Evolution and Human Behavior 35 (2014) 169–175

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination☆

Bastian Schiller *,1, Thomas Baumgartner *,1, Daria Knoch *

Department of Psychology, Social and Affective Neuroscience, University of Basel

ARTICLE INFO

Article history:
Initial receipt 21 February 2013
Final revision received 26 December 2013

Keywords:
Outgroup discrimination
Outgroup hostility
Ingroup favoritism
Third-party punishment
Punishment motives
Intergroup bias

ABSTRACT

Social norms pervade almost every aspect of social interaction. If they are violated, not only legal institutions, but other members of society as well, punish, i.e., inflict costs on the wrongdoer. Sanctioning occurs even when the punishers themselves were not harmed directly and even when it is costly for them. There is evidence for intergroup bias in this third-party punishment: third-parties, who share group membership with victims, punish outgroup perpetrators more harshly than ingroup perpetrators. However, it is unknown whether a discriminatory treatment of outgroup perpetrators (outgroup discrimination) or a preferential treatment of ingroup perpetrators (ingroup favoritism) drives this bias. To answer this question, the punishment of outgroup and ingroup perpetrators must be compared to a baseline, i.e., unaffiliated perpetrators. By applying a costly punishment game, we found stronger punishment of outgroup versus unaffiliated perpetrators and weaker punishment of ingroup versus unaffiliated perpetrators. This demonstrates that both ingroup favoritism and outgroup discrimination drive intergroup bias in third-party punishment of perpetrators that belong to distinct social groups.

1. Introduction

People punish norm violators, i.e., they inflict costs on others in response to wrongdoing (Kurzban, DeScioli, & O'Brien, 2007). This punishment of norm violators has been proposed to play a role in enforcing social norms and thus to be an important mechanism for human cooperation and the regulation of social life. It has been shown that people punish norm violators even though the transgression does not affect them directly and the punishment comes at a cost to them (e.g., Fehr & Fischbacher, 2004). Some have argued that this so-called costly third-party punishment has become an increasingly important mechanism for the maintenance of social order, because second-party punishment (i.e., when the punisher is directly affected by the transgression) has become less effective as community size has increased and repeat interactions between the same people have become less probable (e.g., Marlowe et al., 2008; but see Pedersen, Kurzban, & McCullough, 2013).

When third-parties have to dispense punishment, they are often not objective. In particular, third-parties respond differently to norm violations committed by a member of their group (ingroup) than to those committed by a member of an alien group (outgroup). For example, justice research indicates that the perpetrator's group affiliation strongly influences unaffected third-parties, i.e., they judge an outgroup perpetrator's norm violation more harshly than the same transgression by an ingroup perpetrator (Graham, Weiner, & Zucker, 1997; Sommers & Ellsworth, 2000). In real life situations, third-parties incur personal costs in order to punish norm-violating behavior (e.g., time, energy, money, social exposure). In experiments, making the punishment personally costly instead of using hypothetical punishment decisions and thus measuring actual behavior is crucial in order to study this phenomenon. To date there are only very few studies on the influence of group affiliation on punitive behavior that have used tangible resources. These studies show a clear intergroup bias, namely a harsher punishment of outgroup than ingroup perpetrators who have committed the same norm violation against ingroup victims (Bernhard, Fischbacher, & Fehr, 2006; Baumgartner, Gotte, Gugler, & Fehr, 2012).

To the best of our knowledge, however, there is no answer to the question of what drives this intergroup bias in third-party punishment. A simple observation that outgroup members are punished more strongly compared to ingroup members does not permit the determination of whether this difference is due to outgroup discrimination (a poorer treatment of outgroup members), ingroup favoritism (a preferential treatment of ingroup members), or a combination of both. This is because the studies mentioned above lack a crucial condition in which baseline behavior toward unaffiliated perpetrators is measured. The sources of intergroup bias in third-party punishment can only be determined by examining whether punishment of outgroup and/or ingroup perpetrators significantly

☆ This project was supported by the Swiss National Science Foundation Grant PP00P1-123381 to DK.

* Corresponding authors. Department of Psychology, University of Basel, Birmannsgasse 8, CH-4055 Basel, Switzerland.
E-mail addresses: bastianschiller@web.de (B. Schiller), t.baumgartner@unibas.ch (T. Baumgartner), daria.knoch@unibas.ch (D. Knoch).

[1] The first two authors contributed equally to this work.

1090-5138/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.evolhumbehav.2013.12.006

differs from this baseline punishment. Given that punishment of outgroup perpetrators compared to ingroup perpetrators is typically stronger, three scenarios are possible: (1) Ingroup perpetrators are punished more clemently compared to both unaffiliated and outgroup perpetrators, but punishment of unaffiliated perpetrators and outgroup perpetrators does not differ. This would indicate that the differential punishment is only due to ingroup favoritism. (2) Ingroup and unaffiliated perpetrators are punished equally, while outgroup perpetrators are punished more harshly compared to both ingroup and unaffiliated perpetrators. In this case, outgroup discrimination would be the source of intergroup differences in punishment. (3) Ingroup perpetrators are punished more clemently compared to unaffiliated and outgroup perpetrators, and at the same time outgroup perpetrators are punished more harshly compared to unaffiliated and ingroup perpetrators. This would mean that both ingroup favoritism and outgroup discrimination underlie the differential punishment of in- and outgroup perpetrators.

