fit with our reasoning that people are reluctant to punish. Moreover, the observed difference in size of sanctions also corroborates the notion that people indeed consider choice behavior as more commendable (and thus more rewardable) and less objectionable (and thus less punishable) in public good dilemmas than in common resource dilemmas.

■ General discussion

The use of rewards and punishments is usually proposed as a means to promote cooperative choice behavior (Hardin, 1968; Olson, 1965). Whereas earlier research showed that people are often inclined to provide and impose sanctions (see e.g., Fehr & Gächter, 2002; Yamagishi, 1986; for an overview, see Balliet et al., 2011), no clear comparison has been made between the willingness to use costly rewards versus punishments. In the present chapter, we filled this gap in the literature by demonstrating that the willingness to sanction is not only determined by the type of sanction (reward versus punishment) but is also moderated by the type of social dilemma people face (public good dilemma versus common resource dilemma).

According to the do-no-harm principle, people are reluctant to inflict harm on others, even if the overall benefit outweighs the harm done (e.g., Baron, 1993, 1995; Baron & Jurney, 1993). In line with this principle, we observed that people punished non-cooperation less than they rewarded cooperation. In Experiment 2.1, we found that when people had to decide whether or not they wanted to punish, they sanctioned less often and to a lesser extent than people who had to decide whether or not they wanted to reward. Experiment 2.2 further showed that when people had the opportunity to choose between reward and punishment, the large majority chose to reward and subsequently rewarded to a greater extent than the people who chose to punish. Thus, when in the position to respond to others' choice behavior, people are reluctant to inflict harm on those who do not cooperate (i.e., punish), but they prefer to benefit those who do cooperate (i.e., reward).

Furthermore, we reveal a difference in costly sanctioning across public good dilemmas and common resource dilemmas. In two experiments, we demonstrated that people use punishments less often than rewards (Experiment 2.1) and to a lesser extent (Experiments 2.1 and 2.2) in public good dilemmas compared to common resource dilemmas. In public good dilemmas the property rights are private and in common resource dilemmas they are collective (Van Dijk & Wilke, 1997; see also Van Dijk et al., 2003). Against this background, people are less willing to punish those who do not cooperate by not giving up private property in public good dilemmas than those who do not cooperate by infringing on collective property in common resource dilemmas. Likewise, people are more willing to reward those who cooperate by giving up private property in public good dilemmas than those who cooperate by not

infringing on collective property in common resource dilemmas. Thus, these findings identify the allocation of property rights as an important determinant of people's willingness to reward cooperation and punish non-cooperation.

Implications, limitations, and directions for future research

Although one should always be cautious when generalizing experimental results to practice, two interesting implications may derive from our findings. First of all, the possibility of rewarding should not be overlooked by policymakers when implementing sanction opportunities in real-world social dilemmas. Earlier research indicated that costly rewarding comes with social benefits, while costly punishment has social costs (Kiyonari & Barclay, 2008; Milinski et al., 2002; Rand et al., 2009). In accordance with such secondary sanctioning research, our results (which pertained to first-order sanctioning) revealed that people have a relative preference for administering rewards over administering punishments. Although reward and punishment can both be effective means to enhance cooperation (Balliet et al., 2011), people will generally consider punishing the less appropriate course of action (see also March, 1994; Messick, 1999) and thus use punishments less often and to a lesser extent than rewards. Thus, when the opportunity to punish is not sufficiently used, implementing reward opportunities can be decisive to promote cooperative choice behavior.

The second important implication is that public good dilemmas and common resource dilemmas should not be treated as each other's equals, even in case of identical payoff structures. Although both social dilemmas appeal to the same conflict of interests (Camerer, 2003; Dawes, 1980), prior research already acknowledged that the type of social dilemma certainly has an impact on the people involved. For instance, the type of dilemma influences which social norm people adhere to (Van Dijk & Wilke, 1995), whether they experience social responsibility (Van Dijk & Wilke, 1997), and whether they prefer installing a leader (Rutte & Wilke, 1984; Rutte, Wilke, & Messick, 1987; Van Dijk et al., 2003). Our results extend such findings and emphasize the importance of taking the type of social dilemma (public good dilemma versus common resource dilemma) into consideration when implementing reward or punishment opportunities in real-world social dilemmas. Although people will generally prefer to reward cooperative choice behavior over punishment of non-cooperative choice behavior, this preference will be stronger in public good dilemmas than in common resource dilemmas.

Furthermore, the implications are also relevant for the research on costly sanctioning. For example, the majority of social dilemma research on costly sanctioning only focused on punishment of non-cooperation and was conducted within public good dilemmas (see Balliet et al., 2011). Our results reveal that the type of sanction (reward versus punishment) and the type of social dilemma (public good dilemmas versus common resource dilemmas) both determine people's willingness to sanction. This indicates that one should take both determinants into account when investigating the underlying considerations of costly sanctioning.

At this point, it is relevant to address some limitations of our research, and give some suggestions for future research. In our two studies, we used a third party sanction paradigm in which participants were not dependent on the choice behavior they observed (see Fehr &

Fischbacher, 2004). With this procedure, we ensured that the participants' interpretation of others' choice behavior was not colored by self-interest. At the same time, this procedure may constitute a limitation since one may wonder whether our idea that people are reluctant to use punishments would also be observed in situations in which people are part of the group themselves. People's willingness to punish non-cooperation may be higher when they are personally involved in the social dilemma because in such situations, revenge-like motives might justify the infliction of harm (see e.g., De Quervain et al., 2004). As such, it might be interesting for future research to test whether our finding that people are reluctant to administer punishments can be replicated using a second party sanction paradigm (see Chapter 3).

Another characteristic of our experimental paradigm is that we designed in such a way that we were able to compare people's willingness to administer rewards and punishments directly. Since the cost–consequence-ratio (i.e., 1:3) was identical for rewards and punishments, we merely varied the type of sanction. Thus, it was equally costly for participants to influence a person's outcome either positively (i.e., reward) or negatively (i.e., punishment). Nevertheless, an intrinsic element of rewards and punishments is the fact that the joint outcomes of a group rise with the use of rewards, whereas they drop with the use of punishments. As a result, people's preference to administer rewards over punishments may also reflect a preference to further the collective welfare (Sutter et al., 2010; see also Charness & Rabin, 2002). Such an efficiency perspective can, however, not explain why the preference to administer rewards over punishments is moderated by the type of social dilemma people face. It would, however, be a good idea for future research to further explore the differences in the underlying motives of rewarding versus punishing.

Finally, future research might want to explore to what extent discretion over what type of sanctions people can use may play a role in the use of sanctions. The results of our two studies suggest that while the majority of people generally opt for rewards instead of punishments, a substantial proportion of people may also administer punishments when this is the only type of sanction at their disposal. As we mentioned in the discussion of Experiment 2.1, participants in the punishment conditions may have punished because they felt inclined to reduce the inequality they observed (see e.g., Tricomi et al., 2010), even though this meant they had to inflict harm (i.e., punish). The opportunity to choose between reward and punishment, however, offers people an opportunity to respond to the inequality and at the same time refrain from punishment. It may therefore be interesting for future research to investigate people's willingness to sanction when the inequality is less obvious and they only have discretion over one type of sanction (see Chapters 3 and 4).

Conclusions

The present chapter addresses the importance of distinguishing between the willingness to costly reward and punish. As we were able to demonstrate, people's willingness to sanction is determined by the type of sanction (reward versus punishment) and the type of social dilemma they face (public good dilemmas versus common resource dilemmas). As a result, we hope that the present chapter represents an important step in identifying the determinants

of people's willingness to promote cooperative choice behavior with the use of sanctions. When one knows the determinants of costly sanctioning, one is able to predict how sanction opportunities will be used to promote cooperative choice behavior.

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, prices, 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, Haigner, & 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, Zyzniewski, & 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_{age} = 21.22 years, SD_{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) x 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) x 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 ($ps > .10$, $\eta^2 < .01$). 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 x 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) x 2 (Sanction Type: Punishment versus Reward) ANOVA on harvest decision was performed, which yielded no significant effects ($ps > .10$, $\eta^2 < .01$). 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_{age} = 20.42 years, SD_{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) x 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 α = .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) x 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 ($ps > .10$, $\eta^2 < .01$). 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) x 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 x 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	
	M	*SD*	*M*	*SD*	*M*	*SD*
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) x 2 (Sanction Type: Punishment versus Reward) ANOVA on harvest decision was performed, which yielded no significant effects (*p*s > .10, η^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) x 2 (Sanction Type: Punishment versus Reward) ANOVA on sanction behavior, this covariate was a non-significant predictor of sanction behavior (*p* > .10, η^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_{age} = 20.38 years, SD_{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) x 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 α = .88), felt external accountability (Cronbach's α = .88) and felt entitlement (Cronbach's α = .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 α = .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) x 2 (Accountability: Accountable versus Unaccountable) ANOVA only yielded a significant Responsibility main effect ($F(1,195) = 5.26$, $p = .023$, $\eta^2 = .03$, 90% 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 ($F(1,195) = 0.20$, $p = .656$, $\eta^2 < .01$, CI [.00, .02]) and the Responsibility x Accountability interaction effect ($F(1,195) = 1.14$, $p = .287$, $\eta^2 < .01$, 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	
	M	*SD*	*M*	*SD*	*M*	*SD*
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) x 2 (Accountability: Accountable versus Unaccountable) ANOVA on harvest decision yielded no significant effects ($ps > .10$, $\eta^2 < .01$). 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) x 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) x 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 x 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) x 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) x 2 (Accountability: Accountable versus Unaccountable) ANOVA, which only showed a significant Responsibility main effect (F(1,195) = 7.81, p = .006, η^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

[7] We 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.

[8] We 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).

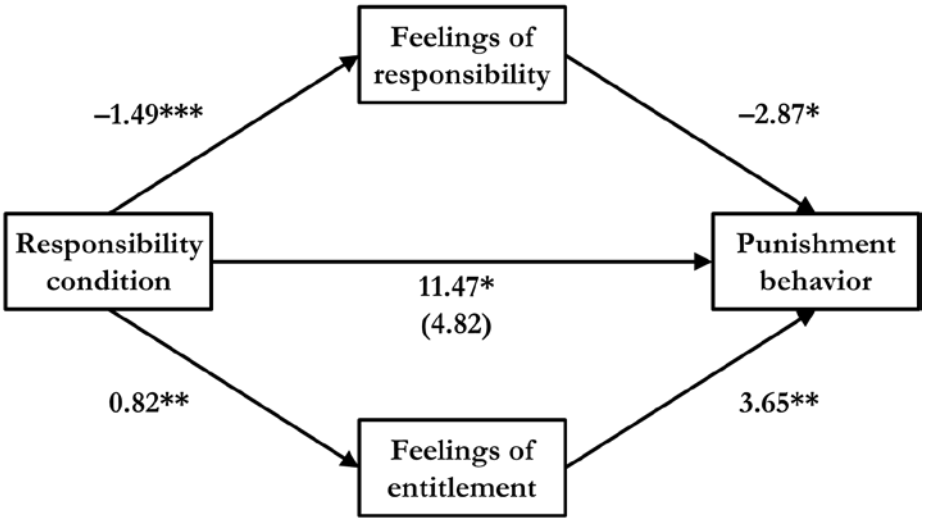


Figure 3.1. Mediation analysis for Experiment 3.3. Values are unstandardized regression coefficients. The parenthetical value is the direct effect of responsibility condition (Individual versus Joint) on punishment behavior, controlling for feelings of responsibility and feelings of entitlement (see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

Chapter 4

The willingness to costly reward cooperation and punish non-cooperation before versus after the choice behavior: Sanctioning the past, the present, or the future

This chapter is based on: Molenmaker, W. E., De Kwaadsteniet, E. W., & Van Dijk, E. (2016). The willingness to costly reward cooperation and punish non-cooperation before versus after the choice behavior: Sanctioning the past, the present, or the future. *Manuscript under review*

■ Abstract

Numerous empirical studies demonstrate that sanctions can promote cooperative choice behavior. However, to successfully implement sanction opportunities it is not only important to know that sanctions can work, it is also important to know under what conditions people are actually willing to sanction. Although people can decide about sanctioning at various moments in time, it either involves a decision before or a decision after others' choice behavior. We argue that people are less willing to sanction choice behavior that may possibly occur in the future (i.e., beforehand) than choice behavior that did actually occur in the past (i.e., afterwards). In two experiments we showed that people sanction less often and to a lesser extent when sanctioning decisions are made before instead of after the choice behavior. These findings corroborate our reasoning that decision timing has an impact on the willingness to employ costly rewards for cooperation and punishments for non-cooperation.

■ Introduction

The implementation of sanctions is often proposed as an effective means to promote cooperative choice behavior. This proposition is supported by numerous empirical studies demonstrating that positive sanctions (i.e., rewards like bonuses, prices, or privileges) and negative sanctions (i.e., punishments like fines, penalties, or restrictions) stimulate cooperation and minimize non-cooperation (e.g., Fehr & Gächter, 2002; Komorita & Barth, 1985; McCusker & Carnevale, 1995; Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009; Sefton, Shupp, & Walker, 2007; Wit & Wilke, 1990; Yamagishi, 1986, 1988; for overviews, see Balliet, Mulder, & Van Lange, 2011; Van Dijk, Molenmaker, & De Kwaadsteniet, 2015). However, to successfully implement sanction opportunities, it is not only important to know that sanctions can work. It is also important to know under what conditions people are actually willing to administer sanctions. After all, sanctions can only show their effect if those in control of rewards and punishments are willing to bear the costs of administering them (i.e., in terms of money, effort and/ or risk). The current research addresses this critical question by focusing on the timing of sanction decisions. That is, we examine whether decision timing has an impact on the willingness to employ rewards for cooperation and punishments for non-cooperation.

Although one can decide to sanction others' choice behavior at various moments in time, it either involves a decision *before* or a decision *after* the choice behavior. Consider, for instance, managers in organizations who have sanctions at their disposal to steer employees' choice behavior in the desired direction. When evaluating the performance of their employees, they can decide afterwards whether employees who furthered the success of the organization should be rewarded (e.g., by giving them a bonus) and whether employees who weakened the success of the organization should be punished (e.g., by cancelling their vacation leave). Whereas in this case the sanctioning decisions are made after the employees' choice behavior, sanctioning decisions can also be made before the choice behavior. For instance, managers can also decide beforehand if those employees who will further the success of the organization should be rewarded and whether those employees who will weaken the success of the organization should be punished. Until now, very little experimental research has been done to investigate whether the willingness to sanction choice behavior *beforehand* differs from the willingness to sanction choice behavior *afterwards*. This is unfortunate since a better understanding of the impact of decision timing not only has a practical relevance for those who implement sanction opportunities, it may also shed new light on how people make sanctioning decisions.

The need for sanctions

To address the timing of sanctioning decisions, one should first understand why authorities may need to implement sanction opportunities. There are many everyday situations (at work, home, or other places) that require us to cooperate with others. Although cooperative choice behavior is beneficial to the collective in these situations, it is not self-evident that individuals will cooperate (Dawes, 1980). To give just one example, employees may be more motivated to further their own careers than to further the success of their organization. Personal interest may thus collide with the collective interest (Hardin, 1968; Olson, 1965; Samuelson, 1954).

Situations in which personal interests conflict with collective interests are generally referred to as social dilemmas (for overviews, see Parks, Joireman, & Van Lange, 2013; Van Lange, Joireman, Parks, & Van Dijk, 2013; Weber, Kopelman, & Messick, 2004). It is within this context of social dilemmas that we investigate the willingness to sanction.

Public transport, medical care, and clean environments are all examples of goods and services that stand or fall with individuals' willingness to provide and maintain them because they can, in fact, be used freely by everyone (Samuelson, 1954). If too many people choose not to contribute to the provision of these public goods, it may eventually be impossible to provide them and all will be worse off. However, public goods provision – which is a specific type of social dilemma called the public good dilemma (Camerer, 2003; Dawes, 1980) – is not only a problem on a societal or global level, it is a problem for groups in general. After all, the performance of groups is usually based on each group members' effort to attain the goals of the group, and if too many group members lack effort (i.e., free-ride) the performance of the group may be jeopardized. Thus, to prevent collective failure, it is often necessary to make cooperation more attractive and non-cooperation less attractive (e.g., Hardin, 1968; Olson, 1965).

Straightforward tools to increase the relative attractiveness of cooperation over non-cooperation are rewards for those who cooperate, and punishments for those who do not cooperate (Messick & Brewer, 1983; Van Lange, Rockenbach, & Yamagishi, 2014). Indeed, several studies have shown that the opportunity to use costly punishments enables people to self-govern social dilemmas (e.g., Fehr & Gächter, 2000, 2002; Gürerk, Irlenbusch, & Rockenbach, 2006; Ostrom, Burger, Field, Norgaard, & Policansky, 1999; Rand et al., 2009; Yamagishi, 1986). For example, Ostrom, Walker, and Gardner (1992) demonstrated that in small groups people especially punish those group members who tend to free-ride on the generosity of others. A costly sanctioning opportunity enhances the level of cooperation in such groups. Furthermore, some people even are willing to punish others' selfishness when direct gains for themselves are absent (Fehr & Fischbacher, 2004; Fehr & Gächter, 2002). In fact, the presence of individuals within a group who are willing to deter non-cooperation with costly punishments (i.e., strong reciprocators) is considered to be a prerequisite for the evolution of cooperation (e.g., Boyd & Richerson, 1992; Fehr & Rockenbach, 2004; Gintis, 2000; Gintis, Bowles, Boyd, & Fehr, 2003; Gintis, Henrich, Bowles, Boyd, & Fehr, 2008).

Whereas prior research demonstrated that people may be willing to use costly punishments for non-cooperation, the use of costly rewards for cooperation has received far less attention (for some exceptions, see e.g., Rand et al., 2009; Sefton et al., 2007). This is remarkable, as rewards are just as effective as punishments in promoting cooperation (e.g., Balliet et al., 2011). In addition, the scarce research done on rewarding revealed that people generally prefer to administer rewards over punishments (Chapters 2 and 3; Molenmaker, De Kwaadsteniet, & Van Dijk, 2014, 2016; also see Molm, 1997; Sutter, Haigner, & Kocher, 2010; Wang, Galinsky, & Murnighan, 2009). Thus, to identify whether the willingness to promote cooperative choice behavior *before* the choice behavior differs from the willingness to promote cooperative choice behavior *after* the choice behavior, one should address both reward of

cooperation and punishment of non-cooperation. In the present research, we therefore take both types of sanctions into consideration and investigate their sensitivity to the timing of sanctioning decisions.

The timing of sanctioning decisions

Why would the timing of sanctioning decisions have an impact on the willingness to costly reward and punish? To answer this question, we draw attention to the fact that a decision beforehand differs markedly from a decision afterwards. One of the most apparent differences between these two moments in time is that afterwards people decide about the sanctioning of choice behavior that has actually taken place in the past, whereas beforehand people decide about the sanctioning of choice behavior that may or may not take place in the future. We argue that this fundamental difference between facing choice behavior that *may possibly occur in the future* (i.e., beforehand) or choice behavior that *did actually occur in the past* (i.e., afterwards) radically alters the decision environment. More importantly, we aim to show that this alteration of the decision environment has an impact on the willingness to employ costly rewards for cooperative choice behavior and costly punishments for non-cooperative choice behavior.

Decision timing alters the decision environment, first and foremost, because in the decision environment beforehand (as opposed to the decision environment afterwards) it is not known yet whether particular choice behavior will actually occur in the future. Research on the disjunction effect (e.g., Shafir, 1994; Shafir, Simonson, & Tversky, 1993; Shafir & Tversky, 1992; Tversky & Shafir, 1992) demonstrated that uncertainty about outcomes may induce nonconsequential reasoning (see also Langer, 1975; Messé & Sivacek, 1979; Quattrone & Tversky, 1984). That is, if the outcome of a particular situation is unknown, people are often reluctant to think through the implications of all possible outcomes (e.g., Tversky & Shafir, 1992) and are less likely to make decisions based on uncertain information than on certain information (Van Dijk & Zeelenberg, 2003). To illustrate this, it is informative to consider an example given by Tversky and Shafir (1992). In one of their studies on the disjunction effect, participants were presented the hypothetical scenario in which they had just taken a qualifying exam and had either passed the exam, failed the exam, or did not know whether they had passed or failed the exam. Next, the willingness to book a vacation to Hawaii was measured. The majority of the participants were willing to book the vacation when they knew that they had passed the exam. The same preference was observed when they had failed the exam. However, when they did not know whether they had passed or failed the exam, only a minority of the participants were willing to book the vacation. Apparently, they reasoned that they could not book the vacation if they did not know their test result. Participants' decisiveness to book the vacation was thus hampered by the uncertainty about the outcome of the exam, which is an example of nonconsequential reasoning. After all, if they would have known their test result, they would have booked the vacation, regardless of whether they had passed or failed the exam. We believe that a similar effect may be observed for sanctioning decisions. That is, people may be less willing to employ costly sanctions if the choice behavior is not known yet (see Van Dijk, De Kwaadsteniet, & Mulder, 2009). Since in the decision environment beforehand others'

actual choice behavior still has to take place – whereas in the decision environment afterwards it did actually take place – we thus argue that people may be less willing to sanction choice behavior beforehand than afterwards.

In addition, there may be another reason as to why decision timing may alter the decision environment. Decision timing may also have an impact on how people experience others' choice behavior. Scholars from various disciplines have proposed that emotions are an important proximate mechanism underlying the willingness to employ sanctions (e.g., Darley & Pittman, 2003; Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007; Fehr & Fischbacher, 2004; Fehr & Gächter, 2002; Pillutla & Murnighan, 1996; Rotemberg, 2008; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003; Seip, Van Dijk, & Rotteveel, 2009; Van't Wout, Kahn, Sanfey, & Aleman, 2006; Wang et al., 2009). For instance, the anger that people experience about selfish peers or unfair choice behavior in general has been identified as a driving force of costly punishment (e.g., De Kwaadsteniet, Rijkhoff, & Van Dijk, 2013; Nelissen & Zeelenberg, 2009; Seip, Van Dijk, & Rotteveel, 2014). The anger experienced when confronted with others' non-cooperative choice behavior in the past does, however, not necessarily resemble the anger associated with thinking about the possibility that non-cooperative choice behavior may occur in the future. After all, how people experience future events may frequently differ in intensity and quality from how they experience present or past events (e.g., Loewenstein, 1996; Loewenstein & Lerner, 2003; Miceli & Castelfranchi, 2015). Inspired by work on the disjunction effect (e.g., Tversky & Shafir, 1992; Van Dijk & Zeelenberg, 2003), research revealed that people experience less intense emotions based on uncertain outcomes than on certain outcomes (Van Dijk & Zeelenberg, 2006; Wang, Li, & Jiang, 2012). Since it is not known yet what choice behavior will actually occur in the decision environment beforehand (as opposed to the decision environment afterwards), one might thus expect that the intensity of the emotions experienced in the decision environment beforehand is lower than in the decision environment afterwards. The fact that people may experience less intense emotions beforehand than afterwards is consistent with the notion that decision environments can be characterized as "cold" when decisions are made about future events and as "hot" when decisions are made about present events (see Loewenstein, 1996; Loewenstein & Schkade, 1999; Wang et al., 2011). The timing of sanctioning decisions may thus also constitute a distinction between hot and cold decision environments, which may be another reason why people may be less willing to sanction choice behavior in the decision environment beforehand than in the decision environment afterwards.

The present research

In this chapter, we examine whether the timing of sanction decisions has an impact on the willingness to reward cooperative choice behavior and punish non-cooperative choice behavior. More specifically, we test the prediction that people are less willing to sanction choice behavior when sanctioning decisions are made before (instead of after) the occurrence of others' choice behavior. In two experiments, we use a third party sanction paradigm in which participants have the opportunity to reward a cooperator or the opportunity to punish a non-cooperator (see Fehr & Fischbacher, 2004; Molenmaker et al., 2014). We manipulate

the timing of the sanction decision by presenting participants either with choice behavior that could possibly occur in the future (i.e., beforehand) or with choice behavior that did actually occur in the past (i.e., afterwards). Subsequently, we measure participants' willingness to sanction that particular choice behavior by having them decide whether to employ a sanction (i.e., choice to sanction) and decide about the size of sanction they employed (i.e., sanction size). We examine both the choice to sanction and the sanction size since both are indicators of the willingness to sanction (Molenmaker et al., 2014, 2016). As outlined in the introduction of this chapter, we argue that there are two reasons why people would be less willing to sanction choice behavior beforehand than afterwards: decision timing may (1) induce nonconsequential reasoning and (2) constitute a hot/cold distinction.

