

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/40167> holds various files of this Leiden University dissertation

Author: Mooijman, Marlon

Title: On the determinants and consequences of punishment goals : power, distrust, and rule compliance

Issue Date: 2016-06-14

On the Determinants and Consequences of Punishment Goals:
Power, Distrust, and Rule Compliance

Marlon Mooijman



**On the Determinants and Consequences of Punishment Goals:
Power, Distrust, and Rule Compliance**

© Marlon Mooijman

Leiden, 2016

ISBN: 978-94-028-0162-0

Printed by: Ipskamp Printing

On the Determinants and Consequences of Punishment Goals:
Power, Distrust, and Rule Compliance

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 dinsdag 14 juni 2016
klokke 15:00 uur
door
Marlon Mooijman
geboren te Monst
1988

Promotores:

Prof. Dr. E. van Dijk	Universiteit Leiden
Prof. Dr. W. W. van Dijk	Universiteit Leiden
Prof. Dr. N. Ellemers	Universiteit Utrecht

Promotiecommissie:

Prof. Dr. E. C. M. Crone	Universiteit Leiden
Prof. Dr. C. K. W. de Dreu	Universiteit van Amsterdam
Prof. Dr. K. van den Bos	Universiteit Utrecht
Prof. Dr. B. Beersma	Vrije Universiteit Amsterdam

Contents

Chapter 1	Introduction	7
Chapter 2	Why leaders punish: A power perspective Studies 2.1a–2.4c	15
Chapter 3	Power, distrust, and the motivation to maintain power Experiments 3.1–3.4	45
Chapter 4	On sanction effectiveness: How and why sanction-goal justifications affect rule compliance Experiments 4.1–4.3	59
Chapter 5	Summary, discussion, and conclusions	83
	References	93
	Dutch summary (<i>Samenvatting</i>)	105
	Acknowledgements (<i>Dankwoord</i>)	109
	Curriculum Vitae (<i>CV</i>)	110
	KLI dissertation series	111

Chapter 1

Introduction

Promoting rule compliance is of vital importance for societal leaders such as managers, policy makers, and politicians. People's willingness to comply with rules and guidelines tends to be positively associated with an organization's financial success, in part because rules tend to promote collective as opposed to individual interests (e.g., coming on time, behaving ethically, prudent use of resources, paying taxes; Akintoye & Tashie, 2013; Parks, Joireman, & Van Lange, 2013). Although the importance of rule compliance can hardly be overstated, rule compliance is not always easy to achieve. Enron, for example, went bankrupt as a consequence of unethical behavior and poor oversight from regulatory agencies (Raul, 2002). Similarly, governments' success in promoting compliance with tax rules varies widely between countries. Tax evasion is estimated to cost Greece up to 25% of its GDP, Germany up to 13.5% of its GDP, and the United States of America up to 8.6% of its GDP (Schneider, 2008; Tax Justice Network, 2011).

Although leaders can promote rule compliance through rewarding rule-abiding behavior (Thaler & Sunstein, 2011), they typically punish rule-breaking behavior with fines, penalties, and prison sentences (Kirchler, Kogler, & Muehlbacher, 2014). For example, judges sentence citizens to jail, managers fire employees for not sticking to rules, and government officials fine businesses for evading taxes. But what makes some leaders more, and other leaders less, effective at promoting rule compliance with such punishments? In this dissertation, I focus on the determinants and consequences of leaders' punishment *goals*. I investigate how and why leaders rely on certain punishment goals, and how and why leaders' reliance on such punishment goals affects punishment effectiveness.

I argue that—with increasing power over others—leaders rely more on punishment goals that are actually suboptimal in promoting rule compliance. I propose that power fosters a distrustful mindset towards people, which increases reliance on deterrence—but not just deserts as a punishment goal. Using deterrence—as opposed to just deserts—as a justification for punishments, in turn, decreases people's willingness to comply with rules because they feel distrusted by the leader. Although power may thus increase leaders' reliance on punishments to deter rule-breaking behavior, paradoxically, this may at times decrease the effectiveness of the punishment.

Punishment Goals

What do leaders aim to achieve with punishment? Scholars have typically classified punishment goals into goals that aim to *deter* future rule-breaking behavior (Bentham, 1789/1988; Hobbes, 1651/1988; Kirchler et al., 2014; Nagin, 1998) or goals that aim to give people their *just deserts* (i.e., give offenders their deserved punishment, thereby achieving justice; Darley, 2009; Kant, 1780/1961).

Although both goals may co-occur, they have different aims. A deterrence goal aims to deter future rule breaking from potential rule breakers and, as such, is prospective rather than retroactive. When having such a goal, leaders should be primarily concerned with deterring future rule breaking instead of achieving retributive justice through punishing (past) rule breakers proportionate to their crime. This approach is most often associated with legal philosopher Jeremy Bentham (1789/1988) who argued that “general prevention ought to be the chief end of punishment, as its real justification” (p. 396). In contrast, a just-deserts goal aims to punish past rule breakers proportionately (i.e., achieve balance between crime and punishment), regardless of the punishment’s ability to deter future rule breaking. A just-deserts goal is thus retroactive rather than prospective. Having this goal, leaders should be primarily concerned with achieving retributive justice through punishing rule breakers proportionate to their crime instead of preventing future rule breaking. This approach is generally associated with moral philosopher Immanuel Kant (1780/1961) who argued that “punishment can never be administered merely as a means to promoting another good” and that “punishment should be pronounced over all criminals proportionate to their internal wickedness” (p. 397).

There are two general reasons why I am interested in understanding the determinants and consequences of leaders’ punishment goals. First, it provides an explanation for why leaders use punishments. What do leaders want to achieve when they impose (potentially life-changing) punishments on others? Second, it provides an explanation for leaders’ (in)effectiveness in promoting rule compliance with punishments. Are the goals that leaders use to justify their punishments beneficial or detrimental for people’s willingness to comply with rules? In this dissertation, I thus investigate punishment goals from both the top-down perspective of the leader (the determinants) and the bottom-up perspective of the people (the consequences). In the following paragraphs, I will first address the role of power as a determinant of punishment goals. I will then address the consequences of punishment goals for people’s willingness to comply with rules. Last, I will integrate the proposed determinants and consequences into an overarching framework with one underlying psychological explanation.

Determinants of punishment goals

Previous research on the psychology of punishment suggests that punishments are typically guided by a just-deserts goal rather than a deterrence goal (Darley, 2009). That is, punishments tend to be aimed at giving rule breakers what they deserve instead of preventing future rule-breaking behavior from these rule breakers (Aharoni & Fridlund, 2011; Carlsmith, 2006; Carlsmith, Darley, & Robinson, 2002; Keller, Oswald, Stucki, & Gollwitzer, 2010). In a recent experiment, for instance, participants still desired a rule breaker to be punished even when

the rule breaker (or other potential rule breakers) could never be deterred from breaking rules (Crocket, Ozdemir, & Fehr, 2014). When assigning punishments, these participants were also shown to be sensitive to factors that are relevant for just-deserts theory (e.g., extenuating circumstances) while being insensitive to factors that are relevant for the deterrence of rule breaking (e.g., publicity of punishment; Carlsmith et al., 2002). Just-deserts punishments are in part preferred over deterrence punishments because giving rule breakers their just deserts through punishment is (emotionally) satisfying (de Quervain et al., 2006; Strobel et al., 2011; Wenzel, Okimoto, Feather, & Platow, 2008) and driven by emotions such as anger (Nelissen & Zeelenberg, 2009; Seip, Van Dijk, & Rotteveel, 2014). This has led some scholars to conclude that: “the just-deserts goal is the psychological foundation of citizens’ desire to punish transgressions” (Darley, 2009, p. 1).

However, this conclusion seems premature in light of the limited amount of research that has been conducted on the psychological determinants of deterrence. The majority of research on punishment goals has focused on the determinants of just-deserts driven punishments, such as anger (Nelissen & Zeelenberg, 2011; Seip et al., 2014), concerns about group members’ status (Okimoto & Wenzel, 2011), and victims’ perceived desires (Gromet, Okimoto, Wenzel, & Darley, 2012), but has left the psychological determinants of deterrence relatively unaddressed. This is surprising since governmental authorities and organizational managers are well documented to use punishments to prevent-and-deter citizens and employees from breaking rules (Butterfield, Trevino, Wade, & Ball, 2005). A major concern among European tax agencies, for example, is to deter citizens from evading taxes with harsh fines and penalties (Kirchler et al., 2014). Moreover, philosophers and legal scholars have long stressed the importance of deterring people from rule breaking with punishments (Bentham, 1789/1988; Hobbes, 1651/1988). But what, then, determines reliance on deterrence as opposed to just deserts as a punishment goal?

Power

Power can be broadly defined as asymmetric control over valuable resources (Anderson & Brion 2014; Magee & Galinsky, 2008). As a result, power entails the ability to reward or punish others by granting or withholding valuable resources (Keltner, Anderson, & Gruenfeld, 2003). To help organizations achieve rule compliance, leaders are often given such power. Government officials can set the height of fines that citizens have to pay when they evade taxes, university professors can control whether and when lower-ranked academics are given tenure, and managers can control employees’ salaries or decide about bonuses. Having power can have pervasive psychological effects on people’s perceptions, emotions, and behaviors (Blader & Chen, 2012; Mooijman, Van Dijk, Ellemers, & Van Dijk, 2015). For instance, studies have shown

that control over resources can benefit power holders—they tend to be less dependent on others (Fiske, 1993; Lee & Tiedens, 2001) and therefore enjoy greater freedom to act according to their personal desires (Galinsky, Gruenfeld, & Magee, 2003; Lammers, Stoker, Jordan, Pollman, & Stapel, 2011). It is easier for power holders to disregard social norms (Keltner, Gruenfeld, & Anderson, 2003) and focus on accomplishing their own goals (Guinote, 2007a; Maner & Mead, 2010), instead of having to devote attention to what others think and feel (Goodwin, Gubin, Fiske, & Yzerbyt, 2000). Power therefore tends to boost people's self-esteem (Wojciszke & Struzynska-Kujalowicz, 2007) and leads them to express more positive—approach related—emotions (e.g., amusement and happiness) and less negative—inhibition related—emotions (e.g., embarrassment and shame; Keltner et al., 2003). As Rucker, Galinsky, and Duboi (2012, p. 353) noted: “the cumulative evidence suggests that power is an omnipresent force whose tentacles reach out and grasp nearly every situation to guide and ultimately shape human behavior”.

Recent research suggests that having power can impact how severely people believe others should be punished for rule-breaking acts (Van Prooijen, Coffeng, & Vermeer, 2014; Wiltermuth & Flynn, 2013). Van Prooijen et al. (2014) demonstrated that power holders punish rule breakers more harshly because power tends to increase people's tendency to perceive rule breaking acts as diagnostic of the rule breaker's personality. Moreover, Wiltermuth and Flynn (2013) demonstrated that power holders punish rule breakers more harshly than non-power holders because power increases the moral clarity with which people perceive morally right acts from morally wrong ones (i.e., power holders view rule-breaking acts as more immoral than non-power holders). Although these previous studies on power and punishments are informative, they do not address the *goals* that power holders strive for when imposing punishments on others. Indeed, what do such power holders aim to achieve with their punishments? The first aim of this dissertation is to examine how and why power affects leaders' punishment goals. Since punishment goals are a vital source of punishment behavior, understanding the effects of power on punishment goals can provide a fundamental understanding of leaders' subsequent punishment behavior (e.g., the type of punishment they tend to use). More specifically, I propose that—through fostering a distrustful mindset towards others—power increases reliance on deterrence, but not on just deserts, as a punishment goal. Power is thus predicted to be an important determinant of leaders' punishment goals through affecting their distrust towards others. The research—consisting of eight experimental studies and a field study—testing this hypothesis is reported in Chapter 2.

Power and distrust

Considering the theorized importance of distrust in explaining why power increases reliance on deterrence—but not just deserts—as a punishment goal, I also examined why having power increases distrust in others. Distrust entails expecting others to break rules that promote cooperation (such as a declaration of business expenses; Mulder, Van Dijk, De Cremer, & Wilke, 2006). Understanding the power-distrust link is important because it gives further insight into the psychological mechanism that underlies leaders' reliance on deterrence as a punishment goal. Powerful leaders tend to be motivated to maintain their power (because of its many benefits; see Fehr, Herz, & Wilkening, 2013) and distrusting others prepares them to counteract behaviors aimed at undermining their power (e.g., through expecting others to break rules; Kramer, 1999). For instance, managers who trust their employees to comply with organizational rules are more likely to fail to take the appropriate actions required to prevent their employees from breaking rules, thereby potentially undermining their own power position. Distrusting their employees to comply with organizational rules, however, increases the likelihood that a manager engages in acts that prevent employees from breaking rules (e.g., introducing more monitoring; Lount & Pettit, 2012). I propose that the power-distrust link is in part explained by leaders' motivation to maintain their power over others. Decreasing the motivation to maintain power may thus attenuate the extent to which power fosters distrust. Because power is hypothesized to increase reliance on deterrence as a punishment goal through increasing distrust, these predictions suggest that powerful leaders' reliance on deterrence as a punishment goal can in part be explained by their motivation to maintain power over others. The research—consisting of three experimental studies—addressing this prediction is reported in Chapter 3.

Consequences of punishment goals

In addition to examining power as a determinant of punishment goals, the aim of this dissertation is to examine the consequences of punishment goals. When leaders justify their punishment behavior as an attempt to deter people or provide people their just deserts, does this affect people's willingness to comply with rules? Inspired by the literature on (perceived) interpersonal justice, I examine whether punishment goals may have direct consequences for people's willingness to comply with rules. Previous research has demonstrated that rule compliance is in part determined by how people feel treated by their leader (i.e., interpersonal justice; Tyler & Lind, 1992). For instance, people's willingness to comply with rules decreases when a leader is perceived to act unjustly through pursuing his or her own interest instead of others' interests (De Cremer & Van Knippenberg, 2002). This is in part because of the “social contract” between people and their leaders; people are willing to comply with leaders' rules and

grant them power as long as leaders ensure that justice is done (i.e., people who cooperate are proportionately rewarded and people who break rules are proportionately punished; see Plato 380 BC/1992). These effects of perceived interpersonal treatment are often independent from the outcome that people (expect to) receive from leaders (Cropanzano, Gillian, & Gilliland, 2007). In other words, it can be the subjective treatment *itself*, regardless of the objective monetary outcome that one expects to receive, that plays a role in people's willingness to comply with rules.

In this dissertation, I examine whether leaders' use of punishment goals as a punishment justification affects the subsequent effectiveness of this punishment. I propose that justifying the use of punishments as an attempt to deter, compared to an attempt to achieve justice, decreases the extent to which a punishment is effective at promoting rule compliance. I further propose that this is explained by people feeling more distrusted by a leader that justifies punishments as an attempt to deter compared to achieve justice. Indeed, distrust may not only underlie reliance on deterrence as a punishment goal, it may also directly influence people's willingness to comply with rules because being distrusted by one's leader may seem unjust and unwarranted. Examining these consequences of punishment-goal justifications allows an understanding of how leaders' reliance on punishment goals (as described briefly above, and in more detail in Chapter 2) affects the subsequent effectiveness of the punishment. This provides an integrative understanding of the consequences of leaders' punishment goals. Is a leader's inclination to rely on certain punishment goals beneficial or detrimental for his or her ability to promote rule compliance through the use of such punishments? The research—consisting of four experiments—addressing the consequences of punishment goals for punishment effectiveness is reported in Chapter 4.

Integrating determinants and consequences of punishment goals

Central to the above-mentioned predictions about the determinants and consequences of punishment goals is the role of distrust. Distrust entails an expectation of malicious intent from others (integrity-based distrust; Kramer, 1999). Distrust thus means expecting others to break rules that promote cooperation (such as rules regarding tax payments or declarations of business expenses; Mulder, Van Dijk, De Cremer, & Wilke, 2006). As described briefly above, distrust is predicted to both underlie leaders' reliance on deterrence as a punishment goal, and decrease the effectiveness of punishments that are justified as an attempt to deter people from rule breaking. Distrust, in other words, is predicted to mediate the effect of power on punishment goals (the determinant) and to mediate the effect of using deterrence as punishment justification on rule compliance (the consequence). More specifically, I predict leaders' *distrust towards people* to mediate

the positive relationship between power and use of deterrence as a punishment goal, and I predict people *feeling distrusted by the leader* to mediate the negative relationship between leaders' use of deterrence as punishment justification and rule compliance. Leaders' distrust towards people may, in other words, be determined by their power, and consequently inferred from their punishment-goal justifications by people, thereby undermining people's willingness to comply with rules. Distrust may thus explain why leaders rely on deterrence as a punishment goal *and* why this reliance on deterrence may decrease punishment effectiveness.

Summary

In sum, in the current dissertation I examine, (a) how and why the power that leaders have affects their distrust in others, (b) how this distrust affects leaders' reliance on punishment goals, (c) how leaders' use of punishment goals as a justification affects the extent to which people feel distrusted by their leader, and (d) how feeling distrusted by their leader affects people's willingness to comply with this leader's rules. Together, addressing these four questions facilitates both a top-down understanding of how and why leaders are inclined to use punishments and a bottom-up understanding of how and why people are willing to comply with rules set by leaders that rely on deterrence or just deserts goal for punishments. In the remainder of this dissertation, I will outline these questions in more detail and provide empirical support for the current set of predictions. I conclude this dissertation by discussing the theoretical and practical implications of my analysis, and the limitations of my research, in Chapter 5. Because Chapters 2, 3, and 4 were written as independent articles that can be read separately, readers may notice some overlap between the chapters. Please also note that throughout the dissertation, the words punishments and sanctions, and leaders and authorities, are used interchangeably.

Chapter 2

Why Leaders Punish: A Power Perspective

This chapter is published as: Mooijman, M., Van Dijk, W., Ellemers, N., & Van Dijk, E. (2015). Why leaders punish: A power perspective. *Journal of Personality and Social Psychology*, *109*, 75–89.

Leaders of all sorts frequently install sanctions to induce compliance with rules (Parks et al., 2013). Managers publicly reprimand employees for bad behavior and policy-makers implement policies with a mandatory minimum for punishments. Yet, punishment is not always effective and can even be counterproductive (Gneezy & Rustichini, 2000; Van Dijk, Mulder, & De Kwaadsteniet, 2014). If punishment can be problematic, what do leaders then aim to achieve with their punishments? In the current paper, we argue that leaders' power affects what they aim to achieve with punishments. Specifically, we propose that power increases the tendency to punish to deter rule breaking instead of give rule breakers their just deserts, and that this power-deterrence link can be explained by power fostering a distrustful mindset.

Addressing this question is important because the managers and policy-makers who design and implement punishments have, by definition, a form of power—that is, they possess and can allocate critical resources. Yet, little is known about how power affects how and why they punish. Answering this fundamental question about power and punishments can provide important insights into the punitive practices of leaders such as managers and policy-makers, and in what ways their punishments are effective in inducing rule-abiding behavior (e.g., Arvey & Ivancevich, 1980; Ball, Trevino, & Sims, 1994). Gaining more insight into how and why power affects punitive practices may therefore help organizations and governmental institutions design more effective sanction policies.

Power and Punishment Motives

Power can be defined as asymmetric control over critical resources (Magee & Galinsky, 2008); that is, power entails the capacity to control the outcomes of others. Power provides the benefit of controlling one's own and others' monetary (e.g., salary), social (e.g., inclusion) or physical (e.g., housing) resources. Importantly, having power goes hand in hand with the ability to implement punishments such as salary cuts, fines, or prison-sentences.

The motives that guide such punishments are typically classified into deterrence motives or just desert motives (Carlsmith et al., 2002). The deterrence motive of punishment reflects the desire to deter people from engaging in uncooperative, rule-breaking behavior (see Bentham, 1789/1988 and Hobbes, 1651/1988). For example, managers can punish employees publicly to set an example and deter future rule breaking by other employees. In contrast, the just deserts motive is mainly concerned with giving offenders their deserved punishment (i.e., their “just deserts”; Darley, 2009). Punishments based on this motive are generally proportionate to the severity of the offence but insensitive to the likelihood of deterring uncooperative, rule-breaking behavior (Carlsmith et al., 2002; Kant, 1780/1952; Vidmar & Miller, 1980). Thus, the deterrence and just deserts motives reflect different punishment goals. The deterrence motive aims to deter

future rule-breaking behavior from all individuals through the installment and implementation of punishments. Instead, the just deserts motive aims to achieve retributive justice by retroactively and proportionally punishing individual rule breakers.

Research has suggested that people generally prefer punishments that restore retributive justice by giving offenders their just deserts (Aharoni & Fridlund, 2011; Carlsmith, 2006; Carlsmith et al., 2002; Darley, 2009). We argue, however, that power increases reliance on deterrence, but not just deserts, as a punishment motive. We propose that power increases distrust, which in turn increases the reliance on deterrence as a punishment motive and increases the use of punishments that are suitable as a deterrent—for instance, punishments that are public or have a mandatory minimum.

Power and Distrust

Asymmetric control over resources benefits power holders. For example, managers can award themselves salaries and bonuses that are bigger than those typically given to employees (Kipnis, Castel, Gergen, & Mauch, 1976). This is possible because power holders are less dependent on others (Fiske, 1993; Lee & Tiedens, 2001) and therefore enjoy greater freedom to act according to their personal desires (Galinsky, et al., 2003; Lammers et al., 2011). Indeed, it is easier for power holders to disregard social norms (Keltner et al., 2003) and focus on accomplishing their own goals (Guinote, 2007a; Maner & Mead, 2010) instead of having to devote attention to what others think and feel (Goodwin et al., 2000). Power, in other words, is beneficial for those who hold it. Power holders are therefore motivated to protect their beneficial position (Fehr et al., 2013; Maner, Gailliot, Butz, & Peruche, 2007) and prevent others from obtaining access to their power (Case & Maner, 2014; Lammers & Stapel, 2011; Mead & Maner, 2012).

One way in which power holders can lose resource control is to be too trusting of others. Trust entails an expectation of benign behavior from others—that is, behavior that furthers the interests of the trustee (Colquitt, Scott, & LePine, 2007; Lewicki, Tomlinson, & Gillespie, 2006). Trust further entails an expectation of the other party's benign intent. Trust can therefore be distinguished from being assured that others take your interests into account. In contrast to trust, assurance is not based on an expectation of benign intent, but on the idea that others further your interest because this is in their self-interest (Yamagishi & Yamagishi, 1994). Thus, trusting others means expecting others to take your interests into account, even when this is not in the other's self-interest. We similarly operationalize trust as an expectation of cooperative intent, which we define as expecting others to further your interest in interest-conflicting situations (cf. Balliet & Van Lange, 2013; Mulder et al., 2006). For instance, in the classic trust game, trusting

others means giving away control over one's money based on the expectation that others will cooperate and give control over this money back at the end of the game. When one's trust is met, one keeps control over the money. However, when one's trust is abused, one loses control over the money to the other (Berg, Dickhaut, & McCabe, 1995). Trusting others to cooperate can then be considered a form of potential resource sharing (i.e., power sharing) that opens up a power holder for loss of resource control (Kramer, 1999; Mayer, Davis, & Schoorman, 1995; Zand, 1997). Indeed, expecting benign intentions from others has been shown to be an important antecedent of opening oneself up to these others' potentially exploitative behaviors (Lount & Pettit, 2012). Managers that trust their employees to stick to organizational rules may fail to take the appropriate actions required to prevent employees' rule-breaking acts that undermine this manager's power position. Because power holders such as managers are motivated to stay in power, we predict that power fosters an individual's expectation that others cannot be trusted to cooperate (i.e., choose other-interest over self-interest).

This prediction rests on the notion that power facilitates goal-directed behavior (Guinote, 2007b) and that an important goal of people in power positions is to maintain their privileged position within the power hierarchy (Williams, 2014; Willis & Guinote, 2011). Research has demonstrated that people exhibit a strong tendency to retain obtained power (Fehr et al., 2013)—a tendency that strongly increases with individuals' increasing degree of loss aversion. This suggests that power is especially desirable once you obtain it and that power holders are strongly motivated to retain this desirable power (i.e., power-endowment effect). Since trusting people's inclination to cooperate increases vulnerability to potentially losing resources and thus power to others, we predict that power holders are likely to distrust others to cooperate. These predictions depart from previous research on how power can make people less averse to losing resources (Anderson & Galinsky, 2006; Inesi, 2010). In contrast to this previous research, the current theorizing revolves around power holders potentially losing resources to others. We suggest that a loss of resources may be threatening to power holders when others are able to gain such resources and hereby undermine a power holder's resource control. As such, we hypothesize that power fosters distrust as a resource-protection strategy.

Distrust and Punishment Motives

We have hypothesized so far that power holders become more distrustful due to their power. We further propose that this increased distrust in turn increases the reliance on deterrence, rather than just deserts, as a punishment motive. We base this prediction on the notion that uncooperative, rule-breaking behavior is expected when distrust is high (Mayer et al., 1995; Mulder et al., 2006). Distrust may therefore increase the belief that people need to be

deterred with punishments from breaking rules that promote cooperation. In contrast, punishments that aim to give offenders their just deserts are more concerned with restoring retributive justice through proportionate punishments. Punitive responses that give offenders their just deserts have been robustly linked to factors that have little to do with decreasing the future likelihood of rule breaking (e.g., moral outrage and concern for social cohesion, see Carlsmith et al., 2002; Tyler & Boeckman, 1997). We therefore predict that distrust increases the reliance on deterrence as a punishment motive whereas it does not affect the reliance on the just deserts motive. This conjecture corresponds with the view of seventeenth-century philosopher Thomas Hobbes (1651/1996), who in his book *Leviathan* argued that—because people generally care little about others—punishments should deter people from breaking rules that promote cooperation. In sum, we propose that power increases reliance on such Hobbesian philosophy; that is, an individual will be more inclined to distrust others and therefore rely more on deterrence as a punishment motive.

Punishment Preferences

How might such an increased reliance on deterrence as a punishment motive affect punishment preferences? We address two ways in which punitive practices are likely to be affected by the increased reliance on deterrence rather than just deserts—that is, public punishments and punishments with a mandatory minimum.

Power holders can use public punishment to deter rule-breaking behavior. Microsoft, for example, prominently displayed its lawsuits against piracy on its website, probably in an attempt to deter piracy through reminding everyone of the enforced rule and ensuing punishment (Xiao & Houser, 2011). Publicly communicating the names of rule-breakers may serve as a similar deterrent; government-officials sometimes implement policies that publicly communicate the names of offenders as a way to show the consequences of rule breaking and hereby deter future rule breaking from others. For instance, a Texas judge recently convicted a drunk driver to publicly wear a sign stating that he killed someone while driving drunk (Texas Judge, 2012). Similarly, both India and parts of Australia recently started publicly naming and shaming offenders in an attempt to deter crime (see Langlois, 2012). Setting a mandatory punishment minimum provides another way in which rule breaking can be deterred (Gabor & Crutcher, 2002). Setting a guaranteed minimum for a punishment may deter rule-breaking behavior through increasing the certainty of punishment severity. In other words, by increasing the extent to which potential offenders can be sure of receiving a fixed punishment upon breaking a rule.