The goal of this study was to investigate the basis of the intergroup bias in third-party punishment of perpetrators that belong to distinct social groups. For this purpose, we used a third-party punishment paradigm with real social groups, real monetary stakes, and unaffiliated perpetrators. Participants in the role of an outside observer, i.e., a third-party, were given the option of punishing people for behaving inappropriately in a previously played Prisoner's Dilemma Game. They could punish – at their own expense – norm-violating behavior of the perpetrator by reducing the violator's payoff (see Fig. 1). We kept the victims' group affiliations constant so that the perpetrators' norm transgressions were always against ingroup members. We focused on norm transgressions against ingroup members because third-parties respond most strongly to them, whereas costly punishment is virtually absent when outgroup victims are involved (e.g., Bernhard et al., 2006; Baumgartner et al., 2012). The norm violators who could be punished were – from the view of the third-party – either outgroup members, ingroup members, or unaffiliated persons.

Just by observing the punishment behavior, one cannot infer the motives underlying punishment. Thus, in order to get an idea about which motives might underlie the sources of intergroup bias in third-party punishment, we additionally measured punishment motives (such as retribution, anticipated punishment satisfaction, improvement of perpetrator's future behavior) known to affect intergroup evaluations, behavior, and punishment decisions (e.g., Carlsmith & Darley, 2002; Keller, Oswald, Stucki, & Gollwitzer, 2010).

In sum, evidence shows that third-parties' willingness to punish perpetrators at personal costs is affected by the perpetrator's group membership. To date, there is no study that investigates the sources of this intergroup bias. Here, we address this question by using a punishment game with real monetary stakes and by adding a condition where punishment of an unaffiliated perpetrator is measured.

2. Methods

2.1. Participants

Eighty-nine healthy participants (48 males, 41 females; age: Mean ± SE = 22.03 ± 0.33) participated. To explore the sources of intergroup bias in third-party punishment, we used naturally occurring social groups. We first solicited contact information of students in lecture halls who were interested to take part in experiments on decision-making. We then contacted these students by e-mail and asked them about their personal interests in several domains (e.g., arts, music, politics, religion, soccer). Finally, we recruited participants who had, on a scale from 1 (very weak) to 5 (very strong), at least medium (= 3) self-reported interest in soccer (N = 39) or in politics (N = 50), because previous studies using these groups have reported strong behavioral intergroup biases (Koopmans & Rebers, 2009; Hein, Silani, Preuschoff, Batson, & Singer, 2010). Participants interacted with either supporters of their own ("ingroup"; IN) or a corresponding rival ("outgroup"; OUT) soccer club/political party or with unaffiliated persons ("unaffiliated", UN).

2.2. Third-party punishment paradigm

Participants in the role of a third-party were given the opportunity to punish the behavior of players who had previously played a Prisoner's Dilemma Game (PDG; conducted online with the software tool Unipark). In the PDG, two players A and B (either ingroup members, outgroup members, or unaffiliated persons) were each endowed with 20 points and each had to decide simultaneously whether to keep all the points or pass them to the other player. Passed points were doubled. Thus, keeping the points equals defection (denoted as D) and passing the points equals cooperation (denoted as C). For example, if player A retained the 20 points while player B transferred the 20 points (behavioral pattern DC), player A earned a total of 60 points (40 points from the transfer plus the initial endowment of 20 points) and player B earned nothing. Note that unfair and egoistic intentions of the perpetrator are less certain in a simultaneous PDG compared to a sequential PDG (e.g., Kurzban et al., 2007) played by a second mover who defects despite the fact that the first mover was cooperative. In the simultaneous PDG the perpetrator may defect because he believes that the other player will defect too. Thus defection in this version of the PDG might also be due to potential mistrust or worries of being exploited. As a consequence, the third-party is unsure about the perpetrator's unfair and egoistic intentions. We chose a simultaneous PDG, because in real-life third-parties often encounter situations, where a perpetrator`s intentions are ambiguous.

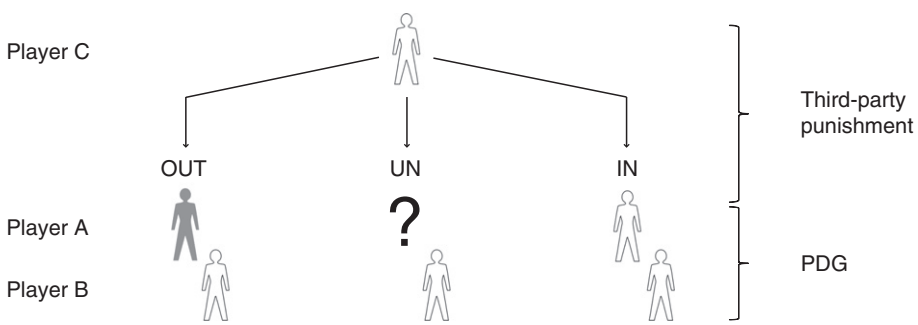

Fig. 1. Depicted is the applied third-party punishment paradigm. Third-parties, depicted as player C, faced real decisions of two players, A and B, who had played a simultaneous Prisoner's Dilemma Game (PDG). Player B was always an ingroup member and player A (who could be punished) was either an outgroup member (OUT; depicted by a gray character), an unaffiliated person (UN; depicted by a question mark), or an ingroup member (IN; depicted by a white character). Participants could punish the behavior of player A by assigning punishment points. Each punishment point spent reduced player A's income by three points.

