Social Psychological and Personality Science
2014, Vol. 5(6) 662-670

DOI: 10.1177/1948550614527115
spps.sagepub.com
SAGE

For Members Only: Ingroup Punishment of Fairness Norm Violations in the Ultimatum Game

Saaid A. Mendoza[1], Sean P. Lane[2], and David M. Amodio[2]

Abstract

Although group membership has many privileges, members are expected to reciprocate those privileges. We tested whether in-group members would be punished more harshly than out-group members for marginal fairness norm violations within ultimatum game bargaining interactions. Participants considered monetary splits (of US$20) from in-group and out-group proposers, which ranged in proportion. Accepting an offer yielded the proposed payout; rejecting it caused each player to earn nothing—a punishment of the proposer at a personal cost. Participants exacted stricter costly punishment on racial in-group than out-group members for marginally unfair offers (Study 1), an effect that was replicated with college group membership and magnified among strong in-group identifiers (Study 2). Importantly, ultimatum game decisions were driven by fairness perceptions rather than proposer evaluations (Study 3), suggesting our effects reflected norm enforcement and not esteem preservation. These findings illuminate a previously unexplored process for maintaining group-based norms that may promote in-group favoritism.

Keywords

in-group favoritism, fairness norms, ultimatum game, costly punishment, black sheep effect

Membership is said to have its privileges. Indeed, when group boundaries are formed, members of one's group are usually given preferential treatment—a pattern of *in-group favoritism* that is believed to underlie most forms of intergroup bias (Brewer, 1999; Tajfel & Turner, 1979; see Hewstone, Rubin, & Willis, 2002 for review). However, group membership may also come at a cost, such that in-group members expect preferential treatment to be reciprocated. We investigated whether in-group members are punished more harshly than out-group members for marginally unfair behavior—a response strategy that may broadly function to maintain prescriptive in-group norms.

In-group Favoritism and Norm Violation

In-group favoritism is driven, in part, by the notion that members of a group support each other (Brewer, 2007). Although research on in-group favoritism has focused primarily on intergroup attitude processes, further implications of in-group favoritism for behavior become evident in the context of social exchanges. In social exchanges, individuals typically expect cooperation and fair treatment from others (Fehr & Fischbacher, 2004; Tyler & Blader, 2000). These expectations are guided by *prescriptive group norms*, which serve to coordinate the actions of individuals in a manner that promotes the group's goals and values (Abrams, Marques, Brown, & Henson, 2000; Jetten, Spears, & Manstead, 1997; Marques, Abrams, & Serodio, 2001).

Interestingly, because prescriptive norms function to maintain group cohesion and promote group interests, these norms should be more salient during *intra*group interactions, relative to *inter*group interactions. An individual would therefore have greater expectations of fairness in an exchange relationship with a fellow group member than with someone from another group (Bernhard, Fischbacher, & Fehr, 2006; Brewer, 1999; Tajfel, Billig, Bundy, & Flament, 1971). Furthermore, when fairness norms are violated, these transgressions are felt more acutely when they come from an in-group than from an out-group member (Valenzuela & Srivastava, 2012). Hence, a consideration of prescriptive norms suggests that in exchange interactions, norm-violating behavior is more likely to be punished when committed by an in-group member than by an out-group member, given that in-group trust and cooperation relies on the enforcement of such norms (Balliet & Van Lange, 2013; Shinada, Yamagishi, & Ohmura, 2004).

[1] Department of Psychology, Amherst College, Amherst, MA, USA
[2] Department of Psychology, New York University, New York, NY, USA

Corresponding Author:
Saaid A. Mendoza, Department of Psychology, Amherst College, 324 Merrill Science, Amherst, MA 01002, USA.
Email: smendoza@amherst.edu

In contrast to the present focus on punishing norm-violating behaviors in reciprocity contexts, it is notable that prior research on the *black sheep effect* has examined how participants evaluate deviant group members in performance settings (see Marques & Paez, 1994 for review). Although well-performing in-group members are evaluated more positively than comparable out-group members, poorly performing in-group members are evaluated more harshly and excluded from the group (Hutchison, Abrams, Gutierrez, & Viki, 2008; Marques, Abrams, Paez, & Martinez-Taboada, 1998; Marques & Yzerbyt, 1988; Marques, Yzerbyt, & Leyens, 1988). Moreover, this pattern of enhanced in-group derogation and exclusion tends to be more pronounced among individuals who are highly identified with their group (Biernat, Vescio, & Billings, 1999; Branscombe, Wann, Noel, & Coleman, 1993). As such, the derogation and exclusion of deviant in-group members who clearly violate group standards of performance are thought to be motivated by a desire to preserve the positive image of the group and, by extension, to protect the esteem of the derogator (Castano, Paladino, Coull, & Yzerbyt, 2002; Eidelman & Biernat, 2003; Marques et al., 1998).