To our knowledge, we are the first to investigate whether the willingness to sanction differs between the decision environment beforehand and the decision environment afterwards. However, it should be noted that the timing of decisions does connect with studies that were specifically aimed at investigating the behavioral validity of two experimental methods frequently used in research on social decision making: the *strategy method* and the *direct-response method* (see e.g., Brandts & Charness, 2011; Brosig, Weimann, & Yang, 2003; Fischbacher, Gächter, & Quercia, 2012; Selten, 1967). The strategy method requires individuals to make precompiled strategies for responding to all feasible choices that others' could possibly make (i.e., decide about multiple possible choices), whereas the direct-response method, by contrast, requires individuals to only respond to others' actual choices (i.e., decide about a single choice). Although these two response methods were not designed to study the impact of decision timing, the strategy method has similarities with situations in which sanction decisions are made beforehand and the direct-response method has similarities with situations in which sanction decisions are made afterwards.

The results of the studies that directly compared the strategy method with the direct-response method are, however, mixed. Some research, for instance, revealed differences between both methods on decisions that involved emotions, especially when the decisions were punitive in nature (for an overview, see Brandts & Charness, 2011). For example, unfair offers were rejected less frequently in strategy response than in direct response (Güth, Huck, & Müller, 2001; Oxoby & McLeish, 2004). Moreover, lower punishment rates in economic interactions were observed in strategy response than in direct response to another's choices (Brandts & Charness, 2003; Brosig et al., 2003). Finally, a study by Falk, Fehr, and Fischbacher (2005) demonstrated that non-cooperation was punished to a lesser extent in strategy response than in direct response, even though the proportion of individuals choosing to punish was equal in both methods. Unfortunately, very few studies compared reward-like decisions in the strategy method versus the direct-response method, and the results of the studies that did were inconsistent (Brandts & Charness, 2011). Research that is indirectly related to rewarding has, for instance, shown that trust seems to be repaid less frequently in strategy response than in direct response (Casari & Cason, 2009; but see Büchner, Coricelli, & Greiner, 2007). However, people were not less willing to reciprocate others' cooperative choice behavior in strategy response than in direct response (Fischbacher et al., 2012; Muller, Sefton, Steinberg, &

Vesterlund, 2008; also see Reuben & Suetens, 2012). Whereas these studies were on trust and reciprocity (and did not address rewarding directly), a study by Brandts and Charness (2003) focused on people's use of actual rewards in economic interactions. Their results indicated that lower reward rates were observed in strategy response than in direct response to another's choices, although this difference was not statistically significant.

Research that directly compared the strategy method with the direct-response method is thus inconclusive about whether the willingness to sanction in strategy response differs from the willingness to sanction in direct response to others' choice behavior (for an overview, see Brandts & Charness, 2011). In addition, despite the similarities between the two response methods and the two timing moments we distinguish (beforehand versus afterwards), there also are important differences. Most importantly, people in the decision environment beforehand do not necessarily have to make full precompiled strategies for all feasible choices that could possibly occur, as is the case with the strategy method. Whereas the strategy method forces people to think through the implications of all possible outcomes, in reality they are often reluctant to do so (e.g., Shafir, 1994; Shafir et al., 1993; Shafir & Tversky, 1992; Tversky & Shafir, 1992), thereby giving rise to the disjunction effect. As such, the above makes apparent that we have to go beyond prior research that compared the strategy method with the direct-response method to examine whether decision timing has an impact on the willingness to reward cooperation and punish non-cooperation.

■ Experiment 4.1

To investigate the timing of both reward and punishment decisions in social dilemmas, we conducted a first experiment in which participants were third party observers of a one-shot public good task (see Fehr & Fischbacher, 2004; Molenmaker et al., 2014). As third party observers, participants themselves were not involved in the public good task but they had the opportunity to reward the group members or the opportunity to punish the group members (sanction type manipulation). These sanctioning decisions either had to be made before or after the group members had made their choices (decision timing manipulation). For exploratory purposes, we also measured the (anticipated) emotional reactions to the choice behavior. Based on our reasoning, we predicted that people would sanction others' choice behavior less often and to a lesser extent when they decided beforehand than afterwards. Furthermore, we explored whether the impact of decision timing would have a different effect on the reward of cooperation than on the punishment of non-cooperation.

Method

Participants and design

We recruited 159 students from a university in the Netherlands (97 women and 62 men; M_{age} = 21.44 years, SD_{age} = 3.74) to participate in an experiment on "group decision making".[1]

[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).

A 2 (Decision Timing: Beforehand versus Directly afterwards) × 2 (Sanction Type: Reward versus Punishment) between-participants factorial design was used.

Procedure

When participants arrived at the laboratory, they were seated in separate cubicles, each containing a personal computer to give instructions and register their responses. Assignment to one of the four conditions was randomly determined by a computer automated procedure. Participants were instructed that they had to perform a joint task with four fellow participants whose identities were unknown to them. The choices they would make in the joint task determined how much extra money they could earn on top of the standard initial participation fee. Participants learned that whether they would actually receive this extra money would be determined by a lottery after the study was conducted.

The participants were instructed that they were randomly assigned to a different role than the other four persons in the joint task (for a similar procedure, see Molenmaker et al., 2014). That is, their role was to observe the other four persons performing a one-shot public good task. Each person in the public good task would be endowed with €10 (which is approximately US $13) that they could either keep for themselves or contribute to a common pool. When contributed to the common pool, the €10 would be multiplied by 1.5 and divided equally among the four persons in the public good task (i.e., each would receive €3.75). Thus, the participants learned that the four persons had to make a dichotomous choice between being cooperative (i.e., contributing the €10 to the common pool) or not being cooperative (i.e., keeping the €10 for themselves). After participants read the instructions about the public good task, we posed four practice questions to ensure that they understood the task. We asked, for example, what would happen if a person would contribute his/her €10 to the common pool. The correct answer was disclosed after answering each question.

After this, participants read the instructions about their own role in the joint task. The instructions explained that they would be endowed with 100 points (worth €0.10 each) per person. In the reward conditions, participants could keep these points for themselves, but they could also assign points as increment points (we never used the word 'reward'). The value of the assigned increment points would be multiplied by 3 and added to the individual outcome of the person concerned. Thus, it would cost the participant €0.10 to increase a group member's outcome with €0.30. The instructions in the punishment conditions were identical, except that they could assign points as decrement points (we also never used the word 'punishment') and the value of the assigned decrement points would be multiplied by 3 and subtracted from the individual outcome of the person concerned. Thus, it would cost the participant €0.10 to decrease a group member's outcome with €0.30 (for a similar procedure, see Molenmaker et al., 2014).

In the instructions about participants' role in the joint task, we also introduced our manipulation of decision timing. In the beforehand conditions, participants learned that they had to decide about assigning points before the others actually decided about contributing their €10. That is, participants would have to compose a binding strategy for responding to the cooperative or non-cooperative choice that the persons could make in the public good

task. In addition, they were informed that they would have to make separate strategies for each person in the public good task.[2] In contrast, participants in the directly afterwards conditions learned that they had to decide about assigning points after their group members had decided about contributing their €10. Thus, they would have to respond to the cooperative or non-cooperative choice that the persons had actually made in the public good task.

Participants in all conditions also learned that the other four persons would be informed beforehand about the presence of a fifth person in the joint task who would have the opportunity to increase (reward conditions) or decrease (punishment conditions) their individual outcomes. Moreover, the instructions in the beforehand conditions also stated that group members would not be informed what this fifth person had decided before they themselves had decided about contributing their €10. Note that we thus merely manipulated when participants would make their sanction decisions, not when their group members would learn about the sanction decisions. In this way, we ruled out the possibility that participants in the beforehand conditions would opt for sanctioning to influence the four persons' choices in the public good task, while participants in the afterwards conditions would not have this opportunity because the choices in the public good task are already made. To ensure comprehension of their role in the joint task, we again posed four practice questions. For example, we asked participants when they would have to decide about assigning points. The correct answer was disclosed after answering each question.

Subsequently, the joint task started and participants were endowed with their first 100 points. In the beforehand conditions, we presented participants with the possibility that a person (named person M, see Footnote 2) would contribute his/her €10 to the common pool in the reward condition or would keep it for his/herself in the punishment condition.[3] We reminded participants that this was a possible choice that person M could make and that their decision would be executed if it would turn out that person M actually made this choice (i.e., their decision was binding). In the reward condition, we first asked whether the participants wanted to assign points as increment points and when they decided to assign increment points to person M, they had to indicate how many increment points they assigned. The procedure in the punishment condition was identical, except that they could assign points as decrement points.

In the directly afterwards conditions, we first asked participants to wait until we could confirm that all four persons had read their instructions and had made their choice in the public good task, which took about a minute. Next, they received the (bogus) feedback that a person

[2] We decided to ensure that the participants would make a sanction decision in response to an identifiable (but anonymized) person in both the beforehand and afterwards conditions (instead of one strategy in the beforehand conditions that would apply to all persons in the public good task) since research by Small and Loewenstein (2003, 2005) has shown that identifiability can influence people's sanctioning decisions.
[3] To keep the beforehand and afterwards conditions as identical as possible, we instructed participants that we would present each feasible choice one by one instead of presenting all feasible choices at once, as in the strategy method (Selten, 1967).

(named M) had contributed his/her €10 to the common pool in the reward condition or had kept it for him/herself in the punishment condition. In response to this actual choice that person M had made, we first asked participants whether they wanted to assign points as increment points in the reward condition or as decrement points in the punishment condition, and if they decided to assign points to person M, they had to indicate how many points they assigned. In all conditions, the maximum number of points participants could assign to person M was 100 points (and the minimum was zero points).

Next, we asked participants about their emotional reactions to the (*possible*) choice by person M. On a 9-point rating scale ranging from 1 (*not at all*) to 9 (*totally*), participants in the directly afterwards conditions indicated to what extent nine statements currently applied to them, whereas participants in the beforehand conditions indicated to what extent they anticipated that these statements would apply to them when later on it would turn out that person M has actually made this choice. To measure participants' positive emotions, we posed besides happiness ("*This choice by person M makes me feel happy*") also four additional positive emotional reactions (i.e., joy, pride, admiration, and elevation). To measure their negative emotions, we posed besides anger ("*This choice by person M makes me feel angry*") also three additional negative emotional reactions (i.e., fury, disappointment, and contempt).

At this point in the experiment, participants in all conditions had only learned about person M's (*possible*) choice in the public good task (see Footnote 2), and were asked about their sanction response and (anticipated) emotional reactions to particularly this cooperative choice behavior in the reward conditions or non-cooperative choice behavior in the punishment conditions (see Footnote 3). Next, participants were informed that the joint task was stopped. Before the participants were thoroughly debriefed and paid (1 course credit or €3 monetary compensation), we first checked our manipulations, the believability and the comprehension of the joint task. Finally, after the experiment was performed by all the participants, ten participants were randomly selected who received their actual earnings from the joint task.

Results

Manipulation checks

The manipulation of sanction type was checked by asking participants whether they could assign increment or decrement points. All participants (100%) answered this question correctly. We checked the manipulation of decision timing by asking participants whether they had to decide about assigning increment points (*decrement points*) before or after the other four persons made their choices in the public good task. All participants except three (98.1%) answered this question correctly.[4] Based on these results we can conclude that our manipulations were successful and we included the data of all 159 participants in the analyses.

[4] The data of these participants were included in the analyses because exclusion of the data did not alter the pattern of results.

Choice to sanction

We started by analyzing the effect of decision timing (Beforehand versus Directly afterwards) and sanction type (Reward versus Punishment) on the proportion of participants choosing to sanction (N = 159). In accordance with our prediction, a binary (Sanction Choice: 0 = not sanctioned, 1 = sanctioned) logistic regression yielded a significant Decision Timing main effect (B = 1.04, SE = 0.47, Wald (df=1) = 4.97, p = .026, Odds Ratio = 2.84, 95% CI [1.13, 7.12]). This main effect indicated that the proportion of participants choosing to sanction beforehand (75.6%) was significantly lower than the proportion of participants choosing to sanction directly afterwards (88.8%). Moreover, the analysis yielded a significant Sanction Type main effect (B = 2.19, SE = 0.58, Wald (df=1) = 14.37, p < .001, Odds Ratio = 8.91, CI [2.87, 27.58]), which showed that the proportion of participants choosing to punish (69.6%) was significantly lower than the proportion of participants choosing to reward (95%).

The impact of decision timing did not differ between reward and punishment, as indicated by the non-significant Decision Timing × Sanction Type interaction effect, B = 0.13, SE = 1.28, Wald (df=1) = 0.01, p = .921, Odds Ratio = 1.14, 95% CI [0.09, 14.07]. A closer inspection of the proportions for reward and punishment (also see Table 4.1 for the frequencies), revealed that only in the punishment condition the proportion of participants choosing to punish beforehand (59%) seems significantly lower than the proportion of participants choosing to punish directly afterwards (80%), $\chi^2(1) = 4.13$, p = .042, Odds Ratio = 2.78, CI [1.02, 7.59]. In the reward condition, the proportion of participants choosing to reward beforehand (92.5%) did not differ significantly from the proportion of participants choosing to reward directly afterwards (97.5%), $\chi^2(1) = 1.05$, p = .305, Odds Ratio = 3.16, CI [0.32, 31.78]. This suggests that the main effect of decision timing on the proportion of participants choosing to sanction is particularly driven by the choice to punish, and less so by the choice to reward. A possible explanation might be that the willingness to reward was so high (Molenmaker et al., 2014, 2016), that a potential difference in the choice to reward between the beforehand and directly afterwards conditions could not occur (i.e., a ceiling effect).

Sanction size

Furthermore, a 2 (Decision Timing: Beforehand versus Directly afterwards) × 2 (Sanction Type: Reward versus Punishment) ANOVA on the number of points (N = 159) yielded a marginal significant Decision Timing main effect ($F(1,155) = 3.06$, p = .082, η^2 = .01, 90% CI [.00, .06]). As predicted, the size of the sanctions administered beforehand (M = 32.72, SD = 31.98) was smaller than the size of the sanctions administered directly afterwards (M = 40.80, SD = 36.59). Furthermore, the analysis yielded a significant Sanction Type main effect ($F(1,155) = 52.90$, p < .001, η^2 = .25, CI [.16, .34]), which showed that the size of the punishments (M = 19.51, SD = 27.19) was significantly smaller than the size of the rewards (M = 53.85, SD = 32.52).

The impact of decision timing did again not differ between reward and punishment, as indicated by non-significant Decision Timing × Sanction Type interaction effect, $F(1,155)$ = 1.09, p = .298, η^2 = .01, 90% CI [.00, .04]. But a closer inspection of the simple effects

Table 4.1. *Number of participants choosing to sanction by Decision timing and Sanction type in Experiments 4.1 and 4.2.*

			Choice to sanction	
Experiment	Decision timing	Sanction type	Yes	No
1	Beforehand	Overall	60	19
		Punishment	23	16
		Reward	37	3
	Afterwards	Overall	71	9
		Punishment	32	8
		Reward	39	1
2	Beforehand	Overall	45	24
		Punishment	15	22
		Reward	30	2
	Afterwards	Overall	127	16
		Punishment	61	9
		Reward	66	7
	Directly afterwards	Overall	68	5
		Punishment	34	2
		Reward	34	3
	Delayed afterwards	Overall	59	11
		Punishment	27	7
		Reward	32	4

for reward and punishment (see also Table 4.2 for the mean number of points and standard deviation per condition) revealed that only the size of the rewards administered beforehand (M = 47.25, SD = 29.00) was significantly smaller than the size of the rewards administered directly afterwards (M = 60.45, SD = 33.82), $F(1,155) = 3.93$, $p = .049$, $\eta^2 = .03$, CI [.00, .08]. In contrast, the size of the punishments administered beforehand (M = 17.82, SD = 28.05) did not differ significantly for the size of the punishments administered directly afterwards (M = 21.15, SD = 26.59), $F(1,155) = 0.19$, $p = .67$, $\eta^2 < .01$, CI [.00, .03]. This seems to indicate that the main effect of decision timing on the number of points assigned to sanction is particularly driven by the size of the rewards, and less so by the size of the punishments. As prior research suggested (Molenmaker et al., 2014), a possible explanation might be that a potential difference in the punishment size between the beforehand and directly afterwards conditions could not occur because participants were very reluctant to punish (i.e., a floor effect).

Emotional reactions

We also analyzed the effects of decision timing (Beforehand versus Directly afterwards) on participants' emotional reactions. Since we gave different feedback in the reward conditions (i.e., a cooperative choice) and punishment conditions (i.e., a non-cooperative choice), we analyzed participants' positive emotions to the cooperative choice (i.e., the reward conditions;

Table 4.2. *Number of points assigned by Decision timing and Sanction type in Experiments 4.1 and 4.2.*

			Sanction size	
Experiment	Decision timing	Sanction type	*M*	*SD*
1	Beforehand	Overall	32.72	31.98
		Punishment	17.82	28.05
		Reward	47.25	29.00
	Afterwards	Overall	40.80	36.59
		Punishment	21.15	26.59
		Reward	60.45	34.82
2	Beforehand	Overall	22.69	31.44
		Punishment	11.43	24.35
		Reward	34.60	33.81
	Afterwards	Overall	42.30	37.04
		Punishment	24.44	27.03
		Reward	59.42	37.40
	Directly afterwards	Overall	44.95	36.11
		Punishment	26.89	28.75
		Reward	62.51	34.07
	Delayed afterwards	Overall	39.54	38.05
		Punishment	21.85	25.25
		Reward	56.25	40.77

$N = 79$) and negative emotions to the non-cooperative choice (i.e., the punishment conditions; $N = 78$) in separate analyses. First, a MANOVA on positive emotions about the cooperative choice showed no significant effect of decision timing on the positive emotional reactions, $V = 0.09$, $F(5,74) = 1.39$, $p = .24$, $\eta_p^2 = .09$. Second, a MANOVA on negative emotions about the non-cooperative choice also showed no significant effect of decision timing on the negative emotional reactions, $V = 0.04$, $F(4,74) = 0.73$, $p = .58$, $\eta_p^2 = .04$. See Table 4.3 for the overall means and standard deviations.

Discussion

The results of Experiment 4.1 provide first evidence for our reasoning that decision timing has an impact on the willingness to employ costly rewards for cooperative choice behavior and costly punishments for non-cooperative choice behavior. The willingness to sanction was lower when participants decided before the choice behavior than when they decided after the choice behavior. More specifically, participants both rewarded and punished less often and to a lesser extent in the decision environment beforehand than in the decision environment afterwards. However, a closer inspection of the results did suggest that the impact on the choice to sanction was particularly driven by punishment and the impact on sanction size was particularly driven by reward. Also note that we found no difference between the anticipated emotional reactions beforehand and the actual emotional reactions directly afterwards, which

Table 4.3. *Emotional reactions to the cooperative or non-cooperative choice in Experiments 4.1 and 4.2.*

		Experiment 4.1		Experiment 4.2	
Feedback	Emotions	*M*	*SD*	*M*	*SD*
Cooperative choice	Happiness	6.64	2.06	6.46	1.76
	Joy	6.50	1.79	6.44	1.77
	Pride	5.77	2.05	6.43	2.12
	Admiration	6.65	1.92	6.77	2.02
	Elevation	3.97	2.47	3.91	2.05
Non-cooperative choice	Anger	3.57	2.23	4.08	2.19
	Fury	2.96	2.20	3.28	1.99
	Disappointment	5.22	2.31	5.16	2.34
	Contempt	4.14	2.33	4.84	2.19

may suggest that a distinction between hot and cold decision environments may not necessarily explain our behavioral findings. To elaborate further on this tentative conclusion, we designed Experiment 4.2.

■ Experiment 4.2

Sanctioning either involves a decision *before* or a decision *after* others' choice behavior. Our first experiment demonstrated, however, that people are not as willing to sanction beforehand as they are willing to sanction afterwards. The aim of Experiment 4.2 was, first of all, to replicate the findings of Experiment 4.1. But although we argue that the willingness to sanction beforehand differs from the willingness to sanction afterwards, this is not the only temporal distinction that may alter the decision environment. In real-life, it is often the case that sanctioning decisions are *not* made directly after the choice behavior, but at a later moment in time. Importantly, deciding directly afterwards versus after a time delay has – besides deciding beforehand versus directly afterwards – also been related to the distinction between hot and cold decision environments (see e.g., Harinck & De Dreu, 2008; Wang et al., 2011). The fact that emotions tend to have a relatively short lifespan (Fridhandler & Averill, 1982) suggests that "hot" decision environments become less emotionally charged after some time has passed, and thus turn into colder decision environments (see Loewenstein, 1996; Loewenstein & Lerner, 2003; Loewenstein & Schkade, 1999). Put differently, the intensity of emotions may be lower after a time delay than directly afterwards (see also Gross, 1998; Ray, Wilhelm, & Gross, 2008). To further examine the impact of decision timing on the willingness to sanction, we therefore not only focused on the willingness to sanction beforehand and directly afterwards, but also on the willingness to sanction after a time delay (i.e., delayed afterwards).

Given that emotions have been identified as possible driver of the willingness to employ sanctions (e.g., De Kwaadsteniet et al., 2013; Fehr & Gächter, 2002; Seip et al., 2014), but also tend to have a relative short lifespan (Fridhandler & Averill, 1982), it may very well be that people are less willing to sanction after a time delay than directly afterwards. Preliminary support

for this reasoning can be found in research by Wang and colleagues (2011) who demonstrated that people react less punitive to others' norm transgressions after a short time delay than when they react directly afterwards. Their research also showed, however, that such an effect is only observed if people are distracted during the time delay, such that they cannot re-arouse the experienced anger by ruminating about the anger-arousing stimulus (see also Fridhandler & Averill, 1982; Gross, 1998; Harinck & De Dreu, 2008; Ray et al., 2008). In Experiment 4.2, we therefore manipulated the time delay by using the same distraction task as Wang and colleagues (2011) used in one of their experiments (i.e., the dots-estimation task; e.g., Gerard & Hoyt, 1974). If people experience intense emotions about others' choice behavior, we expect that they would particularly be willing to sanction in the decision environment directly afterwards, and less so in the decision environment after a time delay (and the decision environment beforehand). Testing this prediction was the second aim of Experiment 4.2.

It is important to note, however, that sanctioning does not necessarily have to be an emotional response due to 'heated tempers', it can also result from deliberate reasoning (see Schroeder, Steel, Woodell, & Bembenek, 2003). Indeed, this would explain why we found no difference in Experiment 4.1 between the anticipated emotional reactions beforehand and the actual emotional reactions directly afterwards. The fact that we did find that the willingness to sanction was lower beforehand than afterwards is consistent with the notion that people are less willing to employ costly sanctions if the choice behavior is not known yet (see Van Dijk et al., 2009), as is the case in the decision environment beforehand. Thus, the observed pattern of results in Experiment 4.1 fits the explanation in terms of nonconsequential reasoning (see also e.g., Shafir, 1994; Shafir & Tversky, 1992; Tversky & Shafir, 1992; Van Dijk & Zeelenberg, 2003), which presupposes more nonconsequential reasoning before rather than after others' choice behavior. A nonconsequential reasoning account does, however, not distinguish between whether decisions are made directly after the choice behavior or only after a time delay (as in both cases, people may be expected to engage in consequential reasoning). An interesting implication of this could be that adding a decision environment with a time delay provides us with a useful paradigm to further illuminate the reasons why people would be less willing to sanction choice behavior beforehand than afterwards.

In sum, in Experiment 4.2 we provided another test of the impact that decision timing may have on the willingness to sanction by focusing on the willingness to sanction beforehand, directly afterwards and after a time delay. Our main prediction is that people would sanction others' choice behavior less often and to a lesser extent when they decide beforehand than when they decide afterwards, thereby replicating the findings of Experiment 4.1. This pattern of results would, first and foremost, fit the explanation in terms of nonconsequential reasoning since this account only distinguishes between a decision environment beforehand and a decision environment afterwards, and not between a decision environment directly afterwards and a decision environment after a time delay. If the hot/cold distinction would have explanatory power for the sanctioning decisions in our research, we may also expect that people would sanction others' choice behavior less often and to a lesser extent when they decide in the decision environment after a time delay than in the decision environment directly afterwards. Moreover, we again explored whether the impact of decision timing would

differ between reward of cooperation and punishment of non-cooperation. In addition to our behavioral measures, we also again measured the (anticipated) emotional reactions to the choice behavior.

