



Universiteit  
Leiden  
The Netherlands

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

**Citation**

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

Version: Not Applicable (or Unknown)

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

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

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

Cover Page



Universiteit Leiden



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

**Author:** Molenmaker, Welmer E.

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

**Issue Date:** 2017-01-19





# Chapter 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 they 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 possibly 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.