Although the black sheep effect is well established, our interest lies in a different yet complementary process for group norm enforcement. Although clear violations of in-group *performance standards* can lead to derogation and/or exclusion from the group, as illustrated by the black sheep effect, *behavioral violations of fairness norms* in a negotiation—especially when the violation is marginal and potentially ambiguous—may elicit an alternative response. Rather than ostracize those who commit slight violations to preserve the group's image, one may choose to punish norm violators to bring their behavior back in line with group-based expectations—a response driven by reciprocity norms rather than self-esteem needs (Balliet & Van Lange, 2013). This should be particularly true within intragroup negotiations where adherence to fairness/cooperation norms is vital to the collective's prosperity (Balliet, Mulder, & Van Lange, 2011; Fehr & Gachter, 2002; Tajfel et al., 1971).

Negotiation research in behavioral economics has revealed that people are often willing to incur a personal cost to enforce fairness norms (Camerer & Thaler, 1995). This so-called "costly punishment" response pattern reflects a short-term cost that nevertheless may serve a long-term benefit (i.e., to maintain cooperative behavior that would benefit the self and/or one's group in future situations; Henrich et al., 2006). Importantly, for the present research, norm enforcers may be selective in whom they choose to punish because of its repercussions: Only in-group members who marginally violate the norm may be worth redeeming through costly punishment. By contrast, those who clearly violate the norm may be perceived as a "lost cause," since their behavior is so deviant from the group's standard. Hence, the costly punishment of in-group violators of fairness norms would reflect a commitment to group coherence and the preservation of in-group favoritism (see also Yamagishi, Jin, & Kiyonari, 1999).

Research Overview

We investigated whether individuals would punish fairness transgressions more strictly when committed by in-group than by out-group members, even at a cost to the self. We tested this hypothesis in the context of the *ultimatum game* (Güth, Schmittberger, & Schwarze, 1982), a bargaining task in which one player (*proposer*) decides how to divide a sum of money with another player (*responder*). If the responder accepts an offer, each player earns the proposed amounts; if the responder rejects the offer, each player earns nothing. According to traditional economic perspectives, "rational" responders should accept all offers because it is in their best economic interest. However, research consistently shows that responders' decisions are also influenced by psychological factors. For example, unfair offers (e.g., 70:30 splits) are typically rejected due to a valued sense of fairness (Bolton & Zwick, 1995; Henrich et al., 2001; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006). Additionally, while adherence to fairness norms is adaptive in repeated interactions, rejection of unfair offers is also observed in anonymous one-shot interactions even though it is costly and yields no direct material benefits (Boyd, Gintis, Bowles, & Richerson, 2003; Güth & Tietz, 1990). Thus, the ultimatum game provided an ideal paradigm for testing our hypothesis regarding differences in the punishment of fairness norm violations by in-group and out-group members.

Study 1

In Study 1, we examined responses to offers from racial in-group and out-group proposers within the ultimatum game. Race represents a salient social identity (Frable, 1997) that permitted a strong initial test of our hypothesis. We predicted that participants would be more sensitive to fairness norm violations committed by in-group than by out-group members as indicated by the stricter use of "costly" punishment in the ultimatum game. As noted previously, this effect was expected to emerge for marginal norm violations, for which punishment could serve to enforce group norms. By contrast, we expected participants to accept very fair offers (e.g., even splits) regardless of the proposer's race. Similarly, we expected participants to reject clearly unfair offers irrespective of race, given that major norm violations would be less amenable to correction through costly punishment. These predictions resemble patterns observed in prior research such that group-based biases most often emerge when other decision criteria are ambiguous (Dovidio & Gaertner, 2000).

Method

Participants and Procedure

Thirty-five White American New York University undergraduates participated for course credit in a study advertised as an interactive decision-making game. This experiment employed a fully within-subjects design.