Method

Participants and design

215 students from a university in the Netherlands (148 women and 67 men; M_{age} = 20.47 years, SD_{age} = 4.09) were recruited to participate in an experiment on "group decision making". We used a 3 (Decision Timing: Beforehand versus Directly afterwards versus Delayed afterwards) × 2 (Sanction Type: Reward versus Punishment) between-participants factorial design.

Procedure

The procedure of Experiment 4.2 was almost identical to the procedure of Experiment 4.1. Thus, the instructions explained that the participants' role was to observe the four other persons performing a one-shot public good task. Whereas participants in Experiment 4.1 learned that the four persons in the public good task had to decide whether they would contribute their endowment of €10 to the common pool or keep it for themselves (i.e., dichotomous choice), participants in Experiment 4.2 learned that the persons had to decide *how many* euros from their endowment of €10 they would contribute to the common pool and *how many* euros they would thus keep for themselves (i.e., a continuous choice). Thus, instead of the dichotomous choice that we used in our first experiment, the persons in the public good task had to determine their degree of (non-)cooperativeness in our second experiment. The (bogus) feedback about the decision that a person (named person M) would possibly make (beforehand conditions) or had actually made (afterwards conditions) was, however, again the choice to contribute all his/her €10 to the common pool in the reward conditions or to keep it all for his/herself in the punishment conditions.

The delayed afterwards conditions were almost identical to the directly afterwards conditions. However, immediately after participants in the delayed afterwards conditions received the feedback (but before we asked them whether they wanted to assign points) they had to perform a dots-estimation task (Gerard & Hoyt, 1974; Sivanathan, Molden, Galinsky, & Ku, 2008; Wang et al., 2011). In this 5 minutes task – which has been shown to decrease the intensity of emotions because it interferes with emotional thoughts (see Wang et al., 2011) – participants had to make a series of estimations about the number of dots that were presented on their computer screen for 5 seconds. The feedback about person M's (*possible*) choice remained on the screen during the task. To rule out the possibility that performing the dots-estimation task itself (and not its distracting nature) would influence our results, participants in the beforehand and directly afterwards conditions also had to perform the dots-estimation task. However, they performed the task both before we presented the feedback and they could make their sanction decision. As such, participants in all conditions performed the dots-estimation task, but only in the delayed afterwards conditions it served as a distracting filler task between the feedback and the sanction decision.

As in Experiment 4.1, participants in all conditions were asked whether they wanted to assign points, and if they decided to assign points to person M, they had to indicate how many points they assigned. The maximum amount of points participants could assign to person M was again 100 points (and the minimum was zero points), and each assigned point would cost the participant €0.10 but increased the personal outcome of person M with €0.30 in the reward conditions or decreased the personal outcome of person M with €0.30 in the punishment conditions.

Results

Manipulation check

The manipulation of sanction type was checked by asking participants whether they could assign increment or decrement points. All participants except one (99.1%) answered this question correctly. The manipulation of decision timing was checked by asking participants whether they had to decide about assigning increment points (*decrement points*) beforehand, directly afterwards, or after a time delay. This question was answered correctly by all participants except fifteen (93%). These fifteen participants who gave an incorrect answer were mainly part of the delayed afterwards conditions (11 participants), which suggests that they did not recognized the filler task as a time delay (see Footnote 4). We conclude that our manipulation of decision timing was successful and included the data of all 215 participants in the analyses.

Choice to sanction

We first performed a binary (Sanction Choice: 0 = not sanctioned, 1 = sanctioned) logistic regression (N = 215) to analyze the effects of decision timing (Beforehand versus Directly afterwards versus Delayed afterwards) and sanction type (Reward versus Punishment) on the proportion of participants choosing to sanction. As predicted, the analysis yielded a significant Decision Timing main effect (Wald (df=2) = 19.70, $p < .001$). Planned contrasts revealed that the proportion of participants choosing to sanction beforehand (62.5%) was significantly lower than the proportion of participants choosing to sanction afterwards (88.8%), regardless of whether they decided directly afterwards or after a delay, B = 1.70, SE = 0.39, Wald (df=1) = 19.29, p = .001, Odds Ratio = 5.50, 95% CI [2.57, 11.76]. In addition, the proportion of participants choosing to sanction directly afterwards (93.2%) did not differ significantly (but marginally) from the proportion of participants choosing to sanction after a delay (84.3%), B = 0.97, SE = 0.58, Wald (df=1) = 2.84, p = .092, Odds Ratio = 2.64, CI [0.85, 8.19]. Moreover, the significant Sanction Type main effect (B = 1.29, SE = 0.40, Wald (df=1) = 10.57, p = .001, Odds Ratio = 3.62, CI [1.67, 7.85]) showed that the proportion of participants choosing to punish (71%) was significantly lower than the proportion of participants choosing to reward (88.9%). Also see Table 4.1 for the frequencies.

Furthermore, the Decision Timing × Sanction Type interaction effect was significant (Wald (df=2) = 6.12, p = .047); not because sanction type altered the difference in proportions of participants choosing to sanction directly afterwards and after delay (B = 1.14, SE = 1.16, Wald (df=1) = 0.95, p = .329, Odds Ratio = 3.11, 95% CI [0.32, 30.42]), but because

sanction type significantly altered the difference in proportions of participants choosing to sanction beforehand and afterwards (B = 2.01, SE = 0.83, Wald (df=1) = 5.93, p = .015, Odds Ratio = 7.48, CI [1.48, 37.84]). That is, the proportion of participants choosing to punish beforehand (40.5%) was significantly lower than the proportion of participants choosing to punish afterwards (87.1%), regardless of whether they decided directly afterwards (94.4%) or after a delay (79.4%), $\chi^2(1) = 25.55$, $p < .001$, Odds Ratio = 9.94, CI [3.81, 25.95]. In contrast, the proportion of participants choosing to reward beforehand (85.7%) did not differ significantly from the proportion of participants choosing to reward afterwards (90.4%), regardless of whether they decided directly afterwards (91.9%) or after a delay (88.9%), $\chi^2(1) = 0.53$, p = .467, Odds Ratio = 1.57, CI [0.46, 5.36]. As in Experiment 4.1, these findings might indicate a ceiling effect (see also Molenmaker et al., 2014, 2016).

Sanction size

Next, we analyzed the effect of decision timing and sanction type on the number of points with a 3 (Decision Timing: Beforehand versus Directly afterwards versus Delayed afterwards) × 2 (Sanction Type: Reward versus Punishment) ANOVA (N = 215), which yielded a significant Decision timing effect, $F(2,209) = 9.07$, $p < .001$, $\eta^2 = .12$, 90% CI [.06 .19]. Planned contrasts revealed that the size of the sanctions administered beforehand (M = 22.69, SD = 31.44) was significantly smaller than the size of the sanctions administered afterwards (M = 42.30, SD = 31.44), regardless of whether they decided directly afterwards or after a delay, $t(209) = 4.11$, $p < .001$, d = 0.59, 95% CI [0.30, 0.88]. In addition, the size of the sanctions administered directly afterwards (M = 44.95, SD = 36.11) did not differ significantly from the size of the sanctions administered after a delay (M = 39.54, SD = 38.05), $t(209) = 1.06$, p = .289, d = 0.18, CI [-0.15, 0.51]. Moreover, the analysis yielded a significant Sanction Type main effect ($F(1,209) = 51.46$, $p < .001$, $\eta^2 = .18$, 90% CI [.11, .26]), which showed that the size of the punishments (M = 19.94, SD = 28.81) was significantly smaller than the size of the rewards (M = 51.38, SD = 37.95).

Furthermore, the impact of decision timing did not differ between reward and punishment, as indicated by the non-significant Decision Timing × Sanction Type interaction effect, $F(2,209) = 0.84$, p = .432, $\eta^2 = .01$, 90% CI [.00, .04]. A closer inspection of the simple contrasts for reward and punishment (see also Table 4.2 for the mean number of points and standard deviation per condition) revealed that the size of the rewards administered beforehand (M = 34.60, SD = 33.81) was significantly smaller than the size of the rewards administered afterwards (M = 59.42, SD = 37.40), regardless of whether they decided directly afterwards (M = 62.51, SD = 34.07) or after a delay (M = 56.25, SD = 40.77), $t(209) = 3.90$, $p < .001$, d = 0.80, 95% CI [0.39, 1.21]. In addition, also the size of the punishments administered beforehand (M = 11.43, SD = 24.53) was significantly smaller than the size of the punishments administered afterwards (M = 24.44, SD = 27.03), regardless of whether they decided directly afterwards (M = 26.89, SD = 28.75) or after a delay (M = 21.85, SD = 25.25), $t(209) = 2.14$, p = .03, d = 0.44, CI [0.03, 0.84]. This demonstrates that the main effect of decision timing on the on the number of points assigned to sanction is driven by both the size

of the rewards and punishments, although the effect seems greater for reward (d = 0.80, CI [0.39, 1.21]) than for punishment (d = 0.44, CI [0.03, 0.84]).

Emotional reactions

The effects of decision timing (Beforehand versus Directly afterwards versus Delayed afterwards) on participants' positive emotions to the cooperative choice (i.e., the reward conditions; N = 108) and negative emotions to the non-cooperative choice (i.e., the punishment conditions; N = 107) were analyzed in separate analyses. First, a MANOVA on positive emotions about the cooperative choice showed no significant (but marginal) effect of decision timing on the positive emotional reactions, V = 0.16, F(10,204) = 1.74, p = .074, η_p^2 = .08. [5] Second, a MANOVA on negative emotions about the non-cooperative choice also showed no significant effect of decision timing on the negative emotional reactions, V = 0.11, F(8,204) = 1.46, p = .17, η_p^2 = .05. See Table 4.3 for the overall means and standard deviations.

Discussion

The results of Experiment 4.2 further corroborated our reasoning that decision timing (beforehand versus afterwards) has an impact on the willingness to employ costly rewards for cooperative choice behavior and costly punishments for non-cooperative choice behavior. That is, participants sanctioned less often and to a lesser extent before others' choice behavior than after others' choice behavior. Experiment 4.2 also showed that the willingness to sanction directly afterwards was not significantly different from the willingness to sanction after a distracting time delay.[6] In addition, the emotional reactions that participants reported were

[5] Although the MANOVA on positive emotions was marginally significant, separate one-way ANOVAs did reveal (marginal) significant effects of decision timing on pride (F(2,105) = 2.64, p = .08, η^2 = .09, 90% CI [.02, .18]) and admiration (F(2,105) = 6.34, p = .003, η^2 = .20, CI [.08, .29]). Planned contrasts showed that the anticipated pride (M = 5.00, SD = 2.03) and admiration (M = 5.89, SD = 2.18) in the beforehand condition was lower than the experienced pride (M = 6.73, SD = 2.12) and admiration (M = 7.19, SD = 1.80) in the afterwards conditions (Pride: t(105) = 2.15, p = .03, d = 0.44, 95% CI [0.33, 0.85]; Admiration: t(105) = 3.29, p = .001, d = 0.68, CI [0.26, 1.09]), whereas the emotions did not differ between the directly afterwards condition and the delayed afterwards conditions (Pride: t(105) = 0.80, p = .43, d = 0.19, CI [-0.27, 0.65]; Admiration: t(105) = 1.33, p = .19, d = 0.31, CI [-0.15, 0.77]).

[6] We ran an additional experiment (Experiment 4.3) with a 2 (Decision Timing: Directly afterwards versus Delayed afterwards) × 2 (Sanction Type: Reward versus Punishment) between-participants factorial design in which we used the game Tetris (e.g., Holmes, James, Coode-Bate, & Deeprose, 2009; Van Dillen, Van der Wal, & Van den Bos, 2012) – instead of the dots-estimation task – as a 3 minutes filler task. The binary (Sanction Choice: 0 = not sanctioned, 1 sanctioned) logistic regression on the proportion of participants choosing to sanction (N = 112) yielded only a significant Sanction Type main effect (B = 2.48, SE = 0.78, Wald (df=1) = 10.14, p = .001, Odds Ratio = 11.92, 95% CI [2.59, 54.79]), which showed that the proportion of participants choosing to punish (69.6%) was significantly lower than the proportion of participants choosing to reward (96.4%). The Decision Timing main effect (B = 0.44, SE = 0.55, Wald (df=1) = 0.65, p = .420, Odds Ratio = 1.55, 95% CI [0.53, 4.51]) and the Decision Timing × Sanction

not different between conditions (as in Experiment 4.1). In the general discussion of this chapter, we further reflect on these findings.

■ General discussion

The decision to reward cooperative choice behavior and to punish non-cooperative choice behavior can be made at various moments in time. However, it either involves a decision *before* or a decision *after* the choice behavior. In the present chapter, we argued and showed that people are less willing to employ costly sanctions when they decide beforehand than when they decide afterwards. In the decision environment beforehand others' actual choice behavior still has to take place in the future, whereas in the decision environment afterwards the choice behavior did actually take place in the past. Research on the disjunction effect revealed that the presence of uncertainty about outcomes may induce nonconsequential reasoning (e.g., Shafir & Tversky, 1992; Tversky & Shafir, 1992; Van Dijk & Zeelenberg, 2003). In accordance with this work, we demonstrated that people are less willing to sanction choice behavior that *may possibly occur in the future* than choice behavior that *did actually occur in the past*. More specifically, people rewarded cooperation and punished non-cooperation less often and to a lesser extent when sanctioning decisions were made before (instead of after) the occurrence of others' choice behavior (Experiments 4.1 and 4.2), regardless of whether they decided directly afterwards or after a time delay (Experiment 4.2). By doing so, we thus revealed that people are less willing to employ costly sanctions if the choice behavior is not known yet.

At this point, it is important to stress that we do not claim that the experience of emotions directly afterwards may never be related to the impact that decision timing can have on the willingness to employ costly sanctions. After all, emotions have been identified as a driving force of sanctioning decisions (e.g., De Kwaadsteniet et al., 2013; Nelissen & Zeelenberg, 2009; Seip et al., 2014), and the emotions that people experience may differ in intensity and quality between the decision environment beforehand and the decision environment afterwards (e.g., Loewenstein, 1996; Loewenstein & Lerner, 2003; Loewenstein & Schkade, 1999; Miceli & Castelfranchi, 2015). The present research revealed, however, that the experience of emotions directly afterwards does not seem to be a necessary precondition of differences in the willingness to employ costly sanctions before versus after others' choice behavior. It would therefore be a good idea for future research to investigate whether the experience of (intense)

Type interaction effect (B = 0.51, SE = 1.56, Wald (df=1) = 0.11, p = .743, Odds Ratio = 1.67, 95% CI [0.79, 35.16]) both were non-significant. The ANOVA on the number of points (N = 112) yielded only a significant Sanction Type main effect (F(1,108) = 57.70, p < .001, η^2 = .35, 90% CI [.23, .45]), which showed that the size of the punishments (M = 15.54, SD = 24.73) was significantly smaller than the size of the rewards (M = 60.61, SD = 36.62). The Decision Timing main effect (F(1,108) = 0.67, p = .415, η^2 < .01, CI [.00, .05]) and the Decision Timing × Sanction Type interaction effect (F(1,108) = 0.27, p = .606, η^2 < .01, CI [.00, .04]) both were non-significant. Thus, also with a different distracting task during the time delay we found no difference in the willingness to costly reward cooperation and punish non-cooperation between the directly afterwards conditions and the delayed afterwards conditions.

emotions would amplify the impact that decision timing (beforehand versus afterwards) has on sanctioning decisions, for example, by experimentally manipulating such emotions (see Seip et al., 2014).

Another point worth mentioning is that, whereas the present research demonstrated that both the willingness to reward cooperation and the willingness to punish non-cooperation are sensitive to the timing of sanctioning decisions, the impact of decision timing is not necessarily identical for both types of sanctions. Prior research has shown that people are generally reluctant to punish non-cooperation and prefer to reward cooperation (Chapters 2 and 3; Molenmaker et al., 2014, 2016; see also Molm, 1997; Sutter et al., 2010; Wang et al., 2009). Consistent with these earlier findings, we also showed that participants punished non-cooperation less often and to a lesser extent than they rewarded cooperation. However, our results – those of Experiment 4.2 in particular – suggest that this relative preference for the use of rewards over punishments may be more pronounced when sanctioning decisions are made beforehand than when they are made afterwards. After all, participants were particularly reluctant to opt for punishing before (as compared to after) the choice behavior, whereas they were very willing to opt for rewarding, both beforehand and afterwards. An interesting direction for future research would therefore be to examine whether people particularly are more reluctant to punish (as compared to reward) others' actual choice behavior they are not certain about yet (see Van Dijk et al., 2009).

Although one should always be cautious when generalizing experimental results to practice, we do want to address an interesting practical implication that may derive from our work. When implementing sanction opportunities in real-life social dilemmas, policymakers should realize that whether people consider sanctioning the appropriate course of action can differ across time (see also March, 1994; Messick, 1999). Whereas people may employ sanctions after the choice behavior has occurred, they may not be that willing to employ them before the choice behavior has occurred. When the opportunity to sanction beforehand is not used sufficiently, implementing the opportunity to sanction afterwards can thus be decisive to promote cooperative choice behavior.

In addition to this practical implication, our work may also contribute to the methodological debate about the behavioral validity of the strategy method and the direct-response method (e.g., Brandts & Charness, 2011; Brosig et al., 2003; Fischbacher et al., 2012; Selten, 1967). Given that there are similarities between the two response methods and the two timing moments that we distinguished in our experiments (beforehand versus afterwards), the insight that people are less willing to sanction choice behavior beforehand (which has not taken place yet, as compared to choice behavior afterwards) may thus also apply to the strategy method. As such, it may very well be that the mixed results of the studies that compared the strategy method with the direct-response method (see Brandts & Charness, 2011) could be explained by the fact that it is not known yet whether particular choice behavior will occur in the future. Future research should investigate whether the impact of uncertainty about the choice behavior is attenuated by the fact that the strategy method may force people to think through the implications of all possible outcomes (see e.g., Shafir, 1994; Shafir et al., 1993; Shafir & Tversky, 1992; Tversky & Shafir, 1992).

Before closing, we also want to discuss two aspects of the experimental paradigm we used. First of all, we used a third party paradigm in our experiments. As third party observers, participants themselves were not involved in the public good dilemma they observed (Fehr & Fischbacher, 2004; Molenmaker et al., 2014). Whereas this procedure eliminated the possibility that participants' interpretation of the choice behavior was colored by self-interest, the willingness to sanction may be higher when they take part in the social dilemma themselves (i.e., a second party paradigm) because in such situations revenge-like motives might drive the infliction of harm (e.g., see De Quervain et al., 2004). As such, the impact of decision timing might be different when people's personal outcomes are affected by others' choice behavior. However, this is an empirical question and should be addressed in future research (see Appendix A).

Another point worth discussing is that we focused on the willingness to sanction in a one-shot interaction. One-shot interactions constitute an appropriate setting to test the impact of decision timing, as this setting eliminates confounds that might arise in repeated interactions (Gächter & Herrmann, 2009). For example, strategic considerations about future interactions do not play a role in one-shot interactions, only whether people consider it the appropriate course of action to sanction others' choice behavior. It would be interesting to examine to what extent decision timing effects persist as people gain experience in repeated interactions or have gained experience in comparable interactions with other people. Does deciding beforehand (as compared to afterwards) have less of an impact on sanctioning decisions when people are more experienced with the dilemma at hand? In addition, it would be interesting to examine whether the effect of decision timing on the willingness to employ sanctions has consequences for the enforcement of cooperative choice behavior. Does the fact that people are less willing to sanction beforehand than afterwards lead to lower levels of cooperation in repeated interactions? A fruitful avenue for future research may thus be to investigate the impact of decision timing on the willingness to sanction in repeated interactions.

Conclusions

The present work substantiates that decision timing (beforehand versus afterwards) has an impact on the willingness to costly sanction. In our research, we demonstrated that people are less willing to sanction beforehand than afterwards, regardless of whether they decide directly afterwards or after a time delay. These findings imply that people are less willing to sanction choice behavior that *may possibly occur in the future* than choice behavior that *did actually occur in the past*. As such, our findings shed new light on the willingness to costly reward cooperative choice behavior and punish non-cooperative choice behavior. At the same time, we provide a better understanding of the use of sanction opportunities to promote cooperative choice behavior.

Chapter 5

General discussion

■ General discussion

Social dilemmas pose a key challenge to groups, organizations, and societies. When collective interests collide with personal interests, it is not self-evident that people opt for the collective interests and mutually cooperate (Olson, 1965; Samuelson, 1954). If too many people do not cooperate and rather opt for their personal interests this can have detrimental consequences for the collective welfare because public goods cannot be provided and common resources become depleted (Hardin, 1968). The use of sanctions has often been suggested as a solution to social dilemmas (e.g., Hardin, 1968; Hobbes, 1651/1991; Olson, 1965; Ostrom, 1990). Whereas positive sanctions (i.e., rewards) for cooperation and negative sanctions (i.e., punishments) for non-cooperation can indeed be effective means to enhance cooperation (for an overview, see Balliet, Mulder, & Van Lange, 2011), a prerequisite for any effect of sanctions is, first and foremost, that people are actually willing to administer them. This important – yet long neglected – question is the central theme of the present dissertation.

As argued in the introductory chapter of this dissertation, it is of critical importance to study the willingness to sanction in social dilemmas. To shed more light on this topic, the present dissertation was aimed at identifying determinants of the (un)willingness to reward cooperation and punish non-cooperation. Hence, three empirical chapters reported the results of a series of experiments that revolved around the question of how willing people are to reward cooperative choice behavior and to punish non-cooperative choice behavior. The objective of this final chapter is to summarize and discuss the main findings of this dissertation and, more importantly, discuss their general implications and elaborate on future research directions.

■ Main findings

The central assumption tested in the present dissertation was that the willingness to reward cooperation differs markedly from the willingness to punish non-cooperation. The use of punishments – in contrast to the use of rewards – implies that one directly inflicts harm on another person. Research on the *do-no-harm principle* demonstrated that, even if the overall benefit outweighs the harm done, people are actually reluctant to inflict harm on others (e.g., Baron, 1993; Baron, 1995; Baron & Jurney, 1993; Baron & Ritov, 1994; Ritov & Baron, 1990; Spranca, Minsk, & Baron, 1991; see also Van Beest, Van Dijk, De Dreu, & Wilke, 2005). Applied to the use of sanctions, this principle thus suggests that people are not as willing to administer punishments as they are willing to administer rewards. An important aim of this dissertation was to examine whether people indeed are less willing to punish non-cooperative choice behavior than to reward cooperative choice behavior.

In all experiments reported in the empirical chapters (except Experiment 3.3), both the willingness to reward cooperation and the willingness to punish non-cooperation were assessed, while at the same time various factors were experimentally manipulated (see Chapters 2-4) or varied across experiments (e.g., the costs of sanctioning, presented feedback, etc.). The results of these experiments consistently showed that people are less willing to punish

non-cooperation than they are willing to reward cooperation. In fact, when people have both sanction means available, they tend to completely refrain from punishing – thereby leaving non-cooperation unpunished – and rather opt for rewarding (Experiment 2.2). To provide further support for the robustness of this general preference for the use of rewards over punishments, I conducted two meta-analyses that not only included the data reported in the empirical chapters of this dissertation, but also the data from experiments not included in these chapters (see Appendix A). The combined results on Choice to sanction (k = 13, n = 2073) and the combined results on Sanction size (k = 13, n = 2056) both revealed a significant overall effect, which demonstrated that people punish non-cooperation less often (Z = 6.24, p < .001, Odds ratio = 3.39, 95% CI [2.22, 5.19]) and to a lesser extent (Z = 8.79, p < .001, d = 0.94, 95% CI [0.71, 1.18]) than they reward cooperation (see Appendix B).

So my findings provide strong evidence that the type of sanction people have at their disposal – either reward or punishment – is a primary determinant of the (un)willingness to sanction. In addition to sanction type, I also tested whether the willingness to sanction is influenced by *what* kinds of (non-)cooperative choice behavior people face (Chapter 2), *how* they can sanction (Chapter 3), and *when* they can sanction (Chapter 4). In doing so, I investigated whether these situational factors are also determinants of the (un)willingness to use rewards and punishments in social dilemmas.