In the third-party punishment paradigm, we ran 6 sessions with 10–20 participants per session. All participants were in the role of a third-party (player C) who was confronted with player A's and player B's decisions in the PDG. Third-parties were informed that each decision was made by a different player A and player B pair. We did not explicitly inform third-parties about whether PDG players had known that other players could punish their behavior (see Fehr & Fischbacher, 2004). In each of the total of 15 trials (see also Supplementary Table 1), player C received an endowment of 10 points which he could use to punish the behavior of one player. We recoded all player C's decisions in such a way that A always refers to the player that C could punish. Each point spent on punishment cost player C 1 point and the sanctioned player A 3 points. Points not used for punishment were paid out to the participants at the end of the experiment (exchange rate 10 points = 1 Swiss Franc, about $ 1). The PDG players had already been paid according to their decisions in the online-PDG and had agreed that their decisions could be used in further experiments. They were then paid a second time according to the decisions of the third-parties in this experiment. In actuality, there was a mild form of deception for third-parties, because we used the average punishment of about 15 third-parties to determine the payments for player A participants. In other words, several third-parties punished the same player A (otherwise we would have needed an enormous amount of participants in the PDG). PDG decisions were selected in such a way that each player C was confronted with the same decision situations, which were presented in pseudo-randomized orders. We included DC (player A defects, player B cooperates) and CC (both players cooperate) trials from the PDG. We focused our analysis on the eight DC trials because in these trials third-parties can execute punishment according to the definition given in the Introduction ("inflict costs on others in response to wrongdoing"). Furthermore, we told third-parties that in each trial they would be informed about the group affiliation of each player. Player B was always an ingroup member and player A (whose income could be reduced) was either (1) an outgroup member (OUT), (2) an unaffiliated person (UN), or (3) an ingroup member (IN). Participants were told that unaffiliated players were neither ingroup nor outgroup members. This is crucial to prevent participants from assuming that they were actually interacting with an in- or an outgroup member, which would contaminate baseline behavior toward unaffiliated persons. Future studies might additionally include a manipulation check to assess how participants actually represent unaffiliated others. Group affiliations and PDG decisions of players A and B were presented on a computer screen both in text (your group/other group/unaffiliated person; keeps points/transfers points) and in a picture (symbol of the political parties/shirts of the soccer clubs; unaffiliated persons were represented by a question mark).

2.3. Punishment motives

After the third-party punishment paradigm, decision situations (DC trials with outgroup, unaffiliated or ingroup perpetrators) were shown again to the participants. They then had to indicate on a scale ranging from 1 (not at all) to 7 (very much) how strongly three punishment motives (retribution, anticipated punishment satisfaction, improvement of perpetrator's future behavior) had affected their decisions to assign punishment points, separately for outgroup, ingroup, and unaffiliated perpetrators.

2.4. Statistical analyses

We calculated third-parties' average points spent to reduce the income of player As for DC and CC trials, broken down into the three groups (OUT, UN, IN). Furthermore, in order to investigate whether inter-individual variance in outgroup discrimination, ingroup favoritism, or both can be explained by inter-individual differences in punishment motives, we calculated differences scores for punishment points as well as for punishment motives for the contrast OUT minus UN and for the contrast UN minus IN (DC trials). We used SPSS 19 (SPSS Inc., Chicago, IL) for data analysis. Please see the results section for details about the statistical tests conducted, including ANOVAs, paired t-tests, correlation analyses, and multiple linear regression analyses ("enter method"). To address the issue of multicollinearity in the regression analyses we further conducted commonality analyses (Nimon, Lewis, Kane, & Haynes, 2008). Commonality analysis determines the unique and common variance contributions of predictors in the regression analyses by calculating distinct regression analyses with all possible combinations of predictors. Results were considered significant at the level of $P < 0.05$ (two-tailed). In case of significant multivariate effects, post hoc paired t-tests were computed using the Bonferroni correction according to Holm (Holm, 1979). Measure ETA^2 is reported as the effect size.

3. Results

First, we checked whether we observe an intergroup bias in third-parties' punishment of norm violators (DC trials). Comparing average punishment of outgroup perpetrators (OUT; Mean ± SE = 3.85 ± 0.34) with that of ingroup perpetrators (IN; Mean ± SE = 1.99 ± 0.30), we found a clear intergroup bias, i.e., outgroup perpetrators were punished more strongly than ingroup perpetrators ($t_{88} = 5.27$, $P < .001$, $ETA^2 = .24$).