From the perspective of deterrence theory, private punishments serve less purpose than public punishments, since private punishments have less potential to deter rule breaking

(Carlsmith et al., 2002; Nagin, 1998). Similarly, raising a mandatory minimum for punishment guarantees that offenders are less likely to think they can “get away” with breaking rules (Nagin, 1998). Just-deserts theory makes less clear predictions about these types of punishments—giving offenders their just deserts revolves around proportionate punishments instead of revenge, which is the desire to make the offender suffer (often *dis*proportionately; see Gerber & Jackson, 2012 for empirical evidence of the distinction between just deserts and revenge). Instead of wanting the offender to suffer disproportionately, the just-deserts motive aims to achieve proportionality between the rule-breaking act and punishment (i.e., balance). Crucially—for just-deserts theory—private punishments are often sufficient because there is little need for the punishment to publicly deter others and private punishments can give offenders their deserved punishment without making them suffer disproportionately (Carlsmith et al., 2002). Moreover, for just-deserts theory, mandatory minimum punishments are often unnecessary. Just-deserts theory predicts that rule-breaking acts have to be punished on the basis of the characteristics of the act, such as extenuating circumstances—no mandatory minimum is required as some offences might go unpunished. Thus, in just-deserts theory there is generally less purpose for public punishments and mandatory minimum punishments than in deterrence theory. Deterrence theory proposes that one should frequently use public punishments and mandatory minimum punishments, as they can be efficient in preventing future rule breaking.

The majority of states in the USA implement some form of mandatory minimum for punishments, even when the public considers such policies harsh and unjust (see “Mandatory Minimum”, 2007). Furthermore, punishments that are implemented publicly are frequently considered to be ineffective or to violate offenders’ privacy (Jabour, 2013). Nevertheless, power holders frequently install and implement such punishments against the wishes and preferences of their employees or the general public. Our theorizing provides an explanation for this phenomenon: Power increases distrust in others and therefore increases the reliance on deterrence as a punishment motive. This increased reliance on deterrence, in turn, facilitates the use of public punishments and punishments with a mandatory minimum.

Research overview

We tested the hypothesized effect of power on deterrence through distrust across nine studies. The predictions and the specific studies that test every prediction are depicted in Figure 2.1. Because every proposed relationship is new and lacks empirical validation from previous research, we decided to first test every prediction separately (i.e., power increasing distrust, distrust increasing deterrence, and power increasing deterrence) in a correlational and causal

manner before testing the full mediation model. This gave us the advantage to increase generalizability and replicability, and to establish the causality of the effects.

First, we describe the six studies that test the hypothesized links between power and distrust, distrust and deterrence, and power and deterrence, respectively. Next, we describe three studies that test the mediating role of distrust. Across all these studies, we used different instantiations of power—from measuring a general sense of power (Studies 2.1a and 2.4a), to power primes (Studies 2.3b and 2.4b) and structural manipulations of power (Studies 2.1b, 2.3a and 2.4c). We also measured (Studies 2.1a, 2.1b, 2.2a, 2.4a, and 2.4b) and manipulated distrust (Studies 2.2b and 2.4c), and measured punishment motives across a number of different situations—from the problem of tax fraud (Studies 2.4a and 2.4b) and academic plagiarism (Study 2.3b) to cooperation in social dilemmas (Studies 2.1b, 2.2b, and 2.4c) and less specific situations (Studies 2.1a, 2.2a, and 2.3a). We also included cross-cultural samples from the USA (Studies 2.1a, 2.2a, 2.4a, and 2.4b), the Netherlands (Studies 2.1b, 2.2b, 2.3b, and 2.4c) and a sample of Western European countries included in the European Value Survey (Study 2.3a).

To demonstrate that the effect of power on deterrence is mediated by distrust specifically, we also assessed the role of victim identification and the importance of group interest. Previous research has demonstrated that power can increase a focus on the welfare of the entire group relative to the individual (Magee & Smith, 2013) and decrease identification with victims (Galinsky et al., 2006; Van Kleef et al., 2008). To rule out that the effect of power on deterrence is explained by one of these variables instead of distrust, we assessed them in Study 2.4c. Moreover, in order to address the extent to which self-reported punishment motives translate into actual punitive practices (see Carlsmith, 2008), we measured the extent to which power holders use public punishments (Study 2.3b) and mandatory minimum sentences (Study 2.4a) as a way to deter rule breaking.

In our studies, we excluded participants who misreported their power position or failed to follow instructions with regard to writing an autobiographical story about power. Consistent with the recommendations of Simmons, Nelson, and Simonsohn (2011), we made sure that every condition had at least 20 participants, although most studies reported have considerably more participants per condition (i.e., more than 40; cf. Simmons, Nelson and Simonsohn, 2013). All items in the studies could be answered on seven-point scales (1 = *disagree completely*; 7 = *agree completely*), unless stated otherwise and all participants provided informed consent and were debriefed after completing the study.

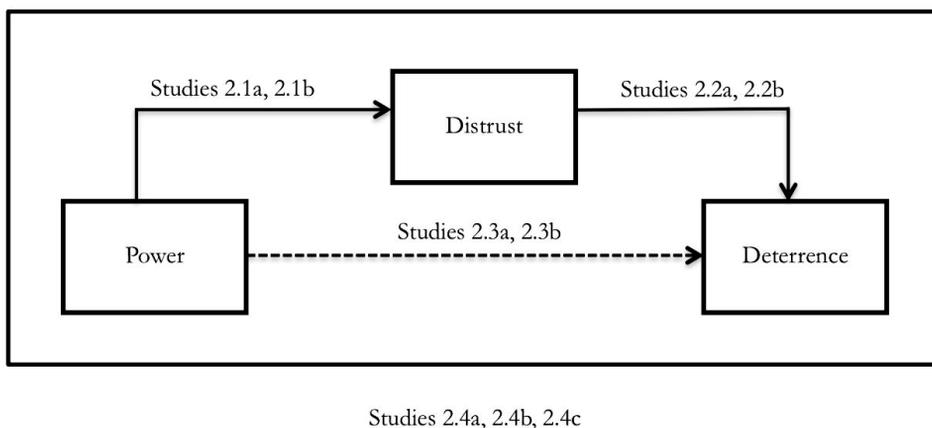


Figure 2.1. Schematic depiction of all predictions. The study numbers indicate which studies test which prediction. The surrounding line with accompanying studies represents tests of the full model.

Studies 2.1a and 2.1b: Power and Distrust

We tested the hypothesized power-distrust relationship in two studies. Specifically, we tested in Study 2.1a whether chronically experiencing power (i.e., trait power) is associated with a general distrust towards others, and we tested in Study 2.1b whether occupying a structural position of power causes people to become more distrustful.

Study 2.1a Method

Participants, design, and procedure. A total of 165 participants (75 males, $M_{\text{age}} = 32.03$ years, $SD_{\text{age}} = 10.88$) were recruited from the Mechanical Turk website (see for a discussion of this platform as a research tool, Buhrmester, Kwang, & Gosling, 2011) and were paid \$0.50 for their participation.

Generalized sense of power. Participants completed the generalized sense of power scale taken from Anderson and Galinsky (2006, see also Anderson, John, & Keltner, 2011). The scale assesses the extent to which people experience power in their everyday lives. Specifically, the scale is comprised of four items measuring high power (e.g., “In my relationships with others, I think I have a great deal of power”; “If I want to, I get to make the decisions”) and four items measuring low power (e.g., “Even if I voice them, my opinions have little sway”; “My ideas and opinions are often ignored”). The low-power items were reverse coded and averaged with the high-power items to form the generalized sense of power scale (cf. Anderson et al., 2011; $M = 4.78$, $SD = 0.85$; Cronbach’s $\alpha = .74$). To disentangle the high- versus low-power subscales and

indicate if the experience of power or powerlessness drives the scale's effect, we also computed separate high- and low-power scales with the items measuring high and low power, respectively. A factor analysis (PCA) confirmed that the high- versus low-power items loaded onto two different components (Eigenvalue = 2.99 and 2.34, respectively) without cross-loadings and with a total of 66.55% variance explained for. Both the low-power subscale ($\alpha = .87$) and high-power subscale ($\alpha = .76$) had sufficient reliability and did not correlate significantly ($r = .05, p = .49$).

General distrust. Participants then completed a seven-item scale measuring their general trust towards others ($M = 4.95, SD = 0.98; \alpha = .81$). We adapted and extended items from the general trust scale (Yamagishi, 1986). Items included, "In general, I believe that people will break the rules whenever they can get away with it", "When it really comes down to it, most people only think about what is good for themselves but forget what is good for others", and "Most people can be trusted to cooperate with others" (reverse-coded).

Results and discussion

A regression analysis in which trust was regressed on power showed that, as predicted, the generalized sense of power scale was negatively associated with trusting others ($\beta = -.17, t(165) = 2.21, p = .028$). We repeated this regression analysis with the four-item scale that measured high power and the four-item scale that measured low power, respectively. Results demonstrated that the high-power subscale was negatively associated with trust ($\beta = -.28, t(165) = 3.71, p < .001$), whereas the low-power subscale was unrelated to trust ($\beta = .021, t(165) = .79, p = .79$). This provides preliminary support for our hypothesis that high power decreases trust. To examine whether power *causes* people to become more distrustful, we conducted Study 2.1b.

Study 2.1b Method

Participants, design, and procedure. Ninety-eight Dutch university students (68 females, $M_{age} = 22.53$ years, $SD_{age} = 3.30$) participated in exchange for €1 and were assigned to one of two conditions (power position vs. neutral position).

Role-based power manipulation. All participants were told that they would play in a "business simulation" in which they could earn extra money by working together with other participants. Depending upon condition, participants were then informed that they were assigned to a powerful manager position or they received no information about their power position (cf. Lammers, Galinsky, Gordijn, & Otten, 2012). As managers, participants were led to believe that they had control over how money would be allocated to the other participants. The other participants (described as "workers") allegedly were unaware of this resource control of the participant.

Interpersonal trust. The business simulation was in reality a social dilemma game. The social dilemma comprised of a collective resource of 30 chips, with each chip worth \$0.10. Every individual group member could take between 0 and 10 chips out of this collective resource. Consistent with previous social dilemma games, money that was left in the collective resource was multiplied by two and divided equally amongst the group members. This way, taking money out of the collective resource only benefitted the individual group members (Van Lange et al., 2013). Previous research has demonstrated that there exists a strong rule (e.g., norm) to cooperate in such dilemma situations (Fehr & Fischbacher, 2004; Fehr & Gächter, 2002). Participants assigned to the managerial position had full control over the money in the common resource. This manipulation is consistent with the definition of power as control over valuable resources (Magee & Galinsky, 2008). Nothing was mentioned about a power position to participants in the control condition.

All participants then indicated to what extent they trusted the other group members to cooperate on a three-item scale. Items included “I think group members cannot be trusted to cooperate”, “Group members will weigh their self-interest more than the interests of others”, and “Group members cannot be trusted to maximize the group’s outcomes” ($\alpha = .64$).

Power manipulation check. Participants indicated whether they occupied a managerial position (yes, no) and how powerful they felt on one item (i.e., “I feel powerful). Six participants wrongly reported their power position and were therefore left out of the analyses. Including or excluding these six participants from the analyses did not change the results of these analyses. Confirming the validity of the power manipulation for those who correctly reported their power position, participants in the high-power condition ($M = 4.07, SD = 1.88$) felt more powerful than participants in the control condition ($M = 3.06, SD = 1.76; t[90] = 2.64, p = .010, d = 0.56$).

Results and discussion

Participants in the high-power condition placed less trust in other participants’ inclination to cooperate ($M = 4.94, SD = 1.14$) than participants in the control condition ($M = 5.55, SD = 0.87; t[90] = 2.77, p = .007, d = 0.58$). Thus, taken together, Studies 2.1a and 2.1b provide converging evidence for the hypothesis that power increases distrust.

Studies 2.2a and 2.2b: Distrust and Deterrence

In the following two studies, we tested the second hypothesized relationship. More specifically, in Study 2.2a we measured participants’ chronic distrustfulness towards others and their desired punishment motives. We expected that participants who are chronically distrustful rely more on deterrence, but not just deserts, as a punishment motive. In Study 2.2b, we

experimentally manipulated distrust and subsequently measured participants' punishment motives. We expected the experimental manipulation of distrust to cause participants to rely more on deterrence, but not just deserts, as a punishment motive.

Study 2.2a Method

Participants, design, and procedure. A total of 125 participants (75 males, $M_{\text{age}} = 32.03$ years, $SD_{\text{age}} = 10.88$) were recruited from the Mechanical Turk website and were paid \$0.50 for their participation.

General trust. Participants completed the same general distrust scale as in Study 2.1a ($M = 4.95$, $SD = 1.01$; $\alpha = .83$).

Punishment motives. Participants indicated their preference for deterrence ($M = 4.82$, $SD = 1.27$; $\alpha = .82$) on a four-item scale (e.g., "In general, punishments should be aimed at deterring crime"; "Punishments should be aimed at preventing crime"). Participants also indicated their preference for just deserts ($M = 4.59$, $SD = 1.33$; $\alpha = .85$) on a four-item scale (e.g., "In general, punishments should give offenders their just deserts"; "Punishments should make offenders pay for their wrongdoings"). These items measuring deterrence and just deserts were adapted from past work examining punishment motives (De Keijser, Van der Leeden, & Jackson, 2002). The deterrence and just deserts motives were moderately correlated ($r = .42$, $p < .001$) and a factor analysis resulted in a two-factor solution with deterrence (Eigenvalue = 3.83) and just deserts (Eigenvalue = 1.65) items loading on two different factors with one cross loading and a total of 68.39% variance accounted for.

Results and discussion

Two separate linear regression analyses regressing deterrence and just deserts on distrust showed that general distrust was positively associated with deterrence ($\beta = .34$, $t(124) = 4.00$, $p < .001$) but not with just deserts ($\beta = .11$, $t(123) = 1.17$, $p = .24$). This provides initial support for the hypothesis that distrust increases reliance on deterrence as a punishment motive. To examine whether distrust *causes* reliance on deterrence, Study 2.2b was conducted.

Study 2.2b Method

Participants, design, and procedure. Eighty-five Dutch university students (71 females, $M_{\text{age}} = 22.40$ years, $SD_{\text{age}} = 5.01$) participated in exchange for €1 and were assigned to one of two conditions (distrust vs. trust).

Trust manipulation. All participants were told that they would participate in a social dilemma (cf. Study 2.1b but nothing was mentioned about a business simulation and no power

position was manipulated). Importantly, it was explained to participants in the distrust condition that “previous social dilemma experiments demonstrated that participants generally cannot be trusted to act cooperatively”, whereas participants in the trust condition were told that “previous social dilemma experiments demonstrated that participants generally can be trusted to act cooperatively”.

Punishment motives. It was then explained that fines are sometimes introduced in social dilemma situations. Such fines decrease the amount of money that participants who act uncooperatively can earn in the social dilemma game. Participants had to indicate whether they thought that in the current social dilemma such fines should deter, or give others their just deserts. More specifically, participants indicated their preference for deterrence ($\alpha = .83$) and just deserts ($\alpha = .87$) on two six-item scales. Sample items for deterrence included, “Fines are needed to prevent group members from acting uncooperatively”, “Fines should deter uncooperative behavior”, and “Fines should act as a deterrent to group members”. Sample items for just deserts included, “Fines are needed to give group members who act uncooperatively their deserved punishment”, “Fines should make uncooperative group members pay for their behavior”, and “Fines should give uncooperative group members their just deserts”. The deterrence and just deserts motives were moderately correlated ($r = .45, p < .001$) and a factor analysis resulted in a two-factor solution with deterrence (Eigenvalue = 5.19) and just deserts (Eigenvalue = 2.47) items loading on two different factors with one cross loading and a total of 63.89% variance accounted for.

Trust manipulation check. Lastly, participants indicated to what extent they distrusted their group members in the social dilemma (i.e., “group members cannot be trusted to act cooperatively”). Confirming the validity of the trust manipulation, participants distrusted group members’ inclination to cooperate more in the distrust ($M = 5.05, SD = 1.39$) compared to the trust condition ($M = 4.37, SD = 1.68; t[83] = 2.02, p = .047, d = .44$).

Results and discussion

Distrustful participants were more likely to indicate that fines should deter uncooperative behavior ($M = 5.35, SD = 0.69$) compared to trusting participants ($M = 4.67, SD = 1.35; t[83] = 2.96, p = .005, d = 0.65$). However, the experimental manipulation of distrust did not affect the extent to which participants relied on just deserts as a punishment motive ($M_{distrust} = 4.35, SD_{distrust} = 1.23; M_{trust} = 4.38, SD_{trust} = 1.32; t[83] = .11, p = .91, d = 0.02$). In sum, Studies 2.2a and 2.2b provide evidence that people increasingly rely on deterrence, but not just deserts, as a punishment motive when they believe others cannot be trusted.

Studies 2.3a and 2.3b: Power and Deterrence

In the third series of studies, we examined the hypothesized power-deterrence relationship. In Study 2.3a, we used data from the European Value Survey to test whether occupying a real-life power position was associated with relying on deterrence as a motive for punishment. We then investigated in Study 2.3b whether experimentally manipulating power through a priming procedure impacts reliance on deterrence, and we tested how reliance on deterrence affects the use of public punishments.

Study 2.3a Method

Participants, design, and procedure. A total of 6,147 participants (3,274 females, $M_{\text{age}} = 42.49$ years, $SD_{\text{age}} = 18.36$) who participated in the European Value Survey from 1981 were analyzed (EVS, 2011). We selected this value survey because it was the only value survey we were able to trace that measured people's power positions *and* punishment motives. This dataset consists of survey-data that was collected through face-to-face interviews in over ten European countries such as the UK and France, totaling more than 12,000 interviews. Importantly, this dataset gave us the advantage of testing the hypothesis that power is associated with more reliance on deterrence as a punishment motive in a representative, large-scale European sample.

Managerial position. Power was determined by categorizing participants into those who occupied a power role and those who did not. Specifically, we coded those who occupied a managerial position as 1 and those who did not occupy a managerial position (i.e., skilled-workers, clerks, and unskilled workers) as 0. We reasoned that occupying a managerial position entails controlling and managing subordinates—that is, having power. In total, 14.5 % of participants occupied a managerial position.

Punishment motives. Punishment motives were assessed with the question “When a person is sentenced by a court of law, what should be the main aim of imprisonment? Participants could choose between four options: a) to re-educate the prisoner, b) to make those who have done wrong pay for it, c) to protect other citizens, and d) to act as a deterrent to others. We analyzed these response options in two ways. First, we created a variable in which we coded the just deserts option (i.e., option b) as 0 and the deterrence option (i.e., option d) as 1 and left the other two options (a and c) out of the analyses. However, aiming to protect citizens through imprisonment can be argued to be similar to deterring rule breakers, since both these motives are focused on preventing further rule breaking (Carlsmith et al., 2002; Darley, Carlsmith, & Robinson, 2000). We therefore created another variable in which the just deserts option (i.e., option b) was again coded as 0 but this time we combined the deterrence option (i.e., option d) with the protect-citizens option (i.e., option c) and coded this as 1.

Results and discussion

Results are presented in Table 2.1. In agreement with our predictions, participants' punishment motives were affected by whether they occupied a managerial position or not. More specifically, for both contrasts (just deserts versus deterrence and just deserts versus deterrence/protecting citizens) managers were more likely to indicate that imprisonment should act as a deterrent compared to non-managers ($\chi^2 = [1, n = 4196] = 31.42, p < .001, d = 0.17$; $\chi^2 = [1, n = 6147] = 72.06, p < .001, d = 0.22$, respectively). This provides preliminary support for the hypothesis that power is associated with more reliance on deterrence. To further demonstrate that it is power¹ that increases reliance on deterrence, we manipulated power in Study 2.3b. Moreover, Study 2.3b also tested whether power increases the use of public punishments as a means to deter rule breaking. This would demonstrate how reliance on deterrence could directly affect power holders' punitive practices.

Table 2.1. Managers and non-managers' preferred imprisonment motive for Study 2.3a.

	Managerial Position	
	Yes	No
Contrast 1:		
Just Deserts	342 (66%)	2280 (77%)
Deterrence	179 (34%)	846 (23%)
Contrast 2:		
Just Deserts	343 (38.5%)	2828 (53.8%)
Deterrence/Protect	549 (61.5%)	2427 (46.2%)

¹The European value survey also contained a question relating to how much "control participants had over their lives" (ranging from 1 "none at all" to 10 "a great deal"). Importantly, managers felt like they had more control over their lives ($M = 6.75, SD = 2.06$) than non-managers ($M = 6.43, SD = 2.28$; $t(11696) = 6.15, p < .001, d = 0.12$). Moreover, this control-over-life item was positively associated with relying on deterrence as a punishment motive ($B = .05, t(9181) = 3.80, p < .001$). This provides additional evidence for our usage of managerial position as a proxy of power and further corroborates the link between power and deterrence.

Study 2.3b Method

Participants, design, and procedure. Forty-eight participants (32 females, $M_{age} = 20.10$ years, $SD_{age} = 1.68$) at a Dutch University participated in exchange for €1 and were randomly assigned to one of the two power conditions (high vs. low power).

Power manipulation. Depending on condition, participants recalled a powerful or powerless event in their lives (cf. Galinsky et al., 2003). Specifically, participants in the high [low] power condition were asked to, “Please recall a particular incident in which you had power over another individual or individuals [someone else had power over you]. By power, we mean a situation in which you [someone else] controlled the ability of another person or persons to get something they wanted, or were in a position to evaluate those individuals [controlled your ability to get something you wanted or was in the situation to evaluate you]. Please describe this situation in which you had [no] power—what happened, how you felt, etc”.

Punishment motives and public punishments. Participants then read a small excerpt about plagiarism. This excerpt was allegedly part of an unrelated experiment. The excerpt was introduced as a survey deployed by the university. It explained that the university had recently adopted a system that automatically tracks down plagiarism in student assignments. It also explained that the university desired the input from students to align their sanction policies with the opinions of students. In reality, the university adopted a system to track down plagiarism, but never deployed such a survey. This was explained to participants during the debriefing at the end of the experiment.

After reading the excerpt, participants indicated on a five-item scale to what extent the specific punishments that offenders (i.e., students that commit plagiarism) receive should be publicly made known to other students. Sample items included, “The punishment should be made known to other students” and “Punishments for plagiarism should be kept private” (reverse-coded; $\alpha = .85$). Participants also indicated on a five-item scale the extent to which the names of offenders should be publicly made known to others students. Sample items included, “The names of offenders should be made known to other students” and “The names of offenders should be kept private” (reverse-coded; $\alpha = .87$). The public-punishment and public-naming scales were moderately correlated ($r = .30, p = .040$) and a factor analysis resulted in a three-factor solution with public naming (Eigenvalue = 4.55) and public punishment (Eigenvalue = 2.51) items loading on two different factors with four items also loading on a third factor (Eigenvalue = 1.05), and a total of 78.49% variance explained for.

Participants then indicated their preference for deterrence as a punishment motive on two items (i.e., “Overall, punishments for plagiarism should be aimed at deterring all students from committing plagiarism” and “Overall, punishments for plagiarism should prevent all

students from committing plagiarism”; Spearman-Brown $\rho = .66$). They also indicated their preference for just deserts as a punishment motive on two items (i.e., “Overall, punishments for plagiarism should be aimed at giving offenders a punishment he/she deserves” and “Overall, punishments for plagiarism should give offenders a proportionate punishment”; $\rho = .97$). The deterrence and just-deserts motives were correlated ($r = .31, p = .030$) and a factor analysis resulted in a two-factor solution with deterrence (Eigenvalue = 2.32) and just deserts (Eigenvalue = 1.12) items loading on two different factors without cross-loadings and a total of 86.02% variance accounted for. Note that for the reliability of two-item scales, the Spearman-Brown statistic is preferred over both the Pearson correlation and Cronbach’s alpha (Eisinga, Te Grotenhuis, & Penzer, 2013).

Power manipulation check. Lastly, participants indicated on a fourteen-item scale to what extent they felt powerful (e.g., “I feel powerful”; “I feel like I can control others”, $\alpha = .94$). Importantly, participants felt more powerful in the high-power condition ($M = 5.29, SD = 0.59$) compared to the low-power condition ($M = 2.70, SD = 0.71; t[46] = 13.57, p < .001, d = 4.00$), hereby confirming the validity of the power manipulation.

Results and discussion

Participants in the high-power condition were more favorable towards implementing public punishments ($M = 3.60, SD = 1.52$) than participants in the low-power condition ($M = 2.61, SD = 1.59; t[46] = 2.20, p = .033, d = 0.65$). They were also more favorable towards publicly naming the offenders ($M = 2.41, SD = 1.42$) than participants in the low-power condition ($M = 1.50, SD = 0.69; t[46] = 2.77, p = .007, d = 0.81$). Importantly, with regard to the underlying motive for such punishments, participants in the high-power condition considered the deterrence of plagiarism more important ($M = 2.97, SD = 1.64$) than participants in the low-power condition ($M = 1.98, SD = 1.09; t[46] = 2.45, p = .018, d = 0.72$). No effect of power, however, was found on the just-deserts motive for punishment. Just deserts was considered of equal importance by participants in the high-power ($M = 2.90, SD = 1.93$) as those in the low-power condition ($M = 2.73, SD = 1.55; t[46] = .34, p = .73, d = 0.10$).

Lastly, two bootstrap analyses with 5,000 resamples indicated that powerful participants’ reliance on deterrence as a punishment motive mediated their increased favorability towards using public punishments (95% CI = [0.022, 0.55]) and public naming (95% CI = [0.029, 0.47]) as a punitive policy measure. More specifically, the significant effect of power on public punishment ($\beta = .31, p = .033$) and public naming ($\beta = .39, p = .007$) decreased to non-significance ($\beta = .19, p = .19$) and marginal significance ($\beta = .23, p = .08$) after deterrence (which was predicted by condition [$\beta = .34, p = .018$] and in itself significantly predicted public

punishment [$\beta = .42, p = .003$] and public naming [$\beta = .53, p < .001$]) was added to the model. Although just deserts predicted public punishment ($\beta = .49, p < .001$) and public naming ($\beta = .49, p < .001$), it did not mediate the effect of power on public punishments (95% CI = [-0.17, 0.31]) and public naming (95% CI = [-0.10, 0.27]). In other words, power increased participants' favorability towards public punishments as a way to deter plagiarism.

These results support the hypothesis that power increases reliance on deterrence. Merely inviting participants to think of a situation in which they experienced power, as opposed to powerlessness, was sufficient to elicit the hypothesized effects. Interestingly, results further demonstrated that reliance on deterrence directly affected power holders' punitive practices by increasing power holders' favorability towards public punishments.

Studies 2.4a, 2.4b, and 2.4c: Testing the Mediation Model

Studies 2.4a, 2.4b, and 2.4c were conducted to test the full theoretical model. Study 2.4a tested the mediation model while measuring people's generalized sense of power, whereas Study 2.4b experimentally manipulated power with a priming procedure. Study 2.4c investigated the power-deterrence link when trust is directly manipulated, while ruling out that power-deterrence link can be explained by power increasing a focus on the group or decreasing identification with those who are potentially victimized by punishment. In addition, Study 2.4a tested the idea that deterrence increases the endorsement of mandatory minimum punishments.

Study 2.4a Method

Participants, design, and procedure. A total of 146 participants (88 males, $M_{age} = 33.01$ years, $SD_{age} = 11.79$) were recruited from the Mechanical Turk website and were paid \$0.50 for their participation.

