



Procedural justice in punishment systems: Inconsistent punishment procedures have detrimental effects on cooperation

Jan-Willem van Prooijen^{1*}, Marcello Gallucci² and Gaby Toeset³

¹Free University Amsterdam, The Netherlands

²University of Milan-Bicocca, Milan, Italy

³Catholic University of Louvain, Belgium

The current research examines a moderator who predicts in what situations punishment can have detrimental effects on cooperation. We hypothesized that when a punishment system is perceived as procedurally unfair, people's cooperation level decreases. Results of two experiments indicated that participants cooperated less in a group-based trust game when punishment was inconsistent between persons (i.e. not all group members would be punished for defection) than when punishment was consistent between persons (i.e. any group member who defected would be punished) or when there was no punishment. These effects were mediated by perceived belongingness. The authors conclude that an unfair punishment system leads people to feel marginalized as a group member, and this prompts them to display less cooperation.

People's concern for others is essential to society. A world in which people only pursue their immediate self-interest with no concern for the collective interest would be a chaotic and dangerous place. To prevent self-interest from flourishing, society has developed ways to monitor people's behaviour and to stimulate cooperation (cooperation is broadly defined here as behaviour that serves mutual interest at the expense of immediate self-interest). For example, society has developed a legal system that prescribes punishment for specific types of self-interested behaviours (e.g. not paying taxes). But how effective are these punishment systems to increase people's cooperation? Although numerous empirical studies have revealed positive effects of punishment systems on cooperation (e.g. Fehr & Gächter, 2002; McCusker & Carnevale, 1995; Yamagishi, 1986), a growing body of research has indicated that the effects of punishment are not unequivocally positive. Several studies have reported that punishment systems can, sometimes, have detrimental effects on cooperation (e.g. De Dreu, Giebels, & Van der Vliet, 1998; Fehr & Rockenbach, 2003; Mulder, Van Dijk,

*Correspondence should be addressed to Dr Jan-Willem van Prooijen, Free University Amsterdam, Department of Social Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands (e-mail: jw.van.prooijen@psy.vu.nl).

De Cremer, & Wilke, 2006). In the current research, we explore the detrimental effects of punishment systems on cooperation.

An illustration of detrimental effects of punishment on cooperation can be observed in a research study by Fehr and Rockenbach (2003). In this study, participants played a decision game (a so-called 'trust game'), in which an investor invested a number of money units (MUs) in a receiver. The invested number of MUs was tripled by the experimenter, and the investor specified a request for a back transfer to the receiver. The receiver then had to choose how many MUs to return to the investor. In this situation, participants were facing a dilemma between self-interest and cooperation; the more MUs the participants returned to the investor, the more they served mutual interest (by showing reciprocity and sharing their MUs with the investor) at the expense of immediate self-interest (by keeping the MUs themselves). Fehr and Rockenbach found that if the investor induced a punishment system (i.e. four MUs subtraction if the receiver did not comply with the request), back transfers of the receivers were significantly lower than if the investor did not induce a punishment system. These results were explained by the idea that the induction of a punishment system communicated a sign of distrust to the receivers, undermining their willingness to cooperate. Thus, in Fehr and Rockenbach's social dilemma situation, a punishment system had detrimental effects on cooperation.

Although it may be clear from the above that punishment systems can have negative effects on cooperation, it is unclear in what specific situations these negative effects occur. The current research is designed to identify a moderator who predicts when punishment systems have negative effects on cooperation: we focus on the extent to which people experience decision-making procedures within a punishment system as fair or unfair. The perceived fairness of decision-making procedures is commonly referred to as *procedural justice* (Thibaut & Walker, 1975). We propose that procedural justice is important to the current purposes, because previous research has indicated that procedural justice is strongly related to cooperation in groups (for overviews, see De Cremer & Tyler, 2005; Tyler & Blader, 2000). In the following, we describe how procedural justice is related to cooperation, link this relation to punishment systems and introduce our research.

Procedural justice, cooperation, and punishment systems

Numerous empirical studies have found that procedural justice predicts cooperation in groups, both in experimental settings (e.g. De Cremer, 2002; De Cremer & Van Knippenberg, 2002; De Cremer & Van Vugt, 2002) and in applied settings (Tyler & Degoe, 1995; Tyler, Degoe, & Smith, 1996). De Cremer and Tyler even noted that 'procedural justice is a key antecedent to cooperation in groups' (cf. De Cremer, 2002; Tyler & Blader, 2000). An influential theoretical model to explain this relation between procedural justice and cooperation is the relational model of authority (Tyler & Lind, 1992; cf. Lind & Tyler, 1988), a model that has emphasized the importance of people's group memberships. People regard group authorities as representative for the group as a whole, the model argues, and as a consequence, the way people are treated by group authorities is informative about people's belongingness in the group (Tyler, 1989, 1994; cf. De Cremer, 2002; Smith, Tyler, Huo, Ortiz, & Lind, 1998; Van Prooijen, Van den Bos, & Wilke, 2004). If people are treated fairly by group authorities, people conclude that the authority considers them as full-fledged members of the group, and this enhances their willingness to cooperate. If people are treated unfairly by group authorities, people