To illuminate the sources of intergroup bias in third-party punishment, we next computed a mixed-model ANOVA with average punishment points as dependent variable, within-subject factor "perpetrator's group affiliation" (OUT, UN, IN) and the two between-subject factors "gender" and "group type" (soccer club, political party). As interaction effects of "perpetrator's group affiliation" with "gender" and "group type" were non-significant (all $P > .69$), we conducted all further analyses over the whole sample. Crucial for the key question of this study, the main effect of "perpetrator's group affiliation" was significant ($F_{2,\ 84} = 14.11$, $P < .001$, $ETA^2 = .25$), indicating a significant impact of the perpetrator's group membership on third-party punishment. Paired t-tests revealed on average a stronger punishment of outgroup perpetrators in comparison to unaffiliated perpetrators ($P < .001$, $ETA^2 = .17$) and a weaker punishment of ingroup perpetrators in comparison to unaffiliated perpetrators ($P = .033$, $ETA^2 = .05$). We thus found that both outgroup discrimination (OUT > UN) and ingroup favoritism (UN > IN) exist in third-parties' reactions to norm violations (UN; Mean ± SE = 2.61 ± 0.31; see Fig. 2).

In CC trials, third-parties spent a small amount of points to reduce the income of cooperative outgroup members (OUT_CC: Mean ± SE = 0.93 ± 0.25), whereas they spent virtually no points to reduce the income of cooperative unaffiliated persons and ingroup members (UN_CC: Mean ± SE = 0.19 ± 0.08; IN_CC: Mean ± SE = 0.11 ± 0.06; see Supplementary Fig. 1). Because a cooperative PDG player has not violated any norm in CC trials, third-parties who spent points to reduce the income of cooperative outgroup members solely aimed at harming the outgroup (Goette & Meier, 2011). The total points spent to reduce the income of cooperative outgroup members in CC trials, however, were very low compared to those spent to reduce the income of defecting outgroup members in DC trials (OUT_DC: Mean ± SE = 3.85 ± 0.34; OUT_CC: Mean ± SE = 0.93 ± 0.25).

To examine whether the discriminatory treatment of outgroup members (OUT > UN) differed between trials in which outgroup members had committed a norm violation (DC trials) and trials in which they had not committed a norm violation (CC trials), we calculated a dependent t-test. We found that the discriminatory treatment of outgroup members was larger if player A had committed a norm violation (OUT_UN in DC trials: Mean ± SE = 1.24 ± 0.29; OUT_UN in CC trials: Mean ± SE = 0.74 ± 0.21; $t_{88} = 1.97$, $p = .052$, $ETA^2 = .04$).

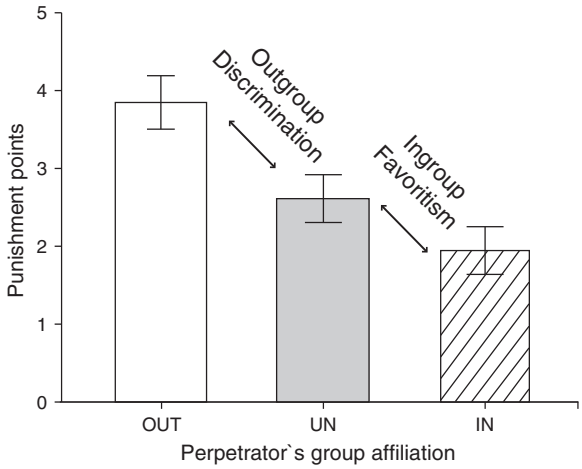

Fig. 2. Depicted are the average punishment points (Mean ± SE) assigned to perpetrators (DC trials, in which player A had defected and player B had cooperated) who were either outgroup members (OUT), unaffiliated persons (UN), or ingroup members (IN). The perpetrator's group affiliation had a significant impact on assigned punishment points ($p < .001$, $ETA^2 = .25$). Third-parties punished outgroup perpetrators more strongly than unaffiliated perpetrators (outgroup discrimination: OUT > UN, $P < .001$, $ETA^2 = .17$), and punished ingroup perpetrators more weakly than unaffiliated perpetrators (ingroup favoritism: UN > IN, $P = .033$, $ETA^2 = .05$).

In a next step, we investigated whether inter-individual variance in outgroup discrimination and/or ingroup favoritism in third-party punishment (DC trials) could be explained by inter-individual differences in punishment motives (for mean values of punishment motives see Supplementary Fig. 2). For that purpose, we calculated two regression and commonality analyses with outgroup discrimination (OUT minus UN) and ingroup favoritism (UN minus IN) as dependent variables and the respective difference scores for punishment motives as independent variables.