After providing informed consent, participants were placed into private computer cubicles. The experimenter explained that the study was part of a collaborative research effort with other schools in the metropolitan area to investigate online decision making. In the current study, they would play an interactive computer game with students at a nearby university (which had a high percentage of African American students).

Participants learned that players select avatars to represent themselves during the game to protect their identity. They viewed a variety of same-sex, human-like avatars that differed in race and appearance and were invited to choose the one that *most closely* resembled them and to name it using their middle name, a nickname, or a three-digit number of their choice. This cover story led participants to believe that avatars representing other players reflected their appearance as well, providing a means to manipulate the race of the interaction partners.

After reviewing the ultimatum game rules, participants learned they had been randomly assigned to the role of "responder" for all interactions and would thus decide whether to accept or reject offers from other players for a hypothetical proportion of US$20. Participants were further informed that they would interact with each player only once during the session. This stipulation obviates participants' concerns about building a reputation yet preserves their ability to enforce fairness norms.

The task began with a generous (US$12/20) "offline" practice trial to ensure participants understood the game's rules. Next, participants viewed a preprogrammed series of attempts for their computer to connect to a server before commencing with their "online" interactions. Each trial began with the presentation of a digital "envelope" icon that opened to reveal the avatar of a same-sex proposer along with his or her offer. Participants considered a series of 10 critical offers from White and Black proposers, which included US$10, US$8, US$7, US$6, or US$4 of the hypothetical US$20 total (six filler offers of different sizes from other types of avatars were also included to obscure the variables of interest). Thus, the task comprised a 2 (Group: White vs. Black) × 5 (Offer Size: US$10, US$8, US$7, US$6, or US$4) within-subjects design. Trial order was randomized and decisions were registered via computer keyboard. The private response context, combined with the use of avatars, ensured participants' anonymity and precluded concerns about social pressure to respond without prejudice. Following task completion, participants were probed for suspicion and fully debriefed.

Results

Accept/reject decisions (coded 1 and 0, respectively) were submitted to a logistic regression with proposer racial group, offer size, and their interaction included as predictors. Analyses were conducted in SAS 9.2 (SAS Institute Inc., 2008) using the GENMOD procedure. Robust standard errors (*SE*s) were calculated using generalized estimating equations (GEE) methods (Liang & Zeger, 1986) to account for the nonindependence and dichotomous nature of the outcome variables. The US$10 offer from a Black proposer was designated as the reference in these equations because our prediction concerned group differences in offer acceptance as a function of their deviation from fairness (i.e., an even US$10 split). Figure 1 illustrates the raw acceptance rates.

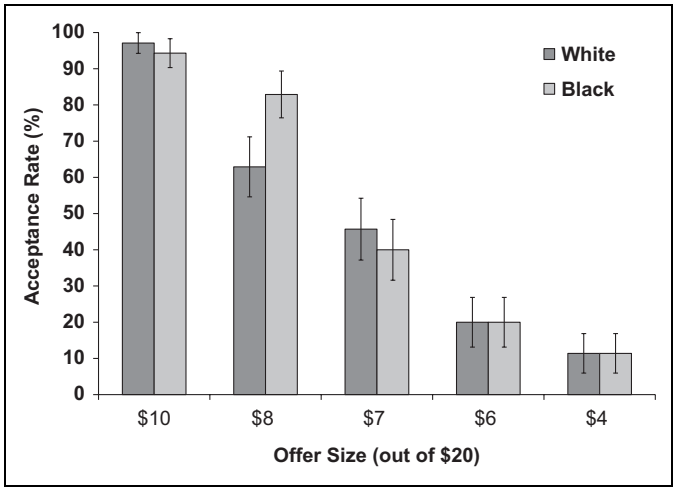

Figure 1. Study 1 acceptance rates depicted as a function of offer size and proposer racial group. Error bars represent ±1 standard error.

We expected clearly fair offers to be accepted and clearly unfair offers to be rejected, regardless of group membership. Indeed, participants were equally likely to accept a US$10 offer from White and Black proposers, odds ratio (OR) = 2.06, Wald = .99, $p = .321$. Relative to the fair US$10 offer, participants were also significantly more likely to reject than accept offers of US$6 and US$4, regardless of the proposer's racial group (see Table S1 in Supplementary Material found at http://spps.sagepub.com/supplemental). This pattern validated the ability of this task to elicit a costly punishment response.