infer that the authority regards them as marginal members of their group, and this decreases their willingness to cooperate. In correspondence with this line of reasoning, research has indicated that the relation between procedural justice and cooperation is mediated by people's sense of belongingness (De Cremer & Van Knippenberg, 2002; Tyler *et al.*, 1996; for an overview, see De Cremer & Tyler, 2005).

Building on these arguments, we infer that procedural justice is related to the effects of punishment systems on cooperation. Punishment is an outcome of a decision-making process (e.g. a court trial), and it stands to reason that the perceived fairness of this process can vary systematically. This latter point can be illuminated by applying Leventhal's (1980) *consistency-between-persons* rule to punishment systems. Leventhal suggested that if procedures are consistent between persons (i.e. apply equally to all litigants), people evaluate procedures as fairer than if procedures are inconsistent between persons (i.e. do not apply equally to all litigants; see also Van Prooijen, Van den Bos, Lind, & Wilke, 2006). In a similar vein, punishment systems that are inconsistent between persons (i.e. systems that prescribe punishment if one individual defects, but not if similar other individuals defect) are likely perceived as less fair than punishment systems that are consistent between persons (i.e. systems that prescribe the same punishment for defection to all individuals).

The current research examines the question of whether the perceived (un)fairness of punishment procedures may predict when punishment has detrimental effects on cooperation. Extending on the relational model (Tyler & Lind, 1992), we suggest that inconsistent punishment systems lead unfairly targeted individuals to feel marginalized as a group member, and this decreased sense of belongingness prompts them to display less cooperation. Indications that a punishment system is procedurally fair (e.g. because it is consistent between persons), however, may communicate to people that, even though defection will be punished, they nevertheless are well-respected group members. Consistency between persons may therefore diminish the previously found detrimental effects of punishment on cooperation, because people do not interpret the punishment system as being targeted at them personally (cf. Fehr & Rockenbach, 2003). As such, we expect that punishment systems have detrimental effects on cooperation if people believe that they are unfairly targeted by the punishment procedures, but not (or less so) if people believe that the punishment system is fair. Furthermore, to confirm that relational models of procedural justice provide a theoretical framework that is applicable to people's reactions to punishment systems, we investigate whether or not unfair punishment procedures lead to a decreased sense of belongingness (cf. De Cremer & Tyler, 2005; Tyler *et al.*, 1996). We expect that the effects of punishment procedures on cooperation are mediated by participants' sense of belongingness.

EXPERIMENT I

To test our line of reasoning, we conducted two experiments in which we modified Fehr and Rockenbach's (2003) trust game in at least one important way: we developed a *group-based trust game*. An investor invested a number of coins in a group of four receivers (participants in our experiments always were one of these receivers) and, additionally, communicated a group-based requested back transfer. Both the investment and the requested back transfer were distributed evenly among the four receivers, and every receiver had to decide individually how many coins to return to the investor. To operationalize procedural justice, we varied the consistency of the punishment system

(Leventhal, 1980). In the consistent punishment conditions, participants were informed that any receiver who would return less than the requested back transfer would be punished with four coins subtraction. In the inconsistent punishment condition of Experiment 1, participants were informed that only the participant (and not the other receivers) would be punished if the participant returned less than the requested back transfer. In a control condition, the investor would not punish any of the receivers. The main dependent variable was participants' cooperation, i.e. the total number of coins returned to the investor.

Method

Participants and design

To test the hypotheses, we induced three punishment conditions (consistent punishment, inconsistent punishment, and no punishment). A total of 120 participants (53 men and 67 women, varying in age from 17 to 42 years) were recruited in the student restaurants of the Free University Amsterdam and were distributed randomly across conditions. The experiment was followed by another unrelated experiment. Together the experiments lasted approximately 30 minutes, and participants were paid 5 euros for participation.

Procedure

Upon entry in the laboratory, participants were led to one of the 15 separate cubicles. In the cubicles, participants found computer equipment, which was used to conduct the experiment and register the data. Participants were informed that the computers in the laboratory were interconnected by means of the computer network. Furthermore, we suggested that five participants were simultaneously conducting the experiment in the cubicles (in reality, all information that participants encountered on the computer screen were pre-programmed; a procedure to which none of the participants objected upon debriefing). Participants would play a decision game that involved all five participants. One participant would be the Investor and the four other participants would form a group of Receivers (Receivers A, B, C, and D). It was suggested that the role each participant would be playing was determined randomly by the computer; in reality, the participant was Receiver A in all conditions.