Regarding outgroup discrimination, the regression model was significant ($F = 5.04$, $P = .003$) and punishment motives accounted for 15.08% of behavioral variance. The difference in retribution motive toward outgroup versus unaffiliated perpetrators significantly explained the individual level of outgroup discrimination ($\beta = 0.30$, $P = .008$) and, on a trend level, the difference in anticipated punishment satisfaction motive toward outgroup versus unaffiliated perpetrators had the same effect ($\beta = 0.19$, $P = .068$). The difference in the motive to improve perpetrator's future behavior made no significant contribution ($\beta = -0.11$, $P = .284$). In other words, the stronger the retribution and anticipated punishment satisfaction motives were toward outgroup perpetrators in comparison to unaffiliated perpetrators, the more harshly third-parties punished outgroup perpetrators in comparison to unaffiliated perpetrators. Commonality analysis showed that retribution uniquely explained 7.45%, anticipated punishment satisfaction uniquely explained 3.40%, and both motives commonly explained 4.23% of variance in outgroup discrimination. Thus both motives shared a substantial part of the explained variance. As seen in the simple correlation coefficients (see Fig. 3), if the overall variance contribution (i.e., unique and common) of anticipated punishment satisfaction was considered, this motive was a significant positive predictor of outgroup discrimination.

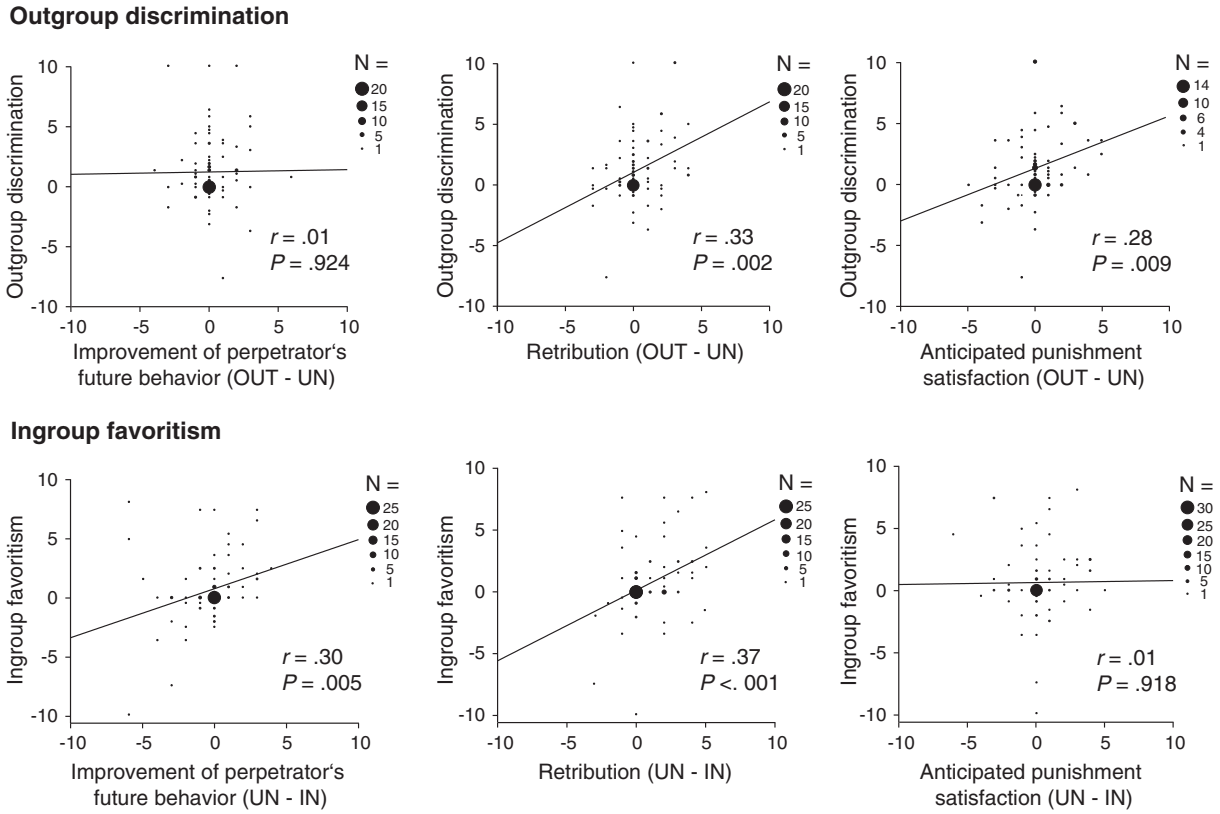

Fig. 3. Shown are scatter-plots demonstrating the relationship between punishment motives ("improvement of perpetrator`s future behavior", "retribution", and "anticipated punishment satisfaction") and outgroup discrimination (upper row), and ingroup favoritism (lower row) toward the perpetrator. The retribution motive ($r = .33$, $P = .002$) and the anticipated punishment satisfaction motive ($r = .28$, $P = .009$) were related to outgroup discrimination (no relationship was observed for the motive to improve perpetrator`s future behavior: $r = .01$, $P = .924$). The retribution motive ($r = .37$, $P < .001$) and the motive to improve perpetrator`s future behavior ($r = .30$, $P = .005$) were related to ingroup favoritism (no relationship was observed for the anticipated punishment satisfaction motive: $r = .01$, $P = .918$). Note that 16 people did not spend punishment points, mainly composing the clusters of people (indicated by the large dots) that had neither outgroup discrimination nor ingroup favoritism effects (and also no differences in the respective punishment motives). Excluding these people from analyses, however, did not significantly change the reported correlations.

