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
Chapter 4

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 (Ellermers, 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; \(M_{\text{age}} = 23.11\) years, \(SD_{\text{age}} = 6.54\)) 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 ($F[1, 82] = 3.73, p = .028, \eta^2_p = .08$). 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 ($M = 5.11, SD = 1.11$; $F[1, 82] = 0.75, p = .48, \eta^2_p = .02$).

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; \(M_{\text{age}} = 32.78\) years, \(SD_{\text{age}} = 10.86\)) 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” (α = .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” (α = .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; α = .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, η² = .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, η² = .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 ($M = 5.84, SD = 1.46; t[124] = 2.56, p = .012, d = 0.47$), 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 $>.30, ps < .01$) to non-significance ($\beta$s $< .10, ps > .10$).
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; 
$M_{\text{age}} = 34.16$ years, $SD_{\text{age}} = 10.45$) 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” (α = .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” (α = .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, η²p = .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, η²p = .35). Participants felt more distrusted in the deterrent 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, η²p = .23). Mandatory rule compliance did not differ in the deterrent condition and combined sanction-justification condition (t[121] = 0.04, p = .97, d = 0.01) but was lower in the deterrent condition and combined sanction-justification condition compared to just-deserts condition (t[123] = 4.27,
Overall, mandatory rule compliance was lowest in the no-sanction condition ($p < .001$, $d > 0.70$).

**Voluntary rule compliance.** The sanction-justification manipulation affected the willingness to voluntary comply with rules ($F[1, 245] = 13.61$, $p < .001$, $\eta^2_p = .14$). Voluntary rule compliance was equal in the deterrence condition and combined-justification condition ($t[121] = 1.81$, $p = .073$, $d = 0.44$) but was lower in the deterrence condition and combined sanction-justification condition compared to the just-deserts condition ($t[123] = 4.43$, $p < .001$, $d = 0.80$). 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 ($F[1, 245] = 8.02$, $p < .001$, $\eta^2_p = .09$). Distrust towards others was equal among the sanction conditions ($p > .45$, $d < 0.12$) but lower in the sanction conditions than in the no-sanction condition ($p < .001$, $d > 0.55$).

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

**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% CI's 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_{mandatory\ 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 < .01) to non-significance (\beta_s < .10, p > .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.

<table>
<thead>
<tr>
<th>Sanction</th>
<th>Conditions</th>
<th>Deterrence</th>
<th>Just Deserts</th>
<th>Deterrence/Just Deserts</th>
<th>No Sanction</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personally</td>
<td></td>
<td>4.50</td>
<td>2.10</td>
<td>3.59</td>
<td>-</td>
<td>3.62</td>
</tr>
<tr>
<td>Targeted</td>
<td></td>
<td>(2.16)</td>
<td>(1.62)</td>
<td>(2.43)</td>
<td>-</td>
<td>(2.21)</td>
</tr>
<tr>
<td>Feeling</td>
<td></td>
<td>4.72</td>
<td>1.94</td>
<td>4.01</td>
<td>2.38</td>
<td>3.27</td>
</tr>
<tr>
<td>Distrusted</td>
<td></td>
<td>(1.76)</td>
<td>(1.11)</td>
<td>(1.96)</td>
<td>(1.37)</td>
<td>(1.95)</td>
</tr>
<tr>
<td>Distrust others</td>
<td></td>
<td>4.54</td>
<td>4.53</td>
<td>4.34</td>
<td>5.53</td>
<td>4.74</td>
</tr>
<tr>
<td>(1.49)</td>
<td></td>
<td>(1.69)</td>
<td>(1.47)</td>
<td>(1.36)</td>
<td>(1.57)</td>
<td></td>
</tr>
<tr>
<td>Legitimacy</td>
<td></td>
<td>3.28</td>
<td>2.28</td>
<td>3.39</td>
<td>3.09</td>
<td>3.00</td>
</tr>
<tr>
<td>(1.93)</td>
<td></td>
<td>(1.62)</td>
<td>(1.00)</td>
<td>(1.76)</td>
<td>(1.87)</td>
<td></td>
</tr>
<tr>
<td>Mandatory</td>
<td></td>
<td>5.21</td>
<td>6.12</td>
<td>5.22</td>
<td>3.91</td>
<td>5.12</td>
</tr>
<tr>
<td>Compliance</td>
<td></td>
<td>(1.41)</td>
<td>(0.94)</td>
<td>(1.60)</td>
<td>(1.79)</td>
<td>(1.66)</td>
</tr>
<tr>
<td>Voluntary</td>
<td></td>
<td>4.87</td>
<td>5.78</td>
<td>4.88</td>
<td>4.09</td>
<td>4.91</td>
</tr>
<tr>
<td>Compliance</td>
<td></td>
<td>(1.63)</td>
<td>(1.05)</td>
<td>(1.61)</td>
<td>(1.59)</td>
<td>(1.59)</td>
</tr>
<tr>
<td>Rule Breaking</td>
<td></td>
<td>2.75</td>
<td>1.63</td>
<td>2.22</td>
<td>3.04</td>
<td>2.41</td>
</tr>
<tr>
<td>(1.68)</td>
<td></td>
<td>(1.12)</td>
<td>(1.57)</td>
<td>(1.53)</td>
<td>(1.57)</td>
<td></td>
</tr>
<tr>
<td>Total Rule</td>
<td></td>
<td>5.10</td>
<td>6.01</td>
<td>5.29</td>
<td>4.35</td>
<td>5.21</td>
</tr>
<tr>
<td>Compliance</td>
<td></td>
<td>(1.47)</td>
<td>(0.88)</td>
<td>(1.47)</td>
<td>(1.43)</td>
<td>(1.47)</td>
</tr>
</tbody>
</table>
Table 4.2. Correlations for Experiment 4.3.

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

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 they 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” (α = .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” (α = .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, η² = .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, η² = .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).
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 ($95\% \text{ CI} = [-16.61, -1.64]$). 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.