After this, the rules of the decision game were outlined to the participant. The game would be played in five subsequent rounds. Each round began with the Investor investing a number of coins in the group of Receivers. The investments could vary throughout the rounds, given that the investor would have a different amount of coins in each round. In each round, the investment would be tripled by the experimenter and then divided equally among the four Receivers. Additionally, the Investor would send a requested back transfer to the group of Receivers. This request was also divided equally into four individual requested back transfers among the four Receivers. After this, every receiver would specify his or her individual back transfer, which could vary between zero and the tripled investment. Participants would not be informed about the back transfers of the other Receivers.

Following the introduction to the decision game, we induced the punishment manipulation. In the punishment conditions, participants were informed that the Investor could impose a punishment on the Receivers: the investor could decide to subtract four coins from a Receiver's total amount of coins if the Receiver returned less

than the requested back transfer to the investor (it was emphasized that these four coins would *not* be added to the Investor's score). In the consistent punishment condition, participants received a message from the investor emphasizing that the Investor would *not* discriminate between Receivers: *every* Receiver who did not return the requested back transfer would be punished with four coins subtraction. In the inconsistent punishment condition, participants received a message from the investor stating that the investor would discriminate between Receivers: *only Receiver A* would be punished with four coins subtraction if receiver A would not return the requested back transfer.¹ In the no punishment condition, participants were not informed that the Investor could impose a punishment on the Receivers.

To enhance comprehension of the procedure, participants received a few calculation examples. Furthermore, participants received two practice questions. If participants gave a wrong answer to these questions, the correct answer was disclosed and explained, and the question was repeated. After this, participants started with the group-based trust game and responded to the investments, and requested back transfers of the investor. In the punishment conditions, participants received a subtraction of four coins every time that they returned less than the requested back transfer. Although the investments and back transfers differed across the five rounds, we made sure that they always aimed for equality: if all receivers would comply with the individual requested back transfers, all receivers and the investor would end up with the same number of coins.²

To assess cooperation, we measured the total number of coins returned by the participants within the five rounds. Following the five decision rounds, we asked participants a number of additional questions. To assess participant's sense of belongingness, we posed the following three questions (1 = *not at all*, 7 = *very much*): 'Do you believe that the investor regards you as a full-fledged member of the group?'; 'Do you believe that the investor respects you?' and 'Do you believe that the investor trusts you?'. These three items were averaged into a reliable belongingness scale ($\alpha = .77$). Furthermore, to assess participants' evaluations of the requested back transfer, we asked the question 'How appropriate do you believe that the requested back transfers were?' (1 = *not very appropriate*, 7 = *very appropriate*). To check whether or not the punishment manipulation successfully varied perceived procedural justice, we asked the following two questions: 'How fair was the way the investor played the decision game?' (1 = *very unfair*, 7 = *very fair*) and 'How just was the way the investor played the decision game?' (1 = *very unjust*, 7 = *very just*). These two items were averaged into a reliable procedural justice scale ($\alpha = .83$). Finally, to more directly check the manipulation, we asked how the investor would punish the receivers if they returned less than the requested back transfer. Participants were given three choice

¹ To obtain a clean and direct operationalization of inconsistent punishment procedures, we made sure not to provide participants with explicit reasons why they were singled out for punishment. After all, providing explanations would confound the punishment manipulation with procedural justifications, which is another operationalization of procedural justice (Folger, Rosenfield, & Robinson, 1983).

² To achieve investments and back transfers that aimed for equality, we determined investments and back transfers by applying the formula $5R = I$ (R being the individual requested back transfers and I being the investment in the group of receivers). To illustrate, if the Investor invested 40 coins in the group of receivers, the individual requested back transfers would be 14. After all, an investment of 40 would give every receiver $(40 \times 3)/4 = 30$ coins. After complying with the request of 14, a receiver would thus end up with 16 coins. If every receiver complied with this request, the investor, in-turn, would also end up with a profit of $(4 \times 14) - 40 = 16$ coins. The calculation examples and practice trials were set up such that participants learned to distinguish between equal and unequal requested back transfers. In the decision games, the Investments in the group (I) and individual requested back transfers (R) were the following: Round 1, $I = 20$, $R = 7$; Round 2, $I = 40$, $R = 14$; Round 3, $I = 44$, $R = 15$; Round 4, $I = 40$, $R = 14$; Round 5, $I = 28$, $R = 10$.

options to answer this question, representing the three conditions in our design. After this question, participants were debriefed, thanked, and paid for their participation.