Regarding ingroup favoritism, the regression model was significant ($F = 6.51$, $P < .001$) and punishment motives explained 18.66% of behavioral variance. The difference in retribution motive toward unaffiliated versus ingroup perpetrators significantly explained the individual level of ingroup favoritism ($\beta = 0.35$, $P = .002$) and, on a trend level, the difference in the motive to improve perpetrator's future behavior toward unaffiliated versus ingroup perpetrators did so as well ($\beta = 0.20$, $P = .059$). The difference in anticipated satisfaction motive made no significant contribution ($\beta = -0.14$, $P = .193$). In other words, the more weakly the retribution motive and the motive to improve perpetrator's future behavior were toward ingroup versus unaffiliated perpetrators, the more weakly third-parties punished ingroup perpetrators in comparison to unaffiliated perpetrators. Commonality analysis showed that retribution uniquely explained 9.89%, improvement of perpetrator's future behavior uniquely explained 3.49%, and both motives commonly explained 5.28% of variance in ingroup favoritism. Thus both motives shared a substantial part of the explained variance. As seen in the simple correlation coefficients (see Fig. 3), if the overall variance contribution (i.e., unique and common) of the motive to improve perpetrator's future behavior was considered, this motive was a significant positive predictor of ingroup favoritism.

4. Discussion

In this study, we showed that both outgroup discrimination and ingroup favoritism drive intergroup bias in costly, third-party punishment of norm violators. More precisely, third-parties, who shared group membership with victims, more strongly punished outgroup perpetrators and more weakly punished ingroup perpetrators in comparison to unaffiliated perpetrators. Moreover, inter-individual variance in outgroup discrimination was found to be explained by the retribution motive and the anticipated punishment satisfaction motive, whereas inter-individual variance in ingroup favoritism was explained by the retribution motive and the motive to improve perpetrator's future behavior.

There is a wealth of literature on intergroup bias, and numerous studies using both classical minimal groups and real social groups have found that people allocate more positive stimuli (e.g., money) and – to a lesser degree (Mummendey & Otten, 1998) – less negative stimuli (e.g., noise) to ingroup than to outgroup members (Sherif, Harvey, White, Hood, & Sherif, 1961; Tajfel, Billig, Bundy, & Flament, 1971; Brewer, 1979; Mummendey et al., 1992; Levine, Prosser, Evans, & Reicher, 2005; Ben-Ner, McCall, Stephane, & Wang, 2009). The majority of these studies, however, have not included unaffiliated interaction partners. As a consequence, despite a fruitful discussion (e.g., Allport, 1954; Brewer, 1999; Hewstone, Rubin, & Willis, 2002), the question of whether "more positive treatment" and "less negative treatment" of the ingroup is due to outgroup discrimination or ingroup favoritism has remained unanswered. In recent years, a few studies have used study designs that allow researchers to determine the sources of intergroup bias. These studies have concluded that intergroup bias is mainly driven by ingroup favoritism, not outgroup discrimination (Ahmed, 2007; Halevy, Bornstein, & Sagiv, 2008; Koopmans & Rebers, 2009; Yamagishi & Mifune, 2009). In contrast to the present study, these studies have examined behavior in the domain of resource distribution and not in the domain of punishment of norm violations. However, this type of punishment is characteristic of many group interactions (DeRidder & Tripathi, 1992; Henrich et al., 2006) so that an investigation of the sources of intergroup bias in this domain seems relevant. Furthermore, our study differs from previous studies on the sources of intergroup bias by investigating members of rival social groups. We found discriminatory treatment of outgroup members, even if outgroup members had cooperated, and this discriminatory treatment was larger if outgroup members had committed a norm violation. These two findings suggest that two features of our study, i.e., investigating members of rival social groups and punishment of norm violations might explain the strong outgroup discrimination effects we observe, in contrast to previous studies (e.g., Koopmans & Rebers, 2009; Yamagishi & Mifune, 2009).

Whereas most studies investigating intergroup bias in the punitive domain have demonstrated harsher punishment of outgroup compared to ingroup perpetrators (Graham et al., 1997; Sommers & Ellsworth, 2000; Bernhard et al., 2006; Baumgartner et al., 2012), some studies have shown more negative reactions toward deviant ingroup compared to outgroup members under certain conditions (e.g., Kerr, Hymes, Anderson, & Weathers, 1995; van Prooijen, 2006; for a review on the so called "black sheep effect" see Marques & Páez, 1994). These studies suggest that if the situation allows third-parties to deny that the perpetrator is fully responsible for wrongdoing, then they give ingroup perpetrators the "benefit of the doubt" (van Prooijen, 2009) and judge them less harshly than outgroup perpetrators. Only if it is impossible for third-parties to deny that the perpetrator is fully responsible do they judge ingroup perpetrators more harshly than outgroup perpetrators. For example, van Prooijen (2006) showed the typical intergroup bias effect when guilt was uncertain — third-parties judged outgroup perpetrators more harshly than ingroup perpetrators. However, this pattern was reversed when guilt was 100% certain. Similarly, in the present study, third-parties could deny that the perpetrator was fully responsible for wrongdoing as it was unclear whether his intentions were unfair and egoistic (see Methods). Thus, third-parties gave the "benefit of the doubt" to ingroup perpetrators and punished them more weakly than outgroup perpetrators. Future studies could include unaffiliated persons and investigate the sources of intergroup bias in situations in which third-parties cannot deny that the perpetrator is fully responsible and harsher treatment of ingroup perpetrators is expected.