Generalized sense of power. To measure participants' generalized sense of power, they first completed six items of the Anderson and Galinsky's Generalized Sense of Power scale used previously in Study 2a (2006; $M = 3.40, SD = 1.21; \alpha = .60$). Consistent with Study 2a, we also computed a low-power and high-power subscale. A factor analysis (PCA) confirmed a two-factor solution with high- (Eigenvalue = 3.33) and low-power (Eigenvalue = 1.32) items loading on two different factors with no cross loadings and a total of 77.49% variance accounted for. Both the low-power subscale ($\alpha = .74$) and high-power subscale ($\alpha = .89$) had sufficient reliability and correlated moderately ($r = -.34$).

General distrust. Next, participants completed the eight-item scale used previously in Study 2a ($M = 4.79, SD = 1.09; \alpha = .87$).

Punishment motives. Participants then read a small excerpt about the enforcement of tax laws in the USA. The excerpt explained that taxes are used to benefit society as a whole but that individual taxpayers might sometimes be tempted to evade taxes. Participants indicated their preference for deterrence ($M = 4.53, SD = 1.40; \alpha = .87$) as a motive for the punishment of tax frauds on a four-item scale. They also indicated their preference for just deserts ($M = 5.54, SD = 1.03; \alpha = .71$) as a motive for the punishment of tax frauds on a three-item scale. Sample items included, “punishments should give offenders their just deserts”, “punishments should make tax frauds pay for their behavior”, “punishments should deter tax fraud”, and “punishments should prevent tax payers from committing tax fraud”. The deterrence and just deserts scales were marginally correlated ($r = .15, p = .060$) but loaded onto two different components (Eigenvalue = 3.03 for deterrence and 1.85 for just deserts) without cross-loadings and a total of 69.65% variance accounted for. Lastly, participants indicated to what extent there should be set a mandatory minimum for sentences and fines that tax frauds receive ($M = 4.60, SD = 1.72$).

Results and discussion

A set of linear regression analyses regressing all variables on the generalized sense of power showed that power was positively associated with generalized distrust ($\beta = .33, t[144] = 4.21, p < .001$), relying on deterrence as a motive for punishments ($\beta = .37, t[144] = 4.51, p < .001$), and a preference for mandatory minimum punishment ($\beta = .16, t[144] = 1.98, p = .050$). Similarly, analyses regressing all variables on generalized distrust showed that distrust was positively associated with relying on deterrence as a punishment motive ($\beta = .35, t[144] = 4.51, p < .001$) and being favorable towards mandatory minimum punishment ($\beta = .33, t[144] = 4.17, p < .001$). However, analyses regressing just desert on power and distrust showed that both power and distrust were not reliably associated with just deserts ($\beta = -.10, t[144] = -1.26, p = .21; \beta = .14, t[144] = 1.73, p = .09$, respectively). Interestingly, we replicated these effects with the high-power subscale. This subscale was associated with generalized distrust ($\beta = .26, t[144] = 3.21, p = .002$), relying on deterrence as a motive for punishments ($\beta = .29, t[144] = 3.69, p < .001$) and mandatory minimum punishments ($\beta = .16, t[144] = 1.99, p = .048$) but not with just deserts ($\beta = -.09, t[144] = 1.07, p = .29$). The low-power subscale, however, showed no association with distrust ($\beta = .12, t[144] = 1.50, p = .14$), deterrence ($\beta = .13, t[144] = 1.55, p = .12$), just deserts ($\beta = .02, t[144] = -.28, p = .78$) or mandatory minimum punishments ($\beta = .02, t[144] = .28, p = .78$). Lastly, deterrence but not just deserts was associated with mandatory minimum punishments ($\beta = .15, t[144] = 1.82, p = .070; \beta = .05, t[144] = .65, p = .52$, respectively).

Mediation analyses. To test whether distrust mediated the effect of power on the deterrence motive and preferences for mandatory minimum punishment, we conducted a set of regression analyses using coefficients from 5,000 resamples bootstrap samples. For the endorsement of deterrence, results showed that the confidence interval for the indirect effect of distrust did not contain zero (95% CI = [0.03, 0.26]). More specifically, adding distrust as a mediator slightly decreased the effect of power on deterrence (from $\beta = .37$, $t[144] = 4.78$, $p < .001$ to $\beta = .29$, $t[144] = 3.56$, $p < .001$). Similarly, a 5,000 resamples bootstrap analysis showed that for the endorsement of a mandatory punishment minimum, the confidence interval for the indirect effect of distrust did not contain zero (95% CI = [0.03, 0.26]). More specifically, adding distrust as a mediator decreased the effect of power on mandatory minimum punishment (from $\beta = .16$, $t[144] = 1.98$, $p = .050$ to $\beta = .06$, $t[144] = .73$, $p = .47$). Thus, the results indicate that distrust mediated the effect of a generalized sense of power on the endorsement of deterrence and mandatory minimum punishments.

Next, we explored to what extent power predicted the implementation of punishments with a mandatory minimum through the mediating effect of distrust predicting deterrence (i.e., power \rightarrow distrust \rightarrow deterrence \rightarrow mandatory punishment minimum). By doing so, we could further validate our reasoning by showing that experiencing power was positively associated with the implementation of punishments with a mandatory minimum through the relationship between distrust and deterrence. We tested this multiple-step model using a macro from Hayes, Preacher and Myers (2011). The overall mediation (95% CI = [0.09, 0.39]), separate effect of distrust (95% CI = [0.023, 0.20]) and the combined effect of distrust through deterrence (95% CI = [0.01, 0.09]) were significant using 5,000 resamples and a 95% bias-corrected confidence interval. The combined effect of distrust through deterrence could be replicated with the high-power subscale (95% CI = [0.01, 0.15]) but not the low-power subscale (95% CI = [-0.02, 0.01]). Thus, people who feel relatively powerful are favorable towards implementing mandatory punishment minimums as a way to deter those they do not trust. In Study 2.4b, we manipulated power to provide support for the causal link between power, distrust and deterrence.

Study 2.4b Method

Participants, design, and procedure. Seventy-three participants from the USA (45 males, $M_{age} = 32.18$ years, $SD_{age} = 12.09$) were recruited from the Mechanical Turk website and participated for \$0.50.

Power manipulation. Participants were informed that they would be completing a study on their attitudes toward tax fraud and were randomly assigned to one of the two power

conditions (high vs. low power). Participants had to write an autobiographical story about themselves experiencing high or low power (cf. Study 2.3b). Five participants completed the study but were not included in the analyses because they either wrote about events unrelated to power or they failed to write more than one word. Importantly, one coder, blind to conditions and hypotheses, rated each story that revolved around power using a seven-point scale measuring how much power the participant reported having. Participants described themselves as more powerful in the high-power stories ($M = 5.41$, $SD = 0.99$) than low-power stories ($M = 2.50$, $SD = 1.02$; $t[66] = 11.94$, $p < .001$, $d = 2.94$), hereby confirming the validity of the power manipulation.

Distrust. Next, participants read an excerpt about tax fraud in the United States (cf. Study 2.4a). Participants' distrust towards taxpayers' tendency to pay their taxes was measured with four items ($\alpha = .80$). Sample items included, "I believe that taxpayers will evade taxes whenever they can get away with it" and "When it really comes down to it, most taxpayers will be tempted to commit tax fraud".

Punishment motives. Subsequently, participants indicated their preference for deterrence on one item (e.g., "A punishment should primarily be aimed at deterring taxpayers") as a motive for punishment. They also indicated their preference for just deserts on one item (e.g., "A punishment should primarily give tax payers their just deserts") as a motive for punishment.

Results and discussion

Participants in the high-power condition were more distrustful towards taxpayers ($M = 4.84$, $SD = 0.98$) than participants in the low-power condition ($M = 4.07$, $SD = 1.56$; $t[66] = 2.40$, $p = .019$, $d = 0.59$). Importantly, with regard to the underlying motive for punishments, participants in the high-power condition considered the deterrence of tax fraud more important ($M = 4.74$, $SD = 1.11$) than participants in the low-power condition ($M = 4.00$, $SD = 1.65$; $t[66] = 2.16$, $p = .035$, $d = 0.53$). The just-deserts motive for punishment, however, was considered less important by participants in the high-power condition ($M = 5.24$, $SD = 1.54$) as compared to participants in the low-power condition ($M = 6.32$, $SD = 1.01$; $t[66] = 3.45$, $p = .001$, $d = 0.85$).

Mediation analysis. To test whether distrust mediated the effect of power on the deterrence motive, we conducted a set of regression analyses using coefficients from 5,000 resamples bootstrap samples. Results showed that the 95% confidence interval for the indirect effect of distrust did not contain zero [0.02, 0.35, see Figure 2.2]. Distrust did not mediate the effect of power on just deserts [-0.13, 0.03]. In Study 2.4c, we aimed to replicate these findings while experimentally manipulating distrust. We also aimed to rule out that the effect of power on

deterrence is caused by lower levels of victim identification or an increased focus on the welfare of the group relative to the individual group member.

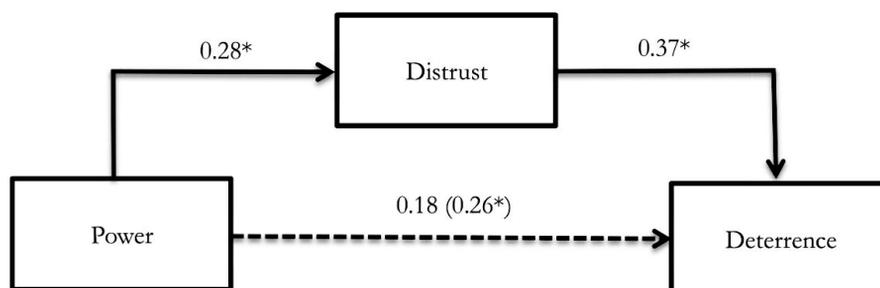


Figure 2.2. Mediation analysis for Study 2.4b. High power is coded as 1, low power as -1. Beta weights are unstandardized, * $p < .05$.

Study 2.4c Method

Participants, design, and procedure. Ninety-four participants from a Dutch university (60 females, $M_{age} = 22.19$ years, $SD_{age} = 2.75$) participated for €1 and were assigned to a 2 (power: high vs. control) \times 2 (trust: distrust vs. control) between-participants design.

Trust manipulation. All participants were told that they would participate in a social dilemma (cf. Study 2.2b). Importantly, it was explained to participants in the distrust condition that “previous social dilemma experiments demonstrated that participants generally cannot be trusted to act cooperatively”, whereas participants in the control condition did not receive this information.

Role-based power manipulation. Depending upon condition, participants were informed that they were assigned to a powerful manager position or they received no information about their power position (cf. Study 2.1b). As managers, the participants were led to believe that they had control over how money would be allocated to other group members. That is, participants assigned to the managerial position had full control over the money in the common resource. Participants were told that the other group members were unaware of the participant’s power.

Punishment motives. Participants were led to believe that they were in this game together with three other participants (cf. Studies 2.1b and 2.2b). Again, there was a collective resource of 30 chips, with each chip worth €0.10 and every individual group member could take

between 0 and 10 chips out of this collective resource. Chips that were not taken by individual group members were multiplied by two and at the end of the game evenly distributed among the group members.

The actual social dilemma was then played. Player A took 7 out of 10, player B took 1 out of 10 and Player C took 10 out of 10 coins out of the collective resource. In other words, it was clear that Player A and C acted uncooperatively. Participants had to indicate whether they thought that Player A and C should be punished to deter selfish behavior, or give them their just deserts. More specifically, participants indicated their preference for deterrence on a four-item scale ($\alpha = .90$) and just deserts on a three-item scale ($\alpha = .92$). Sample items for deterrence included, “They should be punished to prevent uncooperative behavior” and “Uncooperative group members should be fined to deter them”. Sample items for just deserts included, “They should be punished to give them their just deserts” and “Uncooperative group members should pay for what they have done”. The deterrence and just-deserts motives were significantly correlated ($r = .51, p < .001$) and a factor analysis resulted in a two-factor solution with deterrence (Eigenvalue = 4.29) and just deserts (Eigenvalue = 1.41) items loading on two different factors without cross loadings and a total of 81.42% variance accounted for.

Possible alternative explanations. To rule out some alternative explanations for the power-deterrence relationship, we also measured the extent to which participants focused on the welfare of the entire group versus the individual and the extent to which they identified with group members who would be potentially punished. Both constructs could potentially explain the power-deterrence link through power increasing people’s focus on the welfare of the group (cf. Magee & Smith, 2013), or power decreasing identification with the victim (cf. Van Kleef et al., 2008). More specifically, we measured both group orientation ($\alpha = .78$; e.g., “At this moment, I consider the group interest more important than the interest of the individual group member”; “Group-interest should guide decisions to a stronger extent than self-interest”) and victim identification ($\alpha = .95$; e.g., “I feel sorry for group members who will be punished”; “I identify with uncooperative group members”) with four items.

Trust manipulation check. Before group members made their decision, participants also indicated to what extent the group members in the social dilemma could be trusted to act cooperatively on a three-item scale ($\alpha = .89$; e.g., “Group members can generally be trusted to act cooperatively”; “I trust group members to value the collective interest over their own”; “Group members will leave most chips in the common pool”). We conducted an analysis of variance (ANOVA) with power and trust and their interaction as independent variables, and trust as dependent variable. This analysis yielded a (marginally significant) main effect of power ($F[1, 90] = 3.19, p = .077, \eta^2_p = .03$) and a (marginally significant) main effect of trust $F[1, 90] = 3.00, p =$

.087, $\eta_p^2 = .03$). Although no significant interaction effect emerged ($F[1, 90] = 0.95, p = .33, \eta_p^2 = .01$) the means do demonstrate that in the control condition power holders ($M = 4.01, SD = 1.62$) distrusted marginal significantly more than non-power holders ($M = 3.22, SD = 1.78; t[45] = 1.92, p = .060, d = 0.57$) whereas power holders distrusted as much ($M = 4.25, SD = 1.38$) as non-power holders in the distrust condition ($M = 4.03, SD = 1.49; t[45] = 0.53, p = .59, d = 0.15$).

Power manipulation check. At the end of the experiment, participants indicated to what extent they had a position of power on a three-item scale ($\alpha = .91$; “I hold a position of power”, “I feel like I can control others’ outcomes”, and “I feel powerful”). Confirming the validity of the power manipulation, participants felt more powerful in the high-power conditions ($M = 5.12, SD = 1.05$) compared to the control conditions ($M = 3.71, SD = 1.92; F[1, 90] = 17.79, p < .001, \eta_p^2 = .17$). We found no main effect for trust ($F[1, 90] = 0.12, p = .73, \eta_p^2 = .00$) and no interaction effect between power and trust ($F[1, 90] = 1.11, p = .29$).

Results and discussion

We conducted two separate analyses of variances (ANOVA) with power and trust, and their interaction as independent variables, and deterrence and just deserts as dependent variables, respectively (see Figure 2.3). We observed a main effect of power for deterrence ($F[1, 90] = 4.68, p = .033, \eta_p^2 = .05$) but not for just deserts ($F[1, 90] = .46, p = .49, \eta_p^2 = .01$) No main effect of distrust was observed for deterrence ($F[1, 90] = .31, p = .58, \eta_p^2 = .00$) or just deserts ($F[1, 90] = .12, p = .73, \eta_p^2 = .00$) Importantly, an interaction effect between power and distrust emerged for deterrence ($F[1, 90] = 3.69, p = .058, \eta_p^2 = .04$) but not for just deserts ($F[1, 90] = 2.70, p = .11, \eta_p^2 = .03$). More specifically, power only increased reliance on deterrence when nothing was mentioned about the trustworthiness of group members ($t[45] = 2.92, p = .005, d = 0.87$). However, when participants were distrustful, no effect of power on deterrence emerged ($t[45] = .17, p = .87, d = 0.05$).

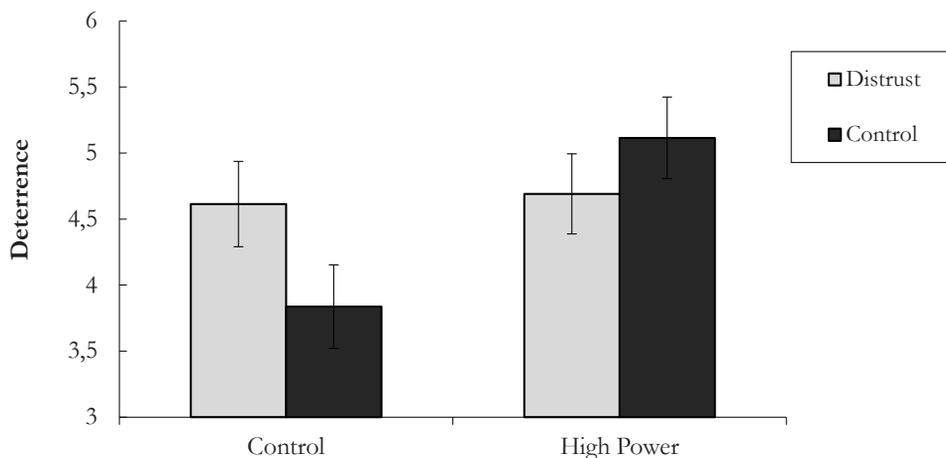


Figure 2.3. Reliance on deterrence as a function of power and trust for Study 2.4c. Error bars represent standard errors.

Possible alternative explanations. Correlations between punishment motives, group focus, and victim identification are reported in Table 2.2. Both deterrence and just deserts correlated positively with victim identification but negatively with being focused on the welfare of the group. We conducted two separate analyses of variances (ANOVA) with power and trust, and their interaction as independent variables, and group focus and victim identification as dependent variables, respectively. We found no main effect of power or distrust, or interaction effect between power and distrust for victim identification ($F[1, 90] = .08, p = .78, \eta_p^2 = .00$; $F[1, 90] = .00, p = .96, \eta_p^2 = .00$ and $F[1, 90] = .03, p = .87, \eta_p^2 = .00$, respectively) or group focus ($F[1, 90] = 1.16, p = .28, \eta_p^2 = .01$; $F[1, 90] = .02, p = .89, \eta_p^2 = .00$ and $F[1, 90] = .38, p = .54, \eta_p^2 = .00$, respectively). In other words, power did not affect victim identification or group focus in the distrust or control condition. The effect of power on deterrence could therefore not be explained by either one of these variables.

These results strongly suggest that distrust is the default for participants with power. Indeed, when trust was low, the relationship between power and deterrence was attenuated; participants in the control condition relied on deterrence to the same extent as participants who occupied a position of power. Distrust, in other words, made participants in the control condition prefer deterrence—just like power holders. Importantly, victim identification and group focus could not explain why power increased reliance on deterrence. Taken together, 2.4a,

2.4b, and 2.4c provide converging support for the hypothesis that distrust plays a unique and pivotal role in explaining why power holders rely on deterrence as a punishment motive.

Table 2.2. Correlation table for Study 2.4c.

	Deterrence	Just Deserts	Victim Identif.	Group Focus
Deterrence	-			
Just Deserts	.51***	-		
Victim Identif.	.21*	.31**	-	
Group Focus	-.14	-.26*	-.30*	-

Note. $N = 94$, * $p < .05$, ** $p < .01$, *** $p < .001$.

General Discussion

Leaders frequently punish to induce compliance with rules (Parks, Joireman, & Van Lange, 2013). Yet, punishment is not always effective and can even be counterproductive (Gneezy & Rustichini, 2000; Van Dijk et al., 2014). If punishment can be problematic, what do leaders then aim to achieve with their punishments?

Nine studies demonstrated that power increased reliance on deterrence as a punishment motive. Specifically, results indicated that power led to an increased reliance on deterrence, but not just deserts, as a punishment motive and, as a consequence, made participants more favorable towards the implementation of punishments that are public or have a mandatory minimum. In addition and importantly, results revealed why this occurred. Power undermined individuals' trust in others, such that the expectation that others are more likely to break the rules accounted for the observed relationship between power and deterrence. Together, these studies provide converging support for the hypothesis that power affects the reliance on deterrence through a decreased trust in others.

Theoretical Implications

The present research makes several important contributions to the literature on power and punishment. First, it provides an understanding why and how those in power punish. It demonstrates that power can lead people to punish to deter. So far, research has demonstrated that power affects the severity of punishments (Van Prooijen, Coffeng, & Vermeer, 2014; Wiltermuth & Flynn, 2013) or has provided suggestions as to why those in power punish others

(Butterfield, Treviño, & Ball, 1996; Butterfield et al., 2005). However, little research to date has provided empirical evidence as to what powerful individuals seek to achieve with such punishments. The current research demonstrates that power increases reliance on deterrence, but not just deserts, as a punishment motive. That is, power does not necessarily decrease a concern with retributive justice, but it increases a concern for deterring rule breaking. These findings are consistent with our reasoning that distrust—elicited by power—increases punishment behavior aimed at deterring rule breaking instead of achieving retributive justice. Note that the current results do not show that power holders consider deterrence more important than just deserts. Instead, the current results demonstrate that due to their power, powerful individuals consider deterrence more important than individuals without power. The advantage of gaining knowledge about the effects of power on punishment motives is that it gives insight into what powerful individuals actually try to achieve with punishments. The current findings imply, for example, that the height of fines and the length of sentences can depend on whether those in power believe that punishments have the potential to deter rule-breaking behavior. Moreover, few studies have yet documented the psychological underpinnings of deterrence. Research on retributive justice has mainly focused on the use of justice as a punishment motive (Darley, 2009). Our work shows that taking the social hierarchy of an organization or society into account when investigating people's punishment motives may provide important additional insights.

Secondly, the current research broadens our understanding of the psychological effects of power and the determinants of interpersonal distrust by demonstrating that power can undermine the trust that individuals have in others, even before others' actual trustworthiness can be inferred from their behavior or other (social) cues. This suggests that powerful individuals in an organization or society (e.g., managers, and policy-makers) tend to approach others (e.g., employees, citizens) in a more distrustful manner. This might seem at odds with research that has documented how the psychological state of status that is often associated with a hierarchical position can increase interpersonal trust (Lount & Pettit, 2012) and acts of justice (Blader & Chen, 2012). Importantly, although power and status often go together, they are conceptually distinct (Magee & Galinsky, 2008). Power refers to asymmetric control over resources, whereas status is the respect and admiration that one has in the eyes of others. Although we did not experimentally distinguish power from status, our findings suggest that power may affect interpersonal trust differently than status.

More specifically, power might impact trust through affecting people's resource-control goals. Research suggests that power holders prefer to strive for goals that maintain and protect their resource control (Fehr et al., 2013; Willis & Guinote, 2011) and that powerful individuals are better able to achieve such control goals than powerless individuals (Magee & Galinsky, 2008;

Smith, Jostmann, Galinsky & Van Dijk, 2008). The current research suggests that distrust might be a consequence of such a heightened resource-control goal, since trust invites potential exploitation and, thus, loss of power. The notion that the motivation to retain power undermines trust in others is further consistent with research showing that power holders' focus on their own resources makes them cynical about others' generous acts (Inesi, Gruenfeld, & Galinsky, 2012). Our theorizing revolves around trust as an expectation of others' cooperative intent. People's (potential) uncooperative behaviors are relevant to power holders because through such acts, power holders can lose resources and thus power. Thus, it seems unlikely that this relationship between power and distrust originates from power making people simply more action-oriented (Galinsky et al., 2003), unless such action is in the service of gaining, maintaining or protecting resources (Keltner et al., 2003). This also implies that power may not undermine trust in others when this trust assessment is irrelevant for the power holder's resource-control (i.e., when the other cannot gain resources through an untrustworthy act; cf. Inesi et al., 2012).

Our theorizing has implications for the emerging literature on people's motivation to acquire and maintain power or status. Individuals vary in the degree to which they are motivated for prestige (status attainment and maintenance) or power (power attainment and maintenance; see Anderson, Willer, Kilduff, & Brown, 2012). The current theorizing and obtained results suggest that distrust and power motivation may be intrinsically linked—being motivated to attain power may foster distrust about others' underlying intentions. When striving for a high ranked power position that is available to only one (or a few) individual(s) it may be functional to assume that others are (also) power hungry, instead of losing your coveted position to those you assumed not to desire it. Being distrustful about others' intentions may in turn *also* foster power motivation as a way to be less dependent on those you expect to be exploitative. Although speculative, these propositions provide a new perspective on the psychology of power motivation and offer suggestions as to why some individuals prefer to strive for respect and admiration, but not power (Anderson et al., 2012; Joseph, Sellers, Newman, & Mehta, 2006). That is, striving for power may be accompanied by viewing others as distrustful, corrupt, and exploitative—a state of mind that might be aversive to many. Striving for status, however, may foster a more trusting mindset, since status is dependent on the respect and admirations of others and distrusting these others is a form of disrespect (Tyler & Blader, 2003).

Practical Implications

Within an organization or society, those individuals who design and implement punishments have, by definition, a form of power. Managers, for example, regularly punish to induce rule compliance (Kerr, 1975; Magee, Kilduff, & Heath, 2011). Similarly, policy-makers are

frequently responsible for designing systems that aim to influence the behavior of people through providing punitive incentives (e.g., fines). The current research indicates that merely due to their power position, such managers and policy-makers may focus more on punishments that deter undesired behavior.

This provides an intriguing speculative psychological explanation for why countries with the most unequal distribution of incomes (i.e., resources) also imprison the most people (see Wilkinson & Pickett, 2007, 2009). For instance, the USA's relatively unequal distribution of incomes compared to other Western countries has been linked to USA's higher imprisonment rates (Wilkinson & Pickett, 2007, 2009). The USA's imprisonment rates (576 people per 100,000) are four and a half time higher than that of the UK (142 people per 100,00) and fourteen times higher than those of Japan (40 people per 100,000)—all of which are rich, industrialized and educated countries (United Nations, 2000). The same pattern can be found for USA states—more unequal states imprison more people than less unequal states (Wilkinson & Pickett, 2007). Interestingly, the relationship between inequality and imprisonment rates does not seem to be explained by unequal countries and states being less educated or people more poor (Wilkinson & Pickett, 2007, 2009). What, then, explains the effect of inequality on imprisonment rates? Some have suggested that citizens demand more severe punishments in unequal societies (Downes & Hansen, 2006). Although this is likely to be partially true, the current research provides an alternative explanation. That is, with increasing power differentials, powerful policy-makers and judges become more concerned with the deterrence of crime. Since relatively unequal societies by definition have greater power differentials, higher imprisonment rates can reflect power holders' increased concern with the deterrence of crime. Inequality might therefore also affect imprisonment rates through its psychological effect on powerful individuals who design and implement punishments. For instance, in the USA, increases in the use of mandatory minimum sentences as a deterrent cause rule breakers to receive longer sentences for minor crimes, hereby increasing imprisonment rates (Chemerinsky, 2004).