Results and discussion

Statistical analysis

In order to provide a coherent analysis of the experimental data, we used a multiple regression approach (Cohen, Cohen, West, & Aiken, 2003), coding the experimental factor into two orthogonal contrasts, according to the hypotheses. We first contrasted the inconsistent punishment condition against the consistent and no punishment conditions (contrast 1 weights: 1 – .5 – .5), and then the consistent punishment condition with the no punishment condition (contrast 2 weights: 0 – 1 1). The first contrast informs us about the specific effect of the inconsistent condition, and the second contrast checks whether consistent punishment differs from the control condition. The use of a regression approach allows us to evaluate the overall effects of the experimental manipulation, the effect of specific comparisons, and the mediational effect of belongingness, using the same type of analysis.

Manipulation checks

On the question directly asking how the investor would punish the receivers, 92.5% of the sample (111 participants) indicated the correct answer. The nine participants that did not give a correct answer were distributed approximately equal across conditions (2 in the consistent punishment condition, 4 in the inconsistent punishment condition, and 3 in the no punishment condition). In the analyses reported below, we included the entire sample (exclusion of the nine participants who reported an incorrect answer yielded similar results).

Punishment conditions exercised a significant effect on the procedural justice scale, $F(2, 117) = 17.73$, $p < .001$, $\omega^2 = .12$. Contrast analysis revealed an effect for the contrast comparing the inconsistent punishment condition with the consistent and no punishment conditions ($B = -1.25$), $t(117) = -5.95$, $p < .001$. Participants in the inconsistent punishment condition ($M = 2.73$, $SD = 1.71$) believed that the way the investor played the decision game was less fair than participants in the consistent punishment condition ($M = 4.64$, $SD = 1.84$) and the no punishment condition ($M = 4.57$, $SD = 1.30$). Procedural justice ratings did not differ significantly between participants in the consistent and no punishment conditions ($B = 0.03$), $t(117) = 0.17$, $p = .864$. These results indicated that the experimental manipulation successfully operationalized participants' perceptions of procedural justice.

Appropriateness of the requested back transfers

The punishment manipulation did not influence responses on the question how appropriate the requested back transfers were, $F(2, 117) = 1.86$, $p < .16$, $\omega^2 = .01$. Furthermore, the overall mean on this measure was significantly higher than the scale mid-point of 4.0 ($M = 4.50$, $SD = 1.57$), $t(119) = 3.51$, $p < .001$. These results suggest that participants in all conditions evaluated the investor's requested back transfers as quite appropriate, as we intended in our experimental paradigm.

Cooperation

The means and standard deviations of the total number of returned coins within the five rounds³ are displayed in Table 1. As a first observation, it can be noted that back transfers were generally lower than the requests (the investor requested a total of 60 coins – see Footnote 2), suggesting that there was some defection in all conditions. The experimental manipulation showed a significant effect on the total back transfers, $F(2, 117) = 9.25$, $p < .001$, $\omega^2 = .12$. Contrast analysis revealed an effect of the contrast comparing the inconsistent punishment condition with the consistent and no punishment conditions ($B = -11.68$), $t(117) = -4.21$, $p < .001$. Participants in the inconsistent punishment condition returned significantly less coins than participants in the consistent and no punishment conditions. Furthermore, participants in the consistent versus no punishment conditions did not differ significantly ($B = -2.12$), $t(117) = -0.88$, $p = .378$. These results supported the hypothesis that inconsistent punishment procedures have detrimental effects on cooperation relative to the consistent and no punishment conditions.

Table 1. Means and standard deviations of cooperation and perceived belongingness as a function of punishment conditions – Experiment 1

Dependent variable	Punishment		
	Inconsistent	Consistent	None
Cooperation ^a	18.25 (21.98)	33.65 (21.23)	37.90 (21.30)
Perceived belongingness	2.77 (1.45)	3.42 (1.44)	3.63 (1.36)

Note. Standard deviations are in parentheses. Higher means indicate higher scores on the dependent variable in question.

^aCoins returned to the investor in the five rounds.

Perceived belongingness

A significant main effect of the experimental manipulation on the belongingness scale was found, $F(2, 117) = 4.05$, $p < .03$, $\omega^2 = .05$. The contrast comparing the inconsistent punishment condition with the consistent and no punishment conditions was significant ($B = -0.50$), $t(117) = -2.76$, $p < .01$. Participants in the inconsistent punishment condition reported lower perceptions of belongingness than participants in the consistent and no punishment conditions (see Table 1). The consistent and no punishment conditions did not differ significantly on this measure ($B = -0.10$), $t(117) = -0.68$, $p = .495$. These results indicated that participants in the inconsistent punishment condition experienced lower levels of belongingness than participants in the consistent and no punishment conditions.