Our results further showed that both outgroup discrimination and ingroup favoritism are associated with the retribution motive. The importance of retribution is in line with findings from justice research indicating that laypersons' punishment judgments are driven primarily by this motive (Carlsmith, 2006; Keller et al., 2010). Our study complements this research in two ways. First, we showed that differences in punishment meted out to outgroup, unaffiliated, and ingroup perpetrators are partially explained by differences in the retribution motive. Second, as punishment decisions in our study had monetary costs for the third-parties, we demonstrated that participants were willing to sacrifice personal interests in order to express their retributive desires. Retribution is thought to be a more backward-looking and affectively-laden "just deserts" motive, whereas utilitarian motives, e.g., improvement of perpetrator's future behavior, are thought to be more forward-looking motive and are associated with more deliberation, e.g., contemplating "what is best for society?" (Carlsmith, 2006; Keller et al., 2010). It therefore seems that negative affective reactions toward perpetrators explain both outgroup discrimination and ingroup favoritism.

Furthermore, outgroup discrimination was associated with another affectively-laden "just deserts" motive, namely the anticipated punishment satisfaction. In contemplating the punishment of outgroup members who had committed norm violations against ingroup members, third-parties might have anticipated a feeling of "sweet revenge" against the rival outgroup. This relates to research which demonstrates that punishment in general (DeQuervain et al., 2004), punishment of outgroup perpetrators (Baumgartner et al., 2012), and outgroup failure (Cikara, Botvinick, & Fiske, 2011) activate areas in the brain's reward circuit.

From an evolutionary perspective, third-party punishment remains puzzling, given that punishment is costly and third-parties have no individual reciprocity or reputation benefits to gain in anonymous, one-shot interactions in the laboratory (e.g., Fehr & Fischbacher, 2004; but see Kurzban et al., 2007). In the present study, third-parties also sacrificed personal resources to inflict costs on persons who had defected against the ingroup. What are the evolutionary roots of differentially punishing outgroup, unaffiliated,

and ingroup perpetrators? Although speculative, we provide two possible explanations for our results, one based on individual selection and one based on group selection. Regarding individual selection, the observed punishment pattern best fits with kin selection theory. Assuming that we have originated from small groups in which average genetic relatedness was high, natural selection might have engineered us with the propensity to treat ingroup members like kin (Pinker, 2012). Following this argumentation, compared to unaffiliated perpetrators, third-parties might treat ingroup perpetrators better and outgroup perpetrators worse because they implicitly assume that their average genetic relatedness is highest with ingroup members, unknown with unaffiliated persons, and lowest with outgroup members. Thus, favoring ingroup perpetrators and discriminating against outgroup perpetrators might yield indirect fitness benefits to third-parties. However, we hasten to add that average genetic relatedness in contemporary large social groups is rather low, casting some doubts on the kin selection explanation. Regarding a possible explanation by group selection, it could be helpful to take the matter of group reputation into account (Bernhard et al., 2006). Although their individual identities remained anonymous, third-parties knew that detailed information about their decisions would be sent to all player As to determine their payments. Therefore, by punishing perpetrators who "attacked" the ingroup third-parties can establish a group reputation that deters future aggression against ingroup fellows. It might be especially important to protect ingroup victims from "attacks" made by rival outgroup members, because such "attacks" might be more common than those of ingroup members or unaffiliated persons. Thus, through strongly punishing outgroup "attackers" third-parties might increase the general security of all ingroup members. Protecting ingroup victims from "attacks" made by other ingroup members is also important and, on the basis of research that emphasizes the adaptive value of punishment for cooperative within-group interactions (e.g., Boyd, Gintis, Bowles, & Richerson, 2003; Shinada, 2009) one could have also expected strong punishment of ingroup members in our study. However, as described in the Method section, our participants might not have been sure whether the intentions of the perpetrator were unfair and egoistic and thus they might not have wanted to risk harming another ingroup member unjustifiably. This is because such unjustified punishment might have detrimental effects on future within-group interactions. In sum, both the preferential treatment of ingroup perpetrators and the discriminatory treatment of outgroup perpetrators by third-parties might provide the group with an evolutionary advantage (for further reading on the evolutionary roots of intergroup bias in third-party punishment, we refer to DeScioli & Kurzban, 2009, 2013).

To conclude, investigating members of real social groups this study shows for the first time that both outgroup discrimination and ingroup favoritism drive the intergroup bias in punitive behavior toward perpetrators shown by third-parties. Recent findings suggest a role of demand effects in third-parties' decisions (e.g., Pedersen et al., 2013). Although we consider it relatively unlikely that third-parties sacrificed large amounts of their money for punishment just to conform to experimenters' expectations, we are not able to fully exclude this possibility. As proposed by Pedersen et al. (2013) future experiments could allow third-parties to also reward behavior, thus diminishing possible demand effects for punishment. Also, one could investigate whether our findings apply to groups that are in more neutral relations and how "group type" (real social groups vs. minimal groups) and "domain of investigation" (resource distribution vs. punishment of norm violations) interact with ingroup favoritism and outgroup discrimination effects.

Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2013.12.006.

References

Ahmed, A. M. (2007). Group identity, social distance and intergroup bias. *Journal of Economic Psychology*, *28*, 324–337.

Allport, G. W. (1954). *The nature of prejudice.* Cambridge, MA: Addison-Wesley.

Baumgartner, T., Gotte, L., Gugler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping*, *33*, 1452–1469.

Ben-Ner, A., McCall, B. P., Stephane, M., & Wang, H. (2009). Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *Journal of Economic Behavior & Organization*, *72*, 153–170.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *442*, 912–915.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 3531–3535.

Brewer, M. (1979). In-group bias in the minimal intergroup situation: A cognitive–motivational analysis. *Psychological Bulletin*, *86*, 307–324.

Brewer, M. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*, 429–444.

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology*, *42*, 437–451.

Carlsmith, K. M., & Darley, J. M. (2002). Why do we punish? Deterrence and just desert motives for punishment. *Journal of Personality and Social Psychology*, *83*, 284–299.

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychological Science*, *22*, 306–313.

DeQuervain, D., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–1258.

DeRidder, R., & Tripathi, R. C. (1992). *Norm violation and intergroup relations.* Oxford, New York: Clarendon Press, Oxford University Press.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*, 281–299.

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, *139*, 477–496.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*, 63–87.

Goette, L., & Meier, S. (2011). Can integration tame conflicts? *Science*, *334*, 1356–1357.

Graham, S., Weiner, B., & Zucker, G. S. (1997). An attributional analysis of punishment goals and public reactions to O J. Simpson. *Personality and Social Psychology Bulletin*, *23*, 331–346.

Halevy, N., Bornstein, G., & Sagiv, L. (2008). "In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm. *Psychological Science*, *19*, 405–411.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, *68*, 149–160.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*, 1767–1770.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistic*, *6*, 65–70.

Keller, L. B., Oswald, M. E., Stucki, I., & Gollwitzer, M. (2010). A closer look at an eye for an eye: Laypersons' punishment decisions are primarily driven by retributive motives. *Social Justice Research*, *23*, 99–116.

Kerr, N. L., Hymes, R. W., Anderson, A. B., & Weathers, J. E. (1995). Defendant–juror similarity and mock juror judgments. *Law and Human Behavior*, *19*, 545–567.

Koopmans, R., & Rebers, S. (2009). Collective action in culturally similar and dissimilar groups: An experiment on parochialism, conditional cooperation, and their linkages. *Evolution and Human Behavior*, *30*, 201–211.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, *28*, 75–84.

Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*, 443–453.

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Ensminger, J., Gurven, M., Gwako, E., Henrich, J., Henrich, N., Lesorogol, C., McElreath, R., & Tracer, D. (2008). More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society B-Biological Sciences*, *275*, 587–590.

Marques, J. M., & Páez, D. (1994). The "black sheep effect": Social categorization, rejection of ingroup deviates, and perception of group variability. In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology, Vol. 5.* (pp. 41–68)New York: Wiley.

Mummendey, A., & Otten, S. (1998). Positive–negative asymmetry in social discrimination. *European Review of Social Psychology*, *9*, 107–143.

Mummendey, A., Simon, B., Dietze, C., Grunert, M., Haeger, G., Kessler, S., Lettgen, S., & Schaferhoff, S. (1992). Categorization is not enough: Intergroup discrimination in negative outcome allocation. *Journal of Experimental Social Psychology*, *28*, 125–144.

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regiression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*, 457–466.

Pedersen, E. J., Kurzban, R., & McCullough, M. E. (2013). Do humans really punish altruistically? A closer look. *Proceedings of the Royal Society B-Biological Sciences*, *280*(1758).

Pinker, S. (2012). The false allure of group selection. An edge original essay. Retrieved from http://www.edge.org/conversation/the-false-allure-of-group-selection.
Sherif, M., Harvey, O. J., White, B. J., Hood, W., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The robbers cave experiment*. Norman, OK: The University Book Exchange.
Shinada, M. (2009). Why do third party punish?: Second-order cooperation of in-group members. *The Japanese Journal of Experimental Psychology*, *48*, 99–110.
Sommers, S. R., & Ellsworth, P. C. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin*, *26*, 1367–1379.
Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization in intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178.
van Prooijen, J. W. (2006). Retributive reactions to suspected offenders: The importance of social categorizations and guilt probability. *Personality and Social Psychology Bulletin*, *32*, 715.
van Prooijen, J. W. (2009). Offenders' social categorization: Ingroup bias or black sheep effect? In M. E. Oswald, S. Bieneck, & J. Hupfeld-Heinemann (Eds.), *Social psychology of punishment of crime* (pp. 211–229). New York: Wiley.
Yamagishi, T., & Mifune, N. (2009). Social exchange and solidarity: In-group love or out-group hate? *Evolution and Human Behavior*, *30*, 229–237.