Relatedly, the increased reliance on deterrence instead of just deserts can also lead punishments and punitive policies to misalign with what employees or citizens consider fair and just (Darley, 2009). This may elicit negative attitudes towards the punishers (e.g., judges, managers; Ball, Treviño, & Sims, 1994) and undermine the effectiveness of their policies through reducing people's compliance with the very rules those punishments try to uphold (Magee et al., 2011; Tannenbaum, Valasek, Knowles, & Ditto, 2013; Tyler, 2006). Indeed, the tendency of power-holders to punish an employee publicly or to implement a mandatory minimum for punishments has in the past been perceived as unjust and counterproductive (Jabour, 2013; Robinson, 2005). Similarly, the current findings also suggest that punishments that aim to deter

are driven by distrust. This may have the consequence that such punishments signal distrust to those who are (potentially) subjected to them. Managers and policy-makers may, in other words, communicate through their punitive policies that rule-breaking behavior is expected and widespread. Ironically, this might have the exact opposite effect of those intended. Punishments that increase distrust have been shown to increase rule-breaking behavior (Mulder et al., 2006). Those in power can therefore create a self-fulfilling prophecy with their punishments. They may feel that people will break the rules when not deterred by punishments, but people increasingly break the rules because punishments signal distrust, leading those in power to further enforce punishments to deter more rule breaking. This cycle of increasing distrust and punishment could undermine people's compliance with the rules that punishments try to uphold and hereby severely undermine cooperation within an organization or society.

Taken together, this suggests that power holders' reliance on deterrence as a punishment motive can have detrimental consequences. Managers and policy-makers should therefore be aware of the effect that power has on their punitive decisions.

Possible limitations and directions for future research

Whereas the current research provides convergent support for our hypotheses across different samples and measures, there are some issues to be noted. For instance, participants were not confronted with severe moral transgressions that evoke strong emotional reactions (e.g., rape or murder). It is possible that giving offenders their just deserts becomes more important in situations that involve severe moral transgressions because the need to restore retributive justice might be higher upon observing a blatant injustice. Power holders might therefore rely more on just deserts as a punishment motive when confronted with a severe moral transgression. Since judges deal with such issues on a regular basis, understanding the effects of power on their punitive decision-making may prove interesting future research. Relatedly, in the current studies scenarios revolved around unknown others with whom participants had little to no interaction. In organizational settings, managers often have to decide on how to punish an employee who they know very well. It could be that familiarity with the rule breaker moderates the extent to which power holders punish to deter. We believe investigating the boundary conditions will prove an interesting avenue for future research.

There may also be additional structural reasons for power holders' increased concern with deterring rule breaking. Since rule compliance can (although imperfectly) be measured but retributive justice often cannot due to its subjective nature, power holders are generally held accountable for how well they can prevent rule breaking instead of provide fair and just punishments. Combined with the psychological effects of power on deterrence, this may further

facilitate reliance on deterrence as a motive for punishment. Moreover, when managers and policy-makers are made responsible and accountable for deterring rule breaking through punishments, they may experience a greater sense of power. This power may further increase their belief that deterrence is needed. For example, a manager might be made responsible for preventing employees' unethical behavior. This can instill a sense of power into the manager that propels further reliance on deterrence as a punishment motive. Power holders are generally also less susceptible to being punished than their less powerful counterparts. This difference between the powerful and powerless may also affect reliance on deterrence and just deserts, and thus affect punishment preferences. Indeed, since power holders are less likely to be punished, a punishment that is proportionate may become less important to power holders. This might further facilitate the use of punishments that deter instead of punishments that provide offenders their just deserts. It should be noted, however, that the current research clearly shows the mediating role of distrust—an effect that cannot be easily explained by power holders' decreased susceptibility to punishment. A decreased susceptibility to being punished is more likely to decrease, instead of increase, distrust towards others since punishment is a threat (Keltner et al., 2003) and feeling threatened can increase distrust towards others (Kramer & Schaffer, 2014). Our results suggest that losing resources may be threatening to power holders when others are able to gain their (lost) resources. Future research could experimentally manipulate whether power holder' trust could lead to resource loss to others (or not), to further verify our theorizing.

In sum, we believe that some of the above factors can further facilitate power holders' reliance on deterrence opposed to just deserts. Considering the problems that can be associated with using deterrence as a punishment motive, we believe that investigating these issues further to be an important direction for future research.

Conclusion

We presented nine studies that examined how power changes why people punish. Across a range of different instantiations of power, measurements of distrust, and rule-breaking acts, we consistently observed that power undermines trust and that this distrust increases the reliance on deterrence as a punishment motive, and facilitates the use of punishments to achieve this (i.e., public punishments; punishments with a mandatory minimum). In doing so, the current work broadens our knowledge about power, punishment motives, and trust. It also provides practical implications for managers and policy-makers about how their power can bias their punitive decisions and, ironically, undermine their effectiveness.

Chapter 3

Power, distrust, and the motivation to maintain power

This chapter is based on: Mooijman, M., Van Dijk, W. W., Van Dijk, E., & Ellemers, N. (manuscript in preparation). Power, distrust, and the motivation to maintain power.

Trust is often described as a “binding force” in organizations—it connects people and binds them together in a pursuit of collective goals and achievements. For instance, when people trust others to be cooperative, they tend to focus less on fulfilling self-interested goals, and are more likely to comply with rules that promote cooperation (e.g., paying taxes, contributing to group effort; Balliet & Van Lange, 2013; Tam, Hewstone, Kenworthy, & Cairns, 2009; Yamagishi, 2011). Although stimulating trust is thus important for organizations—such as teams, corporations, and societies—previous research has demonstrated that the power differences that are often prevalent within such organizations can increase individual’s distrust towards each other. That is, having—versus not having—power over others has been shown to increase individuals’ distrust towards others, thereby motivating behavior aimed at preventing others from being uncooperative (e.g., punish them to prevent uncooperative behavior; see Chapter 2; Mooijman et al., 2015). Individuals’ power is thus a vital source of distrust within organizations.

But what is the reason why power increases distrust towards others? In the current research, we argue that powerful individuals’ distrust can in part be explained by their motivation to maintain power.

Power

Despite the existence of egalitarian forms of organization, the hierarchical organization remains one of the most prevalent (Anderson & Brown, 2012). Central to a hierarchically structured organization is the notion that some individuals have more power than others (e.g., managers having more power than non-managers). Power can be broadly defined as an individual’s asymmetric control over valuable resources (Anderson & Brion 2014; Magee & Galinsky, 2008). That is, power entails the ability to reward or punish others by granting or withholding valuable resources (Keltner et al., 2003). Having such power is often beneficial. For instance, power holders can award themselves higher salaries and bonuses (Kipnis et al., 1976), disregard others’ desires and feelings (Goodwin et al., 2000; Van Kleef et al., 2006), and focus on pursuing their own instead of others’ goals (Galinsky et al., 2003; Guinote, 2007a; Lammers et al., 2011; Maner & Mead, 2010). Furthermore, power boosts people’s self-esteem (Wojciszke & Struzynska-Kujalowicz, 2007) and leads them to express more positive—approach related—emotions (e.g., amusement and happiness) and less negative—inhibition related—emotions (e.g., embarrassment and shame; Anderson et al., 2003).

Given the benefits commonly associated with having power, powerful individuals are motivated to maintain their power. Recent research has shown that after individuals obtain a position that enables them to make group decisions (i.e., power), the majority of them tend to choose to maintain their power position, even when delegating this position to others is in the

interest of the group (Fehr et al., 2013). The tendency to maintain a power position was also shown to become stronger when individuals' degree of loss aversion increased, suggesting that power is especially desirable after one obtains it (i.e., a power-endowment effect; Fehr et al., 2013). The notion that powerful individuals aim to maintain power is corroborated by research showing that—when powerful individuals perceive their power to be threatened—they are more likely to create divisions among less powerful individuals in such a way that it prevents them from forming alliances (Case & Maner, 2014). Moreover, powerful individuals pay more attention (and seek proximity) to individuals who they perceive as a threat to their power (Maner & Meader, 2010). Thus, when individuals gain power, they seem motivated to maintain their power.

Distrust and the motivation to maintain power

Distrust involves an expectation of malicious intent (Kramer, 1999). If you distrust others you thus expect them to pursue their own interests rather than yours (Balliet & Van Lange, 2011). Such distrust may in part be a consequence of power holders' motivation to maintain their power over others. The rationale for this follows from the notion that power holders can lose their power by trusting others too much—trusting others to take your interests into account can be considered a form of resource sharing (i.e., power sharing) that entails a risk for power holders to lose resource control (Kramer, 1999; Mayer et al., 1995; Zand, 1997). For instance, managers that trust their employees to comply with organizational rules may fail to take the appropriate actions required to prevent their employees from breaking rules, thereby potentially undermining their own power position. Distrusting employees to comply with organizational rules, however, increases the likelihood that a manager engages in acts that prevent employees from breaking rules (e.g., introducing more monitoring; Lount & Pettit, 2012). This may, in turn, help the manager to maintain power (since a manager's position depends in part on how employees behave).

More evidence corroborating the role of power motivation in explaining the power-distrust relationship comes from research demonstrating that individuals distrust others more when they are less willing to risk losing their resources to these others (Das & Teng, 2004), and that this distrust often explains why individuals take actions to prevent resource loss (i.e., do not give others access to their resources; Lount & Pettit, 2012). Similarly, economists frequently define trust as involving the risk of losing resources to others (e.g., Berg et al., 1995). Since power holders have, by definition, more resources to lose to others than non-power holders, we propose that the motivation to maintain resource control (i.e., power) explains at least in part why power fosters a distrustful mindset (consistent with Thomas Hobbes' notion that individuals should distrust each other to prevent exploitation; Hobbes, 1651/1988). If our reasoning is

correct then, (a) occupying a high (versus low) power position should increase distrust towards others through increasing the motivation to maintain this power position, (b) individuals with a high (versus low) motivation to maintain power should distrust others more when they occupy a position of power, and (c) occupying a stable power position should attenuate the power-distrust link (to the extent that a stable power position fulfills the motivation to maintain power; see Maner & Mead, 2010).

Overview of current research

We tested these predictions in three experiments. Our first aim in Experiment 3.1 was to replicate our finding from Chapter 2 that power fosters distrust. Our second aim was to test whether power (high versus low) affects distrust through affecting the desire to maintain one's (high- or low-power) position. In Experiment 3.2, we measured how important participants considered obtaining a position of power and tested its interactive effect with the power position that we assigned participants to (i.e., leader versus subordinate). As such, we could test whether the power-distrust link is especially strong for those with a high (versus low) motivation for power. Finally, in Experiment 3.3, we investigated how the stability of participants' power position affected their trust. Consistent with previous research (Maner & Mead 2010), we reasoned that having a stable (versus unstable) power position fulfills the motivation to maintain power, thereby attenuating the power-distrust link (i.e., making individuals less distrustful towards others).

Consistent with recommendations of Simmons et al., (2011), we ensured that each condition had around 30 participants (Experiment 3.1), or more (i.e., more than 45 in Experiments 3.2 and 3.3; cf. Simmons et al., 2013). Unless indicated otherwise, all measured variables were assessed on seven-point scales on which participants could indicate their level of agreement (1 = *disagree completely*, 7 = *agree completely*). All participants provided informed consent and were debriefed, compensated, and thanked for their participation.

Experiment 3.1

In Experiment 3.1, we investigated the extent to which having high (versus low) power affects the trust that individuals have in others. We further investigated how this power-trust link is affected by individuals' motivation to maintain power. Consistent with the conceptualization of power as *control* over critical resources (Magee & Galinsky, 2008), we manipulated the extent to which individuals controlled a large (high power) compared to small (low power) amount of money that they give to others (cf. reward power; see Anderson et al., 2003). We further measured both individuals' motivation to keep control over these resources and their trust in

others. If the power-distrust link is explained by power holders' motivation to maintain power, then having high (versus low) power should increase the motivation to maintain power and subsequently foster distrust in others.

Method

Participants and design. Ninety-seven Dutch university students (77 females; $M_{\text{age}} = 20.75$ years, $SD_{\text{age}} = 2.28$) participated in exchange for €2 and were randomly assigned to either the high-power or low-power condition.

Procedure. Participants were informed that they were part of a five-person group consisting of four "workers" and one "leader". Four group members would be able to earn up to an extra €20 by finding correct words in a scrambled-letter task, whereas the fifth group member would be assigned as the group leader. This group leader could control (part) of the €20 that the other group members could earn. Crucially, group members had to self-report to the group leader the number of correct words they found. The group leader would then be able to reward the group members on the basis of that information. The rule was that group members would be rewarded €0.10 for each correctly found word, but that group leaders could deviate from this rule at their own discretion.

Power manipulation. To determine the leadership position, all participants filled out the Management Assessment Inventory scale (MAI; see Stouten, De Cremer & Van Dijk, 2005). This is a bogus scale consisting of 26 items measuring leadership style (e.g., "a leader should be able to command respect"). Allegedly on the basis of participants' scores on this scale, all participants learned that they were assigned to the group leader position (participants were informed that they scored comparatively "high" on this scale). Participants in the high-power condition were then informed that they controlled €17.50 that they could allocate, whereas participants in the low-power condition were informed that they controlled €2.50 that they could allocate. That is, as group leaders, participants could fully determine how this money was allocated to the other group members (but they could not allocate any money to themselves and the money would only be relevant within the experiment; they would thus not be able to keep the remaining money for themselves or give it to others after the experiment). Participants were told that the money not controlled by them (€2.50 or €17.50, depending on condition) was controlled by the experimenter, who would use it to allocate it to the other group members. This manipulation is consistent with the conceptualization of power as control over valuable resources (see Magee & Galinsky, 2008). Thus, high power entailed more control over critical resources (i.e., money) than low power.

Motivation for maintaining power. Participants then indicated to what extent they were motivated to maintain their power on the following item, “I want to maintain the power I currently have”.

Distrust. Participants further indicated the extent to which they distrusted the other group members to report the correct number of words. Note that group members could try to gain more money from the group leader through lying about the number of words they found (i.e., reporting a higher number of words). Thus through lying, group members could undermine the group leader’s reward power (i.e., the group leader could reward group members less accurately), and undermine the reward-rules of the experiment. Four items were framed positively (i.e., trust) and four items were phrased negatively (i.e., distrust). Trust items included the following, “group members can be trusted”, “group members will be honest when self-reporting”, “group members are not inclined to lie”, and “I think group members are trustworthy”. Distrust items included the following, “group members cannot be trusted”, “group members will be dishonest when self-reporting”, “group members are inclined to lie”, and “I think group members are untrustworthy”. These four trust items were reverse coded and averaged with the four distrust items to form an eight-item distrust scale ($\alpha = .87$).

Power manipulation check. To verify the power manipulation, participants’ sense of power was measured on a five-item scale. Items included the following, “I feel powerful”, “I feel in control of others”, “Others depend on me”, “I think I have power over others”, and “I feel influential” ($\alpha = .87$). Confirming the validity of the power manipulation, high-power leaders felt more powerful ($M = 5.39$, $SD = 0.92$) than low-power leaders ($M = 3.58$, $SD = 1.17$; $t[95] = 8.50$, $p < .001$, $d = 1.75$).

Results

Confirming our predictions, high-power leaders were both more motivated to stay in power ($M = 4.96$, $SD = 1.21$) than low-power leaders ($M = 4.44$, $SD = 1.52$; $t[95] = 1.88$, $p = .063$, $d = 0.39$) and more likely to distrust workers ($M = 4.36$, $SD = 1.12$) than low-power leaders ($M = 3.85$, $SD = 1.14$; $t[95] = 2.23$, $p = .028$, $d = 0.46$). Motivation to maintain power was positively correlated with distrust ($r = .46$, $p < .001$) and a bootstrapping analysis using 5,000 resamples (Hayes & Preacher, 2011) demonstrated that the motivation to maintain power mediated the effect of power on distrust (95% CI = [0.01, 0.024]). Moreover, the significant effect of power on distrust ($\beta = .22$, $t = 2.23$, $p = .028$) was reduced to non-significance ($\beta = .14$, $t = 1.53$, $p = .13$) when power motivation (which in itself still positively predicted distrust, $\beta = .43$, $t = 4.71$, $p < .001$) was added to the model.

Discussion

These results replicate our previous finding that power increases distrust in others (see Chapter 2). Results from Experiment 3.1 also extend these previous findings by demonstrating that this power-distrust link can be explained by power holders' motivation to maintain their power over others. Although consistent with our conceptualization of power, the power manipulation used in Experiment 3.1 did not enable group leaders to punish group members for rule-breaking behavior. In Experiment 3.2, we addressed this shortcoming while providing additional support for the idea that the power-distrust relationship can be partly explained as a means to maintain power.

Experiment 3.2

In Experiment 3.2, we first measured the extent to which individuals considered occupying a power position personally important. If the motivation to maintain power explains why power increases distrust, then powerful individuals who consider having a power position important (high-power motivation) should distrust others more than individuals who do not (low-power motivation; cf. Case & Maner, 2014; Maner & Mead, 2010). Furthermore, we used a power manipulation that entailed the possibility to punish others for untrustworthy behavior (cf. influence; Keltner et al., 2003). Although power and influence are conceptually distinct (i.e., others can decide to not comply with a power holder's request; Anderson & Brion, 2014), power holders often have influence over others through the possibility to punish them for their rule-breaking behavior. In Experiment 3.2 we therefore used a power manipulation that included the possibility for power holders to punish others.

Method

Participants and design. A total of 174 Dutch university students (111 females; $M_{\text{age}} = 20.72$ years, $SD_{\text{age}} = 2.31$) participated in exchange for €2 and were randomly assigned to the high-power or low-power condition. Prior to the power manipulation, we measured the extent to which participants considered obtaining such a power position important.

Procedure. Participants were informed that they would observe a six-person group. These group members could take resources out of a common resource pool for their own benefit (cf. social dilemma game; Molenmaker et al., 2014; Van Lange et al., 2013). Participants learned that they could harvest from a common resource of 300 chips, each worth €0.10. Each individual group member could take between 0 and 60 chips from this resource and chips that were left were multiplied by two and divided equally among the group members. This way, taking chips from the common resource only benefitted individual group members, whereas leaving chips in

the common pool benefitted the entire group (Van Lange et al., 2013). This social dilemma game can be used to measure the extent to which group members trust others to cooperate and leave chips in the common resource pool, while also manipulating the power that group members have in the game (Mooijman et al., 2015; Mulder et al., 2006). Indeed, it was further explained to participants that they—or another participant—would be assigned to a role that entailed either full control or no control over the common resource, respectively. That is, it was explained to all participants that having power entailed having full control over the chips in the common resource pool but that it was either themselves (high-power condition) or a different participant that was given this resource control (low-power condition). Both participants in the high-power condition and the low-power condition would observe the choices made by the other group members in terms of numbers of chips harvested, but only participants in the high-power condition could ultimately reallocate the chips among the group members (and thus punish group members for uncooperative behavior). It was explained to all participants that they would play multiple social dilemma game rounds, in which power positions would be reassigned to participants before every new round. In reality, the experiment stopped after one social dilemma game round.

Importance of power position. Participants indicated whether they considered it personally important to acquire such a power position (yes, no). Forty-five participants (30.6%) considered it personally important to acquire such a power position, whereas one hundred-and-two participants did not (69.4%). Participants were then randomly assigned to either the high-power or no-power condition.

Distrust. Participants indicated the extent to which they distrusted the other group members to refrain from taking chips from the common resource pool. Consistent with Experiment 3.1, four items were framed positively (i.e., trust) and four items were phrased negatively (i.e., distrust). Trust items referred to the following, “group members can be trusted”, “group members will further the interest of the group”, “group members are willing to forego their self-interest”, and “group members are inclined to pursue group interests”. Distrust items included the following, “group members cannot be trusted”, “group members are not willing to forego their self-interest”, “group members consider self-interest more important than group interest”, and “group members will take as many chips from the resource pool as possible”. The four trust items were reverse coded and averaged with the four distrust items to form an eight-item distrust scale ($\alpha = .93$).

Power manipulation check. To verify the power manipulation, participants’ sense of power was measured on a four-item scale. Items referred to the following, “I feel powerful”, “I feel in control of others”, “Others depend on me”, and “I think I have power over others” ($\alpha =$

.82). Confirming the validity of the power manipulation, participants felt more powerful in the high-power condition ($M = 4.96$, $SD = 0.79$) compared to the control condition ($M = 2.33$, $SD = 0.79$; $F[1, 143] = 356.33$, $p < .001$, $\eta_p^2 = .71$). No significant main effect of power importance or interaction effect between power motivation and the power manipulation was found ($F[1, 143] = 0.58$, $p = .45$, $\eta_p^2 = .00$; $F[1, 143] = 1.26$, $p = .26$, $\eta_p^2 = .01$).

Motivation for maintaining position. Lastly, to verify that participants who considered it personally important (compared to not important) to acquire a group leader position would be more motivated to maintain this position throughout the experiment, we measured their motivation to maintain their position with the following item, “I like to maintain the current position throughout the experiment”.

Results

Motivation for maintaining position. A univariate analysis of variance with power position (high, low) and power importance (yes, no) as independent variables, and power motivation as dependent variable yielded a significant main effect of power position ($F[1, 143] = 120.95$, $p < .001$, $\eta_p^2 = .46$) and a significant two-way interaction between the power manipulation and power importance ($F[1, 143] = 6.14$, $p = .014$, $\eta_p^2 = .04$), but no significant main effect of power importance ($F[1, 143] = 0.17$, $p = .89$, $\eta_p^2 = .00$). Participants in the high-power position who considered it personally important to acquire power were more motivated to maintain their high power position ($M = 5.85$, $SD = 1.49$) than participants who did not consider acquiring power important ($M = 4.96$, $SD = 1.58$; $t[71] = 2.17$, $p = .033$, $d = 0.52$). Power importance did not matter for participants in the low-power position ($M = 2.64$, $SD = 1.32$; $M = 2.78$, $SD = 1.21$, respectively; $t[72] = 0.44$, $p = .66$, $d = 0.10$).

Distrust. A univariate analysis of variance with power position (high, low) and power importance (yes, no) as independent variables, and distrust as dependent variable demonstrated a significant main effect of power ($F[1, 143] = 4.01$, $p = .047$, $\eta_p^2 = .04$) and a significant two-way interaction between power position and power importance ($F[1, 143] = 5.47$, $p = .021$, $\eta_p^2 = .04$), but no significant main effect of power importance ($F[1, 143] = 0.58$, $p = .49$, $\eta_p^2 = .00$). As shown in Figure 3.1, occupying a high-power position increased distrust for those high in power importance ($t[43] = 2.49$, $p = .017$, $d = 0.76$), but not for those low in power importance ($t[100] = 0.31$, $p = .76$, $d = 0.01$).

Mediation analysis. A moderated mediation analysis with 5,000 resamples (see Hayes & Preacher, 2011) demonstrated that participants’ motivation to maintain their (high) power position mediated the interaction between the (high, low) power position that participants occupied and how important (yes, no) they considered obtaining this power position (95% CI =

[0.02, 0.41]). That is, the motivation to maintain power explained (in part) why (high versus low) power increased distrust for individuals who considered acquiring the power position important (95% CI = [0.09, 0.10]) but not for participants who did not consider acquiring the high-power position important (95% CI = [-0.85, 0.09]).

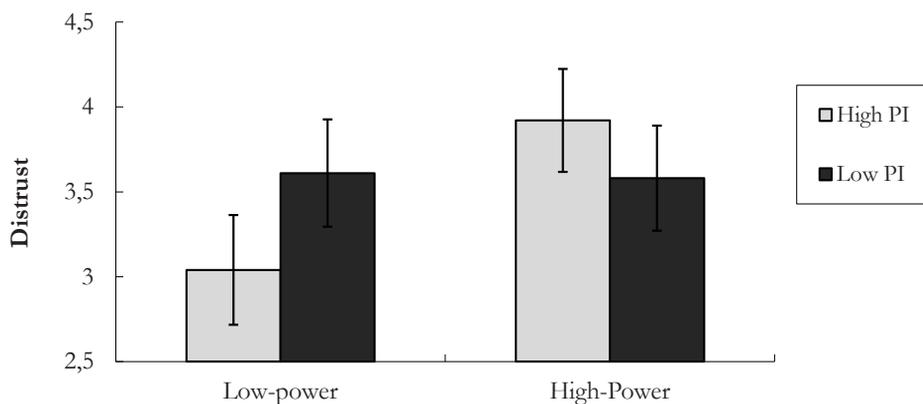


Figure 3.1. Distrust as a function of power and power importance for Experiment 3.2. Error bars indicate standard errors.

Discussion

Experiment 3.2 replicates the finding from Experiment 3.1 that power increases distrust in others (see Chapter 2; Mooijman et al., 2015). Results from Experiment 3.2 also provide additional evidence for the role of the motivation to maintain power. Participants who considered acquiring a power position important (versus not important), and occupied this power position, became more motivated to maintain their power position, and in turn became more distrustful. These findings thus affirm that the effects of power on distrust are in part explained by individual's motivation to maintain a high-power position. Although consistent with our predictions, the measurement of power motivation was relatively indirect and the distribution of participants was unevenly distributed over the two power-importance response options (yes, no). We addressed these issues in Experiment 3.3. If the power-distrust link is explained by power holders' motivation to maintain power, then fulfilling this motivation should attenuate the extent to which power fosters distrust. In Experiment 3.3, we addressed this by directly manipulating power stability. Manipulations of power stability have been used extensively in previous research to fulfill (or exacerbate) behavior aimed at maintaining power (Anderson & Galinsky, 2006; Case & Maner, 2014; Ellemers, Wilke, & Van Knippenberg, 1993; Maner & Mead, 2010). Indeed,

individuals tend to be less afraid of losing their power when their power position is stable compared to unstable. This decreases behavior from power holders that is aimed at maintaining power.

Experiment 3.3

In Experiment 3.3, we manipulated the stability of participants' power position. If the power-distrust relationship is in part explained by power holders' motivation to maintain power, then fulfilling the motivation to maintain power should attenuate the power-distrust link. Indeed, being certain about maintaining one's (high) power position should attenuate actions aimed at maintaining power—that is, decrease distrust towards others. Finding evidence for this conjecture would suggest that creating stable power positions could weaken the extent to which power increases distrust.

Method

Participants and design. Sixty-two Dutch university students (37 females; $M_{\text{age}} = 22.53$ years, $SD_{\text{age}} = 2.88$) participated in exchange for €2 and were randomly assigned to either the stable or unstable power position condition.

Procedure. As in Experiment 3.2, all participants were given control over the allocation of resources in the same social dilemma game (i.e., all participants occupied a high-power position). It was explained to them that they would play multiple social dilemma game rounds.

Power stability. Consistent with previous research on power stability (Maner & Mead, 2010), half of the participants were informed that their power position could randomly change throughout the experiment (i.e., was unstable), whereas the other half of the participants were informed that their power position could not change throughout the experiment (i.e., was stable). To verify this stability manipulation, perceived stability was measured on a three-item scale. That is, “my position seems unstable”, “I feel uncertain about my position”, and “my position might change throughout the experiment” ($\alpha = .64$). Confirming the validity of our manipulation, participants felt that their position was more unstable in the unstable power condition ($M = 5.71$, $SD = 0.81$) compared to the stable power condition ($M = 5.16$, $SD = 1.21$; $t[60] = 2.01$, $p = .040$, $d = 0.52$).

Distrust. Participants indicated the extent to which they distrusted the other group members to refrain from taking chips from the common resource pool on a three-item scale. Items included, “group members cannot be trusted”, “group members will not further the interest of the group”, and “group members are not willing to forego their self-interest” ($\alpha = .91$).

Results and discussion

Consistent with our predictions, participants in the unstable power condition distrusted others more ($M = 4.44$, $SD = 1.14$) than participants in the stable power condition ($M = 3.72$, $SD = 1.58$; $t[60] = 2.06$, $p = .044$, $d = 0.53$).