³ We focused on the total number of returned coins to provide a parsimonious presentation of the results. However, we also analysed the individual rounds separately, and the findings were similar: a MANOVA indicated a significant multivariate effect of the manipulation, $F(10, 228) = 2.55$, $p < .01$. The univariate test was significant for four out of five rounds: Round 1, $F(2, 117) = 3.41$, $p < .04$; Round 2, $F(2, 117) = 6.26$, $p < .01$; Round 3, $F(2, 117) = 11.79$, $p < .001$; Round 4, $F(2, 117) = 6.48$, $p < .01$; and Round 5, $F(2, 117) = 2.12$, $p < .13$. In Experiment 2, we examine whether the lack of significance in Round 5 is robust.

Mediational analysis

We then proceeded to test whether or not the effects of the punishment manipulation on cooperation were mediated by participants' sense of belongingness. In the previous analyses, we noticed that the contrast comparing the inconsistent punishment condition with the consistent and no punishment conditions satisfies the first two conditions of mediation (cf. Baron & Kenny, 1986): first, we have an overall effect of the contrast on cooperation and, second, the contrast has an effect on the mediator, i.e. on perceived belongingness. As for the third condition of mediation, we considered both the contrast of interest and the mediator in a single regression predicting cooperation. When perceived belongingness appears in the regression, the effect of the contrast reduces from $B = -11.68$ to $B = -8.33$, $t(116) = -3.22$, $p = .001$, yielding an indirect effect of -3.34 ($SE = 1.37$). A Sobel's test on the indirect effect revealed a significant mediated effect ($z = -2.44$, $p < .02$). It can be concluded that perceived belongingness partially mediated the detrimental effects of punishment on cooperation.

EXPERIMENT 2

In Experiment 2, we sought to extend the findings of Experiment 1. In Experiment 1, inconsistent punishment was operationalized by singling the participant out for punishment. This situation may lead people to feel personally disliked by the investor, and one may wonder whether similar findings occur when punishment procedures are inconsistent without solely targeting the participant. In Experiment 2, participants in the inconsistent condition were informed that (out of four receivers) both they and two other receivers would be punished if they returned less than the requested back transfer, but that one receiver would not be punished. Thus, instead of unfairly targeting only the participant for punishment, in Experiment 2 another participant was unfairly protected from punishment. As such, this situation sought to mimic punishment systems where a limited number of individuals are able to operate above the law (e.g. totalitarian regimes).

As a second extension, in Experiment 2 we improved our belongingness measure. It might be argued that some of the items tapping belongingness in Experiment 1 assess the construct indirectly (by referring to respect or trust). In Experiment 2, we measured a scale that more directly focused on people's sense of belongingness in the group of receivers. Finally, to enhance mundane realism, in Experiment 2 we made payment and behaviour contingent by rewarding the highest score with a book ticket.

Method**Participants and design**

A total of 54 participants (21 men and 33 women, varying in age from 16 to 32 years) were recruited in the student restaurants of the Free University Amsterdam, and were assigned randomly to one of the three punishment conditions (consistent punishment, inconsistent punishment, and no punishment). The experiment lasted approximately 15–20 minutes, and participants received 2.50 euros for participation.

Procedure

The experimental procedure was the same as in Experiment 1, with a couple of modifications. First, participants were informed that the participant who had the most

coins at the end of round 5 would receive a book certificate of 20 euros (after all participants were run, such a book certificate was awarded to the highest scoring participant in each experimental condition). Second, and more importantly, the inconsistent punishment condition differed from Experiment 1: participants received a message from the investor stating that Receivers A, B, and D (and *not* Receiver C) would be punished with four coins subtraction for defection.

The dependent variables were the same as Experiment 1. However, we modified and improved our measure of belongingness by assessing items that more directly pertained to participants' perceived belongingness to the group of receivers. We asked the following four questions (1 = *not at all*, 7 = *very much*): 'Do you believe the investor regarded you as a full-fledged member of the group?'; 'To what extent do you feel like a full-fledged member of the group?'; 'To what extent do you have the feeling that you and the other receivers form a group?' and 'To what extent does the investor regard you as a worthy member of your group?'. These four items were averaged into a reliable belongingness scale ($\alpha = .83$). After participants completed all questions, they again were debriefed, thanked, and paid.

Results and discussion

The analytical strategy was identical as in Experiment 1. We again used a regression approach and specified contrast 1 to compare the inconsistent condition with the consistent and no punishment conditions, and contrast 2 to compare the consistent and no punishment conditions.