In Chapter 2, I reported results on how the preference for rewarding cooperation over punishing non-cooperation is moderated by whether people face a public good dilemma or a common resource dilemma. Although both social dilemmas refer to the same conflict of interests (i.e., self-interest versus collective interest), and can be structured as each other's equivalents in terms of payoffs, they differ in the way in which the initial property is distributed (Camerer, 2003; Dawes, 1980; Van Dijk & Wilke, 1997, 2000). Whereas the property in public good dilemmas is initially possessed by the people themselves (i.e., private property), the property in common resource dilemmas is initially located in a common resource (i.e., collective property). I hypothesized that people consider choice behavior about giving up private property in public good dilemmas as less objectionable (and thus less punishable) and more commendable (and thus more rewardable) than choice behavior about infringing on collective property in common resource dilemmas. As such, I predicted that people would be less willing to punish non-cooperation and more willing to reward cooperation in the public good dilemma than in the common resource dilemma.

To test this, two experiments were conducted in which participants observed the choice behavior of two persons in a one-shot social dilemma task. This social dilemma context was either presented as a public good dilemma or a common resource dilemma. The feedback participants received indicated that one person displayed a relatively high level of cooperation and the other person displayed a relatively low level of cooperation. When participants had to decide about rewarding or had to decide about punishing (i.e., Experiment 2.1), they punished less often and to a lesser extent than they rewarded in the public good dilemma than in the common resource dilemma. In addition, when they had the opportunity to choose between rewarding and punishing (i.e., Experiment 2.2), the large majority of participants in both social dilemmas chose to reward, but they rewarded to a greater extent in the public good

dilemma than in the common resource dilemma. These findings corroborate the notion that people's willingness to reward cooperative choice behavior and to punish non-cooperative choice behavior is moderated by the type of social dilemma they face (public good dilemmas versus common resource dilemmas).

After having identified social dilemma type as a determinant of the willingness to sanction, I turned my attention to a situational factor that may in fact teach us more about *why* people seem reluctant to use punishments (as compared to the use of rewards). That is, in Chapter 3, I examined the impact of personal responsibility for sanctions on the willingness to administer them. Although the preference for rewarding cooperation over punishing non-cooperation seems to be rooted in the do-no-harm principle (see Baron, 1993, 1995; Baron & Jurney, 1993; Spranca et al., 1991), one may question why people tend to adhere to the do-no-harm principle when making sanctioning decisions. Is this because they generally feel that no harm should be done, even when it is directed at someone who has impaired the collective interests, or is this perhaps because they are the ones *doing* the harm? Prior research on the do-no-harm principle has, for instance, shown that the reluctance to harm is stronger when people are directly (as opposed to indirectly) responsible for the anticipated harm (e.g., Royzman & Baron, 2002) and when people's actions (as opposed to their inactions) have harmful outcomes (e.g., Ritov & Baron, 1990; Ritov & Baron, 1992; Spranca et al., 1991). Therefore, I hypothesized that people's reluctance to punish non-cooperation is a self-restraining tendency that originates from their feeling of personal responsibility for the harm done. As such, I expected that people are reluctant to punish non-cooperation to the extent that the feel personally responsible for the harm done.

Given that people feel less responsible for their actions and often act more aggressively as members of a group than as individual decision makers (Jaffe, Shapir, & Yinon, 1981; Jaffe & Yinon, 1979; Mathes & Kahn, 1975; Meier & Hinsz, 2004), the grouping of individuals was used in three experiments to attenuate the self-restraining impact of the feeling of personal responsibility for the harm done. That is, participants took part in a one-shot common resource task with either an individual sanction opportunity or a joint sanction opportunity implemented. They observed the harvest decision of a group member and subsequently voted about whether or not to sanction (Experiment 3.1) or determined the size of a sanction (Experiments 3.2 and 3.3), either individually or jointly. The results showed that non-cooperation was punished less often and to a lesser extent when people decided as individual decision makers than when they decided as groups, while no such differences were found for the reward of cooperation (Experiments 3.1 and 3.2). Moreover, the attenuating effect of sharing responsibility on the willingness to punish was mediated by felt personal responsibility, even when people could not be held accountable for their actions (Experiment 3.3). Thus, feelings of personal responsibility for the sanctions have a self-restraining impact on the willingness to punish non-cooperative choice behavior, but not on the willingness to reward cooperative choice behavior.

As Chapters 2 and 3 illustrate, situational factors can affect whether people are willing to administer sanctions. Given the practical relevance of such insights, I focused next on a situational factor that particularly has practical relevance for the implementation of sanction

opportunities. In Chapter 4, I studied how the timing of sanction decisions influences the willingness to reward cooperation and punish non-cooperation. Although the decision to sanction others' choice behavior can be made at various moments in time, it either involves a decision before or a decision after the choice behavior. One of the most apparent differences between these two moments in time is that afterwards people decide about the sanctioning of choice behavior that has actually taken place in the past, whereas beforehand people decide about the sanctioning of choice behavior that may or may not take place in the future. Research on the disjunction effect has shown that if the outcome of a particular situation is unknown, people are often reluctant to think through the implications of all possible outcomes (e.g., Tversky & Shafir, 1992) and are less likely to make decisions based on uncertain information than on certain information (Van Dijk & Zeelenberg, 2006). I argued that a similar effect may be observed for sanctioning decisions. That is, I hypothesized that people are less willing to sanction choice behavior that may possibly occur in the future than choice behavior that did actually occur in the past. I therefore tested the prediction that people are less willing to sanction choice behavior when sanctioning decisions are made before (as opposed to after) the occurrence of others' choice behavior.

In two experiments, participants observed another person's choice behavior and had the opportunity to administer a reward or the opportunity to administer a punishment. The timing of the sanction decision was manipulated by presenting participants with choice behavior that could possible occur in the future (i.e., beforehand) or with choice behavior that did actually occur in the past (i.e., afterwards). In line with the prediction, participants rewarded cooperation and punished non-cooperation less often and to a lesser extent when the sanctioning decision was made before (instead of after) the occurrence of others' choice behavior (Experiments 4.1 and 4.2), regardless of whether they decided directly afterwards or after a time delay (Experiments 4.2 and 4.3). Thus, people are less willing to employ sanctions if the choice behavior has not occurred yet. Furthermore, the results suggested that the preference for the use of rewards over punishments may be more pronounced when sanctioning decisions are made beforehand than when they are made afterwards. Participants were particularly reluctant to opt for punishing before (as compared to after) the choice behavior, whereas they were very willing to opt for rewarding, both beforehand and afterwards. My findings thus showed that the timing of sanction decisions affects the willingness to reward cooperative choice behavior and the willingness to punish non-cooperative choice behavior.

In sum, the present dissertation not only identified type of sanction (Reward versus Punishment; see Chapters 2-4) as a primary determinant of the (un)willingness to sanction, but also revealed that *what*, *how*, and *when* people can reward or punish has an influence on their willingness to administer them. More specifically, I demonstrated that the type of social dilemma that people face (Public good dilemma versus Common resource dilemma; see Chapter 2), the extent of personal responsibility that people have for the sanction (Individual responsibility versus Joint responsibility; see Chapter 3), and the timing of the sanctioning decision (Beforehand versus Afterwards; see Chapter 4) also play an important role for the use of sanctions. In doing so, this dissertation provides useful insights on the determinants of the willingness to sanction in social dilemmas.

■ General implications and directions for future research

In the introductory chapter of this dissertation, I presented several reasons why studying the (un)willingness to sanction in social dilemmas is of critical importance. In the remainder of this final chapter, I further reflect on these reasons outlined in Chapter 1 when discussing the general implications of this work and elaborate on future research directions. I evaluate how the present dissertation (1) contributes to a better theoretical understanding of the psychological processes involved in the use of sanctions in social dilemmas, (2) provides fruitful insights about the evolutionary functions of sanctioning, and (3) has practical implications to 'solve' social dilemmas in the real world. In doing so, I put this work in a broader perspective and highlight avenues for future research.

Psychological processes underlying the willingness to sanction

One of the most striking results of this dissertation is that, although social dilemmas may particularly call for punishment of non-cooperation (see Chapter 1), people are actually rather reluctant to punish non-cooperative choice behavior and prefer to reward cooperative choice behavior (Chapters 2-4), even if they could administer the sanctions without any financial cost to themselves (Chapter 3). These findings are in line with the do-no-harm principle (e.g., Baron, 1993, 1995; Baron & Jurney, 1993), which states that people are reluctant to inflict harm on someone to help others. The use of rewards and punishments are both beneficial in the sense that they can enhance cooperation (Balliet et al., 2011), but only punishment – in contrast with reward – implies that one directly inflicts harm to another person. The use of punishments for non-cooperation (but not the use of rewards for cooperation) thus comes with a 'psychological cost'. That is, people seem to be concerned about the moral 'wrongness' of doing harm (see Baron, 1993, 1995, 2012; Baron & Ritov, 2009). As a consequence, people generally consider punishing non-cooperation the less appropriate course of action (see also March, 1994; Messick, 1999), and therefore use punishments less often and to a lesser extent than they use rewards (see Molm, 1997; Sutter, Haigner, & Kocher, 2010; Wang, Galinsky, & Murnighan, 2009). These findings emphasize that the reluctance to harm tends to hamper the willingness to punish non-cooperation, and can even foster the willingness to reward cooperation.

The fact that people are reluctant to harm, however, does not necessarily imply that non-cooperation will never be punished. There are circumstances in which moral concerns about the infliction of harm may be outweighed by strategic considerations (e.g., Yamagishi, 1986; but see Gächter & Herrmann, 2009) or moral sentiments about the norm violation (e.g., Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007; Fehr & Gächter, 2002). For instance, the anger that people may experience about others' unfair choice behavior and the anticipated guilt for the otherwise forgone opportunity to restore justice both have been identified as driving forces of the willingness to punish unfair choice behavior (e.g., Nelissen & Zeelenberg, 2009; Seip, Van Dijk, & Rotteveel, 2014; Wang et al., 2009). Given the willingness to restore justice, on the one hand, and the reluctance to inflict harm, on the other hand, it may very well be that people experience a 'motivational conflict' when they are confronted with others' non-cooperative

choice behavior. Although this is yet an empirical question that should be addressed in future research – for instance, by assessing cardiovascular or neurological indicators of motivational conflict (see Blascovich, 2000; Blascovich & tomaka, 1996; Greene, Nystrom, Engell, Darley, & Cohen, 2004) – the present dissertation does offer an interesting new perspective on how to approach the willingness to sanction in social dilemmas.

As explained in Chapter 1, this present dissertation focuses not only on the type of sanction that people have at their disposal, but also on how situational factors – respectively the type of social dilemma that people face, the extent of personal responsibility they feel for the sanction, and the timing of sanctioning decisions – may affect the willingness to reward cooperation and the willingness to punish non-cooperation. Below I discuss the general implications that can be derived from my research on these determinants.

First, the findings in Chapter 2 identified the type of social dilemma that people face as a moderator of sanctioning behavior. The key difference between public good dilemmas versus common resource dilemmas – and inherent to what defines both social dilemmas – is the way in which the initial property is distributed (e.g., Camerer, 2003; Dawes, 1980; Van Dijk & Wilke, 1995, 1997): either as private property in the public good dilemma or as collective property in the common resource dilemma. People decide about giving up private property in public good dilemmas and about infringing on collective property in common resources. The fact that the preference for rewarding over punishing is more pronounced in the public good dilemma than in the common resource dilemma indicates that these social dilemma types induce distinct moral standards that people use to evaluate others' choice behavior (see also Janoff-Bulman & Carnes, 2013; Janoff-Bulman, Sheikh, & Hepp, 2009). Public good dilemmas seem to induce a prescriptive morality that prescribes *what to do* (i.e., giving up private property) and common resource dilemmas seem to induce a proscriptive morality that proscribes *what not to do* (i.e., not infringing on collective property). It would therefore be a good idea for future research to test the robustness of these induced moral standards, for example, by experimentally manipulating the framing of choice behavior (i.e., give-some versus keep-some and take-some versus leave-some) in both social dilemma types (see also Van Dijk & Wilke, 2000).

Second, Chapter 3 demonstrated that people are not merely concerned about the moral 'wrongness' of inflicting harm, they also are concerned about their own part in it. Specifically, the reluctance to punish non-cooperation (as compared to the willingness to reward cooperation) is particularly strong when people feel personally responsible for the sanction decision. When people feel personally responsible for the anticipated harm, they are more concerned about the punishment they administer (see e.g., Baron & Ritov, 2009; Cushman, Young, & Hauser, 2006; Milgram, 1974; Ritov & Baron, 1992; Spranca et al., 1991). A sense of personal responsibility for the harm done thus is an important reason why people adhere to the do-no-harm principle when they decide about sanctioning. The self-restraining impact that the feeling of personal responsibility has on the willingness to punish, but not on the willingness to reward, may even entail that people rather avoid personal responsibility for punishing non-cooperation. After all, recent studies have shown that people frequently 'outsource' decisions that may harm others, even if this implies that they give up their own

decision rights (Andreoni & Gee, 2012; Bartling & Fischbacher, 2012). In experiments that I conducted recently (and which are not included in the empirical chapters of this dissertation), participants had to choose whether they performed a one-shot public good task in a group in which peer-to-peer punishment was allowed or in a group in which only a specific type of centralized punishment was allowed (e.g., the group could punish by majority vote, a third party could punish, etcetera) and in which they would thus be less personally responsible for the harm done. The results of these experiments showed that there are indeed people willing to delegate their punishment power to a centralized authority (varying from 12% to 35% of the participants), as long as the decision making procedure of this centralized authority is considered fair enough. Although this is still work in progress, these findings provide preliminary support for the reasoning that people may want to avoid being solely responsible for the punishment of non-cooperation. However, more research is still needed to determine whether it is indeed the feeling of personal responsibility that drives the willingness to delegate punishment power to a centralized punishment authority.

Third and finally, the results of Chapter 4 showed that people are less willing to employ sanctions before (as compared to after) others' choice behavior. Whereas beforehand people decide about sanctioning choice behavior that may possibly occur in the future, afterwards they decide about sanctioning choice behavior that did actually occur in the past. The fact that people are not as willing to sanction beforehand as they are willing to sanction afterwards, regardless of whether they decide directly afterwards or after a time delay, indicates that people are reluctant to sanction if the choice behavior is not known yet. Afterwards, however, it does not necessarily have to be the case that people know for certain that others' choice behavior did actually take place in the past. In fact, certainty in social dilemmas is more likely to be the exception than the rule (Van Dijk, Wit, Wilke, & Budescu, 2004; see also De Kwaadsteniet, Van Dijk, Wit, & De Cremer, 2006, 2008, 2010; De Kwaadsteniet, Van Dijk, Wit, De Cremer, & De Rooij, 2007; Van Lange, Ouwerkerk, & Tazelaar, 2002). Given that people are often reluctant to think through the implications of all possible outcomes (e.g., Tversky & Shafir, 1992) and are less likely to make decisions based on uncertain information than on certain information (Van Dijk & Zeelenberg, 2003), it may very well be that the willingness to sanction afterwards is also hampered if people are not certain about others' actual choice behavior. The present findings may thus not only apply to the decision environment beforehand, it may, in fact, also apply to uncertain decision environments in general. An interesting direction for future research would therefore be to examine the impact of uncertainty on the willingness to sanction (Van Dijk, De Kwaadsteniet, & Mulder, 2009). In particular, it would be interesting to see whether people require more certainty to employ punishments than to employ rewards. Unjustly sanctioning would violate the motive to restore justice (e.g., Carlsmith, 2006; Carlsmith, Darley, & Robinson, 2002), which may – due to the do-no-harm principle (see Baron, 1993, 1995) – be considered worse in case of punishment (i.e., unjustly harming others) than in case of reward (i.e., unjustly favoring others).

Evolutionary functions of sanctioning in social dilemmas

Although identifying determinants of the (un)willingness to sanction in social dilemmas provides, in particular, a more in-depth view of the psychological processes underlying the willingness to sanction in social dilemmas, it can also provide fruitful insights for the analysis of the evolutionary functions of sanctioning (Barclay & Kiyonari, 2014; Tinbergen, 1968). Over the last few decades, these evolutionary functions are topic of debate (e.g., Brown & Richerson, 2014; Fehr & Henrich, 2003; Hagen & Hammerstein, 2006; Krasnow, Cosmides, Pedersen, & Tooby, 2012; Krasnow, Delton, Cosmides, & Tooby, 2015; West, El Mouden, & Gardner, 2011; West, Griffin, & Gardner, 2007). From a norm enforcement perspective (e.g., Boyd & Richerson, 1992; Gintis, 2003; Henrich & Henrich, 2007; see also Wilson, 1975), it is assumed that the emergence and maintenance of cooperation norms within large-scaled groups is the ultimate cause of why the willingness to administer costly sanctions has evolved and persisted (Boyd, Gintis, Bowles, & Richerson, 2003; Fehr, Fischbacher, & Gächter, 2002; Fehr & Gächter, 2002; Fehr & Henrich, 2003; Gintis, 2000; Gintis, Bowles, Boyd, & Fehr, 2003; Henrich et al., 2010; Henrich et al., 2006). However, this line of reasoning – which hinges on the assumption that evolution can also select at group level (Wilson, 1975) – is in conflict with the basic principles of evolutionary theory and has therefore received a lot of criticism. Evolutionary psychologists advocated, on the contrary, that the willingness to administer costly sanctions has evolved and persisted to serve social exchange within small-scaled groups (e.g., Krasnow, Delton, Tooby, & Cosmides, 2013; Tooby & Cosmides, 1992; see also Nowak & Sigmund, 1998; Trivers, 1971). Although this may produce behavior that appears irrational when placed in evolutionarily atypical situations such as large-scaled groups (e.g., punishment of strangers in one-shot interactions), this was adaptive in the ancestral social environment (i.e., small-scaled groups) that people evolved in (Krasnow, Delton, Cosmides, & Tooby, 2016; see also Delton, Krasnow, Cosmides, & Tooby, 2011; Kenrick et al., 2009; Todd & Gigerenzer, 2007).

How do my findings fit into the above perspectives and what new insights may derive from it for the analysis of the evolutionary functions of sanctioning in social dilemmas? First of all, the key finding of this dissertation – that people are reluctant to punish non-cooperation and rather reward cooperation – seems at odds with the idea that the willingness to sanction has evolved to enforce cooperation norms within large-scaled groups (i.e., the norm enforcement perspective). Social dilemmas may particularly call for punishment of non-cooperation – and not necessarily for reward of cooperation – because it is the non-cooperative choice behavior that actually jeopardizes the collective welfare (see Chapter 1). However, the results of this dissertation showed that people are less willing to punish non-cooperative choice behavior than to reward cooperative choice behavior (Chapters 2-4), even if they could administer the sanctions without any financial cost to themselves (Chapter 3) and without the possibility that rewards could also serve as indirect punishment for those not rewarded (Chapter 3 and 4). In addition, a theoretical problem with the norm enforcement perspective is that, although group selection (if it exists) would select *between* large-scaled groups that have sufficient group members who are willing to incur the costs to reward cooperation and punish non-cooperation

(i.e., strong reciprocity), natural selection would actually select *within* large-scaled groups against those 'altruistic' group members (West et al., 2007; see Hamilton, 1964).

The findings do align, by contrast, with the notion that the willingness to sanction has evolved to serve social exchange within small-scaled groups (i.e., the social exchange perspective). In small-scaled groups, people had to balance between, on the one hand, administering punishments to deter personally relevant mistreatment and, on the other hand, not punishing too much to maintaining a positive reputation within the group (see Krasnow et al., 2012). Various studies have demonstrated that people who punish non-cooperation – in contrast to those who reward cooperation – not necessarily gain a positive reputation (e.g., Kiyonari & Barclay, 2008; see Barclay & Kiyonari, 2014). In fact, punishers frequently get blamed by others for administering punishments (e.g., Atwater, Waldman, Carey, & Cartier, 2001; Eriksson, Andersson, & Strimling, 2015; Strimling & Eriksson, 2014; Trevino, 1992), and can even get punished in return (Cinyabuguma, Page, & Putterman, 2006; Denant-Boemont, Masclet, & Noussair, 2007; Herrmann, Thöni, & Gächter, 2008; Nikiforakis, 2008). The fact that people are reluctant to punish non-cooperation and willing to reward cooperation, as revealed in this dissertation, thus fits with the social exchange perspective on the evolutionary functions of sanctioning in social dilemmas. Put differently, from the social exchange perspective, it can be argued that not only the willingness to punish non-cooperation has evolved and persisted (Krasnow et al., 2016), but also the reluctance to punish non-cooperation. So the present work suggests that a fruitful new avenue for future research would be to investigate whether the reluctance to punish non-cooperative choice behavior indeed has evolved to maintain a positive reputation and whether retaliation for receiving punishments was, in fact, the force that selected for its design.

Practical implications to 'solve' social dilemmas in real-life

Caution is advised when generalizing experimental findings to practice. However, this dissertation does provide useful insights that may contribute to solving social dilemmas in real-life. Groups, organizations, and societies face many challenges arising from the fact that the collective interests do not coincide with the personal interests of the people belonging to that collective. Although alternatives have been suggested (e.g., Balliet, 2010; Chen, Dang, & Keng-Highberger, 2014; Chen, Pillutla, & Yao, 2009; Wu, Balliet, & Van Lange, 2016), numerous experiments demonstrated that sanctions can be an effective solution to ensure and protect the collective welfare in social dilemmas (for an overview, see Balliet et al., 2011). However, a prerequisite for any effect of sanctions is that those in control of sanctions also consider it the appropriate course of action to administer them. When considering the implementation of sanction opportunities in real world social dilemmas, it thus is important that they understand the conditions under which people will actually use rewards for cooperation and punishments for non-cooperation. I therefore want to discuss some practical implications that derive from the present work.

First and foremost, the possibility of rewarding should not be overlooked in real-life social dilemmas. Even though punishment and reward can both be effective means to

enhance cooperation (Balliet et al., 2011), people usually are not as willing to punish non-cooperation as they are willing reward cooperation (Chapters 2-4). Not only those who can be sanctioned (e.g., Eriksson et al., 2015; Kiyonari & Barclay, 2008; Strimling & Eriksson, 2014), but also those in control of sanction generally consider punishing non-cooperation the less appropriate course of action than rewarding cooperation. To enhance cooperative choice behavior, the implementation of reward opportunities can thus be decisive, especially if punishment opportunities are not used sufficiently. In addition, it is important to realize that the reluctance to use punishment opportunities can also originate from various situational factors, such as the type of social dilemma that people face (Chapter 2), the extent of personal responsibility that they experience for the sanctions (Chapter 3), and at what moment in time they make their sanctioning decisions (Chapter 4). The present dissertation thus provides useful insights about the situational determinants of the willingness to sanction. These insights can help the implementation of effective sanction opportunities and may thereby contribute to the solving real world social dilemmas.

■ Concluding thoughts

Political philosopher Thomas Hobbes (1651/1991) was the first to designate sanctions as solution to solve social dilemmas. Since Toshio Yamagishi's (1986) pioneering work on the effectiveness of punishment systems, numerous experiments were conducted that consistently showed that both reward of cooperation and punishment of non-cooperation can effectively enhance cooperative choice behavior in social dilemmas (for overviews, see Balliet et al., 2011; Van Dijk, Molenmaker, & De Kwaadsteniet, 2015; Van Lange, Rockenbach, & Yamagishi, 2014). In the present dissertation, I broadened the focus to the important – yet long neglected – question of how willing people actually are to sanction in social dilemmas. This is of critical importance, if only because people should first be willing to administer rewards and punishments before they can serve as an effective solution to ensure and protect the collective welfare. The determinants and boundary conditions of the willingness to reward cooperation and punish non-cooperation that I identified in this dissertation reveal that there are not only psychological processes at play that foster sanctioning, but also psychological processes that hamper sanctioning. By taking a closer look at people's (un)willingness to incur the costs of rewarding cooperative choice behavior and punishing non-cooperative choice behavior, this work thus provides a more comprehensive view of the potential that sanctions can have to solve social dilemmas in the real world. I therefore want to end this 'Jerry Springer's final thought moment' by stating that I hope that the present dissertation will inspire fellow scholars to further explore this fascinating topic of the (un)willingness to reward cooperation and punish non-cooperation.

Appendices

■ Appendix A

Supplemental experiment 1

Aim and Design

The aim of this pilot experiment was to investigate whether the anticipation of future interactions would affect the willingness to sanction in a public good dilemma. As third party, participants observed the choice behavior of two persons in a public good task, which either was the only round (one-shot conditions) or was the first of five rounds (multiple-shots conditions), and subsequently had the opportunity to administer increment coins (reward conditions) or decrement coins (punishment conditions). This experiment had a 2 (Sanction type: Punish versus Reward) × 2 (Interactions: One-shot versus Multiple-shots) design with Choice to sanction and Sanction size as dependent variables.