As hypothesized, group effects emerged when proposals were marginally unfair. Specifically, a significant Race × Offer interaction, $F(4, 306) = 2.95$, $p = .021$, revealed that the proposer's race influenced participants' tendency to accept an offer of US$8 compared to US$10, OR = .17, Wald = 3.94, $p = .047$. Simple slope analyses indicated that a US$8 offer was significantly more likely to be rejected than a US$10 offer when the proposer was White, OR = .05, Wald = 7.23, $p = .007$, but not when the proposer was Black, OR = .29, Wald = 1.87, $p = .171$. An additional comparison revealed that participants were more likely to accept a US$8 offer from a Black compared with a White proposer, OR = .35, Wald = 8.01, $p = .005$. Group differences were not observed for the US$7 offers; these were more likely to be rejected than US$10 offers, regardless of the proposer's race (see Table S1 in Supplementary Material; see Online Supplemental Material found at http://spps.sagepub.com/supplemental).

Discussion

In Study 1, we tested the hypothesis that marginal fairness norm violations would be punished more strictly when the

transgressor was an in-group member as compared to an out-group member. Consistent with our predictions, participants accepted clearly fair offers (of US$10) and rejected clearly unfair offers (e.g., of US$4 and US$6), regardless of the proposer's group membership.[1] However, when an offer was potentially perceived as marginally unfair—in this study, when the offer was US$8/US$20—participants were more likely to reject the offer from an in-group member than from an out-group member, even at a "cost" to oneself. Although the game did not involve actual money, participants still responded to offers in a linear manner that reflected thoughtful consideration of their *potential* cost, and this pattern replicates past ultimatum game studies involving real financial losses (Knoch et al., 2006; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Thus, overall fairness concerns were driving punishment responses on the task, with the proposer's group membership playing a role only in ambiguous circumstances as hypothesized.

It is notable that although we expected a group membership effect to emerge for marginally unfair offers, we did not know beforehand whether the US$7 or US$8 offer would be considered unfair to participants. Participants' behavior suggested that a US$8 offer was considered to be a violation of in-group fairness expectations, such that it was rejected more frequently if it came from an in-group than from an out-group member. By contrast, a US$7 offer was considered unfair (i.e., was accepted less than 50%) regardless of the proposer's racial group. Although our interest focused on responses to marginal fairness norm violations, regardless of the specific offer amount, the finding that group membership effects emerged at the US$8 offer level would benefit from replication.

Although we interpreted the Study 1 results as reflecting a process of in-group norm enforcement, we considered alternative explanations. One possibility is that White participants worried about appearing prejudiced toward Black proposers and were therefore more likely to accept a marginally unfair offer from them than from White proposers. However, this explanation is unlikely for a few reasons. First, participants believed their responses were made privately and anonymously—conditions known to mitigate efforts to conceal bias (Plant, Devine, & Brazy, 2003). Second, if participants had tried to "overcorrect" for racial bias, we would expect higher acceptance rates for Black proposers at other offer levels. Nevertheless, we addressed this potential limitation in a follow-up study by using a context in which participants would not be motivated to conceal an out-group prejudice.

Study 2

Study 2 provided a conceptual replication of Study 1 with two important new features. First, group membership was defined by college affiliation rather than by race, in order to rule out the possibility that our initial results may have reflected participants' efforts to appear nonprejudiced toward Black proposers. Second, to bolster our theoretical interpretations, we assessed the strength of participants' college identification and predicted that group differences in the costly punishment of marginally unfair offers would be greater among those reporting stronger in-group identification. We also assessed participants' expectancies for offers from in-group and out-group members to directly examine whether participants indeed expected preferential treatment from in-group members.

Method

Participants and Procedure

One hundred and fifteen Amhert College undergraduates participated in exchange for course credit or US$5. The procedure followed that of Study 1, with two exceptions. First, group membership was defined by college affiliation. The study was described as a collaborative project between the college and a nearby historic rival with similar academic status. Second, we assessed the strength of participants' identification with their college. Before beginning the task, participants completed an 8-item (1 = *strongly disagree*, 7 = *strongly agree*) college-based social identity centrality measure ($\alpha = .85$, $M = 4.75$, $SD = 0.98$; adapted from Sellers, Rowley, Chavous, Shelton, & Smith, 1997) as well as filler questions about previous online experiences to bolster the cover story. Prior to game play, participants were assigned a generic avatar based on their college affiliation and provided a three-digit identification number to establish their anonymity. As in Study 1, participants completed a practice trial to ensure understanding of the game and then considered critical proposals involving the hypothetical division of US$20, with the task comprising a 2 (Group: in-group vs. out-group) × 5 (Offer Size: US$10, US$8, US$7, US$6, or US$4) within-subjects design. Following task completion (to avoid activating fairness goals that could bias responses), participants reported the average amount of money (of US$20) they expected to receive from in-group and out-group proposers. Participants were then probed for suspicion and fully debriefed.