Manipulation checks

On the question directly asking how the investor would punish the receivers, 81% of the sample (44 participants) indicated the correct answer. The 10 participants with an incorrect answer were distributed approximately even across conditions (5 in the consistent punishment condition, 2 in the inconsistent punishment condition, and 3 in the no punishment condition). Exclusion of these participants produced similar results, and we included the entire sample in our analyses.

On the procedural justice scale ($\alpha = .75$), we again found a significant effect of punishment conditions, $F(2, 51) = 4.30$, $p < .02$, $\omega^2 = .11$. The contrast comparing the inconsistent punishment condition with the consistent and no punishment conditions was significant ($B = -0.91$), $t(51) = -2.75$, $p < .01$. Participants in the inconsistent punishment condition ($M = 3.62$, $SD = 2.06$) again believed that the way the investor played the decision game was less fair than in the consistent and no punishment conditions ($M = 4.66$, $SD = 1.69$; and $M = 5.29$, $SD = 1.30$, respectively). The contrast comparing the consistent and no punishment conditions was non-significant ($B = 0.32$), $t(51) = 1.11$, $p = .27$. These results indicated that the manipulation again successfully varied levels of perceived procedural justice.

Appropriateness of the requested back transfers

The punishment manipulation again did not influence participants' responses on the question how appropriate the requested back transfers were, $F < 1$. The overall mean on this measure was significantly higher than the scale mid-point of 4.0 ($M = 4.48$, $SD = 1.60$), $t(53) = 2.21$, $p < .04$. These results revealed that participants in all

conditions regarded the requested back transfers as appropriate, which is in correspondence with Experiment 1.

Cooperation

The punishment manipulation significantly influenced the total number of returned coins⁴ within the five rounds, $F(2, 51) = 4.63$, $p < .02$, $\omega^2 = .12$. The contrast comparing the inconsistent punishment condition with the consistent and no punishment conditions was significant ($B = -13.13$), $t(51) = -2.98$, $p < .01$. As can be seen in Table 2, participants in the inconsistent punishment condition returned significantly less coins than participants in the consistent and no punishment conditions. The second contrast indicated that participants in the consistent and no punishment conditions did not differ in the number of coins that they returned ($B = -1.99$), $t(51) = -0.52$, $p = .61$. These results further supported the hypothesis.

Table 2. Means and standard deviations of cooperation and perceived belongingness as a function of punishment conditions – Experiment 2

Dependent variable	Punishment		
	Inconsistent	Consistent	None
Cooperation ^a	29.06 (27.77)	50.74 (22.87)	46.76 (16.17)
Perceived belongingness	2.83 (1.31)	4.11 (1.38)	4.24 (1.43)

Note. Standard deviations are in parentheses. Higher means indicate higher scores on the dependent variable in question.

^aCoins returned to the investor in the five rounds.

Perceived belongingness

The punishment manipulation also influenced participants' responses on the belongingness measure, $F(2, 51) = 5.68$, $p < .01$, $\omega^2 = .15$. The first contrast was significant ($B = -0.89$), $t(51) = -3.37$, $p < .002$. As indicated by the means displayed in Table 2, participants in the inconsistent punishment condition reported lower perceived belongingness than participants in the consistent and no punishment conditions. The consistent and no punishment conditions did not differ on perceived belongingness, as indicated by the non-significant second contrast ($B = 0.07$), $t(51) = 0.28$, $p = .78$.

Mediational analysis

We then proceeded to test whether perceived belongingness mediated the effect of the first contrast on cooperation. The reported effects of this contrast on cooperation and perceived belongingness satisfied Baron and Kenny's (1986) first two conditions of mediation. When the relevant contrast and perceived belongingness were entered as independent variables in a regression predicting cooperation, the effect of the contrast

⁴In Experiment 2, we again also analysed the five rounds separately. The multivariate test was significant, $F(10, 94) = 2.28$, $p < .02$. Furthermore, univariate tests indicated four out of five rounds to be significant and one (Round 3) to approach significance: Round 1, $F(2, 51) = 3.74$, $p < .04$; Round 2, $F(2, 51) = 5.23$, $p < .01$; Round 3, $F(2, 51) = 2.94$, $p < .07$; Round 4, $F(2, 51) = 4.32$, $p < .02$; and Round 5, $F(2, 51) = 3.20$, $p < .05$.

was reduced to non-significance ($B = -7.03$), $t(51) = -1.57$, $p = .12$. The Sobel test indicated that the mediated effect was significant ($z = -2.31$, $p < .03$). These findings replicate and extend the findings of Experiment 1 by revealing that perceived belongingness mediated the effects of punishment on cooperation.