Discussion

These results provide additional evidence for the hypothesis that the effect of power on distrust is explained by the motivation to maintain this power. Indeed, consistent with the idea that stable power positions decrease the likelihood that individuals engage in behavior aimed at maintaining power (Case & Maner, 2014), these results demonstrate that having a stable (versus unstable) power position attenuated the power-distrust link.

General Discussion

Although promoting trust is vital for powerful individuals such as managers, previous research has demonstrated that power fosters distrust in others (see Chapter 2). In the current research, we investigated the psychological process that underlies this power-distrust link. In three experiments, we demonstrated that, (a) having high (versus low) power increased distrust through increasing the motivation to maintain power, (b) individuals with a high (versus low) motivation to stay in power distrusted others more when they occupied a position of power, and (c) occupying a stable (versus unstable) power position attenuated the power-distrust link. Together, these studies provide converging support for the hypothesis that power increases distrust in part because of power holders' motivation to maintain power.

Implications

The present research makes several contributions to the literature on power and trust. First, it provides an understanding why power increases distrust. Previous research has demonstrated that having power can increase an individual's distrust towards others (Mooijman et al., 2015) but has provided little evidence as to what explains this power-distrust link. The current research provides support for the notion that power holders distrust others (partially) to maintain their own power. This suggests that powerful individuals in an organization (e.g., managers) may approach others in a more distrustful manner to maintain power. This might seem at odds with research that has documented how the psychological state of status that is often associated with a hierarchical position can increase interpersonal trust (Lount & Pettit, 2012). Importantly, although power and status often go together, they are conceptually distinct

(Magee & Galinsky, 2008). Power refers to asymmetric control over resources, whereas status refers to the respect and admiration that one has in the eyes of others. Experiment 3.1 manipulated power between conditions while keeping status differences between these conditions to a minimum (i.e., all participants were leaders). The present research therefore suggests that power affects interpersonal trust differently than status. This observation is consistent with recent research suggesting that power often has different, and even opposing, psychological effects than status (Blader & Chen, 2012).

Second, the current research broadens our understanding of distrust by demonstrating why power increases distrust. Although trust and hierarchy have been much studied topics (Kramer, 1999), few studies have investigated how power holders (dis)trust others. The present research suggests why an organizational hierarchy may create distrust through its power holders. As such, we are—to our knowledge—the first to demonstrate why power fosters distrust. These results have implications for attempts to promote interpersonal trust in organizations and societies. For instance, attempts to promote organizational trust may fail when those who aim to promote them (e.g., managers) tend to distrust others (e.g., employees) in the organization. Indeed, given the influence of power holders in setting exemplary behavior (Mayer, Kuenzi, Greenbaum, Bardes, & Salvador, 2009) and the negative effects associated with interpersonal distrust (e.g., decreased liking, increased tendency for unethical behavior; see Chapter 3; Mulder et al., 2006), power holders' distrust is potentially problematic (e.g., communicates and spreads distrust throughout an organization or society). The current research suggests that attenuating power holders' fear of losing their power may help alleviate some of the negative effects of power on distrust.

Limitations and Future Directions

Whereas the current research provides converging support for our hypotheses, there are several possible limitations to address. For instance, we measured (trust) expectations about others' behavior instead of measuring actual trust-related behaviors (e.g., sharing money with others; see Berg et al., 1995). Although the two are highly correlated (Lount & Pettit, 2012), future research could use the trust game to investigate the effects of power on trust-related behaviors. In the current experiments, power holders were also relatively independent from others (cf. the definition of power; Magee & Galinsky, 2008). However, power holders are in real life often also dependent on their subordinates (e.g., politicians relying on citizens' votes) or higher-ranked officials (e.g., an assistant professor relying on a full professor) for maintaining their power.

Future research could investigate how these two types of dependence affect the power-distrust link. Being dependent on someone else for maintaining one's power may for instance foster trust towards this person, especially for those individuals who are highly motivated to maintain power (consistent with research on system-justification showing that being dependent on others increases the motivation to view these others more positively; Van der Toorn et al., 2015). Indeed, on the basis of the current research, it may be fruitful to investigate ways of making power holders more dependent on others. Moreover, future research could also investigate how to effectively eliminate the power-distrust link. For instance, future research could focus on how power holders' identification with a relevant organization affects their distrust (cf. Gaertner & Dovidio, 2000).

Conclusion

We presented three experiments that examined the power-distrust in more detail. Using different manipulations of power and measurements of power motivation, we consistently observed that distrust increased as a function of the motivation to maintain power. The motivation to maintain power thus explained in part why power holders distrust others. In doing so, the current work broadens our knowledge about power and trust. It also suggests practical implications for managers, leaders, and policy-makers about why their power can make them distrustful about others' intentions.

Chapter 4

On sanction effectiveness: How and why sanction-goal justifications affect rule compliance

This chapter is based on: Mooijman, M., Van Dijk, W. W., Van Dijk, E., & Ellemers, N.
(invited revision at *Journal of Personality and Social Psychology*). On sanction effectiveness: How and
why sanction-goal justifications affect rule compliance.

Authorities frequently use sanctions to promote rule compliance. Judges sentence citizens to jail, managers fire employees for not sticking to ethical rules, and universities expel students for misconduct. Because such sanctions can greatly affect people's lives, authorities often provide a justification for their sanctioning behavior. For instance, judges sentence people to prison with the explicit justification that this is meant to deter future criminal behavior (e.g., see "Colorado Woman", 2015) or that this is meant to give this person their just deserts (e.g., deserved punishment, see "Osama Bin Laden Dead", 2011). Although providing sanction justifications may seem appealing to authorities, we propose that these justifications can in fact influence how effective a sanction will be in promoting future rule compliance. We propose that authorities' sanctions are less effective at promoting rule compliance when they are justified as attempts to deter people from breaking rules as compared to giving people their just deserts. Moreover, we argue that these effects can be attributed to people feeling more distrusted when sanctions are justified as attempts to deter them from breaking rules as compared to giving them their just deserts.

Previous research has mainly focused on the extent to which sanction goals guide sanctioning decisions (Carlsmith, 2006, 2008; Carlsmith et al., 2002; Darley et al., 2000; Gerber & Jackson, 2012), but has left the effects of sanction-goal justifications on rule compliance unaddressed. Examining how and why sanction justifications shape people's willingness to comply with rules can provide valuable insights into how authorities should—and should not—use sanctions. Societal and organizational authorities (e.g., policy-makers, leaders, and managers) tend to justify their use of sanctions by stressing the necessity to deter rule-breaking behavior (Kirchler et al., 2014; Mooijman et al., 2015). Understanding how such justifications affect rule compliance may therefore be helpful in, (a) explaining the (in)effectiveness of real-life sanctions, and (b) suggesting ways to improve the manner in which authorities justify their sanctioning behavior.

Sanction Justifications

Authorities can stress different goals as justification for their use of sanctions. Scholars have typically classified sanction goals into goals that aim to deter future rule-breaking behavior (Bentham, 1789/1988; Hobbes, 1651/1988; Kirchler et al., 2014; Nagin, 1998) versus goals that aim to give people their just deserts (i.e., give offenders their deserved punishment; Darley, 2009; Kant, 1780/1961). Although both goal types may co-occur, they have different aims. The deterrence goal aims to deter future rule breaking from potential rule breakers and, as such, is prospective rather than retroactive. When having this goal, authorities should be primarily concerned with deterring future rule breaking instead of achieving retributive justice through

punishing (past) rule breakers proportionate to their crime. In contrast, the just-deserts goal aims to punish past rule breakers proportionately (i.e., achieve balance between crime and punishment), regardless of the sanction's ability to deter future rule breaking. As such, the just-deserts goal is retroactive rather than prospective. When guided by this goal, authorities should be primarily concerned with achieving retributive justice through punishing rule breakers proportionate to their crime instead of preventing future rule breaking.

While both sanction goals can affect the type and severity of the sanction used (Carlsmith et al., 2002) and thereby influence rule compliance (Ball et al., 1994), we argue that authorities' use of a sanction goal as a *justification* creates an additional source of influence. That is, independently of the type and severity of a sanction, people's willingness to comply with rules may be affected by whether an authority justifies the existence of a sanction as an attempt to deter them or give them their just deserts. We propose that authorities' sanctions are less effective at promoting people's willingness to comply with rules when they are justified as an attempt to deter people from breaking rules, whereas sanctions will be more effective when they are justified as an attempt to give people their just deserts. Moreover, we argue that these effects can be attributed to people feeling more distrusted when sanctions are justified as attempts to deter them from breaking rules as compared to giving them their just deserts.

When sanction justifications may signal distrust

Authorities' aim to deter people from rule breaking or give them their just deserts can result in sanctions that are equivalent in terms of sanction type and severity. However, these sanction goals are not equivalent with regard to their underlying considerations. A central aspect of a deterrence, but not a just-deserts, goal is that sanctions should be aimed at those who are deemed likely to break rules (hence the need to deter them; Nagin, 1998). In other words, authorities that aim to deter rule breaking expect people to break rules in the future (i.e., they distrust them; Mooijman et al., 2015). Whereas authorities that aim to give people their just deserts are indifferent with regard to people's likelihood of breaking rules (i.e., trustworthiness is irrelevant; Carlsmith et al., 2002; Kant, 1780/1961; Mooijman et al., 2015).

People can often infer such intentions and considerations from authorities' decisions. Managers' attempts to incentivize weight-loss with financial sanctions have been shown, for instance, to unintentionally signal negative attitudes towards the overweight (Tannenbaum et al., 2013). Authorities who justify a sanction as an attempt to deter people from rule breaking may therefore be more likely to signal their distrust than authorities who justify a sanction as an attempt to give people their just deserts. Authorities using deterrence justifications may signal that sanctions are needed because people are likely to break rules in the absence of sanctions. The

sanction is then used as a means to deter people's future rule-breaking behavior. In contrast, just-deserts justifications signal that an authority's sanction is aimed at those *who have* broken rules in the past instead of those who *are expected to* break rules in the future. The communicated "breadth" of a just-deserts justified sanction is thus smaller (i.e., targets only rule breakers) than the communicated breadth of a sanction that is justified as an attempt to deter (i.e., targets all potential rule breakers). We argue that just-deserts justifications therefore may signal less distrust than deterrence justifications.

Some evidence corroborating this reasoning comes from research showing that authorities' distrust predicts deterrence—but not just-deserts—justified sanctions (see Chapter 2; Mooijman et al., 2015) and from research showing that people are highly motivated to infer authorities' intentions from their sanction decisions (Fiske, 1993; Keltner et al., 2003). People may thus infer from an authority's deterrence—but not just-deserts—justification that they are expected to have the malicious intention to undermine the interests of the authority (consistent with definitions of distrust, Kramer, 1999; Yamagishi & Yamagishi, 1994; Zand, 1997). In sum, we hypothesize that justifying a sanction as an attempt to deter people from breaking rules, compared to giving people their just deserts, makes people feel more distrusted (Hypothesis 1).

Why feeling distrusted may undermine rule compliance

How might people's rule compliance be affected by their feelings of being distrusted by an authority? Rule compliance is not solely determined by the severity of a sanction or the probability that one receives a sanction (Balliet et al., 2011). Instead, rule compliance is also determined by how people feel treated by the authority (i.e., interpersonal justice; Tyler & Lind, 1992). For instance, people's satisfaction with authorities' decisions decreases when authorities communicate disrespect through, (a) pursuing their own interest instead of the interest of the people (De Cremer, 2002) and (b) using non-transparent and biased procedures (Tyler, 2012). In contrast, authorities that are perceived to pursue the collective interest (Mulder & Nelissen, 2010), show respect for others (Blader & Tyler, 2003), and use transparent and unbiased procedures foster decision acceptance (Tyler, 2012). Importantly, these effects of perceived interpersonal treatment often go beyond the outcome that people (expect to) receive from authorities (Cropanzano et al., 2007). We argue that how people feel treated by the authority that installs and justifies a sanction is also of vital importance for people's willingness to behave according to the authority's rules.

More specifically, we propose that feeling distrusted by an authority undermines people's willingness to comply with the authority's rules. Some evidence for this conjecture follows from the notion that people are motivated to see themselves as trustworthy (Brown, 2012; Sedikides,

Meek, Alicke, & Taylor, 2014; Steele, 1988), and want and expect others to trust them (Ellemers, 2012; Tyler & Lind, 1992). Feeling *distrusted* by an authority is therefore likely to foster the feeling that the authority does not view oneself favorably (e.g., without respect). This perception alone may be sufficient to feel poorly treated, thus undermining one's willingness to abide to this authority's rules. Indeed, perceived interpersonal treatment does not have to revolve around tangible outcomes that one receives, but can also entail subjective assessments of how others view oneself (e.g., respect; Ellemers, Doosje, & Spears, 2004). A perceived lack of trust in one's willingness to comply with relevant guidelines and rules may seem unwarranted when no prior breach of rules was displayed, and may thus seem disrespectful and unjust. We propose, then, that feeling distrusted by an authority undermines people's willingness to comply with this authority's rules.

Although sanctions increase the costs of rule breaking regardless of an authority's justification, we argue that the potential effectiveness of a sanction is decreased by the distrust that people experience when authorities provide a deterrence justification. We hypothesize that authorities' sanctions are less effective at promoting rule compliance when they justify such sanctions as deterring people from rule breaking compared to giving people their just deserts (Hypothesis 2a). We further hypothesize distrust to mediate the negative relationship between deterrence justifications and rule compliance (Hypothesis 2b).

Overview of Current Research

We tested our hypotheses in four experiments in which we, (a) manipulated which justifications authorities provided for their sanctioning behavior, (b) measured the distrust that participants felt, and (c) tested how this affected rule compliance. More specifically, in these experiments we tested how providing participants with a deterrence or just-deserts sanction justification affected their willingness to comply with university rules (Experiment 4.1), their willingness to comply with their manager's rules (Experiments 4.2 and 4.3), and the extent to which they lied to their team leader to further their own self-interest (Experiment 4.4). We tested our hypotheses in both Dutch (Experiments 4.1 and 4.4) and American samples (Experiments 4.2 and 4.3). We also examined how providing both deterrence *and* just-deserts sanction justifications simultaneously, affected participants' willingness to comply with rules (Experiment 4.3). We expected that providing a deterrence justification negatively affects sanction effectiveness, even when it is provided simultaneously with a just-deserts justification. In Experiment 4.2, we examined how providing a sanction justification aimed at deterring others—instead of the participant—affected distrust and rule compliance. We expected that the negative effects of providing a deterrence justification would be eliminated when the sanction was justified

as deterring others—as opposed to the participant—from breaking rules. Last, in all experiments, we measured the extent to which participants perceived the sanction as aimed at themselves. This enabled us to verify our assumption that, compared to just-deserts, deterrence justifications signal that potential—instead of past—rule breakers are targeted by the sanction.

To provide support for the proposed mediating role of distrust and exclude alternative explanations, we assessed participants' group identification, their attitudes towards the authority, their perceived legitimacy of the authority, and their distrust towards others. Previous research has demonstrated that rule compliance can be affected by the extent to which people identify with their group (Ellemers, Gilder, & Haslam, 2004; Tyler & Blader, 2000), the attitudes that they hold towards authorities (Tyler & Blader, 2003), and the extent to which they trust others to comply with rules (Mulder et al., 2006). We assessed these control variables in Experiments 4.1–4.3. All reported mediation analyses were conducted using a bootstrap analysis with 5,000 resamples (Hayes, Preacher, & Meyers, 2011). Consistent with the recommendations of Simmons, Nelson, and Simonsohn (2013), we made sure that every condition had more than fifty participants. Unless indicated otherwise, all measured variables were assessed using seven-point scales, on which participants could indicate their level of agreement (1 = *disagree completely*, 7 = *agree completely*). All participants provided informed consent and were debriefed, compensated, and thanked for their participation.

Experiment 4.1

Experiment 4.1 investigated how a university's sanction justification affects students' feelings of distrust and their willingness to comply with the university's rules and policies. To test this, we manipulated whether a real-life sanctioning system was justified to students as an attempt to deter plagiarism or give students their just deserts. We compared these two sanction-justification conditions to a condition in which no information was given about a sanction. To rule out that decreased group identification instead of distrust mediated the hypothesized effect, we also measured students' identification with their university.

Method

Participants, design, and procedure. Eighty-seven Dutch university students (68 females; $M_{age} = 23.11$ years, $SD_{age} = 6.54$) participated in exchange for €2 and were randomly assigned to one of three punishment conditions (deterrence vs. just deserts vs. no sanction). Two participants indicated being recently caught for plagiarism. Their data were excluded from the reported analyses, although including their data in the analyses did not change the pattern of results.

Sanction justification. All participants were informed that the study assessed students' attitudes toward university policy. More specifically, it was explained how university students might sometimes be tempted to directly copy information from professional articles for their own work. In the deterrence condition and just-deserts condition, it was explained that the policy of their university was to immediately exclude students who committed such plagiarism from their respective courses in order to *deter* students from committing plagiarism, or give students their *just deserts* for committing plagiarism, respectively. As such, the severity of the sanction was held constant across the sanction conditions. In the no sanction condition, no information was given about a sanction.

Distrust. Participants indicated to what extent the policy made them feel distrusted on a four-item scale (adapted from Mooijman et al., 2015). Items included, "I feel distrusted by the university", "I think the university assumes that I want to break the rules", "I believe the university distrusts me", and "I think I am trusted by the university" (reverse-coded; $\alpha = .87$).

Rule compliance. Participants' willingness to comply with university rules was assessed on a three-item scale. Items included, "I feel inclined to stick to university rules", "I feel obliged to behave according to university rules", and "I feel inclined to break university rules" (reverse-coded; $\alpha = .89$).

Group identification. Group identification was measured on a six-item scale (adapted from Tyler & Blader, 2001). Sample items included, "I am proud to be a student of this university", "I share norms and values with the university", and "when I talk about where I study I usually say we rather than they"; $\alpha = .87$).

Manipulation check. As a manipulation check, we measured the extent to which participants perceived the sanction as aimed at them with the following item, "I feel like this sanction is meant for me". Confirming that the deterrence justification signals that the sanction is aimed at potential instead of past rule breakers, participants in the deterrence condition were more likely to perceive the sanction as aimed at themselves ($M = 4.86$, $SD = 1.46$) than participants in the just-deserts condition ($M = 3.72$, $SD = 0.59$; $t[56] = 3.89$, $p < .001$, $d = 1.04$).

Results

Distrust. The sanction-justification manipulation affected the extent to which students felt distrusted by the university ($F[1, 82] = 3.81$, $p = .026$, $\eta^2_p = .09$). Perceived distrust was higher in the deterrence condition ($M = 3.54$, $SD = 1.33$) compared to the just-deserts condition ($M = 2.57$, $SD = 1.35$; $t[55] = 2.72$, $p = .009$, $d = 0.73$) and higher, but not significantly, compared to the no-sanction condition ($M = 3.02$, $SD = 1.28$; $t[55] = 1.50$, $p = .14$, $d = 0.40$).

The no-sanction and just-deserts condition did not differ from each other ($t[54] = 1.27, p = .21, d = 0.35$).

Rule compliance. The sanction-justification manipulation also influenced students' willingness to comply with university rules ($F[1, 82] = 3.73, p = .028, \eta_p^2 = .08$). Rule compliance was lower in the deterrence condition ($M = 5.76, SD = 1.16$) than in the just-deserts condition ($M = 6.40, SD = 1.12; t[55] = 2.14, p = .037, d = 0.58$) and equally low as in the no-sanction condition ($M = 5.58, SD = 1.28; t[55] = 0.54, p = .59, d = 0.14$). Rule compliance was higher in the just-deserts condition compared to the no-sanction condition ($t[54] = 2.56, p = .013, d = 0.69$).

Group identification. The sanction-justification manipulation did not affect students' identification with the university ($M = 5.11, SD = 1.11; F[1, 82] = 0.75, p = .48, \eta_p^2 = .02$).

Mediation analyses. Using a bootstrap analysis procedure with 5,000 resamples (Hayes & Preacher, 2011), we tested whether group identification or distrust mediated the effect of the sanction-justification manipulation on rule compliance. Although group identification was positively correlated with rule compliance (Pearson's $r = .23, p = .035$), bootstrap analyses showed that it did not mediate the effect of the sanction-justification manipulation on rule compliance (95% CI = [-0.44, 0.04]). Instead, feeling distrusted was negatively correlated with rule compliance ($r = -.39, p < .001$), and bootstrap analyses demonstrated that this distrust mediated the effect of the sanction-justification manipulation on the decrease in rule compliance (95% CI = [-0.67, -0.06])—even after adding group identification as additional mediator (95% CI = [-0.74, -0.05]) or after adding group identification as a covariate (95% CI = [-0.77, -0.05]). Moreover, the significant effect of sanction justifications on rule-compliance ($\beta = -.28, t = -2.14, p = .037$) was reduced to non-significance ($\beta = -.17, t = -1.25, p = .22$) when distrust (which in itself still negatively predicted rule compliance; $\beta = -.33, t = -2.47, p = .017$) was added to the model.

Discussion

Consistent with Hypothesis 1, results demonstrate that students felt more distrusted when their university's sanction policy was justified as an attempt to deter students from rule breaking compared to giving those who broke the rule their just deserts. Consistent with both Hypotheses 2a and 2b, results further demonstrate that this distrust undermined the extent to which students were willing to comply with university rules (i.e., distrust mediated the effect of the deterrence justification on rule compliance). In fact, providing a deterrence justification for a sanction was as (in)effective in promoting a willingness to comply with university rules as providing no information about the possibility of receiving a sanction for breaking rules.

Sanctions that were justified as giving students their just deserts were more effective at promoting students' willingness to comply with rules. Moreover, identification with the university could not explain these results.

Experiment 4.2

Experiment 4.2 investigated how sanctions that are explicitly justified as an attempt to deter oneself (compared to others) from rule breaking affects distrust and rule compliance. If deterrence justifications foster distrust through signaling that one is considered a potential rule breaker then deterrence justifications aimed at *oneself* should increase distrust, whereas deterrence justifications aimed *at others* should attenuate distrust. To test this, we manipulated whether the sanction was justified as deterring a group of people of which the participant was—or was not—a part of. We compared these two conditions to a condition in which the sanction was justified as giving participants' just deserts and a condition in which no information was given about the sanction. We predicted that a sanction signals more distrust (and thus is less effective in promoting a willingness to comply with rules) when it is justified as deterring oneself—compared to others—from rule breaking, and compared to a sanction that is justified as an attempt to give rule breakers their just deserts. We tested our predictions in an employee-supervisor context while measuring attitudes towards the supervisor. This enabled us to rule out that negative attitudes towards the authority instead of distrust mediated the hypothesized effect.

Method

Participants, design, and procedure. A total of 245 American participants (147 males; $M_{age} = 32.78$ years, $SD_{age} = 10.86$) were recruited from the Mechanical Turk website and were randomly assigned to one of four sanction conditions (deterrence-self vs. deterrence-other vs. just deserts vs. no sanction). Participants received \$1 for their participation.

Sanction justifications. All participants were asked to imagine themselves as a *sales employee* of a company called Big City Electronics (BCE). The scenario explained that BCE employs both sales and administrative employees and that recently office-supplies (e.g., binders, staplers) were stolen from BCE's stock. The scenario continued by explaining that the supervisor reacted by introducing a sanction. This sanction was described as a "substantial pay-cut" for anyone caught stealing. Importantly, the supervisor justified this sanction as either meant to deter *sales employees* from theft (deterrence-self condition), deterring *administrative employees* from theft (deterrence-other condition), or giving rule-breaking employees their just deserts (just-deserts condition). In the no-sanction condition, no information was given about a sanction.

Distrust. Feeling distrusted was measured on a six-item scale. Consistent with Experiment 4.1, we measured the extent to which the participants felt distrusted by the authority (i.e., supervisor). Items included, “I feel distrusted by the supervisor”, “I think the supervisor assumes I want to break the rules”, “The supervisor does not trust me”, “I think the supervisor believes I want to steal office supplies”, “I believe the supervisor considers me a potential rule breaker”, and “I feel like the supervisor assumes I have bad intentions” ($\alpha = .96$).

Rule compliance. Participants’ willingness to comply with the rules was measured on a five-item scale (adapted from Tyler & Blader, 2005). Items included, “I feel inclined to slack off towards the end of the day”, “I feel inclined to come late if the supervisor doesn’t find out”, “I feel inclined to undermine the supervisor’s rules”, “I feel inclined to not do my best at work”, and “I feel inclined to find ways to undermine the supervisor” ($\alpha = .95$).

Attitudes towards supervisor. We measured participants’ attitudes towards the supervisor on a four-item scale. Items included, “I like this supervisor”, “I have a positive feeling about this supervisor”, “I tend to view this supervisor positively”, and “I dislike this supervisor” (reverse-coded; $\alpha = .93$).

Manipulation check. Participants indicated on one item to what extent they felt personally targeted by the sanction (i.e., “I feel like this sanction is targeted at me”). The sanction-justification manipulation affected the extent to which participants felt personally targeted by the sanction ($F[1, 241] = 11.44, p < .001, \eta_p^2 = .13$). Confirming the validity of our manipulation, participants in the deterrence-self condition felt more personally targeted by the sanction ($M = 4.08, SD = 2.29$) than in the deterrence-other condition ($M = 2.78, SD = 1.83; t[124] = 3.53, p = .001, d = 0.63$) and just-deserts condition ($M = 2.63, SD = 1.98; t[120] = 3.74, p < .001, d = 0.69$). The just-deserts condition and deterrence-other conditions did not differ ($t[120] = 0.43, p = .66, d = 0.08$).

Results

Distrust. The sanction-justification manipulation affected the extent to which participants felt distrusted by the supervisor ($F[1, 241] = 18.21, p < .001, \eta_p^2 = .19$). Perceived distrust was higher in the deterrence-self ($M = 4.61, SD = 1.80$) compared to the deterrence-other condition ($M = 2.88, SD = 1.69; t[124] = 5.55, p < .001, d = 0.99$), just-deserts condition ($M = 2.76, SD = 1.89; t[120] = 5.52, p < .001, d = 1.01$), and no-sanction condition ($M = 2.58, SD = 1.55; t[121] = 6.66, p < .001, d = 1.21$). The deterrence-other condition did not differ from the just-deserts condition or no-sanction condition ($t[120] = 0.36, p = .72, d = 0.06; t[121] = 1.01, p = .32, d = 0.18$, respectively). Lastly, the just-deserts condition and no-sanction condition did not differ from each other ($t[117] = 0.56, p = 0.58, d = 0.10$).

Rule compliance. The sanction-justification manipulation influenced participants' willingness to comply with rules ($F[1, 241] = 4.43, p = .005, \eta^2_p = .05$). Willingness to comply with rules was lower in the deterrence-self ($M = 5.06, SD = 1.89$) compared to the deterrence-other condition ($M = 5.73, SD = 1.65; t[124] = 2.10, p = .038, d = 0.38$), just-deserts condition ($M = 5.84, SD = 1.46; t[120] = 2.56, p = .012, d = 0.47$), and equally low as in the no-sanction condition ($M = 4.95, SD = 1.70; t[121] = 0.34, p = .73, d = 0.06$). Moreover, rule compliance was higher in the deterrence-other condition and just-deserts condition compared to the no-sanction condition ($t[121] = 2.56, p = .012, d = 0.47; t[117] = 3.08, p = .003, d = 0.57$, respectively). The deterrence-other condition and just-deserts condition did not differ significantly ($t[120] = 0.43, p = .67, d = 0.08$).