Results

We tested three main predictions. First, we predicted that participants would be more likely to reject marginally unfair US$8 offers from in-group than from out-group college students, replicating Study 1. Second, we predicted that college identity strength would moderate this effect. Third, we predicted that participants, particularly those with stronger in-group identification, would report an expectation of larger offers from in-group than out-group members.

Group Effect on Costly Punishment

A GEE analysis identical to that in Study 1 was conducted in which proposer group, offer size, and their interactions were included as predictors of the accept/reject decisions. Figure 2 illustrates the raw acceptance rates. As in Study 1, participants did not differ in their tendency to accept a fair US$10 offer from in-group and out-group proposers, OR = 3.05, Wald = 1.85, $p = .175$. Participants were also more likely to reject the

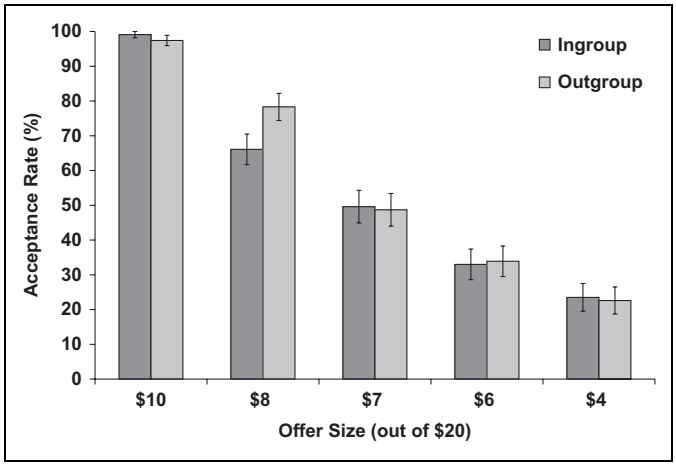

Figure 2. Study 2 acceptance rates depicted as a function of offer size and proposer college group. Error bars represent ± 1 standard error.

US$4, US$6, and US$7 offers relative to the US$10 (all $ps < .01$), regardless of proposer group (all interactions $ps > .165$). As in Study 1, the omnibus Group × Offer interaction was significant, $F(4, 1017) = 3.14$, $p = .014$, as was the specific Group × US$8 offer size contrast, OR = .18, Wald = 4.33, $p = .037$. A simple effect test confirmed that participants were more likely to reject a US$8 offer from an in-group than an out-group proposer, OR = .54, Wald = 5.90, $p = .015$, replicating our group effect from Study 1. There were no group differences at the other offer levels (all $ps > .175$)

In-Group Identification Effects

If the costly punishment of in-group proposers reflects an effort to enforce group norms, then the effect should be larger among participants who are more strongly identified with and, consequently, invested in their group. We tested whether in-group identity moderated the effect of group membership on responses to US$8/20 offers, relative to the other offers (see Table S2 in Supplementary Material found at http://spps.sagepub.com/supplemental).

A logistic regression revealed a significant two-way interaction between proposer college group and US$8 offer acceptance, OR = .51, Wald = 9.61, $p = .002$. Importantly, the three-way interaction was also significant, OR = .68, Wald = 4.08, $p = .044$, indicating that college identification moderated responses to the US$8 offer from in-group and out-group proposers. As predicted, participants with stronger (+1 *SD*) in-group identification were more inclined, OR = .35, Wald = −3.21, $p = .001$, to differentially enforce fairness norms than those with weaker (−1 *SD*) identification, OR = .74, Wald = −1.27, $p = .205$.