GENERAL DISCUSSION

Results of both experiments indicated that people cooperate less if punishment procedures are inconsistent between persons than if punishment procedures are consistent between persons. This difference was attributable to a decrease in cooperation in the inconsistent punishment condition, as indicated by the finding that the consistent punishment condition did not differ from the control condition. This latter finding points to a difference between Fehr and Rockenbach's (2003) trust game, where people cooperated less in the punishment than no punishment conditions, and the present group-based trust game: in our consistent condition punishment applied to an entire group, increasing the legitimacy of the punishment system and decreasing the likelihood that receivers believe that the system targets them personally. Furthermore, the effects of (in)consistency between persons on cooperation were mediated by participants' perceived belongingness. If punishment procedures were inconsistent between persons, participants reported that they felt marginalized as a group member, and as a consequence, they cooperated less to the requests of the investor. Overall, these results provide support for the notion that Leventhal's (1980) inconsistency operationalization of procedural injustice influences the relation between punishment systems and cooperation, in ways that can be predicted from the present theoretical analysis.

The mediational analyses were included to obtain evidence that relational models of justice are applicable to understand people's reactions to punishment systems. However, mediational analyses inherently are correlational in nature, and it might be argued that the present effects are attributable to other variables that are closely correlated with the mediator. Of course, it is impossible to fully exclude this possibility in any correlational analysis. Nevertheless, in the present study we regard our belongingness explanation as plausible, particularly because we can support the mediational role of belongingness also by taking a different (non-statistical) view of mediation. Our experiments bring additional evidence of the causal link between procedural justice and perceived belongingness, which can be coupled with the well-established causal effect of belongingness on cooperation (De Cremer & Tyler, 2005; Tyler *et al.*, 1996). The link between these two sets of results qualifies our reasoning as an 'experimental causal-chain' proof of mediation, which recently has been advocated as an alternative method to establish evidence of mediation (Spencer, Zanna, & Fong, 2005).

Having said this, it might be speculated that belongingness is not the only factor contributing to the procedure effect on cooperation. One plausible possibility is that inconsistent punishment procedures produce negative emotions such as anger in receivers. Procedural injustice is associated with negative emotions (Weiss, Suckow, & Cropanzano, 1999), and these negative emotions and corresponding negative impressions of the investor may prompt receivers to display negative reciprocity in their back transfers. Based on our mediational analysis, it is difficult to justify negative reciprocity as an alternative explanation to belongingness, although it may represent

an additional mechanism supporting the link between injustice, belongingness, and cooperation. Future research would do well to test this possibility. Furthermore, it may be the case that inconsistent punishment procedures elicit a sense of relative deprivation because participants hold lower outcome expectancies than other receivers. These lowered outcome expectancies are inherent to inconsistent procedures, and procedural injustice is generally associated with negative outcome expectancies (Thibaut & Walker, 1975). Given previous empirical findings, it is likely that participants' cooperation levels were shaped more strongly by concerns about procedural than distributive justice. Previous research robustly indicated that peoples' concern for procedural justice outweighs their concern for distributive justice (Tyler, 1989, 1994), a preference that is often explained by the link between procedural justice and belongingness (De Cremer, 2002; Tyler, 1994). Nevertheless, future research could usefully extend the present studies using more individual-oriented procedures (e.g. denying a receiver's opportunities to voice an opinion) that may less directly promote suspicions of relative deprivation among participants.

In our study, we found evidence for negative effects of punishment on cooperation if procedures were unfair (cf. De Dreu *et al.*, 1998; Fehr & Rockenbach, 2003), but we did not find evidence for positive effects of punishment on cooperation if procedures were fair. How can we reconcile this finding with previous studies that did find positive effects of punishment on cooperation (e.g. Fehr & Gächter, 2002; Yamagishi, 1986)? The most likely explanation for this can be found in the fact that in the present trust game recipients only interacted with the investor, and their outcomes did not directly depend on other recipients' willingness to cooperate. Such interdependence between own outcomes and others' cooperation may be a necessary precondition for positive effects of punishment to occur. Yamagishi noted that one of the main reasons why punishment can be beneficial to cooperation is because it establishes trust that other group members cooperate as well. In correspondence with this, positive effects of punishment on cooperation are typically found in public good dilemmas where the outcome of each individual group member depends directly on other member's willingness to cooperate (e.g. Fehr & Gächter, 2002). As such, people may desire punishment systems when it is important to them that other people support the collective interest at the expense of immediate self-interest.