Results

The 2 (Sanction type) × 2 (Interactions) binary logistic regression on Sanction choice (N = 122) yielded only a significant Sanction type main effect (B = 1.39, SE = 0.51, Wald (df=1) = 7.33, p = .007, Odds Ratio = 3.40, 95% CI [1.47, 10.91]), which indicated that the proportion of participants choosing to punish (69.4%) was smaller than the proportion of participants choosing to reward (90%). The Interactions main effect (B = 0.32, SE = 0.47, Wald (df=1) = 0.48, p = .487, Odds Ratio = 1.38, CI [0.55, 3.45]) and the Sanction type × Interactions interaction effect (B = -0.62, SE = 1.06, Wald (df=1) = 0.34, p = .562, Odds Ratio = 1.85, CI [0.23, 14.84]) both were non-significant. The 2 (Sanction type) × 2 (Interactions) ANOVA on Sanction size (N = 122) yielded a significant Interactions main effect (F(1,118) = 3.94, p = .05, η^2 = .03, 90% CI [.00, .09]) and a significant Sanction type main effect (F(1,118) = 13.55, p < .001, η^2 = .10, CI [.03, .19]), which indicated that the size of the punishments (M = 16.45, SD = 24.34) was significantly smaller than the size of the rewards (M = 33.02, SD = 25.82). The Sanction type × Interactions interaction effect (F(1,118) = 0.25, p = .621, η^2 < .01, CI [.00, .04]) was non-significant.

Supplemental experiment 2

Aim and Design

The aim of this pilot experiment was to investigate the willingness to sanction in a common resource dilemma. As third party, participants observed the choice behavior of two persons in a one-shot common resource task, and subsequently had the opportunity to administer increment coins (reward conditions) or decrement coins (punishment conditions). This experiment had a One-factor (Sanction type: Punish versus Reward) design with Choice to sanction and Sanction size as dependent variables.

Results

The Chi-squared test on Choice to sanction (N = 83) showed that the proportion of participants choosing to punish (66.7%) was smaller than the proportion of participants choosing to reward

(85.4%), $\chi^2(1) = 3.97$, $p = .046$, Odds Ratio = 2.92, 95% CI [0.99, 8.57]. The One-way ANOVA on Sanction size ($N = 83$) showed that the size of the punishments ($M = 15.93$, $SD = 17.57$) was significantly smaller than the size of the rewards ($M = 26.29$, $SD = 21.79$), $t(81) = 2.39$, $p = .019$, $\eta^2 = .07$, 90% CI [.01, .17].

Supplemental experiment 3

Aim and Design

The aim of this experiment was to investigate whether advisors' lack of personal responsibility would affect their willingness to advice for sanctioning, as compared to administrators willingness to administer the sanctions. Participants observed the choice behavior of a group member in a common resource task, and subsequently had the opportunity to administer (administrator conditions) or advice an administrator about the administration of (advisor conditions) increment points (reward conditions) or decrement points (punishment conditions). This experiment had a 2 (Sanction type: Punish versus Reward) × 2 (Responsibility: Administrator versus Advisor) design with Choice to sanction and Sanction size as dependent variables.

Results

The 2 (Sanction type) × 2 (Responsibility) binary logistic regression on Sanction choice ($N = 157$) yielded a non-significant Sanction type main effect ($B = 0.17$, $SE = 0.55$, Wald (df=1) = 0.92, $p = .76$, Odds Ratio = 0.85, 95% CI [0.29, 2.49]), which indicated that the proportion of participants choosing to punish (91.1%) did not differ from the proportion of participants choosing to reward (89.7%). The Responsibility main effect ($B = 0.32$, $SE = 0.47$, Wald (df=1) = 0.48, $p = .487$, Odds Ratio = 1.38, CI [0.55, 3.45]) and the Sanction type × Responsibility interaction effect ($B = 0.28$, $SE = 1.11$, Wald (df=1) = 0.06, $p = .80$, Odds Ratio = 1.32, CI [0.15, 11.63]) were also non-significant. The 2 (Sanction type) × 2 (Responsibility) ANOVA on Sanction size ($N = 157$) yielded a non-significant Sanction type main effect ($F(1,153) = 1.08$, $p = .299$, $\eta^2 = .01$, CI [.00, .04]), which indicated that the size of the punishments ($M = 40.24$, $SD = 26.79$) did not differ from the size of the rewards ($M = 45.21$, $SD = 32.30$). The Responsibility main effect ($F(1,153) = 2.54$, $p = .113$, $\eta^2 = .02$, CI [.00, .06]) and Sanction type × Responsibility interaction effect ($F(1,153) = 0.01$, $p = .93$, $\eta^2 < .01$, CI [.00, .00]) were also non-significant.

Supplemental experiment 4

Aim and Design

The aim of this experiment was to investigate whether the timing of sanctioning decisions would affect the willingness to sanction when people themselves are involved in the public good dilemma. Participants observed the choice behavior of a group member in a public good task, and had the opportunity to administer increment points (reward conditions) or decrement points (punishment conditions), either before they received feedback (beforehand

conditions), directly after they received feedback (directly afterwards conditions), or after a time delay (delayed afterwards conditions). This experiment had a 2 (Sanction type: Punish versus Reward) × 3 (Decision timing: Beforehand versus Directly afterwards versus Delayed afterwards) design with Choice to sanction and Sanction size as dependent variables.

Results

The 2 (Sanction type) × 3 (Decision timing) binary logistic regression on Choice to sanction (N = 197) only yielded a significant Sanction type main effect (B = 1.74, SE = 0.44, Wald (df=1) = 15.84, p < .001, Odds Ratio = 5.68, 95% CI [2.41, 13.36]), while controlling for Donations (B = 0.08, SE = 0.06, Wald (df=1) = 1.57, p = .210, Odds Ratio = 1.08, CI [0.96, 1.22]), which indicated that the proportion of participants choosing to punish (68.4%) was smaller than the proportion of participants choosing to reward (91.9%). The Decision timing main effect (Wald (df=2) = 2.50, p = .287) and the Sanction type × Decision timing interaction effect (Wald (df=2) = 2.11, p = .348) were non-significant. The 2 (Sanction type) × 3 (Decision timing) ANOVA on Sanction size (N = 197) yielded a significant Decision timing main effect (F(1,190) = 3.96, p = .021, η^2 = .06, 90% CI [.01, .11]) and a significant Sanction type main effect (F(1,190) = 76.87, p < .001, η^2 = .28, CI [.19, .36]), while controlling for Donations (F(1,190) = 7.36, p = .007, η^2 = .03, CI [.00, .07]), which indicated that the size of the punishments (M = 22.04, SD = 27.67) was significantly smaller than the size of the rewards (M = 60.85, SD = 36.89). The Sanction type × Decision timing effect (F(1,190) = 0.10, p = .907, η^2 < .01, CI [.00, .01) was non-significant.

Supplemental experiment 5

Aim and Design

The aim of this pilot experiment was to investigate whether involvement in a public good dilemma would affect the willingness to sanction. Participants observed the choice behavior of a group member in a public good task, while they themselves either were involved in the task (second party conditions) or were not involved in the task (third party conditions), and subsequently had the opportunity to administer increment coins (reward conditions) or decrement coins (punishment conditions). This experiment had a 2 (Sanction type: Punish versus Reward) × 2 (Party type: Second party versus Third party) design with Choice to sanction and Sanction size as dependent variables.

Results

The 2 (Sanction type) × 2 (Party type) binary logistic regression on Choice to sanction (N = 156) yielded only a significant Sanction type main effect (B = 2.16, SE = 0.46, Wald (df=1) = 22.43, p < .001, Odds Ratio = 8.70, 95% CI [3.55, 21.29]), which indicated that the proportion of participants choosing to punish (53.9%) was smaller than the proportion of participants choosing to reward (91%). The Party type main effect (B = 0.78, SE = 0.39, Wald (df=1) = 0.04, p = .844, Odds Ratio = 1.08, CI [0.50, 2.34]) and the Sanction type

× Party type interaction effect (B = 0.32, SE = 1.92, Wald (df=1) = 0.12, p = .731, Odds Ratio = 1.37, CI [0.23, 8.32]) both were non-significant. The 2 (Sanction type) × 2 (Party type) ANOVA on Sanction size (N = 156) yielded only a significant Sanction type main effect (F(1,152) = 108.33, p < .001, η^2 = .41, 90% CI [.32, .49]), which indicated that the size of the punishments (M = 8.04, SD = 11.30) was significantly smaller than the size of the rewards (M = 49.28, SD = 33.11). The Party type main effect (F(1,152) = 0.42, p = .521, η^2 < .01, CI [.00, .03]) and the Sanction type × Party type effect (F(1,152) = 1.49, p = .224, η^2 = .01, CI [.00, .04) both were non-significant.

Supplemental experiment 6

Aim and Design

The aim of this pilot experiment was to investigate whether outcome dependence in a public good dilemma would affect the willingness to sanction. As third party, participants observed the choice behavior of group members in a public good task, while they either were dependent on the outcome of the task (third party dependence conditions) or were not dependent on the outcome of the task (third party independence conditions), and subsequently had the opportunity to administer increment coins (reward conditions) or decrement coins (punishment conditions). This experiment had a 2 (Party type: Third party dependence versus Third party independence) × 2 (Feedback: High cooperator versus Low cooperator) mixed design with repeated measures on the latter factor and Choice to reward or punish and Reward or punish size as dependent variables.

Results

The 2 (Party type) × 2 (Feedback) mixed binary logistic regression on Choice to reward or punish (N = 71) yielded only a significant Feedback main effect (B = -3.95, SE = 0.95, Wald (df=1) = 17.24, p < .001, Odds ratio = 7.46, 95% CI [3.36, 16.54]), which indicated that the proportion of participants choosing to punish the low cooperator (40%) was smaller than the proportion of participants choosing to reward the high cooperator (80%). The Party type main effect was non-significant (B = 0.71, SE = 0.50, Wald (df=1) = 2.00, p = .157, Odds ratio = 3.28, CI [0.62, 17.44]). The 2 (Party type) × 2 (Feedback) mixed ANOVA on Reward (positive value) or Punish (negative value) size (N = 75) yielded a marginal significant Party type main effect (F(1,53) = 3.60, p = .062, η^2 = .05, 90% CI [.00, .15]) and a significant Feedback main effect (F(1,73) = 43.61, p < .001, η^2 = .60, CI [.47, .68]), which indicated that the size of the punishments for the low cooperator (M = 16.71, SD = 28.06) was significantly smaller than the size of the rewards for the high cooperator (M = 42.68, SD = 36.49). The Party type × Feedback interaction effect (F(1,53) = 0.05, p = .821, η^2 < .01, CI [.00, .03]) was non-significant.

Supplemental experiment 7

Aim and Design

The aim of this experiment was to investigate whether involvement and outcome dependence in a public good dilemma would affect the willingness to sanction. Participants observed

the choice behavior of a group member in a public good task, while they themselves either were involved in the task (second party conditions), were dependent on the outcome of the task (third party dependence conditions), or were not dependent on the outcome of the task (third party independence conditions), and subsequently had the opportunity to administer increment coins (reward conditions) or decrement coins (punishment conditions). This experiment had a 2 (Sanction type: Punish versus Reward) × 3 (Party type: Second party versus Third party dependence versus Third party independence) design with Choice to sanction and Sanction size as dependent variables.

Results

The 2 (Sanction type) × 3 (Party type) binary logistic regression on Choice to sanction (N = 284) only yielded a significant Sanction type main effect (B = 1.48, SE = 0.31, Wald (df=1) = 22.42, p < .001, Odds Ratio = 4.40, 95% CI [2.38, 8.12]), which indicated that the proportion of participants choosing to punish (62.5%) was smaller than the proportion of participants choosing to reward (97.9%). The Party type main effect (Wald (df=2) = 3.36, p = .186) and the Sanction type × Party type interaction effect (Wald (df=2) = 0.52, p = .771) were non-significant. The 2 (Sanction type) × 3 (Party type) ANOVA on Sanction size (N = 284) yielded a significant Party type main effect (F(1,278) = 3.44, p = .033, η^2 = .04, 90% CI [.01, .08]) and a significant Sanction type main effect (F(1,278) = 47.14, p < .001, η^2 = .14, CI [.08, .20]), which indicated that the size of the punishments (M = 29.99, SD = 33.51) was significantly smaller than the size of the rewards (M = 56.79, SD = 32.45). The Sanction type × Party type effect (F(1,278) = 0.90, p = .407, η^2 = .01, CI [.00, .03) was non-significant.

APP

Supplemental experiment 8

Aim and Design

The aim of this pilot experiment was to investigate whether the experience of shame would affect the willingness to sanction in a public good dilemma. After a shame inducing task, participants observed the choice behavior of a group member in a public good task and subsequently had the opportunity to administer increment coins (reward conditions) or decrement coins (punishment conditions). This experiment had a 2 (Sanction type: Punish versus Reward) × 2 (Shame versus Control) design with Choice to sanction and Sanction size as dependent variables.

Results

The 2 (Sanction type) × 2 (Emotion condition) binary logistic regression on Choice to sanction (N = 148) yielded only a marginal significant Sanction type main effect (B = 0.86, SE = 0.45, Wald (df=1) = 3.60, p = .058, Odds Ratio = 2.35, 95% CI [0.97, 5.69]), which indicated that the proportion of participants choosing to punish (75.7%) was smaller than the proportion of participants choosing to reward (87.8%). The Emotion condition main effect (B = 0.66, SE = 0.44, Wald (df=1) = 2.23, p = .135, Odds Ratio = 1.94, CI [0.81, 4.63]) and the Sanction type × Emotion condition interaction effect (B = 1.12, SE = 1.00,

Wald (df=1) = 1.24, p = .266, Odds Ratio = 3.04, CI [0.43, 21.62]) both were non-significant. The 2 (Sanction type) × 2 (Emotion condition) ANOVA on Sanction size (N = 148) yielded only a significant Sanction type main effect (F(1,144) = 36.40, p < .001, η^2 = .20, 90% CI [.11, .29]), which indicated that the size of the punishments (M = 1.68, SD = 1.73) was significantly smaller than the size of the rewards (M = 4.11, SD = 2.98). The Emotion condition main effect (F(1,144) = 0.22, p = .64, η^2 < .01, CI [.00, .03]) and the Sanction type × Emotion condition effect (F(1,144) < 0.01, p = .947, η^2 < .01, CI [.00, .00) both were non-significant.

Supplemental experiment 9

Aim and Design

The aim of this pilot experiment was to investigate whether the experience of guilt would affect the willingness to sanction in a public good dilemma. After a guilt inducing task, participants observed the choice behavior of a group member in a public good task and subsequently had the opportunity to administer increment coins (reward conditions) or decrement coins (punishment conditions). This experiment had a 2 (Sanction type: Punish versus Reward) × 2 (Guilt versus Control) design with Choice to sanction and Sanction size as dependent variables.

Results

The 2 (Sanction type) × 2 (Emotion condition) binary logistic regression on Choice to sanction (N = 147) yielded only a significant Sanction type main effect (B = 1.74, SE = 0.49, Wald (df=1) = 12.51, p < .001, Odds Ratio = 5.70, 95% CI [2.17, 14.94]), which indicated that the proportion of participants choosing to punish (66.2%) was smaller than the proportion of participants choosing to reward (91.8%). The Emotion condition main effect (B = -0.01, SE = 0.43, Wald (df=1) = 0.00, p = .987, Odds Ratio = .993, CI [0.43, 2.29]) and the Sanction type × Emotion condition interaction effect (B = 0.87, SE = 1.03, Wald (df=1) = 0.71, p = .399, Odds Ratio = 2.38, CI [0.32, 17.75]) both were non-significant. The 2 (Sanction type) × 2 (Emotion condition) ANOVA on Sanction size (N = 147) yielded a significant Emotion condition main effect (F(1,143) = 5.37, p = .022, η^2 = .03, 90% CI [.00, .09]), a significant Sanction type × Emotion condition effect (F(1,143) = 4.78, p = .03, η^2 = .03, CI [.00, .08), and a significant Sanction type main effect (F(1,143) = 30.96, p < .001, η^2 = .09, CI [.08, .26]), which indicated that the size of the punishments (M = 2.00, SD = 2.26) was significantly smaller than the size of the rewards (M = 4.26, SD = 2.85).

■ Appendix B

Two separate meta-analyses were conducted to estimate the combined overall effect of Sanction type (Reward versus Punishment) on Choice to sanction and on Sanction size. The data were taken from experiments reported in the empirical chapters of this present dissertation (Chapters 2-4) and from experiments – conducted by the author – not included in the empirical chapters of this present dissertation (Appendix A). The inclusion criteria's were that (1) sanction type was manipulated between participants – thereby excluding Experiment 2.2, Experiment 3.3, and Supplemental Experiment 6 – and that (2) Choice to sanction (k = 13, n = 2073) and/or Sanction size (k = 13, n = 2056) was measured.

Meta-analytic procedures

For the meta-analysis on Choice to sanction, the Odds ratio statistic was used as measure of effect size. The Odds ratios were calculated in the *Meta-Essentials* workbook using the frequencies of participants choosing to sanction along with the cell sizes (Van Rhee, Suurmond, & Hak, 2015). For the meta-analysis on Sanction size, the Cohen's *d* statistic was used as measure of effect size. The Cohen's *d*'s were calculated in the *Meta-Essentials* workbook using the *F* score or *t* value along with the cell sizes (Van Rhee et al., 2015).

Since sanction type was in most of the studies not the only factor that was manipulated between participants, it was assumed that the effect of Sanction type on Choice to sanction and Sanction size will have systematic variation (i.e., heterogeneity). To estimate the average effect sizes for Choice to sanction and Sanction size, random-effects models were therefore used since these models assume that effect sizes are sampled from a population of varying effect sizes (e.g., Hedges & Vevea, 1998). Finally, the meta-analyses were conducted using the inverse variance weighting method, both in the *Meta Essentials* workbooks (Van Rhee et al., 2015).

Results

Choice to sanction

The effect size estimations and their 95% confidence intervals of the experiments used in the meta-analysis on Choice to sanction (k = 13) are reported in Table 6.1. In accordance with my prediction, Sanction type had a medium-sized overall effect (see also Chen, Cohen, & Chen, 2010) on Choice to sanction (Z = 6.24, p < .001, Odds ratio = 3.39, 95% CI [2.22, 5.19]), which indicated that non-cooperation was punished less often than cooperation was rewarded. As expected, there is heterogeneity in the distribution of effect sizes (T^2 = .27, T = .52, I^2 = 58.46%). Thus, the combined overall effect of Sanction type on Choice to sanction should not be treated as the 'true' effect size (see Hak, Van Rhee, & Suurmond, 2016).

Sanction size

The effect size estimations and their 95% confidence intervals of the experiments used in the meta-analysis on Sanction size (k = 13) are reported in Table 6.2. In accordance with my

Table 6.1. *Odds ratios and their 95% Confidence Intervals per Experiment*

	Odds Ratio	95% Confidence Intervals	
Experiment 2.1	4.98	1.83	13.59
Experiment 3.1	1.12	0.57	2.21
Experiment 4.1	2.31	1.07	4.96
Experiment 4.2	3.67	1.73	7.80
Experiment 4.3 – Footnote	11.77	2.53	54.84
Supplemental Experiment 1	3.98	1.45	10.94
Supplemental Experiment 2	2.92	0.98	8.71
Supplemental Experiment 3	0.85	0.29	2.49
Supplemental Experiment 4	5.26	2.26	12.24
Supplemental Experiment 5	8.69	3.53	21.43
Supplemental Experiment 7	4.34	2.35	8.00
Supplemental Experiment 8	2.32	0.96	5.62
Supplemental Experiment 9	5.70	2.15	15.06

prediction, Sanction type had a large-sized overall effect (see also Cohen, 1988) on Sanction size ($Z = 8.79$, $p < .001$, $d = 0.94$, 95% CI [0.71, 1.18]), which indicated that non-cooperation was punished to a lesser extent than cooperation was rewarded. As expected, there is heterogeneity in the distribution of effect sizes ($T^2 = .11$, $T = .33$, $I^2 = 79.35\%$). Thus, the combined overall effect of Sanction type on Sanction size should not be treated as the 'true' effect size (see Hak et al., 2016).

Table 6.2. *Cohen's d's and their 95% Confidence Intervals per Experiment*

	Cohen's *d*	95% Confidence Intervals	
Experiment 2.1	1.00	0.62	1.39
Experiment 3.2	0.73	0.41	1.06
Experiment 4.1	1.15	0.82	1.49
Experiment 4.2	0.98	0.69	1.26
Experiment 4.3 – Footnote	1.44	1.02	1.86
Supplemental Experiment 1	0.67	0.30	1.03
Supplemental Experiment 2	0.52	0.08	0.97
Supplemental Experiment 3	0.17	-0.15	0.48
Supplemental Experiment 4	1.25	0.94	1.56
Supplemental Experiment 5	1.67	1.30	2.03
Supplemental Experiment 7	0.81	0.57	1.06
Supplemental Experiment 8	0.99	0.65	1.34
Supplemental Experiment 9	0.92	0.58	1.26


サロンパス
もんじゃ
MUSEE Channel
MUSÉE
PLATINUM
2014.03.24
洋服の青山
洋服の青山
書盛堂書

References

■ References

Abbink, K., Bolton, G. E., Sadrieh, A., & Tang, F. F. (2001). Adaptive learning versus punishment in ultimatum bargaining. *Games and Economic Behavior, 37*(1), 1-25. doi: 10.1006/game.2000.0837

Abbink, K., Irlenbusch, B., & Renner, E. (2000). The moonlighting game - An experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization, 42*(2), 265-277. doi: 10.1016/s0167-2681(00)00089-5

Andreoni, J., & Gee, L. K. (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics, 96*(11-12), 1036-1046. doi: 10.1016/j.jpubeco.2012.08.003

Atwater, L. E., Waldman, D. A., Carey, J. A., & Cartier, P. (2001). Recipient and observer reactions to discipline: Are managers experiencing wishful thinking? *Journal of Organizational Behavior, 22*(3), 249-270. doi: 10.1002/job.67

Baldwin, D. A. (1971). The power of positive sanctions. *World Politics, 24*, 19-38. doi: 10.2307/2009705

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution, 54*(1), 39-57. doi: 10.1177/0022002709352443

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin, 137*(4), 594-615. doi: 10.1037/a0023489

Bandura, A., Underwood, B., & Fromson, M. E. (1975). Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality, 9*(4), 253-269. doi: 10.1016/0092-6566(75)90001-X

Barclay, P., & Kiyonari, T. (2014). Why sanction? Functional causes of punishment and reward. In P. A. M. Van Lange, B. Rockenbach & T. Yamagishi (Eds.), *Reward and punishment in social dilemmas.* (pp. 182-196). New York, NY US: Oxford University Press. doi: 10.1093/acprof:oso/9780199300730.003.0010

Barclay, P., & Van Vugt, M. (2015). The evolutionary pscyhology of human pro-sociality: Adaptations, byproducts, and mistakes. In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford Handbook of Prosocial Behavior* (pp. 37-60). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780195399813.013.029

Baron, J. (1993). Heuristics and biases in equity judgments: A utilitarian approach. In B. A. Mellers & J. Baron (Eds.), *Psychological perspectives on justice: Theory and applications.* (pp. 109-137). New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511552069.007

Baron, J. (1995). Blind justice - Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making, 8*(2), 71-83. doi: 10.1002/bdm.3960080202

Baron, J. (2012). Where do nonutilitarian moral rules come from? In J. I. Krueger (Ed.), *Social judgment and decision making*. (pp. 261-277). New York, NY: Psychology Press.

Baron, J., & Jurney, J. (1993). Norms against voting for coerced reform. *Journal of Personality and Social Psychology, 64*(3), 347-355. doi: 10.1037/0022-3514.64.3.347

Baron, J., & Ritov, I. (1994). Reference points and omission bias. *Organizational Behavior and Human Decision Processes, 59*(3), 475-498. doi: 10.1006/obhd.1994.1070

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74-85. doi: 10.1016/j.obhdp.2004.03.003

Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. In D. M. Bartels, B. C. W., L. J. Skitka & D. L. Medin (Eds.), *Moral judgment and decision making* (pp. 133-167). San Diego, CA: Elsevier Academic Press. doi: 10.1016/S0079-7421(08)00404-0

Bartling, B., & Fischbacher, U. (2012). Shifting the blame: On delegation and responsibility. *Review of Economic Studies, 79*(1), 67-87. doi: 10.1093/restud/rdr023

Bazerman, M. H., White, S. B., & Loewenstein, G. F. (1995). Perceptions of fairness in interpersonal and individual choice situations. *Current Directions in Psychological Science, 4*(2), 39-43. doi: 10.1111/1467-8721.ep10770996

Blascovich, J. (2000). Using physiological indexes of psychological processes in social psychological research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods and personality psychology* (pp. 117-137). Cambridge, UK: Cambridge University Press.