Attitudes towards supervisor. The sanction-justification manipulation influenced participants' attitudes towards the supervisor ($F[1, 241] = 10.71, p < .001, \eta^2_p = .12$). Participants in the deterrence-self condition liked the supervisor less ($M = 3.79, SD = 1.46$) than participants in the deterrence-other condition ($M = 4.46, SD = 1.53; t[124] = 2.52, p = .013, d = 0.46$) and just-deserts condition ($M = 4.86, SD = 1.61; t[120] = 3.85, p < .001, d = 0.70$) but equal compared to the no-sanction condition ($M = 3.40, SD = 1.60; t[121] = 1.39, p = .17, d = 0.25$). Participants in the just-deserts condition and deterrence-other condition liked the supervisor more than participants in the no-sanction condition ($t[117] = 4.95, p < .001, d = 0.92; t[121] = 3.74, p < .001, d = 0.68$, respectively). The deterrence-other condition and just-deserts condition did not differ significantly ($t[120] = 1.40, p = .16, d = 0.25$).

Mediation analyses. Using a bootstrap analysis procedure with 5,000 resamples (Hayes & Preacher, 2011), we tested whether attitudes towards the supervisor or distrust mediated the effect of the sanction-justification manipulation on rule compliance. Positive attitudes towards the supervisor correlated positively with rule compliance ($r = .52, p < .001$), but results from the bootstrap analyses showed that attitudes towards the supervisor did not mediate the effect of the sanction-justification manipulation on rule compliance (95% CI = [-0.01, 0.14]). However, distrust was negatively correlated with rule compliance ($r = -.52, p < .001$), and results from the bootstrap analyses showed that distrust mediated the overall effect of the sanction-justification manipulation on rule compliance (95% CI = [-0.34, -0.14]), even after controlling for attitudes towards the supervisor (95% CI = [-0.31, -0.12]) or after adding attitudes towards the supervisor as an additional mediator (95% CI = [-.26, -.09]). This indirect effect of distrust was also significant for the deterrence-self versus just-deserts contrast (95% CI = [-0.76, -0.29]) and the deterrence-self versus deterrence-other contrast (95% CI = [-1.65, -0.69]). More specifically, for both contrasts perceived distrust decreased the significant effect of the sanction-justification manipulation on rule compliance ($\beta_s > .30, p_s < .01$) to non-significance ($\beta_s < .10, p_s > .10$).

Discussion

Replicating Experiment 4.1, these results demonstrate that participants felt more distrusted by their supervisor when this supervisor justified a sanction as deterring participants from rule breaking compared to giving participants their just deserts (Hypothesis 1). Results further demonstrate that this distrust decreased the extent to which participants were willing to comply with rules (Hypotheses 2a and 2b). Crucially, the extent to which a deterrence justification fostered distrust and decreased sanction effectiveness was exacerbated when it was aimed at oneself, but attenuated when it was aimed at others. This corroborates our assumption that—compared to just-deserts justifications—deterrence justifications are more likely to foster distrust through signaling that one is considered a potential rule breaker. Moreover, no differences in distrust were observed when the sanction was justified as deterring others compared to giving participants their just deserts. This provides additional support for our reasoning that just-deserts justifications are less likely to elicit distrust and undermine sanction effectiveness, partially because such sanctions are not perceived to target the self.

Experiment 4.3

The previous two experiments provide converging support for our hypotheses but used the deterrence or just-deserts sanction justifications as mutually exclusive—that is, a sanction was justified as either aimed at deterrence or just deserts. In reality, however, these motives can be, and often are, combined. Experiment 4.3 aims to address this issue by investigating the effect of providing simultaneously a deterrence *and* just-deserts justification. Because deterrence justifications signal that one is considered a potential rule breaker, we predicted that the presence of a deterrence justification still negatively affects rule compliance even when it is provided simultaneously with a just-deserts justification. To further rule out rival explanations, we also measured the extent to which participants distrusted other group members and the extent to which participants perceived the authority as legitimate (Mulder et al., 2006; Tyler, 1990). Lastly, consistent with the literature on organizational and legal rule compliance (O'Reilly & Chatman, 1986; Tyler, 1990) we distinguished between mandatory and voluntary rule compliance, and rule breaking. This distinction is relevant since authorities not only want people to follow (mandatory) rules and prevent rule breaking, but they also aim to promote voluntary acceptance of rules (i.e., increase rule-following behavior when behavior is not being monitored).

Method

Participants, design, and procedure. A total of 249 American participants (146 males; $M_{age} = 34.16$ years, $SD_{age} = 10.45$) were recruited from the Mechanical Turk website and randomly assigned to one of four sanction conditions (deterrence vs. just deserts vs. deterrence/just deserts vs. no sanction). Participants received \$1.50 for their participation.

Sanction justifications. Participants were confronted with the same scenario as in Experiment 3.2. The sanction was explained as either deterring employees from theft (deterrence condition), giving employees their just deserts (just-deserts condition), or focused on both deterring employees and giving employees their just deserts (deterrence/just-deserts condition). In the combined sanction-justification condition, the order of the deterrence and just-deserts explanation was counterbalanced. In the no-sanction condition, no information was given about a sanction.

Distrust towards self. Perceived distrust towards oneself was measured with the same six-item scale as was used in Experiment 4.2 ($\alpha = .96$).

Rule compliance. The rule compliance scales we included were used and validated in previous research (see Tyler & Blader, 2005).

Mandatory rule compliance. Mandatory rule compliance was measured on a four-item scale (adapted from Tyler & Blader, 2005). Items included, “I am inclined to use company rules to guide everything I do on the job”, “I am inclined to seek information about appropriate company policies before acting”, “I feel inclined to follow company policies and rules about how to do my job”, and “I feel inclined to comply with all organizational rules and policies” ($\alpha = .93$).

Voluntary rule compliance. Voluntary rule compliance was assessed on a five-item scale (adapted from Tyler & Blader, 2005). Items included, “I am inclined to follow organizational rules and policies when no one knows whether I did”, “I am inclined to implement the supervisor’s decisions even when he will not know whether I did”, “I am inclined to follow organizational rules and policies without questioning them”, “I am inclined to do what the supervisor expects of me, even when I don’t think its important”, and “I am inclined to happily accept all decisions made by the supervisor” ($\alpha = .92$).

Rule breaking. Rule breaking was measured on a five-item scale (adapted from Tyler & Blader, 2005). Items included, “I am inclined to find ways to undermine the supervisor”, “I am inclined to slack off towards the end of the day”, “I am inclined to come late to work”, “I am inclined to neglect to follow work rules or the instructions of the supervisor”, and “I am inclined to not do my best at work” ($\alpha = .94$).

Distrust towards others. Consistent with Mulder et al. (2006), distrust towards the other employees was measured on a four-item scale. Items included, “I feel like I cannot trust the other

employees”, “I think the other employees are tempted to break the rules”, “I feel like the other employees cannot be trusted”, and “I think the other employees are tempted to take office supplies home” ($\alpha = .94$)

Legitimacy. Legitimacy was measured on a three-item scale. Items included, “The supervisor is overextending his authority with this decision”, “The supervisor does not have the right to make these decisions”, and “I don’t think the supervisor has the authority to make such decisions” ($\alpha = .94$).

Manipulation check. Participants indicated on one item to what extent they felt personally targeted by the sanction (i.e., “I feel like this sanction is targeted at me”). The sanction-justification manipulation affected the extent to which participants felt personally targeted by the sanction ($F[1, 245] = 18.59, p < .001, \eta^2_p = .19$). Confirming the validity of our manipulation, participants in the deterrence condition felt more personally targeted by the sanction than in the just-deserts condition ($t[123] = 7.04, p < .001, d = 1.27$). Moreover, participants felt more feeling personally targeted in the combined-justification condition than in the just-deserts condition ($t[122] = 4.04, p < .001, d = 1.27$) but less compared to the deterrence condition ($t[121] = 2.19, p = .030, d = 0.40$).

Results

The means and standard deviations are reported in Table 4.1. Correlations are reported in Table 4.2.

Distrust towards self. Overall, the sanction-justification manipulation affected feelings of distrust ($F[1, 245] = 44.39, p < .001, \eta^2_p = .35$). Participants felt more distrusted in the deterrence than in the combined sanction-justification condition ($t[121] = 1.93, p = .056, d = 0.35$), just-deserts condition ($t[123] = 10.59, p < .001, d = 1.91$), and no-sanction condition ($t[123] = 8.34, p < .001, d = 1.51$). Moreover, distrust was higher in the combined-justification condition than in the just-deserts condition ($t[122] = 7.49, p < .001, d = 1.36$) and no-sanction condition ($t[122] = 5.62, p < .001, d = 1.02$). Lastly, participants in the just deserts condition felt (marginally significant) less distrusted than participants in the no sanction condition ($t[124] = 1.95, p = .054, d = 0.35$).

Rule compliance.

Mandatory rule compliance. The sanction-justification manipulation also influenced the willingness to comply with mandatory rules ($F[1, 245] = 23.86, p < .001, \eta^2_p = .23$). Mandatory rule compliance did not differ in the deterrence condition and combined sanction-justification condition ($t[121] = 0.04, p = .97, d = 0.01$) but was lower in the deterrence condition and combined sanction-justification condition compared to just-deserts condition ($t[123] = 4.27,$

$p < .001$, $d = 0.77$; $t[122] = 3.84$, $p < .001$, $d = 0.69$). Overall, mandatory rule compliance was lowest in the no-sanction condition ($ps < .001$, $ds > 0.70$).

Voluntary rule compliance. The sanction-justification manipulation affected the willingness to voluntarily comply with rules ($F[1, 245] = 13.61$, $p < .001$, $\eta_p^2 = .14$). Voluntary rule compliance was equal in the deterrence condition and combined-justification condition ($t[121] = 0.05$, $p = .96$, $d = 0.01$) but was lower in the deterrence condition and combined sanction-justification condition compared to the just-deserts condition ($t[123] = 3.74$, $p < .001$, $d = 0.67$; $t[122] = 3.71$, $p < .001$, $d = 0.67$). Overall, voluntary rule compliance was lowest in the no-sanction condition ($ps < .01$, $ds > 0.50$).

Rule breaking. The sanction-justification manipulation affected the willingness to break rules ($F[1, 245] = 11.01$, $p < .001$, $\eta_p^2 = .12$). Rule breaking was (marginally) higher in the deterrence condition compared to the combined sanction-justification condition ($t[121] = 1.81$, $p = .073$, $d = 0.44$) and just-deserts condition ($t[123] = 4.43$, $p < .001$, $d = 0.80$) but equal to the no-sanction condition ($t[123] = 0.99$, $p = .32$, $d = 0.18$). Moreover, rule breaking was higher in the combined-justification condition compared to the just-deserts condition ($t[122] = 2.45$, $p = .016$, $d = 0.44$) but lower compared to no-sanction condition ($t[122] = 2.93$, $p = .004$, $d = 0.53$, respectively). Rule breaking was lower in the just-deserts condition compared to the no-sanction condition ($t[124] = 5.92$, $p < .001$, $d = 1.06$).

Distrust towards others. The sanction-justification manipulation affected distrust towards others ($F[1, 245] = 8.02$, $p < .001$, $\eta_p^2 = .09$). Distrust towards others was equal among the sanction conditions ($ps > .45$, $ds < 0.12$) but lower in the sanction conditions than in the no-sanction condition ($ps < .001$, $ds > 0.55$).

Legitimacy. The sanction-justification manipulation affected authority legitimacy ($F[1, 245] = 4.76$, $p = .003$, $\eta_p^2 = .06$). Legitimacy was higher in the just-deserts condition compared to the other conditions ($ps < .01$, $ds > 0.49$). The deterrence, combined justification, and no-sanction condition did not differ ($ps > .39$, $ds < 0.15$).

Mediation analyses. Using a bootstrap analysis procedure with 5,000 resamples (Hayes & Preacher, 2011), we tested whether legitimacy, distrust towards others, or perceived distrust towards self mediated the effect of the sanction-justification manipulation on rule compliance. Results from the bootstrap analyses demonstrated that distrust towards others and legitimacy did not mediate the overall effect (or the contrast effects between the significant conditions) of sanction justifications on rule compliance (all 95% CIs fell between -0.20 and 0.09 without zero in the 95% CI interval). However, results from the bootstrap analyses demonstrated that perceived distrust towards the self mediated the overall effect of sanction-justification manipulation on the three forms of rule compliance (95% $CI_{mandatory\ compliance} = [-0.09, -0.01]$; 95%

$CI_{voluntary\ compliance} = [-0.12, -0.01]$; 95% $CI_{rule\ breaking} = [0.02, 0.17]$), even after controlling for both participants' distrust towards other employees and the perceived legitimacy of the supervisor. This was similar for the deterrence versus just-deserts contrast (95% $CI_{mandatory\ compliance} = [0.40, 1.35]$; 95% $CI_{voluntary\ compliance} = [0.53, 1.50]$; 95% $CI_{rule\ breaking} = [-2.10, -1.01]$) and combined-justification condition versus just-deserts contrast (95% $CI_{mandatory\ compliance} = [-0.98, -0.014]$; 95% $CI_{voluntary\ compliance} = [-0.99, -0.18]$; 95% $CI_{rule\ breaking} = [0.54, 1.37]$). More specifically, for both contrasts, adding distrust to the model decreased the significant effect of the sanction-justification manipulation on the three forms of rule compliance ($\beta_s > .19, p_s < .01$) to non-significance ($\beta_s < .10, p_s > .20$).

Discussion

Replicating and extending results from Experiments 4.1 and 4.2, results from Experiment 4.3 demonstrate that participants felt more distrusted when, (a) sanctions were justified as an attempt to deter rule breaking compared to giving people their just deserts, and (b) sanctions were justified as both deterring people *and* providing them with their just deserts compared to only giving people their just deserts. Thus, the presence of a deterrence justification negatively affected rule compliance even when it was provided simultaneously with a just-deserts justification. This provides additional support for Hypothesis 1. Perceived distrust towards the self undermined the willingness to comply with both mandatory and voluntary rules, and it increased a willingness to break rules. This provides additional support for Hypotheses 2a and 2b. Lastly, the observed effects of sanction justifications on rule compliance were not explained by participants' perceptions of the authority's legitimacy or participants' distrust towards others (e.g., other employees). As such, results from Experiment 4.3 confirm the (negative) power of justifying a sanction as an attempt to deter rule breaking—even when coupled with a just-deserts justification, participants felt distrusted and sanction effectiveness was decreased. Lastly, these results indicate that the (negative) effect of deterrence-induced distrust is not specific to one form of rule compliance, but affects a wide range of compliance outcomes that are relevant for authorities.

Table 4.1. Means and standard deviations as a function of condition for Experiment 4.3.

	Sanction		Conditions		Total
	Deterrence	Just Deserts	Deterrence/Just Deserts	No Sanction	
Personally Targeted	4.50 (2.16)	2.10 (1.62)	3.59 (2.43)	-	3.62 (2.21)
Feeling Distrusted	4.72 (1.76)	1.94 (1.11)	4.01 (1.96)	2.38 (1.37)	3.27 (1.95)
Distrust others	4.54 (1.49)	4.53 (1.69)	4.34 (1.47)	5.53 (1.36)	4.74 (1.57)
Legitimacy	3.28 (1.93)	2.28 (1.62)	3.39 (1.00)	3.09 (1.76)	3.00 (1.87)
Mandatory Compliance	5.21 (1.41)	6.12 (0.94)	5.22 (1.60)	3.91 (1.79)	5.12 (1.66)
Voluntary Compliance	4.87 (1.63)	5.78 (1.05)	4.88 (1.61)	4.09 (1.59)	4.91 (1.59)
Rule Breaking	2.75 (1.68)	1.63 (1.12)	2.22 (1.57)	3.04 (1.53)	2.41 (1.57)
Total Rule Compliance	5.10 (1.47)	6.01 (0.88)	5.29 (1.47)	4.35 (1.43)	5.21 (1.47)

Table 4.2. Correlations for Experiment 4.3.

Measure	1	2	3	4	5	6
1. Personally Targeted	-					
2. Feeling Distrusted	.49***	-				
3. Distrust towards others	.00	-.05	-			
4. Supervisor Legitimacy	.21***	.37***	.88***	-		
5. Mandatory Rule Compliance	-.19***	-.14*	-.17**	-.24***	-	
6. Voluntary Rule Compliance	-.25***	-.24***	-.12*	-.28***	.88***	-
7. Rule Breaking	.29***	.42***	.20***	.37***	-.70***	-.68***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Experiment 4.4

The previous three experiments demonstrated that deterrence justifications negatively affect self-reported rule compliance. Although highly informative, these experiments have not yet demonstrated the effects of sanction justifications on behavior. Experiment 4.4 was conducted to demonstrate the effect of sanction justifications on a *behavioral* measure of rule compliance. More specifically, we designed an experimental game in which participants could lie to an authority to evade a rule (i.e., by misreporting their earned revenue). To make sure that the established effects are not strictly due to the use of the words deterrence and just deserts, we justified sanctions as *preventing* participants from rule breaking (consistent with what deterrence aims to achieve, see Carlsmith et al., 2002) or giving participants their *deserved* punishment for breaking rules (consistent with what just deserts aims to achieve, see Carlsmith et al., 2002).

Method

Participants, design, and procedure. A total of 104 participants at Leiden University (91 females; $M_{age} = 18.99$ years, $SD_{age} = 1.44$) were randomly assigned to one of three sanction conditions (deterrence vs. just deserts vs. no sanction). Participants received €2 for their participation.

Rule compliance. Consistent with previous work on experimental tax games (Bilotkach, 2006), participants were told that they were randomly assigned to be a “worker” in a work team consisting of eight members, while one other group member was randomly assigned to be the team leader. Workers could earn extra money by finding correct words amongst scrambled letters. The rule was that forty percent of the money would have to be paid to the group leader. The team leader had to evenly redistribute this money amongst all team members such that all group members could share in a part of the revenue (i.e., similar to taxes that have to be paid to a government).

Although participants were told that the money they would earn was contingent on the number of words they found (and could thus vary across participants, depending on their productivity), all participants received €1.50—regardless of the amount of words correctly noted. Crucially, workers then had to *self-report* the amount of money had earned to the team leader. Participants were told that the team leader was able to verify if this self-reported amount was correct for only two workers (i.e., partial monitoring from an authority). As such, participants had the possibility to lie about the amount of money they earned, and in doing so both share in the collective revenue while keeping more of their own revenue. Thus, participants that report €1.50 to the team leader fully comply with the rules, whereas lower self-reported amounts reflect less rule compliance.

Sanction justifications. In the sanction conditions, the team leader decided to fine those who were caught lying by decreasing the money they earned with the task by €1. The type and severity of the sanction was thus held constant. This sanction was either justified as preventing team members from misreporting their revenue (deterrence condition) or as giving team members that misreport their revenue their deserved punishment (just-deserts condition). No information was given about a sanction in the no-sanction condition.

Distrust. Perceived distrust was measured with the following two items, “I feel distrusted by the team leader” and “I feel like the team leader does not trust me” ($\alpha = .95$). Thus, consistent with the previous three experiments, we measured the extent to which the participants felt distrusted by the authority (i.e., team leader).

Manipulation check. Feeling personally targeted by the sanction was measured with the following three items, “I feel like the sanction is targeted at me”, “I perceive the sanction as meant for me”, and “This sanction is meant for me” ($\alpha = .67$). Confirming the validity of our manipulation, participants in the deterrence condition felt (marginally significant) more personally targeted by the sanction ($M = 4.36, SD = 1.79$) than participants in the just deserts condition ($M = 3.85, SD = 1.26; t[68] = 1.77, p = .082, d = 0.43$).

Results

Distrust. Overall, the sanction-justification manipulation affected feelings of distrust ($F[1, 101] = 5.09, p = .008, \eta_p^2 = .09$). Participants felt more distrusted in the deterrence ($M = 4.43, SD = 1.79$) than in the just-deserts condition ($M = 3.53, SD = 1.43; t[68] = 2.78, p = .007, d = 0.67$) and no-sanction condition ($M = 3.60, SD = 1.45; t[67] = 2.72, p = .008, d = 0.62$). Lastly, participants in the just-deserts condition felt equally distrusted as participants in the no-sanction condition ($t[67] = .17, p = .87, d = 0.04$).

Rule compliance. On average (regardless of sanction-justification condition), participants lied to the team leader about the money they earned ($M = 88.72, SD = 33.99; t[103] = 18.38, p < .001, d = 3.62$). Participants’ lying depended on the sanction-justification manipulation ($F[1, 101] = 3.47, p = .035, \eta_p^2 = .06$). As shown in Figure 4.1, participants were more likely to lie about the amount of money they earned in the deterrence condition compared to the just-deserts condition ($t[68] = 2.51, p = .014, d = 0.61$). Lying did not differ between the deterrence condition and no-sanction condition ($t[67] = .22, p = .83, d = 0.05$) but was less likely in the just deserts condition compared to no-sanction condition ($t[67] = 2.40, p = .019, d = 0.59$).

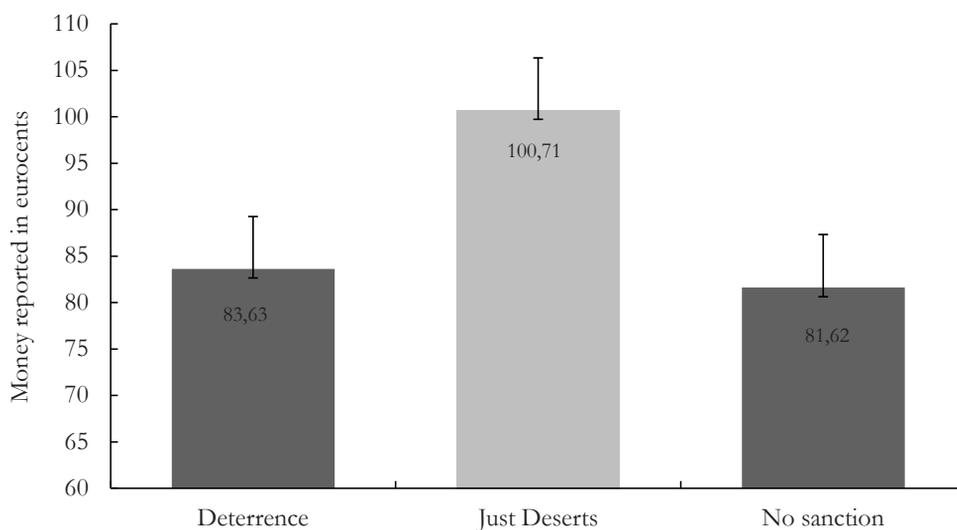


Figure 4.1. Rule compliance (indexed by lying about one's revenue) as a function of sanction condition for Experiment 4.4. Error bars indicate standard errors.

Mediation analyses. Feeling distrusted was negatively correlated with rule compliance ($r = -.44, p = .005$) and mediated the overall effect of sanction-justification condition on rule compliance (95% CI = [-16.61, -1.64]). Moreover, adding distrust to the model decreased the significant effect of the sanction-justification manipulation on rule compliance ($\beta = .30, p = .014$) to non-significance ($\beta = .17, p = .13$).

Discussion

Experiment 4.4 provided a behavioral measure of rule compliance. Participants were more likely to lie to their team leader when the sanction was justified as aimed to deter lying compared to give those who misreport their revenue their just deserts (cf. Hypothesis 2a). Consistent with Hypothesis 1 and 2b and the previous three experiments, this effect was explained by participants feeling more distrusted when the sanction was justified as aiming to deter instead of giving participants their just deserts.

General Discussion

We presented four experiments that examined how sanction justifications affect sanction effectiveness. Across different manipulations, scenarios, and sanctions, we consistently observed that people feel more distrusted when sanctions are justified as attempts to deter them from

breaking rules as compared to giving them their just deserts. This distrust is further shown to decrease people's willingness to comply with the rules of their university (Experiment 4.1), decrease their willingness to stick to the rules set by a manager (Experiments 4.2 and 4.3), and increase the extent to which people lie to their team leader to further their own self-interest (Experiment 4.4). These results strongly suggest that justifying a sanction as an attempt to deter people from breaking rules—compared to giving people their just deserts—makes people feel more distrusted (cf. Hypothesis 1) which decreases the effectiveness of a sanction (cf. Hypotheses 2a and 2b).

Theoretical and Practical Implications

The present set of studies makes several contributions to the literature on rule compliance. Previous research has mainly focused on the extent to which people use deterrence and just-deserts goals in guiding their sanction decisions (Carlsmith, 2006; Carlsmith et al., 2002; Darley et al., 2000; Greber & Jackson, 2012; Mooijman et al., 2015). The present research is to our knowledge the first to demonstrate the effects of using such goals as a *justification* on perceived distrust and rule compliance. The reported studies demonstrate that using sanction goals as a justification can lead an authority to signal the sanction's underlying considerations to the public. Sanction goals are therefore not only "hidden" motivations that drive punitive-sanction decisions (cf. Carlsmith et al., 2002; Mooijman et al., 2015); as justifications they are also highly relevant for influencing the effectiveness of sanctions. Interestingly, the present research demonstrates that sanction goals do not only influence rule compliance through affecting the severity and type of sanction. Rather, sanction goals can have a distinct and independent influence, regardless of sanction type and severity. The underlying considerations that an authority signals to others through sanction justifications are therefore highly relevant for the subsequent effectiveness of the sanction.

The way authorities affect others' tendency to comply with the rules is thus more subtle than some might think. Sanctions are not just means to increase the costs and decrease the benefits of rule breaking (cf. Nagin, 1998). Rather, sanctions are driven by philosophies and goals (Bentham, 1789/1988; Hobbes, 1651/1988; Kant, 1780/1961) that can directly affect the public. Previous research has demonstrated that powerful authorities are inclined to rely on deterrence as a sanction goal because they distrust others (Mooijman et al., 2015). The current research implies that the public is able to infer this distrust from sanctions justified as an attempt to deter rule breaking. This reinforces the notion that authorities should be careful with how they justify their sanctions. Indeed, the perceived distrust that was elicited by deterrence justifications played a unique role in undermining rule compliance. The average (negative) correlation across studies

between perceived authority distrust and rule compliance was .38, suggesting that feeling distrusted has a moderate influence on rule compliance (Cohen, 1988; Preacher & Kelley, 2011). Furthermore, this rule-undermining effect of feeling distrusted was independent from other important variables that have been shown to be relevant for cooperation and rule compliance, such as group identification (Ellemers et al., 2004; Tyler & Blader, 2000), attitudes towards authorities (Tyler & Blader, 2003), and distrust towards others (Mulder et al., 2006). The current research thus demonstrates that perceived authority distrust is of vital importance for an authority's ability to promote compliance with cooperative rules.

The current studies have direct practical relevance for authorities—judges, policy makers, and managers should be aware of the consequences that a deterrence justification can have for sanction effectiveness. To stimulate rule compliance it may be better to, (a) emphasize that sanctions are meant to give people their just deserts or (b) use deterrence justifications as a general motivation while emphasizing one's trust in specific individuals (e.g., target others with the sanction). Such sanctions foster rule compliance without making people feel (too) distrusted. Although this advice seems straightforward, it may be harder to achieve than one may think. Recent research has demonstrated that power increases people's reliance on deterrence as a goal for punishment (Mooijman et al., 2015). That is, power increases distrust towards others, which increases reliance on deterrence—but not just deserts—as a goal for punishment. Powerful authorities may thus ironically be the least inclined to emphasize the just-deserts aspects of a sanction, even though this may be the most effective course of action. However, the current research suggests that deterrence justifications can be effective when they are coupled with an affirmation of trust (e.g., are targeted at others).