The above arguments suggest an interesting, and empirically testable hypothesis. It could be worthwhile to find out how the perceived fairness of punishment systems influences cooperation in social dilemma situations where punishment has been found to exert positive effects on cooperation, such as public good dilemmas (e.g. Fehr & Gächter, 2002). It is possible that the extent to which people experience a sense of belongingness is important to group functioning, particularly in a situation as mutually interdependent as a public good dilemma. Given that the effects of procedural justice on cooperation are (at least partly) mediated by these perceptions of belongingness (De Cremer & Van Knippenberg, 2002; Tyler *et al.*, 1996), procedural fairness of punishment systems might predict cooperation in public good dilemmas. Furthermore, in public good dilemmas punishment can serve the function of establishing trust in other group members' willingness to cooperate (Yamagishi, 1986). It may therefore be the case that in public good dilemmas, fair punishment procedures may have beneficial effects on cooperation when contrasted with unfair or no punishment procedures, which provides a fruitful avenue for future research.

To conclude, the present study examined the influence of procedurally fair versus unfair punishment systems on people's willingness to cooperate. The results have revealed that unfair punishment procedures undermine people's cooperation because they decrease people's sense of belongingness in the group. As such, it can be concluded here that procedural injustice is an important factor to predict detrimental effects of punishment on cooperation.

Acknowledgements

We thank Marieke Roskes for her help with Experiment 2.

References

- Baron, R. B., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.
- De Cremer, D. (2002). Respect and cooperation in social dilemmas: The importance of feeling included. *Personality and Social Psychology Bulletin*, 28, 1335–1341.
- De Cremer, D., & Tyler, T. R. (2005). Managing group behavior: The interplay between procedural justice, sense of self, and cooperation. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 151–218). New York: Academic Press.
- De Cremer, D., & Van Knippenberg, D. (2002). How do leaders promote cooperation? The effects of charisma and procedural fairness. *Journal of Applied Psychology*, 87, 858–866.
- De Cremer, D., & Van Vugt, M. (2002). Intergroup and intragroup aspects of leadership in social dilemmas: A relational model of cooperation. *Journal of Experimental Social Psychology*, 38, 126–136.
- De Dreu, C. K. W., Giebels, E., & Van de Vliert, E. (1998). Social motives and trust in integrative negotiation: The disruptive effects of punitive capability. *Journal of Applied Psychology*, 83, 408–422.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.
- Fehr, E., & Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422, 137–140.
- Folger, R., Rosenfield, D., & Robinson, T. (1983). Relative deprivation and procedural justifications. *Journal of Personality and Social Psychology*, 45, 268–273.
- Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances on theory and research* (pp. 27–54). New York: Plenum.
- Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. New York: Plenum.
- McCusker, C., & Carnevale, P. J. (1995). Framing in resource dilemmas: Loss aversion and the moderating effects of sanctions. *Organizational Behavior and Human Decision Processes*, 61, 190–201.
- 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.
- Smith, H. J., Tyler, T. R., Huo, Y. J., Ortiz, D. J., & Lind, E. A. (1998). The self-relevant implications of the group-value model: Group membership, self-worth, and treatment quality. *Journal of Experimental Social Psychology*, 34, 470–493.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89, 845–851.

- Thibaut, J., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.
- Tyler, T. R. (1989). The psychology of procedural justice: A test of the group-value model. *Journal of Personality and Social Psychology*, 57, 830-838.
- Tyler, T. R. (1994). Psychological models of the justice motive: Antecedents of distributive and procedural justice. *Journal of Personality and Social Psychology*, 5, 850-863.
- Tyler, T. R., & Blader, S. L. (2000). *Cooperation in groups: Procedural justice, social identity, and behavioral engagement*. Philadelphia, PA: Psychology Press.
- Tyler, T. R., & Degoey, P. (1995). Collective restraint in social dilemmas: Procedural justice and social identification effects on support for authorities. *Journal of Personality and Social Psychology*, 69, 482-497.
- Tyler, T. R., Degoey, P., & Smith, H. (1996). Understanding why the justice of group procedures matters: A test of the psychological dynamics of the group-value model. *Journal of Personality and Social Psychology*, 70, 913-930.
- Tyler, T. R., & Lind, E. A. (1992). A relational model of authority in groups. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 115-292). San Diego, CA: Academic Press.
- Van Prooijen, J. -W., Van den Bos, K., Lind, E. A., & Wilke, H. A. M. (2006). How do people react to negative procedures? On the moderating role of authority's biased attitudes. *Journal of Experimental Social Psychology*, 42, 632-645.
- Van Prooijen, J. -W., Van den Bos, K., & Wilke, H. A. M. (2004). Group belongingness and procedural justice: Social inclusion and exclusion by peers affects the psychology of voice. *Journal of Personality and Social Psychology*, 87, 66-79.
- Weiss, H. M., Suckow, K., & Cropanzano, R. (1999). Effects of justice conditions on discrete emotions. *Journal of Applied Psychology*, 84, 786-794.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110-116.

Received 24 January 2006; revised version received 22 May 2007