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 1-51). San Diego, CA: Academic Press Inc. doi: 10.1016/S0065-2601(08)60235-X

Blau, P. M. (1964). *Exchange and power in social life*. New York, NY: Wiley.

Bone, J. E., & Raihani, N. J. (2015). Human punishment is motivated by both a desire for revenge and a desire for equality. *Evolution and Human Behavior, 36*(4), 323-330. doi: 10.1016/j.evolhumbehav.2015.02.002

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America, 100*(6), 3531-3535. doi: 10.1073/pnas.0630443100

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13*(3), 171-195. doi: 10.1016/0162-3095(92)90032-y

Brandts, J., & Charness, G. (2003). Truth or consequences: An experiment. *Management Science, 49*(1), 116-130. doi: 10.1287/mnsc.49.1.116.12755

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics, 14*(3), 375-398. doi: 10.1007/s10683-011-9272-x

Brosig, J., Weimann, J., & Yang, C.-L. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics, 6*(1), 75-90. doi: 10.1023/A:1024204826499

Brosig, J., Weimann, J., & Yang, C.-L. (2004). Communication, reputation, and punishment in sequential bargaining experiments. *Journal of Institutional and Theoretical Economics-Zeitschrift Fur Die Gesamte Staatswissenschaft, 160*(4), 576-606. doi: 10.1628/0932456042776140

Brown, G. R., & Richerson, P. J. (2014). Applying evolutionary theory to human behaviour: Past differences and current debates. *Journal of Bioeconomics, 16*(2), 105-128. doi: 10.1007/s10818-013-9166-4

Büchner, S., Coricelli, G., & Greiner, B. (2007). Self-centered and other-regarding behavior in the solidarity game. *Journal of Economic Behavior & Organization, 62*(2), 293-303. doi: 10.1016/j.jebo.2004.12.006

Buss, A. H. (1961). *The psychology of aggression*. New York, NY: John Wiley & Sons. doi: 10.1037/11160-000

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. New York, NY: Russell Sage Foundation.

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology, 42*(4), 437-451. doi: 10.1016/j.jesp.2005.06.007

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*(2), 284-299. doi: 10.1037//0022-3514.83.2.284

Casari, M., & Cason, T. N. (2009). The strategy method lowers measured trustworthy behavior. *Economics Letters, 103*(3), 157-159. doi: 10.1016/j.econlet.2009.03.012

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics, 117*(3), 817-869. doi: 10.1162/003355302760193904

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation, 39*, 860-864. doi: 10.1080/03610911003650383

Chen, X.-P., Dang, C. T., & Keng-Highberger, F. (2014). Broadening the motivation to cooperate: Revisiting the role of sanctions in social dilemmas. In P. A. M. Van Lange, B. Rockenbach & T. Yamagishi (Eds.), *Reward and punishment in social dilemmas.* (pp. 115-132). New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199300730.003.0007

Chen, X.-P., Pillutla, M. M., & Yao, X. (2009). Unintended consequences of cooperation inducing and maintaining mechanisms in public goods dilemmas: Sanctions and moral appeals. *Group Processes & Intergroup Relations, 12*(2), 241-255. doi: 10.1177/1368430208098783

Cherry, T. L. (2001). Mental accounting and other-regarding behavior: Evidence from the lab. *Journal of Economic Psychology, 22*(5), 605-615.

Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics, 9*(3), 265-279. doi: 10.1007/s10683-006-9127-z

Cohen, J. D. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General, 143*(6), 2279-2286. doi: 10.1037/xge0000018

Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An interdisciplinary review. *Journal of Management, 31*(6), 874-900. doi: 10.1177/0149206305279602

Crowe, B. L. (1969). The tragedy of the commons revisited. *Science, 116*(3909), 1103-1107. doi: 10.1126/science.166.3909.1103

Cubitt, R. P., Drouvelis, M., & Gächter, S. (2011). Framing and free riding: emotional responses and punishment in social dilemma games. *Experimental Economics, 14*(2), 254-272. doi: 10.1007/s10683-010-9266-0

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment testing three principles of harm. *Psychological Science, 17*(12), 1082-1089. doi: 10.1111/j.1467-9280.2006.01834.x

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8*(4), 377-383. doi: 10.1037/h0025589

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review, 7*(4), 324-336. doi: 10.1207/S15327957PSPR0704_05

Darwin, C. (1859/1962). *The origin of species*. New York, NY: Collier Books.

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature, 446*(7137), 794-796. doi: 10.1038/nature05651

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169-193. doi: 10.1146/annurev.ps.31.020180.001125

De Cremer, D., & Van Dijk, E. (2005). When and why leaders put themselves first: Leader behaviour in resource allocations as a function of feeling entitled. *European Journal of Social Psychology, 35*(4), 553-563. doi: 10.1002/ejsp.260

De Kwaadsteniet, E. W., Rijkhoff, S. A. M., & Van Dijk, E. (2013). Equality as a benchmark for third-party punishment and reward: The moderating role of uncertainty in social dilemmas. *Organizational Behavior and Human Decision Processes, 120*(2), 251-259. doi: 10.1016/j.obhdp.2012.06.007

De Kwaadsteniet, E. W., Van Dijk, E., Wit, A. P., & De Cremer, D. (2006). Social dilemmas as strong versus weak situations: Social value orientations and tacit coordination under resource size uncertainty. *Journal of Experimental Social Psychology, 42*(4), 509-516. doi: 10.1016/j.jesp.2005.06.004

De Kwaadsteniet, E. W., Van Dijk, E., Wit, A. P., & De Cremer, D. (2008). 'How many of us are there?': Group size uncertainty and social value orientations in common resource dilemmas. *Group Processes & Intergroup Relations, 11*(3), 387-399. doi: 10.1177/1368430208090649

De Kwaadsteniet, E. W., Van Dijk, E., Wit, A. P., & De Cremer, D. (2010). Anger and retribution after collective overuse: The role of blaming and environmental uncertainty in social dilemmas. *Personality and Social Psychology Bulletin, 36*(1), 59-70. doi: 10.1177/0146167209352192

De Kwaadsteniet, E. W., Van Dijk, E., Wit, A. P., De Cremer, D., & De Rooij, M. (2007). Justifying decisions in social dilemmas: Justification pressures and tacit coordination under environmental uncertainty. *Personality and Social Psychology Bulletin, 33*(12), 1648-1660. doi: 10.1177/0146167207307490

De Quervain, D. J. F., Fischbacher, U., Treyer, V., Schelthammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science, 305*(5688), 1254-1258. doi: 10.1126/science.1100735

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*(6), 627-668. doi: 10.1037/0033-2909.125.6.627

Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227. doi: 10.1207/S15327965PLI1104_01

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences of the United States of America, 108*(32), 13335-13340. doi: 10.1073/pnas.1102131108

Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory, 33*(1), 145-167. doi: 10.1007/s00199-007-0212-0

Edney, J. J., & Harper, C. S. (1978). The commons dilemma: A review of contributions from psychology. *Environmental Managment, 2*(6), 491-507. doi: 10.1007/BF01866708

Effron, D. A., & Miller, D. T. (2015). Do as I say, not as I've done: Suffering for a misdeed reduces the hypocrisy of advising others against it. *Organizational Behavior and Human Decision Processes, 131*, 16-32. doi: 10.1016/j.obhdp.2015.07.004

Eisenberger, R., Lynch, P., Aselage, J., & Rohdieck, S. (2004). Who takes the most revenge? Individual differences in negative reciprocity norm endorsement. *Personality and Social Psychology Bulletin, 30*(6), 789-799. doi: 10.1177/0146167204264047

Eriksson, K., Andersson, P. A., & Strimling, P. (2015). Moderators of the disapproval of peer punishment. *Group Processes & Intergroup Relations, 19*(2), 152-168. doi: 10.1177/1368430215583519

Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica, 73*(6), 2017-2030. doi: 10.1111/j.1468-0262.2005.00644.x

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63-87. doi: 10.1016/s1090-5138(04)00005-4

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature, 13*(1), 1-25. doi: 10.1007/s12110-002-1012-7

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980-994. doi: 10.1257/aer.90.4.980

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137-140. doi: 10.1038/415137a

Fehr, E., & Henrich, J. (2003). Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation*. (pp. 55-82). Cambridge, MA: MIT Press.

Fehr, E., & Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. *Current Opinion in Neurobiology, 14*(6), 784-790. doi: 10.1016/j.conb.2004.10.007

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*(3), 817-868. doi: 10.1162/003355399556151

Fershtman, C., & Gneezy, U. (2001). Strategic delegation: An experiment. *Rand Journal of Economics, 32*(2), 352-368. doi: 10.2307/2696414

Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of de-individuation in a group. *The Journal of Abnormal and Social Psychology, 47*(2), 382-389. doi: 10.1037/h0057906

Fischbacher, U., Gächter, S., & Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology, 33*(4), 897-913. doi: 10.1016/j.joep.2012.04.002

Forsyth, D. R., Zyzniewski, L. E., & Giammanco, C. A. (2002). Responsibility diffusion in cooperative collectives. *Personality and Social Psychology Bulletin, 28*(1), 54-65. doi: 10.1177/0146167202281005

Fox, D. R. (1985). Psychology, ideology, utopia, and the commons. *American Psychologist, 40*(1), 48-58. doi: 10.1037/0003-066X.40.1.48

Fridhandler, B. M., & Averill, J. R. (1982). Temporal dimensions of anger: An exploration of time and emotion. In J. R. Averill (Ed.), *Anger and Aggression* (pp. 253-280). New York, NY: Springer-Verlag. doi: 10.1007/978-1-4612-5743-1_12

Gächter, S., & Herrmann, B. (2009). Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1518), 791-806. doi: 10.1098/rstb.2008.0275

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science, 322*(5907), 1510-1510. doi: 10.1126/science.1164744

Gerard, H. B., & Hoyt, M. F. (1974). Distinctiveness of social categorization and attitutde toward ingroup members. *Journal of Personality and Social Psychology, 29*(6), 836-842. doi: 10.1037/h0036204

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology, 206*(2), 169-179. doi: 10.1006/jtbi.2000.2111

Gintis, H. (2003). The Hitchhiker's guide to altruism: Gene-culture coevolution, an the internalization of norms. *Journal of Theoretical Biology, 220*(4), 407-418. doi: 10.1006/jtbi.2003.3104

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24*(3), 153-172. doi: 10.1016/s1090-5138(02)00157-5

Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research, 21*(2), 241-253. doi: 10.1007/s11211-008-0067-y

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies, 29*(1), 1-17. doi: 10.1086/468061

Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology, 48*(1), 250-256. doi: 10.1016/j.jesp.2011.08.006

Gouldner, A. W. (1960). The norm of reciprocity - A preliminary statement. *American Sociological Review, 25*(2), 161-178. doi: 10.2307/2092623

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389-400. doi: 10.1016/j.neuron.2004.09.027

Greenland, S., Schwartzbaum, J. A., & Finkle, W. D. (2000). Problems due to small samples and sparse data in conditional logistic regression analysis. *American Journal of Epidemiology, 151*(5), 531-539. doi: 10.1093/oxfordjournals.aje.a010240

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271-299. doi: 10.1037/1089-2680.2.3.271

Gürerk, O., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science, 312*(5770), 108-111. doi: 10.1126/science.1123633

Güth, W., Huck, S., & Müller, W. (2001). The relevance of equal splits in ultimatum games. *Games and Economic Behavior, 37*(1), 161-169. doi: 10.1006/game.2000.0829

Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology, 69*(3), 339-348. doi: 10.1016/j.tpb.2005.09.005

Hak, T., Van Rhee, H. J., & Suurmond, R. (2016). How to interpret results of meta-analysis. (Version 1.0). Rotterdam, The Netherlands: Erasmus Rotterdam Institute of Managment. Retrieved from www.erim.eur.nl/research-support/meta-essentials/downloads.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology, 7*, 1-52. doi: 10.1016/0022-5193(64)90038-4

Hamman, J. R., Loewenstein, G. F., & Weber, R. A. (2010). Self-interest through delegation: An additional rationale for the principal-agent relationship. *American Economic Review, 100*(4), 1826-1846. doi: 10.1257/aer.100.4.1826

Hardin, G. (1968). The tragedy of the commons. *Science, 162*(3859), 1243-1248. doi: 10.1126/science.162.3859.1243

Harinck, F., & De Dreu, C. K. W. (2008). Take a break! or not? The impact of mindsets during breaks on negotiation processes and outcomes. *Journal of Experimental Social Psychology, 44*(2), 397-404. doi: 10.1016/j.jesp.2006.12.009

Hart, J. W., Bridgett, D. J., & Karau, S. J. (2001). Coworker ability and effort as determinants of individual effort on a collective task. *Group Dynamics: Theory, Research, and Practice, 5*(3), 181-190. doi: 10.1037/1089-2699.5.3.181

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486-504. doi: 10.1037/1082-989X.3.4.486

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*(5972), 1480-1484. doi: 10.1126/science.1182238

Henrich, J., & Henrich, N. (2007). *Why humans cooperate: A cultural and evolutionary explanation*. New York, NY: Oxford University Press.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Ziker, J. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767-1770. doi: 10.1126/science.1127333

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science, 319*(5868), 1362-1367. doi: 10.1126/science.1153808

Hobbes, T. (1651/1991). Leviathan. Cambridge, UK: Cambridge University Press.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property-rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*(3), 346-380. doi: 10.1006/game.1994.1056

Holmes, E. A., James, E. L., Coode-Bate, T., & Deeprose, C. (2009). Can playing the computer game "Tetris" reduce the build-up of flashbacks for trauma? A proposal from cognitive science. *PLoS ONE, 4*(1), 1-6. doi: 10.1371/journal.pone.0004153

Hornsey, M. J., & Imani, A. (2004). Criticizing groups from the inside and the outside: An identity perspective on the intergroup sensitivity effect. *Personality and Social Psychology Bulletin, 30*(3), 365-383. doi: 10.1177/0146167203261295

Hornsey, M. J., Trembath, M., & Gunthorpe, S. (2004). 'You can criticize because you care': Identity attachment, constructiveness, and the intergroup sensitivity effect. *European Journal of Social Psychology, 34*(5), 499-518. doi: 10.1002/ejsp.212

Insko, C. A., Pinkley, R. L., Hoyle, R. H., Dalton, B., Hong, G. Y., Slim, R. M., . . . Thibaut, J. (1987). Individual versus group discontinuity: The role intergroup contact. *Journal of Experimental Social Psychology, 23*(3), 250-267. doi: 10.1016/0022-1031(87)90035-7

Insko, C. A., Schopler, J., Gaertner, L., Wildschut, T., Kozar, R., Pinter, B., . . . Montoya, M. R. (2001). Interindividual–intergroup discontinuity reduction through the anticipation of future interaction. *Journal of Personality and Social Psychology, 80*(1), 95-111. doi: 10.1037/0022-3514.80.1.95

Jaffe, Y., Shapir, N., & Yinon, Y. (1981). Aggression and its escalation. *Journal of Cross-Cultural Psychology, 12*(1), 21-36. doi: 10.1177/0022022181121002

Jaffe, Y., & Yinon, Y. (1979). Retaliatory aggression in individuals and groups. *European Journal of Social Psychology, 9*(2), 177-186. doi: 10.1002/ejsp.2420090206

Janoff-Bulman, R., & Carnes, N. C. (2013). Surveying the moral landscape: Moral motives and group-based moralities. *Personality and Social Psychology Review, 17*(3), 219-236. doi: 10.1177/1088868313480274

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology, 96*(3), 521-537. doi: 10.1037/a0013779

Jewell, N. P. (1984). Small-sample bias of point estimators of the odds ratio from matched sets. *Biometrics, 40*(2), 421-435. doi: 10.2307/2531395

Kahneman, D., Diener, E., & Schwarz, N. (1999). *Well-being: The foundations of hedonic psychology*. New York, NY: Russell Sage Foundation.

Kameda, T., Tsukasaki, T., Hastie, R., & Berg, N. (2011). Democracy under uncertainty: The wisdom of crowds and the free-rider problem in group decision making. *Psychological Review, 118*(1), 76-96. doi: 10.1037/a0020699

Kamei, K., Putterman, L., & Tyran, J.-R. (2014). State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Experimental Economics, 18*(1), 38-65. doi: 10.1007/s10683-014-9405-0

Kenrick, D. T., Griskevicius, V., Sundie, J. M., Li, N. P., Li, Y. J., & Neuberg, S. L. (2009). Deep rationality: The evolutionary economics of decision making. *Social Cognition, 27*, 764-785. doi: 10.1521/soco.2009.27.5.764

Kerr, N. L., Rumble, A. C., Park, E. S., Ouwerkerk, J. W., Parks, C. D., Gallucci, M., & Van Lange, P. A. M. (2009). "How many bad apples does it take to spoil the whole barrel?": Social exclusion and toleration for bad apples. *Journal of Experimental Social Psychology, 45*(4), 603-613. doi: 10.1016/j.jesp.2009.02.017

Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology, 95*(4), 826-842. doi: 10.1037/a0011381

Kogan, N., & Wallach, M. (1967). Group risk taking as a function of members' anxiety and defensiveness levels. *Journal of Personality, 35*(1), 50-63. doi: 10.1111/j.1467-6494.1967.tb01415.x

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology, 24*, 183-214. doi: 10.1146/annurev.soc.24.1.183

Komorita, S. S., & Barth, J. M. (1985). Components of reward in social dilemmas. *Journal of Personality and Social Psychology, 48*(2), 364-373. doi: 10.1037//0022-3514.48.2.364

Komorita, S. S., & Parks, C. D. (1995). Interpersonal relations: Mixed-motive interaction. *Annual Review of Psychology, 46*, 183-207. doi: 10.1146/annurev.ps.46.020195.001151

Krasnow, M. M., Cosmides, L., Pedersen, E. J., & Tooby, J. (2012). What are punishment and reputation for? *Plos One, 7*(9), e45662. doi: 10.1371/jounral.pone.0045662

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2015). Group cooperation without group selection: Modest punishment can recruit much cooperation. *Plos One, 10*(4), e0124561. doi: 10.1371/journal.pone.0124561

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science, 27*(3), 405-418. doi: 10.1177/0956797615624469

Krasnow, M. M., Delton, A. W., Tooby, J., & Cosmides, L. (2013). Meeting now suggests we will meet again: Implications for debates on the evolution of cooperation. *Scientific Reports, 3*, 1747. doi: 10.1038/srep01747

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior, 28*(2), 75-84. doi: 10.1016/j.evolhumbehav.2006.06.001

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*(2), 311-328. doi: 10.1037/0022-3514.32.2.311

Latané, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology, 10*(3), 215-221. doi: 10.1037/h0026570

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin, 89*(2), 308-324. doi: 10.1037/0033-2909.89.2.308

Le Bon, G. (1903). *The Crowd: A study of the poular mind.* London, UK: Unwin. doi: 10.1037/10878-000

Leknes, S., & Tracey, L. (2008). A common neurobiology for pain and pleasure. *Nature Reviews Neuroscience, 9*(4), 314-320. doi: 10.1038/nrn2333

Leliveld, M. C., Van Dijk, E., & Van Beest, I. (2012). Punishing and compensating others at your own expense: The role of empathic concern on reactions to distributive injustice. *European Journal of Social Psychology, 42*(2), 135-140. doi: 10.1002/ejsp.872

Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin, 24*(6), 563-574. doi: 10.1177/0146167298246001

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*(2), 255-275. doi: 10.1037/0033-2909.125.2.255

Lewin, K. (1951). *Field theory in social science*. New York, NY: Harper & Row.

Li, P., Jia, S., Feng, T., Liu, Q., Suo, T., & Li, H. (2010). The influence of the diffusion of responsibility effect on outcome evaluations: Electrophysiological evidence from an ERP study. *Neuroimage, 52*(4), 1727-1733. doi: 10.1016/j.neuroimage.2010.04.275

Loewenstein, G. F. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65*(3), 272-292. doi: 10.1006/obhd.1996.0028

Loewenstein, G. F., & Lerner, J. S. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 619-642). Oxford, UK: Oxford University Press.

Loewenstein, G. F., & Schkade, D. (1999). Wouldn't it be nice? Predicting future feelings. In D. Kahneman & E. Diener (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 85-105). New York, NY: Russell Sage Foundation.

Lynn, M., & Oldenquist, A. (1986). Egoistic and nonegoistic motives in social dilemmas. *American Psychologist, 41*(5), 529-534. doi: 10.1037/0003-066X.41.5.529

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614. doi: 10.1146/annurev.psych.58.110405.085542

March, J. G. (1994). *A primer of decision making: How decisions happen*. New York, NY: Free Press.

Markussen, T., Putterman, L., & Tyran, J. R. (2014). Self-organization for collective action: An experimental study of voting on sanction regimes. *Review of Economic Studies, 81*(1), 301-324. doi: 10.1093/restud/rdt022

Mathes, E. W., & Kahn, A. (1975). Diffusion of responsibility and extreme behavior. *Journal of Personality and Social Psychology, 31*(5), 881-886. doi: 10.1037/h0076695

McCallum, D. M., Harring, K., Gilmore, R., Drenan, S., Chase, J. P., Insko, C. A., & Thibaut, J. (1985). Competition and cooperation between groups and between individuals. *Journal of Experimental Social Psychology, 21*(4), 301-320. doi: 10.1016/0022-1031(85)90032-0

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*(2), 190-201. doi: 10.1006/obhd.1995.1015

McGregor, H. A., Lieberman, J. D., Greenberg, J., Solomon, S., Arndt, J., Simon, L., & Pyszczynski, T. (1998). Terror management and aggression: Evidence that mortality salience motivates aggression against worldview-threatening others. *Journal of Personality and Social Psychology, 74*(3), 590-605. doi: 10.1037//0022-3514.74.3.590

Meier, B. P., & Hinsz, V. B. (2004). A comparison of human aggression committed by groups and individuals: An interindividual-intergroup discontinuity. *Journal of Experimental Social Psychology, 40*(4), 551-559. doi: 10.1016/j.jesp.2003.11.002

Messé, L. A., & Sivacek, J. M. (1979). Predictions of others' responses in a mixed-motive game: Self-justification or false consensus? *Journal of Personality and Social Psychology, 37*(4), 602-607. doi: 10.1037/0022-3514.37.4.602

Messick, D. M. (1999). Alternative logics for decision making in social settings. *Journal of Economic Behavior & Organization, 39*(1), 11-28. doi: 10.1016/s0167-2681(99)00023-2

Messick, D. M., & Brewer, M. B. (1983). Solving social dilemmas: A review. *Review of personality and social psychology, 4*, 11-44.

Miceli, M., & Castelfranchi, C. (2015). *Expectancy and Emotion*. Oxford, UK: Oxford University Press.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York, NY: Harper & Row.

Milgram, S., & Toch, H. (1969). Collective behavior: Crowds and social motivations. In G. Lindzey & E. Aronson (Eds.), *The Handbook of Social Psychology* (Vol. 4). Reading, MA: Addison-Wesley.

Milinski, M., Semmann, D., & Krambeck, H. J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature, 415*(6870), 424-426. doi: 10.1038/415424a

Miller, D. T. (1999). The norm of self-interest. *American Psychologist, 54*(12), 1053-1060. doi: 10.1037/0003-066X.54.12.1053

Miller, D. T., & Effron, D. A. (2010). Psychological license: When it is needed and how it functions. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology*. San Diago, CA: Academic Press/Elsevier. doi: 10.1016/S0065-2601(10)43003-8

Miller, D. T., Effron, D. A., & Zak, S. V. (2009). From moral outrage to social protest: The role of psychological standing. In D. R. Bobocel, A. C. Kay, M. P. Zanna & J. M. Olson (Eds.), *The psychology of justice and legitimacy*. (Vol. 11, pp. 103-123). New York, NY: Psychology Press.