Limitations and Future Directions

The current research supports our hypotheses across different samples and measures, but there are some issues to be noted. For instance, we did not focus on individuals who have already broken rules. It is possible that these individuals do feel personally targeted by the sanction and distrusted (although this distrust would be partially justified). Although potentially interesting, the majority of people tend to be rule abiding or at least perceive themselves as such (Brown, 2012; Sedikides et al., 2014). Justifying a sanction as a deterrent is therefore still likely to make them feel distrusted and thereby undermine sanction effectiveness. It is also possible that individuals who are more trusting of authorities are more, or less, likely to feel distrusted by the authority. In other words, chronic trait distrust may moderate the relationship between deterrence justifications and perceived distrust. Although we did not measure this in the current studies, we believe this to be an interesting avenue for further exploration.

Sanction goals are typically classified as deterrence or just-deserts goals (Carlsmith et al., 2002). Previous research has mainly focused on the extent to which people use these two goals in guiding their sanction decisions (Carlsmith et al., 2002; Carlsmith, 2006, 2008; Darley et al., 2000; Greber & Jackson, 2012). The present research has been consistent with this approach since we examined how these two goals independently or interactively affect sanction effectiveness. However, future research could also examine how using less known and less used sanction goals as a justification (e.g., incapacitation or rehabilitation; Darley et al., 2000) affects people's rule compliance. Lastly, because the focus in the current chapter was solely on how authorities justify their sanctioning behavior, future research could also investigate how people infer sanction goals from sanctions. Although this lies outside of the scope of the current reported studies, our theorizing and reported studies can provide a meaningful theoretical framework for formulating predictions about sanction-inferred goals. For instance, inferring that a sanction is meant to deter rule breaking can make people feel distrusted, and thus undermine rule compliance. Future research could investigate when sanctions are perceived to reflect a deterrence or just-deserts goal.

Conclusion

Authorities frequently use sanctions to promote rule compliance and often provide a justification for their use of such sanctions. Although providing sanction justifications may seem appealing, the current chapter demonstrates that these justifications can in fact influence how effective a sanction will be in promoting future rule compliance. Four experiments demonstrate that sanctions are less effective at promoting rule compliance when they are justified as deterring people from rule breaking compared to giving people their just deserts. These effects could be attributed to people feeling more distrusted when sanctions were justified as deterring them from rule breaking instead of giving them their just deserts. This suggests that although authorities have been shown to rely on deterrence as a sanction justification (Mooijman et al., 2015), this may—paradoxically—undermine the effectiveness of sanctions.

Chapter 5

Summary, Discussion, and Conclusions

Leaders have a vital interest in successfully promoting rule compliance. Organizations such as corporations and countries perform better financially when people behave according to the rules, in part because rules tend to be promote collective interests rather than individual interests (e.g., paying taxes, Akintoye, & Tashie, 2013; Parks et al., 2013). Part of the reason for this is that—when people behave according to the rules—organizations can refrain from investing resources in preventing rule breaking, thereby freeing up resources for other worthwhile investments (Kramer & Cook, 2007). Considering the importance of rule compliance, the success of leaders in promoting rule compliance with punishments (e.g., fines, penalties, or prison sentences) is surprisingly mixed (Martinez, Moyano, McCaffrey, & Oliva, 2013; Tax Justice Network, 2011). What makes some leaders more, and other leaders less, effective at promoting rule compliance? In this dissertation, I focused on the determinants and consequences of leaders' punishment goals. That is, I investigated how power affects punishment goals, and how punishment goals affect punishment effectiveness. This facilitates both a top-down understanding of leaders' punishment goals (Chapters 2 and 3) and a bottom-up understanding of how people's willingness to comply with rules is affected by leaders' punishment goals (Chapter 4). This final chapter provides an overview of this research and its theoretical and practical implications.

Summary of Findings

The studies described in Chapter 2 were conducted to provide a better understanding of punishment-goal determinants. As shown in Chapter 2, results derived from eight experiments and a European Value Survey demonstrated that power increased reliance on deterrence—but not just deserts—as a punishment goal. This power-deterrence link was explained by power fostering a distrustful mind set, and was generalizable across different measurements and inductions of power (i.e., chronic trait power, autobiographical power manipulations, manipulations of structural positions of power, and real-life managerial positions). As further shown in Chapter 2, the reliance on deterrence as a punishment goal translated into a preference for punishments that can deter people from rule breaking (i.e., public punishments and mandatory minimum for punishments).

Although these findings strongly suggest that leaders are inclined to punish to deter rule breaking, it did yet not demonstrate how relying on a deterrence or just deserts goal for punishment affects the effectiveness of the punishment. In other words, it remained unclear to what extent leaders' reliance on deterrence as a punishment goal affected the extent to which they were (in)effective at promoting rule compliance with punishments. In Chapter 4, I addressed this shortcoming by investigating the relationships between punishment-goal justifications and rule

compliance. As described in Chapter 4, results derived from four experiments demonstrated that leaders decreased people's willingness to comply with their rules when they justified their punishments as an attempt to deter people from rule breaking compared to an attempt to give people their just deserts. This (negative) deterrence-rule compliance relationship was explained by people feeling more distrusted by leaders that justified their punishments as an attempt to deter people compared to an attempt to give people their just deserts. Consistent with distrust being an important determinant of deterrence in Chapter 2, Chapter 4 thus demonstrated that—through leaders signaling their distrust—leaders undermined rule compliance when they relied on deterrence (rather than just deserts) as a punishment justification. Punishment effectiveness, in other words, was affected by the goals that leaders used to justify their punishments.

Integrating Chapters 2 and 4, this strongly suggests that—because of their power—leaders often rely on a punishment goal that is relatively ineffective at promoting rule compliance. Central to this observation was the role of distrust. Distrust explained why leaders relied on deterrence as a punishment goal (i.e., through distrusting others) *and* why reliance on deterrence as a punishment goal backfired (i.e., through people feeling distrusted by the leader). But what is the reason that power increased distrust towards others (and by doing so increased reliance on a relatively ineffective punishment goal)? In Chapter 3, I addressed these questions by investigating why power fosters distrust—and thus increases reliance on a relatively ineffective punishment goal—and how this distrust-link can be attenuated. Results from three experiments demonstrated that distrust can be explained by power holders' motivation to maintain power. That is, when individuals gained power they were shown to be more motivated to distrust others to maintain their power over others. Moreover, individuals who were more strongly motivated to maintain power distrusted others more when they gained power, and occupying a stable (versus unstable) power position attenuated the power-distrust effect. This strongly suggests that the reliance of leaders on deterrence as a punishment goal, and its relative ineffectiveness in promoting rule compliance, is in part due to power holders' motivation to maintain power over others.

Theoretical Implications and Contributions

Determinants of punishment goals

One of the most prominent questions for leaders is how to promote rule compliance. As far back as the 17th and 18th century, philosophers such as Thomas Hobbes (1651/1988) and Jeremy Bentham (1789/1988) already argued that—because people are inclined to break rules—leaders should use punishments to deter rule-breaking behavior. To this day, many corporations and government institutions follow this line of reasoning and use harsh fines and penalties to deter employees and citizens from breaking their rules (Kirchler et al., 2014). The results

presented in this dissertation demonstrate the psychological principles and assumptions that underlie this reasoning, and the consequences that follow from it. As shown in Chapter 2, relying on punishments as a means to deter others is a hallmark of powerful leaders such as top managers, policy makers and politicians. Indeed, reliance on deterrence, but not just deserts, as a punishment goal was determined by the power position that one occupied. Power thus increases leaders' reliance on punishments aimed at the deterrence of rule breaking.

These findings are in contrast with previous research that suggested the importance of just-deserts over deterrence as a goal for punishing (Darley, 2009). Concluding that the just-deserts goal is more important than the deterrence goal is therefore premature. Rather, punishing to deter people from rule breaking seems to be elicited by different concerns than punishing others to give them their just deserts. Previous research on the just-deserts goal has focused mainly on the punishment of the individual offender (micro-level perspective) rather than the punishment of the individual in its broader context such as the effects on societal norms and values (macro-level perspective). Some may thus argue that the currently observed relationship between power and deterrence is explained in part by such societal-level considerations instead of specific offender-focused considerations (consistent with the notion that power holders think more broadly and abstractly; Magee & Smith, 2013). However, Rucker, Polifroni, Tetlock, and Scott (2004) demonstrated that perceptions of societal threat (in terms of low conviction rates of criminals) increased people's desire for harsh punishments through increasing the desire to give criminals their just deserts, but not through increasing the desire to prevent future criminal behavior. Moreover, the studies described in Chapter 2 did not show that macro-level considerations—such as the collective welfare of the group—mediated the effect of power on deterrence. As such, macro-level considerations seem unfit to explain the observed findings of power on deterrence.

Instead, the current findings are perfectly consistent with Thomas Hobbes' and Jeremy Bentham's assumption that others cannot be trusted to comply with rules that promote cooperation. Indeed, they explicitly mention in their writings that punishments should deter rule breaking primarily because people care little about the rules (and would thus break rules when punishments are absent). Power thus seems to increase reliance on such "Hobbesian" and "Benthamian" philosophies; power holders tend to think that deterrence is needed because others cannot be trusted to comply with rules. As shown in Chapter 3, power holders' distrust towards others is explained by power holders' motivation to maintain power over others, and the reliance on deterrence through distrust can therefore be construed as part of the motivation to maintain power. The effect of power on deterrence is, at least in the current dissertation, thus not determined by macro-level considerations but by power holders' desire to maintain their own

power. This provides a novel perspective on the psychology of punishment goals, as punishments can also be guided by power-protecting considerations, instead of more other-oriented concerns such as the desire to restore justice (Darley, 2009), the desire to make offenders suffer (Carlsmith, Wilson, & Gilbert, 2008), or the desire to help victims (Gromet et al., 2012). This dissertation therefore makes an important contribution to the psychology of punishment goals by providing a specific psychological understanding of leaders' punishment goals.

Consequences of punishment goals

Another important contribution of the current dissertation is that it couples the determinants of punishment goals to its consequences for rule compliance. More specifically, I investigated how and why leaders' reliance on punishment goals previously emphasized by philosophers such as Thomas Hobbes (1651/1988) and Immanuel Kant (1780/1961), affects the effectiveness of the punishment in terms of promoting rule compliance. As shown in Chapter 4, leaders that justified their punishments as an attempt to deter people from rule breaking (cf. Thomas Hobbes) were less effective at promoting rule compliance than leaders that justified their punishments as an attempt to give people their just deserts (cf. Immanuel Kant). These findings show that understanding the psychology of punishment goals is not only important for understanding determinants of punishments; it also fosters a specific understanding of how punishment goals affect punishment effectiveness. More specifically, Chapter 4 demonstrated that people felt more distrusted by leaders that justified their punishments as an attempt to deter them from rule breaking compared to provide them with their just deserts, and that this perceived distrust mediated the effect of punishment-goal justifications on rule compliance. This finding is consistent with the findings from Chapter 2 that distrust underlies leaders' reliance on deterrence—but not just deserts—as a punishment goal, and the more general notion that leaders can signal such underlying considerations to others through their behavior. For instance, in a series of experiments, McKenzie, Liersch, and Finkelstein (2009) demonstrated that people perceive the fact that they are by default “not an organ donor” as revealing the implicit recommendation of society's policy makers. Moreover, in three experiments conducted by Tannenbaum, Valasek, Knowles, and Ditto (2013), participants accurately inferred policy makers' negative attitudes towards the overweight from policy makers' decisions to increase (or decrease) health care premiums for overweight employees. The findings presented in this dissertation thus demonstrate that people can similarly, and accurately, infer leaders' distrust from their punishment-goal justifications.

Furthermore, the observed rule-undermining effect of feeling distrusted by a leader is consistent with the literature on interpersonal justice. This literature states that the perceived interpersonal treatment of people by their leaders is an important determinant of people's willingness to comply with leaders' rules (e.g., relational model of authority and the group-value model of procedural justice; Tyler & Lind, 1992). The assumption in this literature is that people comply with leaders' rules in part because of the "social contract" between people and their leaders; people are willing to comply with leaders' rules as long as leaders ensure that justice is done (i.e., people who cooperate are rewarded and people who break rules are punished; see Plato 380 BC/1992). The findings described in Chapter 4 are consistent with this notion. A perceived lack of trust from leaders may seem unwarranted when no prior breach of rules was displayed, and may thus seem unjust, thereby undermining rule compliance. Interestingly, this notion of the social contract is also central to deterrence theory and just-deserts theory (Hobbes, 1651/1988; Kant, 1780/1961). Both theories differ, however, in how leaders should uphold the social contract—by deterring rule breaking with punishments or by giving rule breakers their just deserts with punishments? The current findings strongly suggest that, compared to deterrence, leaders using just-deserts as a justification for their punishments best uphold the social contract, as demonstrated in people's increased willingness to comply with rules. This dissertation therefore makes an important contribution to understanding the consequences of punishment goals by providing a specific psychological understanding to how and why using punishment goals as a justification affects punishment effectiveness.

Integrating determinants and consequences of punishment goals

By integrating both determinants and consequences of punishment goals, this dissertation also addresses an important limitation of previous research on punishment goals. Previous research has largely neglected to investigate the consequences of punishment goals and was therefore unable to connect the determinants of punishment goals to their consequences. Indeed, are leaders' punishment goals beneficial or detrimental for how effective they are at promoting rule compliance? In the current dissertation, I demonstrated that leaders' power increases reliance on deterrence as a punishment goal through increasing distrust towards others *and* that relying on deterrence as a punishment justification is relatively ineffective at promoting rule compliance because it signals leaders' distrust towards others. Distrust is thus shown to directly connect leaders' reliance on deterrence as a punishment goal to the subsequent ineffectiveness of the punishment. Moreover, I demonstrated in Chapter 3 that this distrust is in part explained by leaders' concern to maintain power, thereby demonstrating that leaders' reliance on deterrence as a (relatively ineffective) punishment goal is part of leaders' concern for maintaining power. Thus,

although power increases leaders' reliance on punishments to deter rule-breaking behavior, this can paradoxically stimulate instead of prevent rule-breaking behavior.

This demonstrates that it is fruitful to theorize both on how and why leaders punish, and how and why such punishments affect rule compliance. An integrative understanding of leaders' punishment behavior and subsequent effectiveness, for instance, sheds light onto why informing leaders about how they should punish does not always directly translate into behavioral change. Even if leaders are explicitly told that they should justify their punishments as an attempt to achieve just deserts, and not as an attempt to deter rule breaking, their power still edges them towards reliance on deterrence as a punishment goal. Especially in the long run, when leaders' strategies are subject to unexpected problems and crises, leaders may fall back on their psychological tendency to deter rule breaking with punishments, thereby promoting rule compliance with punishments relatively ineffectively. This strongly suggests that it is leaders' power and subsequent distrust towards others, that is a potential problem for their ability to promote rule compliance effectively. This dissertation therefore makes an important contribution to an integrated understanding of leaders' punishment behavior and peoples' reaction to such punishment behavior. That is, by providing a specific psychological understanding of how and why leaders' inclination to rely on specific punishment goals affects people's willingness to comply with the leader's rules.

Practical Implications and Contributions

Powerful leaders frequently punish others for their rule-breaking behavior. Leaders therefore need to understand how their power affects their punishment goals, and how reliance on such punishment goals affects people's willingness to comply with rules. The current dissertation demonstrates that the power hierarchy of an organization can increase (or decrease) leaders' reliance on the deterrence goal for punishment. Especially in organizations where rule compliance is of great importance (e.g., financial institutions), leaders are often given a great deal of power over others. After the recent financial crisis in 2008, for instance, many financial institutions hired rule-compliance officers who were given the power to oversee bank manager's decisions, create rules, and guidelines, and enforce rules with punishments (Trevino, den Nieuwenboer, Kreiner, & Bishop, 2014; Trevino, den Nieuwenboer, & Kisch-Gephart, 2014). Ironically, the power that such individuals have can thus impede the extent to which they are effective at promoting compliance with such rules. This strongly suggests that organizations that afford a great deal of power to leaders make these individuals relatively ineffective at promoting rule compliance. That is, leaders' power and reliance on deterrence as a relatively ineffective punishment goal are positively related—decreasing leaders' power thus decreases reliance on

deterrence as a punishment goal. Moreover, the findings described in Chapter 3 hint at the notion that decreasing the extent to which leaders feel threatened (e.g., feel stable and certain about their power position) may also decrease their reliance on deterrence as a punishment goal. An interesting implication may thus be to hire temporary leaders to enforce rules through punishments (Andreoni & Gee, 2012). These “guns for hire” know that they will have to leave their job after a certain amount of time, thereby reducing their feelings of power instability and possibly reducing their reliance on deterrence as a punishment goal.

These implications are important because leaders are frequently told to achieve behavioral change through creating a shared vision and providing inspiration to others (i.e., transformational leadership; Bass, 1985). However, leaders still have to enforce rules with punishments, as inspirational leadership is not always sufficient for promoting rule compliance (Wyld, 2013). Although creating “flatter” organizations (with smaller power differences between people) may seem like a possible solution to counteract the negative effects of power on punishment effectiveness, power differences within organizations are unlikely to disappear any time soon. Part of the reason is that the existence of a power hierarchy can be functional in terms of creating clarity with regard to “who has to do what”, thereby increasing the extent to which people are effective at their respective jobs (Anderson & Brown, 2010). Leaders should thus at least be aware of the negative effects of using deterrence as a punishment goal and should also be motivated to create explicit strategies towards communicating about punishments in terms of justice. Leaders should in particular be aware of how their own distrust towards others affects their punishment goals, and others’ willingness to comply with rules.

Even though leaders’ own distrust towards others may at times be hard to attenuate, leaders should also realize that the consequences of such distrust might be contrary to their own interest. Although leaders’ distrust may at times protect them from losing power to others in the short term, the rule-undermining effect of their distrust is unlikely to be an optimal long-term strategy. Leaders’ power positions are often dependent on their ability to successfully promote rule compliance in the long run (e.g., politicians can fail to get reelected if crime rates increase). Consistent with the notion that changes are easiest implemented when such changes appeal to people’s self-interest (Parks et al., 2013), leaders should be aware of the fact that their own power position might be best maintained by using punishments aimed at giving rule breakers their just deserts and thereby achieve justice.

Limitations and Future Research

Although the studies that I have presented in this dissertation are consistent with my hypotheses and deploy a variation of different methodologies, there are still some methodological

limitations that should be addressed. For instance, one of these limitations is the experimental nature of many of the empirical data reported in this dissertation. With one exception (i.e., data from the European Value Survey in Chapter 2; Study 2.3a), all data were derived from carefully constructed laboratory experiments. Such experiments provided advantages. First, it provided the advantage of being able to manipulate specific variables of interest (e.g., power; punishment justifications) while excluding alternative explanations (e.g., income differences between power positions; pre-existing animosity between people and their leaders). Secondly, it allowed me to draw causal inferences between variables, thereby moving beyond reporting simple correlations. Although the advantages of such an experimental approach are profound, it still begs the question to what extent the observed effects of power on deterrence, and deterrence on rule compliance are reflected in more dynamic and real-life settings. For instance, future research could investigate how the justifications that rule-compliance officers in financial institutions provide for their actions affects their effectiveness in promoting rule compliance. Future research could also investigate how tax-compliance officers' justifications for monitoring (and possibly punishing) taxpayers affect the relationship between the tax-compliance officer and taxpayer, and consequently taxpayers' willingness to comply with tax rules. Although these studies may be challenging, the current dissertation provides a specific theoretical framework for exploring these interesting dynamics.

Consistent with the above, future research should also apply a more dynamic and interactive approach to understanding the relationships between leadership, punishment, and rule compliance. Future research could, for instance, investigate the role of status, its effects on punishment goals and subsequent punishment effectiveness. More specifically, leaders' perceptions of status may increase a desire for positive group functioning and cohesion (in part because leaders' status is dependent on others' positive evaluations; see Blader & Chen, 2012), which may increase reliance on just deserts as a punishment goal. But how do people react to such status-induced punishment? And do leaders, in turn, adjust their own punishment behavior when they receive feedback about people's reactions to their punishments? And do such adjustments lead to more (or less) effective punishments? Without an understanding of both leaders' behavior and followers' reactions to leaders' behavior, answering such questions is difficult. Future research should therefore apply a more interactive research approach to understanding leaders' punishment (effectiveness).

Lastly, future research could also benefit from a more functionalist-approach to the study of power (i.e., whether power holders act in ways that is functional for maintaining their power). In recent years, we have seen an increase in the number of studies on the social-cognitive effects of power (e.g., effects of the experience of power on perspective taking; Galinsky et al., 2006),

but few studies have been conducted on how power holders behave to maintain their own power (but see Case & Maner, 2014; Maner & Mead, 2010). Indeed, do some effects of power result from strategic considerations to maintain power? The current dissertation strongly suggests that power can affect relevant behavioral outcomes as part of a strategy to maintain power, thereby implying that strategic considerations play a role for power holders. Since individuals occupying real-life power positions are often reluctant to let go of such positions (Fehr et al., 2013), a further understanding of the psychology of power from a functional perspective seems worthwhile.

Conclusion

Based on the findings in this dissertation, it seems safe to conclude that understanding the determinants and consequences of leaders' punishment goals is important. Together, the findings reported in Chapters 2, 3, and 4 on the role of power, distrust, punishment goals, and punishment effectiveness provide new insights into safeguarding leaders against relying on suboptimal punishment goals. These findings also provide new evidence as to what punishment goals leaders should use to promote rule compliance most effectively.

References

- Aharoni, E., & Fridlund, A. J. (2011). Punishment without reason: Isolating retribution in lay punishment of criminal offenders. *Psychology, Public Policy, and the Law*, *18*, 599–625.
- Akintoye, I. R., & Tashie, G. A. (2013). The effect of tax compliance on economic growth and development in Nigeria. *British Journal of Arts and Social Sciences*, *11*, 222–231.
- Anderson, C., & Brion, S. (2014). Perspectives on power in organizations. *Annual review of Organizational Psychology and Organizational Behavior*, *1*, 67–97.
- Anderson, C., & Brown, C. (2010). The functions and dysfunctions of hierarchy. *Research in Organizational Behavior*, *30*, 55–89.
- Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and the proclivity for risk. *European Journal of Social Psychology*, *36*, 511–536.
- Anderson, C., & John, O. P., & Keltner, D. (2011). The personal sense of power. *Journal of Personality*, *80*, 313–344.
- Anderson, C., Willer, R., Kilduff, G., & Brown, C. L. (2012). The drive to defer: Why some individuals prefer lower status. *Journal of Personality and Social Psychology*, *102*, 1077–1088.
- Andreoni, A., & Gee, L. K. (2012). Guns for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, *96*, 1036–1046.
- Arvey, R. D., & Ivancevich, J. M. (1980). Punishment in organizations: A review, propositions, and research suggestions. *Academy of Management Review*, *5*, 123–132.
- Ball, G. A., Treviño, L. K., & Sims, H. P. (1994). Just and unjust punishment: Influences on subordinate performance. *Academy of Management Journal*, *37*, 299–323.
- Balliet, D., & Van Lange, P. A. M. (2012). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*, 1090–1112.
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.
- Bentham, J. (1789/1988). *An introduction to the principles of morals and legislation*, New York, NY: Prometheus Books.
- Berg, J., Dickaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122–142.
- Bilotkach, V. (2006). A tax evasion-bribery game: Experimental evidence from Ukraine. *The European Journal of Comparative Economics*, *3*, 31–49.
- Blader, S., & Chen, Y. (2012). Differentiating the effects of status and power: A justice perspective. *Journal of Personality and Social Psychology*, *102*, 994–1014.
- Blader, S. L. & Tyler, T. R. (2003). A four component model of procedural justice: Defining the meaning of a "fair" process. *Personality and Social Psychology Bulletin*, *29*, 747-758.
- Brown, J. D. (2012). Understanding the "Better than Average" effect: Motives (still) matter.

- Personality and Social Psychology Bulletin*, 38, 209–219.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3–5.
- Butterfield, K. D., Treviño, L. K., & Ball, G. A. (1996). Punishment from the manager's perspective: A grounded investigation and inductive model. *Academy of Management Journal*, 39, 1479–1512.
- Butterfield, K. D., Treviño, L. K., Wade, K. J., & Ball, G. A. (2005). Organizational punishment from the manager's perspective: An exploratory study. *Journal of Managerial Issues*, 17, 363–382.
- Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology*, 42, 437–451.
- Carlsmith, K. M. (2008). On justifying punishment: The discrepancy between words and actions. *Social Justice Research*, 21, 119–137.
- Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Social and Personality Psychology*, 83, 284–299.
- Carlsmith, K. M., Wilson, T. D., & Gilbert, D. T. (2008). The paradoxical consequences of revenge. *Journal of Personality and Social Psychology*, 95, 1316–1324.
- Case, C. R., & Maner, J. K. (2014). Divide and conquer: When and why leaders undermine the cohesive fabric of their group. *Journal of Personality and Social Psychology*, 107, 1033–1050.
- Chemerinsky, E. (2004). Life in prison for shoplifting: Cruel and unusual punishment. *Human Rights*, 31, 11–13.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. New York, NY: Academic Press.
- Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, 92, 909–927.
- Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General*, 143, 2279–2286.
- Cropanzano, R., Bowen, D. E., & Gilliland, S. W. (2007). The management of organizational justice. *Academy of Management Perspective*, 4, 34–48.
- Darley, J. M. (2009). Morality in the law: The psychological foundations of citizens' desires to punish transgressions. *Annual Review of Law and Social Science*, 5, 1–23.
- Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior*, 24, 659–684.