Fairness Expectations

To examine whether participants expected preferential treatment from in-group members, we asked them to report the anticipated average offer (of US$20) from in-group and out-group proposers. Indeed, participants expected higher offers from in-group members (M = US$8.61, $SD = 1.91$) than from out-group members (M = US$7.52, $SD = 2.11$), paired-$t(114) = 7.59$, $p < .001$, $d = .54$, with the means suggesting that the US$8 offer represented a meaningful threshold for participants' consideration of proposals.[2] Furthermore, college identification was correlated with the in-group–out-group expected disparity in offer size, $r(113) = .27$, $p = .003$, consistent with the hypothesis that group identity influences one's expectations of fairness.

Discussion

The results of Study 2 replicated those of Study 1, providing further support for our hypothesis that in-group members are punished more strictly than out-group members when they marginally violate fairness norms. The finding that this effect was enhanced among participants with stronger in-group identification bolsters our interpretation that the effect is linked to social identity. Furthermore, participants' self-reported expectations of offers closely matched their behavior, suggesting that the greater tendency to reject US$8 offers (of US$20) from in-group members than from out-group members reflected group-based fairness concerns.

Although the significance of the US$8 offer was exploratory in Study 1, it was confirmed in Study 2 with a larger sample of different participants. By using generic college avatars, Study 2 also ruled out the possibility that the observed effect was driven by concerns about appearing racially prejudiced. Unlike White Americans' racial identity, expressions of college-based pride and dominance are likely to be encouraged rather than proscribed. Together, these findings clarify the contributions of Study 1 and offer converging evidence that the harsher costly punishment of in-group members is driven by fairness concerns and the enforcement of in-group norms.

It is worth noting that across studies, offers below US$8 were rejected at a similar rate for both in-group and out-group members. This pattern appears inconsistent with the black sheep effect. Specifically, that literature would suggest that increasingly unfair in-group members would be evaluated with increasing negativity, relative to comparable offers from out-group members, such that the intergroup discrepancy would be most evident for US$4 and US$6 offers. However, this pattern was not observed in Study 1 or 2, consistent with the view that ultimatum game responses reflect reactions to unfair treatment rather than negative evaluations of the proposer. Nevertheless, to directly support this perspective and more clearly distinguish our findings from the black sheep effect, we conducted a third study.

Study 3

Study 3 tested the proposal, suggested by much prior research, that ultimatum game decisions are driven by fairness concerns rather than by proposer derogation. This study was not designed to test our main hypothesis regarding intergroup responses; rather, it was designed to provide a clean and focused test of a key assumption underlying our interpretations of Study 1 and 2 (i.e., that rejection decisions primarily reflect

fairness concerns). Although we expected that proposers of fair (US$10) offers would be evaluated more favorably than all other proposers, we did not expect increasingly unfair offers to elicit increasing derogation of the proposer (as might be predicted by the black sheep effect). More importantly, we hypothesized that rejection decisions would be more strongly predicted by fairness perceptions, and their associated negative emotional reactions, than by evaluations of the proposer.

Method

Participants and Procedure

Amazon MTurk participants ($N = 111$) considered a single random offer amount (US$4, US$6, US$7, US$8, or US$10 of US$20) from an unidentified proposer. After providing their accept/reject decision, they used a 7-point scale (1 = *not at all*, 7 = *very*) to rate the offer's fairness, their emotional reaction to the offer (*happy*, *angry*, *and irritated*), their perceptions of the proposer (*competitive, selfish, inconsiderate, cold, friendly, kind, considerate, warm, likable*, *competent, intelligent, respectable, and worthy*) as well as their preference for the proposer as a future teammate (a social exclusion index). Emotion and perception items were averaged and appropriately reverse-coded to form indices of *anger* (3 items; $\alpha = .90$), *likability* (9 items; $\alpha = .96$), and *competence* (4 items; $\alpha = .87$).

Results and Discussion

A preliminary analysis examined the effect of offer size on acceptance rate in order to establish the basic validity of the task for assessing "costly punishment" behavior. Consistent with prior ultimatum game research, smaller offers were associated with lower acceptance rates, OR = .62, $p < .001$. This effect paralleled the linear trend observed in Studies 1 and 2 ($OR_{Study1} = .40$; $OR_{Study2} = .50$).