Miller, D. T., & Ratner, R. K. (1996). The power of the myth of self-interest. In L. Montada & M. J. Lerner (Eds.), *Current societal concerns about justice*. New York, NY: Plenum Press. doi: 10.1007/978-1-4757-9927-9_3

Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology, 74*(1), 53-62. doi: 10.1037/0022-3514.74.1.53

Molenmaker, W. E., De Kwaadsteniet, E. W., & Van Dijk, E. (2014). On the willingness to costly reward cooperation and punish non-cooperation: The moderating role of type of social dilemma. *Organizational Behavior and Human Decision Processes, 125*(2), 175-183. doi: 10.1016/j.obhdp.2014.09.005

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

Molenmaker, W. E., De Kwaadsteniet, E. W., & Van Dijk, E. (2016). The willingness to costly reward cooperation and punish non-cooperation before versus after the choice behavior: Sanctioning the past, the present or the future. *Manuscript under review.*

Molm, L. D. (1997). Risk and power use: Constraints on the use of coercion in exchange. *American Sociological Review, 62*(1), 113-133. doi: 10.2307/2657455

Mooijman, M., Van Dijk, W. W., Ellemers, N., & Van Dijk, E. (2015). Why leaders punish: A power perspective. *Journal of Personality and Social Psychology, 109*(1), 75-89. doi: 10.1037/pspi0000021

Muehlbacher, S., & Kirchler, E. (2009). Origin of endowments in public good games: The impact of effort on contributions. *Journal of Neuroscience, Psychology, and Economics, 2*(1), 59-67.

Mulcahy, N. J., & Call, J. (2006). Apes save tools for future use. *Science, 312*(5776), 1038-1040. doi: 10.1126/science.1125456

Mulder, L. B., Van Dijk, E., & De Cremer, D. (2009). When sanctions that can be evaded still work: The role of trust in leaders. *Social Influence, 4*(2), 122-137. doi: 10.1080/15534510802469156

Mulder, L. B., Van Dijk, E., De Cremer, D., & Wilke, H. A. M. (2006a). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology, 42*(2), 147-162. doi: 10.1016/j.jesp.2005.03.002

Mulder, L. B., Van Dijk, E., De Cremer, D., & Wilke, H. A. M. (2006b). When sanctions fail to increase cooperation in social dilemmas: Considering the presence of an alternative option to defect. *Personality and Social Psychology Bulletin, 32*(10), 1312-1324. doi: 10.1177/0146167206289978

Mulder, L. B., Van Dijk, E., Wilke, H. A. M., & De Cremer, D. (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. doi: 10.1016/j.joep.2004.12.007

Muller, L., Sefton, M., Steinberg, R., & Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary contribution experiments. *Journal of Economic Behavior & Organization, 67*(3-4), 782-793. doi: 10.1016/j.jebo.2007.09.001

Mynatt, C., & Sherman, S. J. (1975). Responsibility attribution in groups and individuals: A direct test of the diffusion of responsibility hypothesis. *Journal of Personality and Social Psychology, 32*(6), 1111-1118. doi: 10.1037/0022-3514.32.6.1111

Nash, J. F. (1950). Equilibrium points in N-person games. *Proceedings of the National Academy of Sciences of the United States of America, 36*(1), 48-49. doi: 10.1073/pnas.36.1.48

Nelissen, R. M. A., & Zeelenberg, M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. *Judgment and Decision Making, 4*(7), 543-553.

Nemes, S., Jonasson, J. M., Genell, A., & Steineck, G. (2009). Bias in odds ratios by logistic regression modelling and sample size. *Bmc Medical Research Methodology, 9*, 5. doi: 10.1186/1471-2288-9-56

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92*(1-2), 91-112. doi: 10.1016/j.jpubeco.2007.04.008

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*(6685), 573-577. doi: 10.1038/31225

O'Reilly, C. A., & Puffer, S. M. (1989). The impact of rewards and punishments in social context: A laboratory and field experiment. *Journal of Occupational Psychology, 62*(1), 41-53. doi: 10.1111/j.2044-8325.1989.tb00476.x

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review, 46*(8), 1423-1437. doi: 10.1016/s0014-2921(01)00176-3

Oliver, P. (1980). Rewards and punishments as selective incentives for collective action: Theoretical investigations. *Journal of Sociology, 85*(6), 1356-1375. doi: 10.1086/227168

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511807763

Ostrom, E., Burger, J., Field, C. B., Norgaard, R. B., & Policansky, D. (1999). Sustainability - Revisiting the commons: Local lessons, global challenges. *Science, 284*(5412), 278-282. doi: 10.1126/science.284.5412.278

Ostrom, E., Walker, J., & Gardner, R. (1992). Coverants with and without a sword: Self-goverance is possible. *American Political Science Review, 86*(2), 404-417. doi: 10.2307/1964229

Oxoby, R. J., & McLeish, K. N. (2004). Sequential decision and strategy vector methods in ultimatum bargaining: Evidence on the strength of other-regarding behavior. *Economics Letters, 84*(3), 399-405. doi: 10.1016/j.econlet.2004.03.011

Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization, 65*(3-4), 703-713.

Parks, C. D., Joireman, J., & Van Lange, P. A. M. (2013). Cooperation, trust, and antagonism: How public goods are promoted. *Psychological Science in the Public Interest, 14*(3), 119-165. doi: doi: 10.1177/1529100612474436

Parks, C. D., & Stone, A. B. (2010). The desire to expel unselfish members from the group. *Journal of Personality and Social Psychology, 99*(2), 303-310. doi: 10.1037/a0018403

Pennington, J., & Schlenker, B. R. (1999). Accountability for consequential decisions: Justifying ethical judgments to audiences. *Personality and Social Psychology Bulletin, 25*(9), 1067-1081. doi: 10.1177/01461672992512001

Piazza, J., & Bering, J. M. (2008). The effects of perceived anonymity on altruistic punishment. *Evolutionary Psychology, 6*(3), 487-501. doi: 10.1177/147470490800600314

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes, 68*(3), 208-224. doi: 10.1006/obhd.1996.0100

Poundstone, W. (1992). *Prisoner's Dilemma*. New York, NY: Oxford University Press.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891. doi: 10.3758/BRM.40.3.879

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology, 28*, 363-392. doi: 10.1146/annurev.ps.28.020177.002051

Putterman, L. (2014). When punishment supports cooperation: Insights from voluntary contribution experiments. In P. A. M. Van Lange, B. Rockenbach & T. Yamagishi (Eds.), *Reward and punishment in social dilemmas* (pp. 17-33). New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199300730.003.0002

Putterman, L., Tyran, J. R., & Kamei, K. (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics, 95*(9-10), 1213-1222. doi: 10.1016/j.jpubeco.2011.05.001

Quattrone, G. A., & Tversky, A. (1984). Casual versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology, 46*(2), 237-248. doi: 10.1037/0022-3514.46.2.237

Rand, D. G., Armao, J. J., Nakamaru, M., & Ohtsuki, H. (2010). Anti-social punishment can prevent the co-evolution of punishment and cooperation. *Journal of Theoretical Biology, 265*(4), 624-632. doi: 10.1016/j.jtbi.2010.06.010

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science, 325*(5945), 1272-1275. doi: 10.1126/science.1177418

Rand, D. G., & Nowak, M. A. (2011). The evolution of antisocial punishment in optional public goods games. *Nature Communications, 2*, 7. doi: 10.1038/ncomms1442

Rapoport, A., & Au, W. T. (2001). Bonus and penalty in common pool resource dilemmas under uncertainty. *Organizational Behavior and Human Decision Processes, 85*(1), 135-165. doi: 10.1006/obhd.2000.2935

Ratner, R. K., & Miller, D. T. (2001). The norm of self-interest and its effects on social action. *Journal of Personality and Social Psychology, 81*(1), 5. doi: 10.1037/0022-3514.81.1.5

Ray, R. D., Wilhelm, F. H., & Gross, J. J. (2008). All in the mind's eye? Anger rumination and reappraisal. *Journal of Personality and Social Psychology, 94*(1), 133-145. doi: 10.1037/0022-3514.94.1.133

Reuben, E., & Suetens, S. (2012). Revisiting strategic versus non-strategic cooperation. *Experimental Economics, 15*(1), 24-43. doi: 10.1007/s10683-011-9286-4

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making, 3*(4), 263-277. doi: 10.1002/bdm.3960030404

Ritov, I., & Baron, J. (1992). Status-quo and omission biases. *Journal of Risk and Uncertainty, 5*(1), 49-61. doi: 10.1007/BF00208786

Rockenbach, B., & Milinski, M. (2011). To qualify as a social partner, humans hide severe punishment, although their observed cooperativeness is decisive. *Proceedings of the National Academy of Sciences of the United States of America, 108*(45), 1-6. doi: 10.1073/pnas.1108996108

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York, NY: Mcgraw-Hill.

Rotemberg, J. J. (2008). Minimally acceptable altruism and the ultimatum game. *Journal of Economic Behavior & Organization, 66*(3-4), 457-476. doi: 10.1016/j.jebo.2006.06.008

Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research, 15*(2), 165-184. doi: 10.1023/A:1019923923537

Rutte, C. G., & Wilke, H. A. M. (1984). Social dilemmas and leadership. *European Journal of Social Psychology, 14*(1), 105-121. doi: 10.1002/ejsp.2420140109

Rutte, C. G., Wilke, H. A. M., & Messick, D. M. (1987). Scarcity or abundance caused by people or the environment as determinants of behavior in the resource dilemma. *Journal of Experimental Social Psychology, 23*(3), 208-216. doi: 10.1016/0022-1031(87)90032-1

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Samuelson, P. A. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics, 36*(4), 387-289. doi: 10.2307/1925895

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755-1758. doi: 10.1126/science.1082976

Schlenker, B. R. (1986). Self-identification: Toward an integration of the private and public self. In R. F. Baumeister (Ed.), *Public self and private self* (pp. 21-62). New York, NY: Springer. doi: 10.1007/978-1-4613-9564-5_2

Schlenker, B. R., Britt, T. W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological Review, 101*(4), 632-652. doi: 10.1037/0033-295X.101.4.632

Schlenker, B. R., Weigold, M. F., & Doherty, K. (1991). Coping with accountability: Self-identification and evaluative reckonings. In C. R. Snyder & D. R. Forsyth (Eds.), *Handbook of social and clinical psychology: The health perspective* (pp. 96-115). Elmsford, NY: Pergamon Press.

Schopler, J., Insko, C. A., Drigotas, S. M., Wieselquist, J., Pemberton, M. B., & Cox, C. (1995). The role of identifiability in the reduction of interindividual intergroup discontinuity. *Journal of Experimental Social Psychology, 31*(6), 553-574. doi: 10.1006/jesp.1995.1025

Schroeder, D. A., Steel, J. E., Woodell, A. J., & Bembenek, A. F. (2003). Justice within social dilemmas. *Personality and Social Psychology Review, 7*(4), 374-387. doi: 10.1207/s15327957pspr0704_09

Scott, M. B., & Lyman, S. M. (1968). Accounts. *American Sociological Review, 33*(2), 46-62. doi: 10.2307/2092239

Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry, 45*(4), 671-690. doi: 10.1111/j.1465-7295.2007.00051.x

Seip, E. C., Van Dijk, W. W., & Rotteveel, M. (2009). On hotheads and dirty harries: The primacy of anger in altruistic punishment. *Annals of the New York Academy of Sciences, 1167*, 190-196. doi: 10.1111/j.1749-6632.2009.04503.x

Seip, E. C., Van Dijk, W. W., & Rotteveel, M. (2014). Anger motivates costly punishment of unfair behavior. *Motivation and Emotion, 38*(4), 578-588. doi: 10.1007/s11031-014-9395-4

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens in rahmen eines oligopolexperiments. In H. Sauermann (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung* (pp. 136-168). Tübingen, GR: Mohr.

Semin, G. R., & Manstead, A. S. R. (1983). *The accountability of conduct: A social psychological analysis*. New York, NY: Academic Press.

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition, 50*(1-3), 403-430. doi: 10.1016/0010-0277(94)90038-8

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition, 49*(1), 11-36. doi: 10.1016/0010-0277(93)90034-S

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology, 24*(4), 449-474. doi: 10.1016/0010-0285(92)90015-T

Shaver, K. G. (1975). *An introduction to attribution processes*. Cambridge, MA: Winthrop.

Shaver, K. G. (1985). *The attribution of blame*. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4612-5094-4

Sherif, M., Harvey, O. J., White, J., Hood, W., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The robber's cave experiment*. Norman, OK: Institute of Intergroup Relations.

Shinada, M., & Yamagishi, T. (2007). Bringing back Leviathan in social dilemmas. In A. Biel, D. Eek, T. Gärling & M. Gustafsson (Eds.), *New issues and paradigms in research on social dilemmas* (pp. 93-123). New York, NY: Springer.

Sivanathan, N., Molden, D. C., Galinsky, A. D., & Ku, G. (2008). The promise and peril of self-affirmation in de-escalation of commitment. *Organizational Behavior and Human Decision Processes, 107*(1), 1-14. doi: 10.1016/j.obhdp.2007.12.004

Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology, 71*(3), 549. doi: 10.1037/0022-3514.71.3.549

Small, D. A., & Loewenstein, G. F. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty, 26*(1), 5-16. doi: 10.1023/a:1022299422219

Small, D. A., & Loewenstein, G. F. (2005). The devil you know: The effects of identifiability on punishment. *Journal of Behavioral Decision Making, 18*(5), 311-318. doi: 10.1002/bdm.507

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*(1), 76-105. doi: 10.1016/0022-1031(91)90011-T

Strimling, P., & Eriksson, K. (2014). Regulating the regulation: Norms about punishment. In P. A. M. Van Lange, B. Rockenbach & T. Yamagishi (Eds.), *Reward and punishment in social dilemmas* (pp. 52-67). New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199300730.001.0001

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies, 77*(4), 1540-1566. doi: 10.1111/j.1467-937X.2010.00608.x

Sylwester, K., Herrmann, B., & Bryson, J. J. (2013). Homo homini lupus? Explaining antisocial punishment. *Journal of Neuroscience, Psychology, and Economics, 6*(3), 167-188. doi: 10.1037/npe0000009

Taylor, M. (1982). *Community, anarchy and liberty*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511607875

Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly, 44*(4), 684-707. doi: 10.2307/2667052

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. *Advances in Experimental Social Psychology, 25*(3), 331-376. doi: 10.1016/S0065-2601(08)60287-7

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. Oxford, UK: John Wiley.

Tinbergen, N. (1968). On war and peace in animals and man. *Science, 160*(3835), 1411-1418. doi: 10.1126/science.160.3835.1411

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science, 16*, 167-171. doi: 10.1111/j.1467-8721.2007.00497.x

Tooby, J., & Cosmides, L. (1992). The psychology foundations of culture. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of cuture* (pp. 19-136). New York, NY: Oxford University Press.

Trevino, L. K. (1992). The social effects of punishment in organizations: A justice perspective. *Academy of Management Review, 17*(4), 647-676. doi: 10.2307/258803

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature, 463*(7284), 1089-U1109. doi: 10.1038/nature08785

Trivers, R. L. (1971). Evolution of reciprocal altruism. *Quarterly Review of Biology, 46*(1), 35-57. doi: 10.1086/406755

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science, 3*(5), 305-309. doi: 10.1111/j.1467-9280.1992.tb00678.x

Van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Experimental Brain Research, 169*(4), 564-568. doi: 10.1007/s00221-006-0346-5

Van Beest, I., Carter-Sowell, A. R., Van Dijk, E., & Williams, K. D. (2012). Groups being ostracized by groups: Is the pain shared, is recovery quicker, and are groups more likely to be aggressive? *Group Dynamics-Theory Research and Practice, 16*(4), 241-254. doi: 10.1037/a0030104

Van Beest, I., Van Dijk, E., De Dreu, C. K. W., & Wilke, H. A. M. (2005). Do-no-harm in coalition formation: Why losses inhibit exclusion and promote fairness cognitions. *Journal of Experimental Social Psychology, 41*(6), 609-617. doi: 10.1016/j.jesp.2005.01.002

Van Dijk, E., De Kwaadsteniet, E. W., & Mulder, L. B. (2009). *How certain do we need to be to punish and reward in social dilemmas?* Paper presented at the 13th international conference on social dilemmas, Kyoto, Japan.

Van Dijk, E., Molenmaker, W. E., & De Kwaadsteniet, E. W. (2015). Promoting cooperation in social dilemmas: The use of sanctions. *Current Opinion in Psychology, 6*, 118-122. doi: 10.1016/j.copsyc.2015.07.006

Van Dijk, E., & Wilke, H. A. M. (1995). Coordination rules in asymmetric social dilemmas - A comparison between public good dilemmas and resource dilemmas. *Journal of Experimental Social Psychology, 31*(1), 1-27. doi: 10.1006/jesp.1995.1001

Van Dijk, E., & Wilke, H. A. M. (1997). Is it mine or is it ours? Framing property rights and decision making in social dilemmas. *Organizational Behavior and Human Decision Processes, 71*(2), 195-209. doi: 10.1006/obhd.1997.2718

Van Dijk, E., & Wilke, H. A. M. (2000). Decision-induced focusing in social dilemmas: Give-some, keep-some, take-some, and leave-some dilemmas. *Journal of Personality and Social Psychology, 78*(1), 92-104. doi: 10.1037//0022-3514.78.1.92

Van Dijk, E., Wilke, H. A. M., & Wit, A. P. (2003). Preferences for leadership in social dilemmas: Public good dilemmas versus common resource dilemmas. *Journal of Experimental Social Psychology, 39*(2), 170-176. doi: 10.1016/s0022-1031(02)00518-8

Van Dijk, E., Wit, A. P., Wilke, H. A. M., & Budescu, D. V. (2004). What we know (and do not know) about the effects of uncertainty on behavior in social dilemmas. In R. Suleiman, D. V. Budescu, I. Fischer & D. M. Messick (Eds.), *Contemporary psychological research on social dilemmas* (pp. 315-331). New York, NY: Cambridge University Press.

Van Dijk, E., & Zeelenberg, M. (2003). The discounting of ambiguous information in economic decision making. *Journal of Behavioral Decision Making, 16*, 341-352. doi: 10.1002/bdm.450

Van Dijk, E., & Zeelenberg, M. (2006). The dampening effect of uncertainty on positive and negative emotions. *Journal of Behavioral Decision Making, 19*, 171-176. doi: 10.1002/bdm.504

Van Dillen, L. F., Van der Wal, R. C., & Van den Bos, K. (2012). On the role of attention and emotion in morality: Attentional control modulates unrelated disgust in moral judgments. *Personality and Social Psychology Bulletin, 38*(9), 1222-1231. doi: 10.1177/0146167212448485

Van Lange, P. A. M., De Cremer, D., Van Dijk, E., & Van Vugt, M. (2007). Self-interest and beyond: Basic principles of social interaction. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2 ed., pp. 540-561). New York, NY: Guilford Press.

Van Lange, P. A. M., Joireman, J. A., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes, 120*(2), 125-141. doi: 10.1016/j.obhdp.2012.11.003

Van Lange, P. A. M., Ouwerkerk, J. W., & Tazelaar, M. J. A. (2002). How to overcome the detrimental effects of noise in social interaction: The benefits of generosity. *Journal of Personality and Social Psychology, 82*(5), 768-780. doi: 10.1037//0022-3514.82.5.768

Van Lange, P. A. M., Rockenbach, B., & Yamagishi, T. (2014). *Reward and punishment in social dilemmas*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199300730.001.0001

Van Rhee, H. J., Suurmond, R., & Hak, T. (2015). User manual for Meta-Essentials: Workbooks for meta-analysis. (Version 1). Rotterdam, The Netherlands: Erasmus Research Institute of Management. Retrieved from www.erim.eur.nl/research-support/meta-essentials.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Walker, J. M., & Halloran, M. A. (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics, 7*(3), 235-247. doi: 10.1023/b:exec.0000040559.08652.51

Wallach, M. A., & Kogan, N. (1965). The roles of information, discussion, and consensus in group risk taking. *Journal of Experimental Social Psychology, 1*(1), 1-19. doi: 10.1016/0022-1031(65)90034-X

Wallach, M. A., Kogan, N., & Bem, D. J. (1962). Group influence on individual risk taking. *The Journal of Abnormal and Social Psychology, 65*(2), 75-86. doi: 10.1037/h0044376

Wallach, M. A., Kogan, N., & Bem, D. J. (1964). Diffusion of responsibility and level of risk taking in groups. *The Journal of Abnormal and Social Psychology, 68*(3), 263-274. doi: 10.1037/h0042190

Wang, C. S., Galinsky, A. D., & Murnighan, J. K. (2009). Bad drives psychological reactions, but good propels behavior: Responses to honesty and deception. *Psychological Science, 20*(5), 634-644. doi: 10.1111/j.1467-9280.2009.02344.x

Wang, C. S., Sivanathan, N., Narayanan, J., Ganegoda, D. B., Bauer, M., Bodenhausen, G. V., & Murnighan, J. K. (2011). Retribution and emotional regulation: The effects of time delay in angry economic interactions. *Organizational Behavior and Human Decision Processes, 116*(1), 46-54. doi: 10.1016/j.obhdp.2011.05.007

Wang, Z.-J., Li, S., & Jiang, C.-M. (2012). Emotional response in a disjuction condition. *Journal of Economic Psychology, 33*(1), 71-78. doi: 10.1016/j.joep.2011.08.009

Weber, J. M., Kopelman, S., & Messick, D. M. (2004). A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review, 8*(3), 281-307. doi: 10.1207/s15327957pspr0803_4

Weiss, D. M., & Sachs, J. (1991). Persuasive strategies used by preschool children. *Discourse Processes, 14*(1), 55-72. doi: 10.1080/01638539109544774

West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior, 32*(4), 231-262. doi: 10.1016/j.evolhumbehav.2010.08.001

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology, 20*(2), 415-432. doi: 10.1111/j.1420-9101.2006.01258.x

Wildschut, T., Pinter, B., Vevea, J. L., Insko, C. A., & Schopler, J. (2003). Beyond the group mind: A quantitative review of the interindividual-intergroup discontinuity effect. *Psychological Bulletin, 129*(5), 698-722. doi: 10.1037/0033-2909.129.5.698

Wilson, D. S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences of the United States of America, 72*(1), 143-146. doi: 10.1073/pnas.72.1.143

Wit, A. P., & Wilke, H. A. M. (1990). The presentation of rewards and punishments in a simulated social dilemma. *Social Behaviour, 5*(4), 231-245.

Wu, J., Balliet, D., & Van Lange, P. A. M. (2016). Gossip versus punishment: The efficiency of reputation to promote and maintain cooperation. *Scientific Reports, 6*, 23919. doi: 10.1038/srep23919

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*(1), 110-116. doi: 10.1037/0022-3514.51.1.110

Yamagishi, T. (1988). The provision of a sanctioning system in the United-States and Japan. *Social Psychology Quarterly, 51*(3), 265-271. doi: 10.2307/2786924

Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. *Nebraska Symposium on Motivation, 17*, 237-307.

Summary in Dutch

(Nederlandse samenvatting)

■ De (on)bereidheid om te sanctioneren in sociale dilemma's

De grootste uitdaging voor iedere samenleving is het waarborgen van het collectieve welzijn. Deze uitdaging komt voort uit het feit dat het gemeenschappelijk belang van een samenleving niet noodzakelijkerwijs overeenkomt met het eigenbelang van de mensen binnen deze samenleving. Situaties die draaien om een dergelijk belangenconflict worden *sociale dilemma's* genoemd (Dawes, 1980). Mensen worden geregeld geconfronteerd met sociale dilemma's waarin ze moeten kiezen tussen het gemeenschappelijk belang (coöperatief gedrag) of hun eigenbelang (non-coöperatief gedrag). Denk bijvoorbeeld aan de gemeenschappelijke voorzieningen waar men over het algemeen onbeperkt toegang toe heeft, zoals schoon drinkwater, openbaar vervoer, medische zorg, elektriciteit, openbare wegen, natuurparken, et cetera. Ondanks dat het in ieders belang is om dergelijke voorzieningen aan te bieden en te behouden, is het niet vanzelfsprekend dat mensen hieraan zullen bijdragen. Voor een individu is het namelijk aantrekkelijker om uitsluitend gebruik te maken van dergelijke gemeenschappelijke voorzieningen (Olson, 1965; Samuelson, 1954). Zo is het bijvoorbeeld erg verleidelijk om lang te douchen. Dit persoonlijk gemak is echter in strijd met het gemeenschappelijk belang van water- en energiebesparing en kan er in het ergste geval, als er teveel schoon drinkwater wordt verbruikt, zelfs toe leiden dat deze gemeenschappelijke voorziening uitgeput raakt (zie Hardin, 1968). Het nastreven van het eigenbelang ten koste van het gemeenschappelijk belang kan dus desastreuze gevolgen hebben voor het collectieve welzijn van een samenleving.