- Das, T. K., & Teng, B. (2004). The risk-based view of trust: A conceptual framework. *Journal of Business and Psychology, 19*, 85–116.
- De Cremer, D., & Van Knippenberg, D. (2002). How do leaders promote cooperation? The effects of charisma and procedural fairness. *Journal of Applied Psychology, 5*, 858–866.
- De Keijser, J. W., Van Der Leeden, R., & Jackson, J. L. (2002). From moral theory to penal attitudes and back: A theoretically integrated modeling approach. *Behavioral Sciences and the Law, 20*, 317–335.
- de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science, 305*, 1254–1258.
- Downes, D., & Hansen, K. (2006). *Welfare and punishment: The relationship between welfare spending and imprisonment*. London: Crime and Society Foundation.
- Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health, 58*, 637–642.
- Ellemers, N. (2012). The Group Self. *Science, 336*, 848–852.
- Ellemers, N., De Gilder, D., & Haslam, S. A. (2004). Motivating individuals and groups at work: A social identity perspective on leadership and group performance. *Academy of Management Review, 29*, 459–478.
- Ellemers, N., Doosje, B., & Spears, R. (2004). Sources of respect: Effects of being liked by ingroups and outgroups. *European Journal of Social Psychology, 34*, 1551–172.
- Ellemers, N., Wilke, H., & Van Knippenberg, A. (1993). Effects of the legitimacy of low group or individual status on individual and collective identity enhancement strategies. *Journal of Personality and Social Psychology, 64*, 766–778.
- EVS (2011). European Values Study 1981: Integrated Dataset (EVS 2011). *GESIS Data Archive, Cologne*. ZA4438 Data file version 3.0.0.
- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*, 63–87.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140.
- Fehr, E., Herz, H., & Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. *American Economic Review, 103*, 1325–1359.
- Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford University Press.
- Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist, 48*, 621–628.
- Gabor, T., & Crutcher, N. (2002). *Mandatory Minimum Penalties: Their Effects on Crime, Sentencing Disparities, and Justice System Expenditures*. Ottawa: Justice Canada.
- Gaertner, S. L., & Dovidio, J. F. (2000). Reducing intergroup bias: *The common ingroup-*

- identity model*. Philadelphia, PA: The Psychology Press.
- Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, *85*, 453–466.
- Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science*, *17*, 1068–1074.
- Gerber, M. M., & Jackson, J. (2013). Retribution as revenge and retribution as just deserts. *Social Justice Research*, *26*, 61–80.
- Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies*, *29*, 1–17.
- Goodwin, S. A., Gubin, A., Fiske, S. T., & Yzerbyt, V. Y. (2000). Power can bias impression processes: Stereotyping subordinates by default and by design. *Group Processes & Intergroup Relations*, *3*, 227–256.
- Gromet, D. M., Okimoto, T. G., Wenzel, M., & Darley, J. (2012). A victim-centered approach to justice? Victim satisfaction effects on third-party punishments. *Law and Human Behavior*, *36*, 375–389.
- Guinote, A. (2007a). Behavioral variability and the situated focus theory of power. *European Review of Social Psychology*, *18*, 256–295.
- Guinote, A. (2007b). Power and goal pursuit. *Personality and Social Psychology Bulletin*, *33*, 1076–1087.
- Hayes, A. F., Preacher, K. J., & Myers, T. A. (2011). Mediation and the estimation of indirect effects in political communication research. In E. P. Bucy & R. Lance Holbert (Eds.), *Sourcebook for political communication research: Methods, measures, and analytical techniques*. (pp. 434–465). New York: Routledge.
- Hobbes, T. (1651/1996). *Leviathan*. New York: Oxford University Press.
- Inesi, E. M. (2010). Power and loss aversion. *Organizational Behavior and Human Decision Processes*, *112*, 58–69.
- Inesi, E. M., Gruenfeld, D. H., & Galinsky, A. D. (2012). How power corrupts relationships: Cynical attributions for others' generous acts. *Journal of Experimental Social Psychology*, *48*, 795–803.
- Jabour, B. (2013, September 27). Queensland plan to 'name and shame' young offenders under fire. *The Guardian*. Retrieved January 20, 2014, from <http://www.theguardian.com/world/2013/sep/27/crime-australia-queensland>.
- Josephs, R. A., Sellers, J. G., Newman, M. L., & Mehta, P. H. (2006). The mismatch effect: When testosterone and status are at odds. *Journal of Personality and Social Psychology*, *90*, 999–1013.
- Kant, I. (1952). The science of right. (W. Hastie, Trans.). In R. Hutchins (Ed.), *Great books of*

- the western world* (pp. 397–446). Edinburgh: T. & T. Clark. (Original work published 1790).
- Keller, L. B., Oswald, M. E., Stucki, I., & Gollwitzer, M. (2010). A closer look at an eye for an eye: Layperson's punishment decisions are primarily driven by retributive motives. *Social Justice Research, 23*, 99–116.
- Keltner, D., Anderson, C., & Gruenfeld, D. H. (2003). Power, approach and inhibition. *Psychological Review, 23*, 265–284.
- Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal, 18*, 769–783.
- Kipnis, D., Castel, P. J., Gergen, M., & Mauch, D. (1976). Metamorphic effects of power. *Journal of Applied Psychology, 61*, 127–135.
- Kirchler, E., Kogler, C., & Muelbacher, S. (2014). Cooperative tax compliance: From deterrence to deference. *Current Directions in Psychological Science, 23*, 87–92.
- Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology, 50*, 984–993.
- Kramer, R. M., & Schaffer, J. (2014). Misconnecting the dots: Origins and dynamics of intergroup paranoia. In J.-W. van Prooijen & P. A. M. Van Lange (Eds.), *Power, politics and paranoia: Why people are suspicious of their leaders*. New York: Cambridge University Press.
- Lammers, J., & Stapel, D. A. (2009). How power influences moral thinking. *Journal of Personality and Social Psychology, 97*, 279–289.
- Lammers, J., Galinsky, A. D., Gordijn, E. H., & Otten, S. (2012). Power increases social distance. *Social and Personality Science, 3*, 282–290.
- Lammers, J., Stoker, J. I., Jordan, J., Pollmann, M. M. H., & Stapel, D. A. (2011). Power increases infidelity among men and women. *Psychological Science, 22*, 1191–1197.
- Langlois, J. (2012). India to name and shame rapists. *Global Post*. Retrieved January 20, 2014 From <http://www.globalpost.com/dispatch/news/regions/asiapacific/india/121227/india-name-and-shame-rapists>.
- Lee, F., & Tiedens, L. Z. (2001). Is it lonely at the top? Independence and interdependence of power-holders. In B. M. Staw & R. I. Sutton (Eds.), *Research in organizational behavior* (pp. 43–91). Greenwich, CT: JAI Press.
- Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management, 32*, 991–1022.
- Lount, R. B. Jr., & Petitt, N. C. (2012). The social context of trust: The role of status. *Organizational Behavior and Human Decision Processes, 117*, 15–23.

- Magee, J. C., & Galinsky, A. D. (2008). Social hierarchy: The self-reinforcing nature of power and status. *Academy of Management Annals*, 2, 351–398.
- Magee, J. C., & Smith, P. K. (2013). The social distance theory of power. *Personality and Social Psychology Review*, 17, 158–186.
- Magee, J. C., Kilduff, G. J., & Heath, C. (2011). On the folly of principal's power: Managerial psychology as a cause of bad incentives. *Research in Organizational Behavior*, 31, 25–41.
- Mandatory Minimum Terms Result In Harsh Sentencing (2007). Retrieved January 20, 2014, from [http://www.uscourts.gov/News/NewsView/07-06 26/Mandatory_Minimum_Terms_Result_In_Harsh_Sentencing.aspx](http://www.uscourts.gov/News/NewsView/07-06%20Mandatory_Minimum_Terms_Result_In_Harsh_Sentencing.aspx).
- Maner, J. K., Gailliot, J. K., Butz, D., & Peruche, B. M. (2007). Power, risk and the status-quo: Does power promote riskier or more conservative decision-making? *Personality and Social Psychology Bulletin*, 33, 451–462.
- Maner, J. K., & Mead, N. (2010). The essential tension between leadership and power: When leaders sacrifice group goals for the sake of self-interest. *Journal of Personality and Social Psychology*, 99, 482–497.
- Martinez, M. (2015, January 24). Colorado woman gets 4 years for wanting to join ISIS. *CNN*. Retrieved, January 28 from <http://edition.cnn.com/2015/01/23/us/colorado-woman-isis-sentencing>.
- Martínez-Moyano, I. D., McCafrey, P. D., & Oliva, R. (2013). Drift and adjustment in organizational rule compliance: Explaining the ‘regulatory pendulum’ in financial markets. *Organization Science*, 25, 321–338.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709–734.
- Mayer, D. M., Kuenzi, M., Greenbaum, R., Bardes, M., & Salvador, R. (2009). How low does ethical leadership flow? Test of a trickle-down model. *Organizational Behavior and Human Decision Processes*, 108, 1–13.
- McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, 17, 414–420.
- Mead, N. L., & Maner, J. K. (2012). On keeping your enemies close: Powerful leaders seek proximity to in-group power-threats. *Journal of Personality and Social Psychology*, 102, 576–591.
- Molenaar, 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, 175–183.

- Mooijman, M., & Stern, C. (2016, in press). When perspective taking creates a motivational threat: The case of conservatism, same-sexual behavior, and anti-gay attitudes. *Personality and Social Psychology Bulletin*.
- Mooijman, M., & Van Dijk, W. W. (2015). The self in moral judgment: How self-affirmation affects the moral condemnation of harmless sexual taboo violations. *Cognition and Emotion*, 7, 1334–1342.
- 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, 75–89.
- Mooijman, M., Van Dijk, W. W., Van Dijk, E., & Ellemers, N. (under revision, 2015). On sanction effectiveness: How and why sanction-goal justifications affect rule compliance. *Journal of Personality and Social Psychology*.
- Mooijman, M., Van Dijk, W. W., Van Dijk, E., & Ellemers, N. (under review). Power, distrust, and the motivation to maintain power.
- Mulder, L. B., Van Dijk, E., De Cremer, D., & Wilke, H. A. M. (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology*, 42, 147–162.
- Mulder, L., & Nelissen, R. (2010). When rules really make a difference: The effect of cooperation rules and self-sacrificing leadership on moral norms in social dilemmas. *Journal of Business Ethics*, 95, 57–72.
- Nagin, D. (1998). Criminal deterrence research at the outset of the twenty-first century. In M. Tonry (Ed.), *Crime and justice: A review of research* (Vol. 23, pp. 1–42). Chicago: University of Chicago Press.
- Nelissen, 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, 543–553.
- Okimoto, T., & Wenzel, M. (2011). Third-party punishment and symbolic intragroup status. *Journal of Experimental Social Psychology*, 47, 709–718.
- O'Reilly, C. A., & Chatman, J. A. (1986). Organizational commitment and psychological attachment. *Journal of Applied Psychology*, 71, 492–499.
- 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, 119–165.
- Philips, M. (2011, May 02). Osama Bin Laden Dead. Retrieved, April 22nd from <https://www.whitehouse.gov/blog/2011/05/02/osama-bin-laden-dead>.
- Plato (380 BC/1992). *Republic*. (Trans. G. M. A. Grube, Revised by C. D. C. Reeve) Indianapolis,

IN: Hackett Publishing Company.

- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*, 93–115.
- Raul, A. C. (2002, October 11). In era of broken rules, society breaks. *Los Angeles Times*. Retrieved, October 7th from <http://articles.latimes.com/2002/oct/11/opinion/oe-raul11>.
- Robinson, P. H. (2005). *Law without justice: Why criminal law doesn't give people what they deserve*. Oxford, Oxford University Press.
- Rucker, D. D., Galinsky, A. D., & DuBois, D. (2012). Power and consumer behaviour: How power shapes who and what consumers value. *Journal of Consumer Psychology, 22*, 352–368.
- Rucker, D., Polifroni, M. D., Tetlock, P., & Scott, A. (2004). On the assignment of punishment: The impact of general-societal threat and the moderating role of severity. *Personality and Social Psychology Bulletin, 30*, 673–684.
- Schneider, F. (2008). Size and development of the shadow economy in Germany, Austria and other OECD countries: Some preliminary findings. *La Revue Economique, 15*.
- Sedikides, C., Meek, R., Alicke, M. D., & Taylor, S. (2014) Behind bars but above the bar: Prisoners consider themselves more prosocial than non-prisoners. *British Journal of Social Psychology, 53*, 396–403.
- Seip, E. C., Van Dijk, W. W., & Rotteveel, M. (2014). Anger motivates costly punishment of unfair behavior. *Motivation and Emotion, 38*, 578–588.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). *Life after p-hacking*. Paper presented at the Fourteenth Annual Meeting of the Society for Personality and Social Psychology, New Orleans, LA, USA.
- Smith, P. K., & Trope, Y. (2006). You focus on the forest when you're in charge of the trees: Power priming and abstract information processing. *Journal of Personality and Social Psychology, 90*, 578–596.
- Smith, P. K., Jostmann, N. B., Galinsky, A. D., & van Dijk, W. (2008). Lacking power impairs executive functions. *Psychological Science, 19*, 441–447.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21*, 261–302.
- Stouten, J., De Cremer, D., & Van Dijk, E. (2005). I'm doing the best I can (for myself):

- Leadership and variance of harvesting in resource dilemmas. *Group Dynamics: Theory, Research, and Practice*, 9, 205–211.
- Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., & Kirsch, P. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *NeuroImage*, 54, 671–80.
- Tam, T., Hewstone, M., Kenworthy, J., & Cairns, E. (2009). Intergroup trust in Northern Ireland. *Personality and Social Psychology Bulletin*, 35, 45–59.
- Tannenbaum, D., Valasek, C. J., Knowles, E. D., & Ditto, P. H. (2013). Incentivizing wellness in the workplace: Sticks (not carrots) send stigmatizing signals. *Psychological Science*, 24, 1512–1522.
- Tax Justice Network (2011). The costs of tax abuse 2011. Retrieved from <http://www.taxjustice.net/2014/04/01/cost-tax-abuse-2011>.
- Texas judge orders convicted drunk driver to public humiliation (2012). *Fox News*. Retrieved from <http://www.foxnews.com/us/2012/04/21/texas-judge-orders-convicted-drunk-driver-to-public-humiliation>.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge*. Yale University Press: New Haven.
- Treviño, L. K., Den Nieuwenboer, N. A., & Kish-Gephart, J. J. (2014). (Un)ethical behavior in organizations. *Annual Review of Psychology*, 65, 635–660.
- Treviño, L. K., Den Nieuwenboer, N. A., Kreiner, G. E., & Bishop, D. (2014). Legitimizing the legitimate: The challenges of ethics and compliance work in organizations. *Organizational Behavior and Human Decision Processes*, 123, 186–205.
- Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimization. *Annual Review of Psychology*, 57, 375–400.
- Tyler, T. R. (2012). Justice theories. In P. Van Lange, A. Kruglanski, and T. Higgins (Eds.). *Handbook of Theories of Social Psychology*. Thousand Oaks, CA: Sage.
- Tyler, T. R., & Boeckmann, R. (1997). Three strikes and your out, but why? The psychology of public support for punishing rule breakers. *Law & Society Review*, 31, 237–265.
- Tyler, T. R., & Blader, S. L. (2000). *Cooperation in groups: Procedural justice, social identity, and behavioral engagement*. Philadelphia, PA: Psychology Press.
- Tyler, T. R., & Blader, S. L. (2001). Identity and prosocial behavior in groups. *Group Processes and Intergroup Relations*, 4, 207–226.
- Tyler, T. R., & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review*, 7, 349–361.
- Tyler, T. R., & Blader, S. L. (2005). Can businesses effectively regulate employee conduct? The

- antecedents of rule following in work settings. *Academy of Management Journal*, *48*, 1143–1158.
- Tyler, T. R., & Lind, E. A. (1992). A relational model of authority in groups. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 25, pp. 115–191). New York, NY: Oxford University Press.
- Van der Toorn, J., Feinberg, M., Jost, J. T., Kay, A. C., Tyler, T. R., Willer, R., & Wilmoth, C. (2015). A sense of powerlessness fosters system justification: Implications for the legitimation of authority, hierarchy, and government. *Political Psychology*, *36*, 93–110.
- Van Dijk, E., Mulder, L. B., De Kwaadsteniet, E. W. (2014). For the common good? The use of sanctions in social dilemmas. In P. A. M., Van Lange., B. Rockenbach., & T. Yamagishi. (Eds.), *Reward and punishment in social dilemmas* (pp. 70–85). New York, Oxford University Press.
- Van Kleef, G. A., Oveis, C., Van der Löwe, L., Luogokan, A., Goetz, J., & Keltner, D. (2008). Power, distress and compassion: Turning a blind eye to the suffering of others. *Psychological Science*, *19*, 1315–1322.
- 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*, 125–141.
- Van Prooijen, J-W., Coffeng, J., & Vermeer, M. (2014). Power and retributive justice: How trait information influences the fairness of punishment among power holders. *Journal of Experimental Social Psychology*, *50*, 190–201.
- Vidmar, N., & Miller, D. T. (1980). Social psychological processes underlying attitudes toward legal punishment. *Law & Society Review*, *14*, 565–602.
- Wenzel, M., Okimoto, T. G., Feather, N. T., & Platow, M. J. (2008). Retributive and restorative justice. *Law and Human Behavior*, *32*, 375–389.
- Wilkinson, R. G., & Pickett, K. E. (2007). The problems of relative deprivation: Why some societies do better than others. *Social Science and Medicine*, *65*, 1965–1978.
- Wilkinson, R. G., & Pickett, K. E. (2009). *The spirit level: Why equality is better for everyone*. London: Penguin Books.
- Williams, M. J. (2014). Serving the self from the seat of power: Goals and threats predict leaders' self-interested behavior. *Journal of Management*, *40*, 1365–1395.
- Willis, G. B., & Guinote, A. (2012). The effects of social power on goal content and goal striving: A situated perspective. *Social and Personality Psychology Compass*, *5*, 706–719.
- Wiltermuth, S., & Flynn, F. (2013). Power, moral clarity and punishment in the workplace. *Academy of Management Journal*, *4*, 1002–1023.

- Wojciszke, B., & Kujalowicz–Struzynska, A. (2007). Power influences self-esteem. *Social Cognition, 25*, 472–494.
- Wyld, D. C. (2013). Transformational leadership: When is it redundant? *Academy of Management Journal, 27*, 2011–2013.
- Xiao, E., & Houser, D. (2011). Punish in public. *Journal of Public Economics, 95*, 1006–1017.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*, 110–116.
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*, 129–166.
- Zand, D. (1997). *The leadership triad: Knowledge, trust, and power*. New York: Oxford University Press.

Dutch Summary (Samenvatting)

Het handhaven van regels is een primaire taak van leiders in de politiek of het bedrijfsleven. De mate waarin mensen zich aan de regels houden correleert namelijk sterk met het financiële succes van een land of organisatie. Dit komt doordat regels normaal gesproken collectieve (belasting betalen, op tijd op het werk komen) in plaats van individuele belangen (belasting ontwijken, te laat op het werk komen) benadrukken (Akintoye & Tashie, 2013; Parks et al., 2013). Hoewel regel handhaving dus van belang is, is het voor leiders niet altijd even makkelijk om mensen zich aan de regels te laten houden. Het Amerikaanse bedrijf Enron ging bijvoorbeeld failliet omdat werknemers zich niet aan ethische regels hielden (Raul, 2002). Ook verschillen landen in grote mate van elkaar in hoe succesvol ze zijn in het ophalen van voldoende belastinggeld (e.g., belastingontwijking wordt voor Griekenland geschat op 25% van het BNP, terwijl dit voor de Verenigde Staten 8.6% is; Tax Justice Research, 2011). Maar wat maakt een leider dan (in)effectief in het handhaven van regels? Dit proefschrift probeert een antwoord op deze vraag te geven door de determinanten en gevolgen van strafdoelen te onderzoeken. Dat wil zeggen, wat voor specifiek doel hebben leiders voor ogen wanneer ze anderen straffen, en hoe effectief zijn leiders in het stimuleren van regel-avelend gedrag wanneer ze dit strafdoel nastreven? Het antwoord op deze vragen geeft een essentieel inzicht in de manier waarop leiders geneigd zijn (in)effectief te straffen.

Ik heb in Hoofdstuk 2 onderzocht hoe het hebben van macht van invloed is op het doel dat leiders proberen te bereiken met een straf (i.e., macht als determinant van strafdoelen). Macht wordt over het algemeen gedefinieerd als het hebben van controle over waardevolle bronnen (i.e., zoals managers controle hebben over het salaris van werknemers; Keltner et al., 2003; Magee & Galinsky, 2008). Ik heb de invloed van macht op het gebruik van strafdoelen onderzocht door in een veld studie en 8 laboratorium experimenten aan mensen te vragen wat het doel van een straf volgens hen zou moeten zijn. Zou een straf vooral regel-overtredend gedrag van mensen moeten voorkomen, of zou een straf vooral mensen die de regels overtreden hun verdiende loon moeten geven? De straf ging of over studenten die uitgesloten zouden worden van een vak voor het plegen van plagiaat, burgers die een boete zouden krijgen voor belastingontwijking, of team leden die uit een groep gezet zouden worden voor het oneerlijk verdelen van (collectief beheerd) geld.

Voordat ik deze vragen stelde manipuleerde ik in vier van de negen studies ook of mensen wel of geen machtspositie bekleedden, vroeg ik in vier van de negen studies in hoeverre mensen zich machtig of niet machtig voelden, en vroeg ik in een veldstudie of mensen wel of geen positie bekleedden als manager. Ook vroeg ik aan mensen in hoeverre ze anderen wantrouwden. Dachten ze dat studenten geneigd waren om plagiaat te plegen, burgers geneigd waren om belasting te ontduiken, en team leden geneigd waren om geld oneerlijk te verdelen?

Resultaten lieten zien dat mensen met veel (versus weinig) macht eerder aangaven dat een straf bedoeld zou moeten zijn om te voorkomen dat mensen zich niet aan de regels houden, in plaats van mensen hun verdiende loon te geven. Dit werd verklaard doordat hoe machtiger mensen zich voelden, hoe meer ze anderen wantrouwden. Doordat leiders over het algemeen meer macht hebben dan hun volgers, suggereert dit dus sterk dat leiders de neiging hebben te straffen om regel-overtredend gedrag te voorkomen. Ook suggereren de bevindingen uit Hoofdstuk 2 dat dit komt doordat het hebben van macht het wantrouwen in anderen vergroot.

Leiders zijn dus door hun macht geneigd een straf te gebruiken om regel-overtredend gedrag te voorkomen. Maar hoe beïnvloedt dit de effectiviteit van de straf? In Hoofdstuk 4 heb ik onderzocht hoe en waarom het gebruik van strafdoelen als rechtvaardiging van een straf (i.e., een rechter die aangeeft dat een celstraf bedoeld is om te voorkomen dat de dader nog een keer in de fout gaat, of om deze dader een verdiende straf te geven), de effectiviteit van de straf beïnvloedt. Ik heb dit onderzocht door in vier experimenten te manipuleren of een straf door een universiteit, team leider of manager gerechtvaardigd werd als een poging om regel-overtredend gedrag van mensen te voorkomen of om mensen die regels breken hun verdiende loon te geven. Dit ging respectievelijk om een straf voor, (a) het plegen van plagiaat door studenten, (b) het stelen van kantoorgoederen door werknemers, en (c) het oneerlijk verdelen van collectief beheerd geld door groepsleden. Ook vroeg ik aan mensen in hoeverre ze zich gewantrouwd voelden door de leider die een straf rechtvaardigde als een poging om regel-overtredend gedrag te voorkomen of om mensen hun verdiende loon te geven. Resultaten lieten zien dat mensen minder geneigd waren zich aan de regels te houden als ze een leider hadden die een straf rechtvaardigde als een poging om te voorkomen dat ze regels zouden breken vergeleken met het geven van een verdiende straf aan regelovertraders. Dit suggereert dus sterk dat leiders door hun macht strafdoelen gebruiken die relatief ineffectief zijn. Hoe kunnen we dit voorkomen? De bevindingen uit Hoofdstukken 2 en 4 suggereren dat het wantrouwen van leiders in anderen een belangrijk onderdeel van dit ineffektieve straffen is. Dit wantrouwen vergroot de neiging van leiders om straffen te gebruiken om regel-overtredend gedrag te voorkomen *en* verminderd de effectiviteit van de straf (doordat mensen zich gewantrouwd voelen).

In Hoofdstuk 3 heb ik onderzocht hoe het wantrouwen van leiders in anderen verkleind kan worden. In drie experimenten manipuleerde ik of mensen wel of geen machtspositie hadden, vroeg ik in hoeverre ze anderen wantrouwden en in hoeverre ze hun positie wilden behouden. Ook manipuleerde ik in een experiment in hoeverre de positie van mensen kon veranderen tijdens het experiment. Ik manipuleerde, met andere woorden, ook de (in)stabiliteit van machtsposities. Resultaten lieten zien dat mensen anderen meer wantrouwden wanneer ze een machtspositie hadden (versus geen machtspositie), en dat dit verklaard werd doordat

machthebbers graag hun eigen machtspositie wilden behouden (en vertrouwen dus als een manier zagen om macht te verliezen). Resultaten lieten verder zien dat het effect van macht op wantrouwen werd verkleind wanneer mensen met macht het gevoel hadden dat ze hun macht niet konden verliezen aan anderen (doordat ze een stabiele machtspositie hadden). Dit is dus consistent met het idee dat macht wantrouwen vergroot omdat machthebbers hun eigen macht willen behouden.

De bevindingen uit dit proefschrift laten zien dat leiders door hun macht geneigd zijn te straffen om regel-overtredend gedrag te voorkomen, en dat het gebruik van dit strafdoel niet optimaal is voor de effectiviteit van de straf. De neiging van leiders om dit ineffectieve strafdoel te gebruiken kan verminderd worden door leiders zich niet bedreigd te laten voelen in hun machtspositie. De macht die leiders hebben zorgt er dus soms voor dat leiders minder effectief zijn in het handhaven van regels met straffen. Hoewel macht vaak aan leiders wordt gegeven zodat ze de regels kunnen handhaven met straffen, is dit dus niet altijd de meest effectieve strategie.

Acknowledgements (Dankwoord)

Een proefschrift is meer dan alleen wetenschappelijk onderzoek. Een proefschrift is ook een optelsom van jarenlang balans zoeken tussen werk en privé. Dat is bij dit proefschrift niet anders. De steun van familie, vrienden, en promotoren heeft het vinden van de juiste balans tussen werk, familie, vrienden, en de lang afstandsrelatie met Hayley een stuk gemakkelijker gemaakt. Eric, Wilco, en Naomi, bedankt voor het vertrouwen dat jullie in mij hebben gehad, de ruimte die jullie mij hebben gegeven, en het advies dat ik van jullie heb gekregen. Drie hoogleraren als promotoren was een uitdaging maar jullie vulden elkaar goed aan. Ik heb de samenwerking met jullie altijd als zeer plezierig ervaren. Ik ben dan ook trots om onderdeel te mogen uitmaken van de Leiden familie. Mijn collega's, bedankt voor de leuke ASPO congressen. Jojanneke, bedankt voor je hulp, immer goede advies en openstaande kantoordeur (waar ik je regelmatig kwam lastigvallen). Dylan, mijn kleine grote broer, zonder jou had ik überhaupt niet eens nagedacht over een carrière in de wetenschap. Bedankt voor het inspireren van mijn leven en ik voel me bevoorrecht dat ik je enige vriend mag zijn. Mijn lieve ouders, jullie hebben altijd alles gedaan om het leven voor me gemakkelijk te maken. Bedankt voor het geven van het goede voorbeeld. Hard werken loont. Hayley Marie, 4 jaar een lange afstandsrelatie volhouden met mij mag een wereldprestatie heten. We hebben 1, 6, en 8 uur tijdsverschil overleefd en we hebben elkaar soms 6 maanden niet kunnen zien. Die lange afstandsrelatie komt nu tot een einde. Eindelijk kan het volgende avontuur beginnen.

Curriculum Vitae



Marlon Mooijman was born on August 6, 1988 in Monster, The Netherlands. After graduating from ISW college in 's Gravenzande, in 2005, he obtained bachelor degrees in economics at The Hague University (2009) and psychology at Leiden University (2010). He subsequently obtained a research-master degree in social and organizational psychology at Leiden University in 2012, and visited the psychology department at New York University as a visiting scholar. Following the completion of his research-master degree in 2012, he started his PhD project under the supervision of Professor Eric van Dijk, Professor Wilco van Dijk, and Professor Naomi Ellemers at Leiden University in 2012. He is currently a Post-Doctoral Researcher at the University of Southern California, USA.

Kurt Lewin Institute Dissertation 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: Felice 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: Power, Distrust, and Rule Compliance*