De meest voor de hand liggende manier om het collectieve welzijn te waarborgen is het aantrekkelijker maken van coöperatief gedrag en onaantrekkelijker maken van non-coöperatief gedrag. Een effectieve manier om deze relatieve aantrekkelijkheid te wijzigen is het gebruik van sancties. In de afgelopen decennia hebben tal van onderzoeken laten zien dat positieve sancties (beloningen) voor coöperatie en negatieve sancties (bestraffingen) voor non-coöperatie ervoor kunnen zorgen dat mensen hun eigenbelang opzij zetten ten gunste van het gemeenschappelijk belang (Balliet, Mulder, & Van lange, 2011). Een cruciaal aspect dat ten grondslag ligt aan dit positieve effect van sancties, te weten de bereidheid om coöperatief gedrag te belonen en non-coöperatief gedrag te bestraffen, is het centrale thema van dit proefschrift. In Hoofdstuk 1 beargumenteer ik dat het niet vanzelfsprekend is dat mensen gebruik maken van sancties en dat het van cruciaal belang is om de (on)bereidheid te bestuderen voor ons inzicht in de betrokken psychologische mechanismes, evolutionaire functies en praktische toepasbaarheid. In dit proefschrift identificeer ik vervolgens verschillende factoren die een rol spelen bij de bereidheid om te sanctioneren in sociale dilemma's. Centraal staat de vergelijking tussen de bereidheid om coöperatie te belonen en non-coöperatie te straffen (Hoofdstukken 2–4). Daarnaast heb ik mij gericht op *wat* voor soort (non-)coöperatief gedrag men kan sanctioneren (Hoofdstuk 2), *hoe* men kan sanctioneren (Hoofdstuk 3) en *wanneer* men kan sanctioneren (Hoofdstuk 4). Samengevat toont dit proefschrift aan dat er niet alleen psychologische processen zijn die de bereidheid om te sanctioneren bevorderen, maar dat er ook psychologische processen zijn die de bereidheid om te sanctioneren juist belemmeren (zie Hoofdstuk 5).

■ Belonen van coöperatie versus bestraffen van non-coöperatie

Het gebruik van bestraffingen – in tegenstelling tot het gebruik van beloningen – betekent dat men een ander directe schade toebrengt. Onderzoek naar het *do-no-harm* principe heeft laten zien dat mensen, zelfs als het algehele voordeel opweegt tegen de toegebrachte schade, terughoudend zijn met het toebrengen van schade aan anderen (bv. Baron, 1993, 1995). Toegepast op het gebruik van sancties in sociale dilemma's, suggereert dit principe dat mensen minder bereid zullen zijn om non-coöperatief gedrag te bestraffen dan om coöperatief gedrag te belonen. Een belangrijk doel van dit proefschrift is het testen van deze centrale assumptie.

In alle experimenten gerapporteerd in de empirische hoofdstukken van dit proefschrift (behalve Experiment 3.3) heb ik zowel de bereidheid om coöperatie te belonen als de bereidheid om non-coöperatie te straffen gemeten, terwijl er tegelijkertijd verschillende factoren experimenteel waren gemanipuleerd of van elkaar verschilden tussen de experimenten (bijv., de kosten om te sanctioneren, gepresenteerde feedback, et cetera). De resultaten van deze experimenten laten consequent zien dat mensen minder bereid zijn om non-coöperatie te bestraffen dan om coöperatie te belonen. Wanneer mensen beschikken over zowel bestraffingen als beloningen, hebben ze zelfs de neiging om volledig af te zien van straffen – waardoor non-coöperatie ongestraft blijft – en kiezen ze er liever voor om coöperatie te belonen (Experiment 2.2).

Om verdere steun te verschaffen voor de robuustheid van deze algemene voorkeur voor het gebruik van beloningen boven bestraffingen, heb ik twee meta-analyses uitgevoerd. Deze meta-analyses bevatten niet alleen de data van de experimenten die zijn gerapporteerd in de empirische hoofdstukken van dit proefschrift, maar ook de data van experimenten die hierin niet zijn gerapporteerd (zie Hoofdstuk 5 en Appendices A en B). De resultaten van deze meta-analyses laten zien dat mensen non-coöperatief gedrag minder vaak én in mindere mate bestraffen dan dat ze coöperatief gedrag belonen. Op basis van deze resultaten kan dus geconcludeerd worden dat het type sanctie dat mensen tot hun beschikking hebben (beloning of bestraffing) een zeer bepalende factor is voor de bereidheid om te sanctioneren.

■ De modererende rol van type sociaal dilemma

In Hoofdstuk 2 – over wat voor soort (non-)coöperatief gedrag men kan sanctioneren – heb ik onderzocht of de voorkeur voor het belonen van coöperatie boven het bestraffen van non-coöperatie gemodereerd wordt doordat mensen geconfronteerd worden met ofwel een *public good dilemma*, ofwel een *common resource dilemma* (Molenmaker, De Kwaadsteniet, & Van Dijk, 2014). Ondanks dat beide sociale dilemma's verwijzen naar hetzelfde belangenconflict (d.w.z., gemeenschappelijk belang versus eigenbelang) en als elkaars equivalenten gestructureerd kunnen worden (in termen van uitbetalingen), verschillen ze van elkaar in de wijze waarop het initiële eigendom is verdeeld (Dawes, 1980; Van Dijk & Wilke, 1997, 2000). Terwijl het initiële eigendom in public good dilemma's in het bezit is van mensen zelf (eigenbezit), is het initiële eigendom in common resource dilemma's gelegen in een gemeenschappelijke bron (gemeenschappelijk bezit). In dit hoofdstuk beargumenteer ik

dat mensen keuzegedrag over het *opgeven van eigenbezit* in een public good dilemma minder verwerpelijk en meer lovenswaardig vinden dan keuzegedrag over het *inbreuk maken op gemeenschappelijk bezit* in een common resource dilemma. Daarom verwacht ik dat mensen minder bereid zijn om non-coöperatie te bestraffen en meer bereid zijn om coöperatie te belonen in een public good dilemma dan in een common resource dilemma.

Om dit te onderzoeken, heb ik twee experimenten uitgevoerd waarin participanten het keuzegedrag observeerden van twee personen in een sociaal dilemma taak. De sociaal-dilemma-context in deze taak was ofwel gepresenteerd als public good dilemma, ofwel als common resource dilemma. De feedback die participanten ontvingen was dat de ene persoon een relatief hoog niveau en de andere persoon een relatief laag niveau van coöperatie vertoonde. Wanneer participanten alleen de keuze hadden om te belonen of alleen de keuze hadden om te straffen (Experiment 2.1), straften ze minder vaak en in mindere mate (dan dat ze beloonden) in een public good dilemma dan in een common resource dilemma. Wanneer participanten bovendien de mogelijkheid hadden om te kiezen tussen belonen en straffen (Experiment 2.2), koos de meerderheid er in beide sociale dilemma's voor om te belonen, maar beloonden ze evengoed in grotere mate in een public good dilemma dan in een common resource dilemma. Deze bevindingen bevestigen de opvatting dat de bereidheid om coöperatief gedrag te belonen en non-coöperatief gedrag te bestraffen gemodereerd wordt door het type sociaal dilemma waarmee mensen geconfronteerd worden.

■ De impact van persoonlijke verantwoordelijkheid

In Hoofdstuk 3 – over hoe men kan sanctioneren – heb ik onderzocht of *persoonlijke verantwoordelijkheid* een impact heeft op de (on)bereidheid om te straffen en te belonen (Molenmaker, De Kwaadsteniet, & Van Dijk, 2016). Ondanks dat de voorkeur voor het belonen van coöperatie boven het bestraffen van non-coöperatie geworteld lijkt te zijn in het do-no-harm principe, kan men zich afvragen waarom mensen geneigd zijn zich aan het do-no-harm principe te houden bij het nemen van sanctiebeslissingen. Is dit omdat ze vinden dat een ander überhaupt geen schade mag worden aangedaan, zelfs wanneer het gericht is op iemand die de gemeenschappelijke belangen heeft geschaad, of is het misschien omdat *zijzelf* degenen zijn die de schade toebrengen? Eerder onderzoek naar het do-no-harm principe heeft namelijk laten zien dat de terughoudendheid om iemand te schaden sterker is wanneer mensen direct (in tegenstelling tot indirect) verantwoordelijk zijn voor de verwachte schade (Royzman & Baron, 2002) en wanneer hun handelen (in tegenstelling tot hun niet-handelen) de schade veroorzaakt (Ritov & Baron, 1990). In dit hoofdstuk beargumenteer ik dat de terughoudendheid van mensen om non-coöperatief gedrag te bestraffen een zelf-beperkende neiging is, die afkomstig is van het gevoel van persoonlijke verantwoordelijkheid voor het schaden van anderen. Daarom verwacht ik dat mensen terughoudender zijn om non-coöperatie te bestraffen naarmate ze zich meer persoonlijk verantwoordelijk voelen voor de toegebrachte schade.

Aangezien mensen zich minder verantwoordelijk voelen voor hun handelen en zich agressiever gedragen als lid van een groep dan als individuele beslissers (Jaffe, Shapir, & Yinon, 1981; Mathes & Kahn, 1975; Meier & Hinsz, 2004), heb ik in drie experimenten

het groeperen van individuen als methode gebruikt om de zelf-beperkende invloed van het gevoel van persoonlijke verantwoordelijkheid voor de toegebrachte schade af te zwakken. Participanten voerden als groep een common resource taak uit met ofwel de mogelijkheid om individueel, ofwel de mogelijkheid om gezamenlijk te sanctioneren. Ze observeerden het keuzegedrag van een medegroepslid. Vervolgens stemden ze of er al niet dan gesanctioneerd moest worden (Experiment 3.1) of bepaalden ze hoe groot de sanctie moest zijn (Experimenten 3.2 en 3.3), ofwel individueel, ofwel gezamenlijk. De resultaten laten zien dat non-coöperatie minder vaak en in mindere mate wordt bestraft wanneer mensen als individu in plaats van als groep beslissen, terwijl dergelijke verschillen niet werden gevonden voor het belonen van coöperatie (Experimenten 3.1 en 3.2). Bovendien werd het verzwakkende effect van gedeelde verantwoordelijkheid op de bereidheid om te straffen gemedieerd door de ervoeren persoonlijke verantwoordelijkheid, zelfs wanneer mensen niet ter verantwoording konden worden geroepen voor hun strafbeslissing (Experiment 3.3). De gevoelens van persoonlijke verantwoordelijkheid voor sancties hebben dus een zelf-beperkende invloed op de bereidheid om non-coöperatief gedrag te bestraffen, maar niet op de bereidheid om coöperatief gedrag te belonen.

■ De timing van sanctiebeslissingen

In Hoofdstuk 4 – over wanneer men kan sanctioneren – heb ik onderzocht of de bereidheid om te belonen en te straffen beïnvloed wordt door de *timing van sanctiebeslissingen* (Molenmaker, De Kwaadsteniet, & Van Dijk, 2016). Ondanks dat de beslissing om het gedrag van anderen te sanctioneren op verschillende momenten in de tijd gemaakt kan worden is het ofwel een beslissing *vooraf*, ofwel een beslissing *achteraf*. Een van de meest duidelijke verschillen tussen deze twee tijdsmomenten is dat mensen achteraf beslissen over het sanctioneren van gedrag dat daadwerkelijk heeft plaatsgevonden in het verleden, terwijl mensen vooraf beslissen over het sanctioneren van gedrag dat al dan niet plaats zal vinden in de toekomst. Onderzoek naar het disjunctie-effect heeft laten zien dat wanneer de uitkomst van een bepaalde situatie onbekend is, mensen het vaak nalaten om de implicaties van alle mogelijke uitkomsten te doordenken (Tversky & Shafir, 1992) en minder geneigd zijn om beslissingen te nemen op basis van onzekere informatie dan op basis van zekere informatie (Van Dijk & Zeelenberg, 2006). Een soortgelijk effect kan mogelijk worden waargenomen voor sanctiebeslissingen. In dit hoofdstuk beargumenteer ik dat mensen minder bereid zijn om gedrag te sanctioneren dat mogelijk plaatsvindt in de toekomst dan om gedrag te sanctioneren dat daadwerkelijk heeft plaatsgevonden in het verleden. Ik heb daarom de voorspelling getoetst of mensen minder bereid zijn om te sanctioneren als de sanctiebeslissing voor (in tegenstelling tot na) het gedrag van anderen gemaakt wordt.

In twee experimenten observeerden participanten het keuzegedrag van een andere persoon en hadden ze de mogelijkheid om een beloning uit te delen of de mogelijkheid om een straf uit te delen. De timing van deze sanctiebeslissing was gemanipuleerd door participanten keuzegedrag te tonen dat mogelijk plaats kon vinden in de toekomst (de sanctiebeslissing werd dus voorafgaand gemaakt) of dat echt plaats had gevonden in het

verleden (de sanctiebeslissing werd dus naderhand gemaakt). In lijn met mijn voorspelling waren participanten inderdaad minder vaak en in minder mate bereid om coöperatie te belonen en om non-coöperatie te bestraffen wanneer ze deze sanctiebeslissing vooraf in plaats van achteraf maakten (Experimenten 4.1 en 4.2), ongeacht of dit direct achteraf of op een later moment was (Experimenten 4.2 en 4.3). Kortom, mensen zijn minder bereid om sancties te gebruiken als het keuzegedrag nog niet heeft plaatsgevonden. Bovendien laten de resultaten zien dat de voorkeur voor het gebruik van beloningen boven bestraffingen sterker is wanneer de sanctiebeslissing vooraf gemaakt wordt. Participanten waren met name terughoudend met het vooraf bestraffen van non-coöperatie, terwijl ze erg bereid waren om coöperatie te belonen, zowel vooraf als achteraf. Mijn bevindingen laten dus zien dat de timing van sanctiebeslissingen invloed heeft op de bereidheid om coöperatief gedrag te belonen en non-coöperatief gedrag te bestraffen.

■ Conclusies

In dit proefschrift heb ik niet alleen laten zien dat sanctietype (beloning versus bestraffing) een zeer bepalende factor is voor de (on)bereidheid om te sanctioneren, maar heb ik ook aangetoond dat *wat*, *hoe*, en *wanneer* mensen kunnen belonen of straffen invloed heeft op de bereidheid om deze sancties te gebruiken. Dat wil zeggen, ik heb laten zien dat het type sociaal dilemma waarmee mensen geconfronteerd worden (public good dilemma versus common resource dilemma), de mate van persoonlijke verantwoordelijkheid die mensen voor de sanctie voelen (individuele verantwoordelijkheid versus gezamenlijke verantwoordelijkheid) en de timing van sanctiebeslissingen (vooraf versus achteraf) ook een belangrijke rol spelen bij het gebruik van sancties. Zodoende verschaft dit proefschrift een uitgebreid inzicht in de factoren die bepalend zijn voor de bereidheid om te sanctioneren in sociale dilemma's.

Acknowledgments

(Dankwoord)

■ Acknowledgments (Dankwoord)

In het wielrennen – de mooiste sport die er is – ontvangt alleen die renner een prijs die als eerste over de finish komt of bovenaan in het klassement eindigt. Toch is er doorgaans een heel team van renners voor nodig om de betreffende renner in winnende positie te krijgen. Bij promoveren is dat niet anders. Ook daar pronkt uiteindelijk maar één naam op de voorkant van een proefschrift, terwijl er een heel peloton van mensen nodig was om dit voor elkaar te krijgen. Mijn dank gaat dan ook uit naar iedereen die op enigerwijs heeft geholpen bij de totstandkoming van dit proefschrift. Er zijn echter enkele mensen die ik in het bijzonder wil bedanken.

Beginnend bij mijn (co-)promotoren, die van onmisbare waarde zijn geweest voor zowel de vervaardiging van dit proefschrift, als voor mijn persoonlijke ontwikkeling. *Eric*, jouw wetenschappelijk inzicht en scherpe geest zijn enorm inspirerend en uitdagend. Dank voor het vertrouwen en de vrijheid die je mij gegeven hebt. *Erik*, ik ben vereerd jouw eerste AiO te zijn. Met je kritische blik, oog voor detail en treffende feedback (middels de befaamde kringeltjes) heb je niet alleen dit proefschrift naar een veel hoger niveau getild, maar mij ook geleerd nog scherper en nauwkeuriger te zijn. Verder wil ik al mijn Leidse (oud-)collega's bedanken voor hun behulpzaamheid, betrokkenheid en gezelligheid. Daarbij gaat een speciale dank uit naar mijn paranimfen *Félice* en *Marlon*, het is een voorrecht jullie aan mijn zijde te hebben tijdens de verdediging. I also owe a large amount of gratitude to *Toko Kiyonari*, *Toshio Yamagishi*, and their students, who so warmly welcomed me to Tokyo; visiting you was a memorable and inspiring experience.

Als laatst wil ik al mijn vrienden en (schoon)familie bedanken, niet alleen voor jullie steun en belangstelling de afgelopen jaren, maar ook zeker voor de nodige afleiding en ontspanning. *Pa* en *ma*, ik had me geen fijnere ouders kunnen wensen. Jullie hebben me altijd de kans gegeven om mezelf verder te ontwikkelen, me gestimuleerd om te doen wat ik graag wil doen en me geleerd om niet zomaar op te geven. Daarnaast stonden jullie altijd en op elk mogelijke manier voor me klaar. Dit proefschrift draag ik op aan jullie. *Jouke*, ik bewonder jou om wie je bent en de veerkracht die jij en *Branka* hebben getoond. Als we de leegte zin moeten geven, dan is het de les om zoveel mogelijk uit het leven te halen als erin zit. Dit proefschrift draag ik daarom ook op aan *Eliza*, die de kans niet heeft gekregen haar vleugels verder uit te slaan. Lieve *Tamar*, tot slot, jouw liefde, steun en geduld vervullen mij met geluk. Bedankt dat je mijn leven zoveel leuker maakt. Samen met jou ben ik compleet.

The "Kurt Lewin Institute Dissertation Series" started in 1997. Since 2014 the following dissertations have been published in this series:

2014-01: Marijn Stok: *Eating by the Norm: The Influence of Social Norms on Young People's Eating Behavior*
2014-02: Michèlle Bal: *Making Sense of Injustice: Benign and Derogatory Reactions to Innocent Victims*
2014-03: Nicoletta Dimitrova: *Rethinking errors: How error-handling strategy affects our thoughts and others' thoughts about us*
2014-04: Namkje Koudenburg: *Conversational Flow: The Emergence and Regulation of Solidarity through social interaction*
2014-05: Thomas Sitser: *Predicting sales performance: Strengthening the personality – job performance linkage*
2014-06: Goda Perlaviciute: *Goal-driven evaluations of sustainable products*
2014-07: Said Shafa: In the eyes of others: *The role of honor concerns in explaining and preventing insult-elicited aggression*
2014-08: Félice van Nunspeet: *Neural correlates of the motivation to be moral*
2014-09: Anne Fetsje Sluis: *Towards a virtuous society: Virtues as potential instruments to enhance*
2014-10: Gerdien de Vries: *Pitfalls in the Communication about CO2 Capture and Storage*
2014-11: Thecla Brakel: *The effects of social comparison information on cancer survivors' quality of life: A field-experimental intervention approach*
2014-12: Hans Marien: *Understanding and Motivating Human Control: Outcome and Reward Information in Action*
2014-13: Daniel Alink: *Public Trust: Expectancies, Beliefs, and Behavior*
2014-14: Linda Daphne Muusses: *How Internet use may affect our relationships: Characteristics of Internet use and personal and relational wellbeing*
2014-15: Hillie Aaldering: *Parochial and universal cooperation in intergroup conflicts*
2014-16: Martijn Keizer: *Do norms matter? The role of normative considerations as predictors of pro-environmental behavior*
2015-01: Maartje Elshout: *Vengeance*
2015-02: Seval Gündemir: *The Minority Glass Ceiling Hypothesis: Exploring Reasons and Remedies for the Underrepresentation of Racial-ethnic Minorities in Leadership Positions*
2015-03: Dagmar Beudeker: *On regulatory focus and performance in organizational environments*
2015-04: Charlotte Koot: *Making up your mind about a complex technology: An investigation into factors that help or hinder the achievement of cognitive closure about CCS*
2015-05: Marco van Bommel: *The Reputable Bystander: The Role of Reputation in Activating or Deactivating Bystanders*
2015-06: Kira O. McCabe: *The Role of Personality in the Pursuit of Context-Specific Goals*
2015-07: Wiebren Jansen: *Social inclusion in diverse work settings*
2015-08: Xiaoqian Li: *As time goes by: Studies on the subjective perception of the speed by which time passes*

2015-09: Aukje Verhoeven: *Facilitating food-related planning. Applying metacognition, cue-monitoring, and implementation intentions*

2015-10: Jasper de Groot: *Chemosignaling Emotions: What a Smell can Tell*

2015-11: Hedy Greijdanus: *Intragroup Communication in Intergroup Conflict: Influences on Social Perception and Cognition*

2015-12: Bart de Vos: *Communicating Anger and Contempt in Intergroup Conflict: Exploring their Relational Functions*

2015-13: Gerdientje Danner: *Psychological Availability. How work experiences spill over into daily family interactions*

2015-14: Hannah Nohlen: *Solving ambivalence in context. The experience and resolution of attitudinal ambivalence*

2015-15: Stacey Sanders: *Unearthing the Moral Emotive Compass: Exploring the Paths to (Un)Ethical Leadership*

2015-16: Marc Heerdink: *Regulating deviance with emotions: Emotional expressions as signals of acceptance and rejection*

2015-17: Danny Taufik: *"Can you feel it" The role of feelings in explaining pro-environmental behavior*

2015-18: Sarah Elbert: *Auditory information and its parameters in health persuasion. The development of a tailored smartphone application to support behavior change*

2016-01: Anna van 't Veer: *Effortless morality — cognitive and affective processes in deception and its detection*

2016-02: Thijs Bouman: *Threat by association: How distant events can affect local intergroup relations*

2016-03: Tim Theeboom: *Workplace coaching: Processes and effects*

2016-04: Sabine Strofer: *Deceptive intent: Physiological reactions in different interpersonal contexts*

2016-05: Caspar van Lissa: *Exercising Empathy: The Role of Adolescents' Developing Empathy in Conflicts with Parents*

2016-06: Marlon Mooijman: *On the determinants and consequences of punishment goals: The role of power, distrust, and rule compliance*

2016-07: Niels van Doesum: *Social mindfulness*

2016-08: Leonie Venhoeven: *A look on the bright side of an environmentally-friendly life: Whether and why acting environmentally-friendly can contribute to well-being*

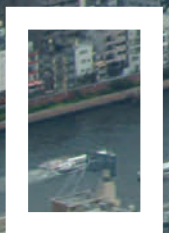
2016-09: Florien Cramwinckel: *The social dynamics of morality*

2016-10: Junhui Wu: *Understanding Human Cooperation: The Psychology of Gossip, Reputation, and Life History*

2016-11: Elise C. Seip: *Desire for vengeance. An emotion-based approach to revenge*

2016-12: Welmer E. Molenmaker: *The (un)willingness to reward cooperation and punish non-cooperation*

The central theme of this dissertation is the (un)willingness to reward cooperation and punish non-cooperation. Whereas rewards and punishments can be effective means to enhance cooperation in social dilemmas, a prerequisite for any effect of sanctions is that people are willing to administer them. In the present work, I shed more light on this important – yet long neglected – topic. The aim is twofold: (1) identifying determinants of the use of sanctions in social dilemmas, and (2) testing the central proposition that people are not as willing to punish non-cooperative choice behavior as they are willing to reward cooperative choice behavior. The results of this dissertation show that the type of sanction people have at their disposal – either reward or punishment – is as primary determinant of the willingness to sanction. In addition to sanction type, I argue and demonstrate that the type of social dilemma people face (Public good dilemma versus Common resource dilemma), the extent of personal responsibility people have for the sanction (Individual responsibility versus Joint responsibility), and the timing of the sanctioning decision (Beforehand versus Afterwards) are also important determinants of the (un)willingness to sanction in social dilemmas. These findings reveal that there are not only psychological processes at play that foster sanctioning, but also psychological processes that hamper sanctioning. By taking a closer look at people's (un)willingness to incur the costs of rewarding cooperative choice behavior and punishing non-cooperative choice behavior, this work thus provides a more comprehensive view of the potential that sanctions can have to solve social dilemmas in the real world.



kurtlewinsinstituut

Dissertatiereeks
Kurt Lewin Instituut 2016-12