The primary goal of Study 3 was to test whether this typical pattern of costly punishment was associated with participants' perceived fairness and subjective anger, as shown by previous ultimatum game research on norm enforcement, or with a devaluation of the proposer, which would be predicted if our aforementioned findings reflected a form of the black sheep effect. As expected, increasingly unfair offers predicted lower perceived fairness, $F(4, 106) = 41.25$, $p < .001$, and greater reactive anger, $F(4, 106) = 22.90$, $p < .001$, at each offer level, consistent with past ultimatum game findings (Knoch et al., 2006; Sanfey et al., 2003; van't Wout, Kahn, Sanfey, & Aleman, 2006).

An offer size effect also emerged for perceived likability, $F(4, 106) = 31.03$, $p < .001$, competence, $F(4, 106) = 10.37$, $p < .001$, and wanting to include the proposer on one's team, $F(4, 106) = 4.69$, $p = .002$. However, whereas the offer size effects were linear for ratings of fairness, $B = .75$, $SE = .07$, $p < .001$, and anger, $B = -.62$, $SE = .07$, $p < .001$, these effects were driven solely by proposers of US$10 offers being considered the most likable, competent, and desirable as a teammate. There were no differences in proposer's evaluations between US$4, US$6, US$7, and US$8 offers (all pairwise $ps > .125$; see Figure S1 in Supplementary Material found at http://spps.sagepub.com/supplemental), suggesting that the tendency to reject lower offers did not reflect proposer derogation.

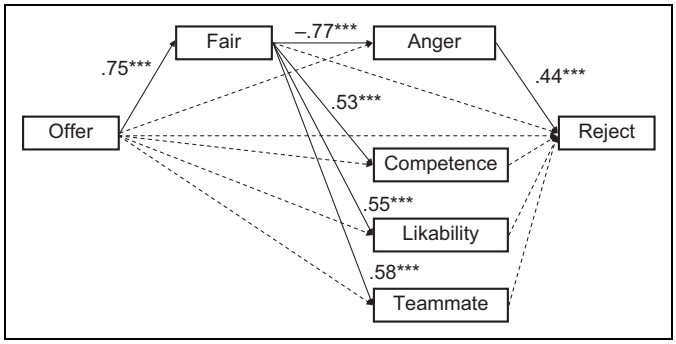

Figure 3. Study 3 path diagram of variables.
Note: Dashed paths are not statistically significant (all ps > .140) and constraining them to be 0 does not worsen the fit ($\times 2(6) = 9.26$, $p = .160$). The fit of the constrained model was satisfactory (CFI = .986, TLI = .951, RMSEA = .070).

To more directly test our predictions regarding ultimatum game decisions, we fit a path model in which fairness perceptions, emotional reactions (anger), and proposer evaluations (competence, likability, and team inclusion) were included as potential mediators through which offers were accepted/rejected (see Figure 3). Consistent with our interpretations, and with prior research, the only reliable predictor of rejection decisions was participants' feelings of anger, which in turn was driven by their offer fairness perceptions (indirect path: Offer→Fairness→Anger→Reject = −.61, $p < .001$). Importantly, offer rejections were not reliably associated with proposer evaluations. These data support the interpretation that the rejection decisions observed in Study 1 and 2 were driven by participants' reactions to fairness norm violations and not their desire to derogate proposers, further distinguishing the present findings from the black sheep effect.

General Discussion

Although in-group favoritism is typically characterized by its benefits to in-group members, our research reveals that in-group members are also punished more harshly when they violate fairness norms in a negotiation. Across two studies, we found that in-group members are held to a higher standard of fair behavior than out-group members in negotiation contexts, presumably because fairness norms operate more strongly within groups than between groups. In Study 1, White participants punished White bargaining partners more harshly than Black partners for offers that deviated slightly from an equal distribution of money. Study 2 replicated this pattern in the context of college identity: Participants were more likely to reject marginally unfair offers from fellow students than from students of a rival school, and this effect was more pronounced among strong in-group identifiers. This pattern was corroborated by participants' self-reported expectation that in-group

members would be more generous in their offers than out-group members. Importantly, Study 3 confirmed that ultimatum game decisions were driven by fairness perceptions rather than by negative proposer evaluations (as would be predicted by the black sheep effect). Together, these findings demonstrate that in-group members are held to a higher standard of fairness in negotiations and are consequently punished more strictly than out-group members for violating it.

On the surface, our findings may seem counterintuitive, as intergroup bias is typically grounded in a preference for the in-group (Allport, 1954; Brewer, 1999). However, we propose that the observed pattern reflects a higher level structural strategy for maintaining group preferences, rather than an interpersonal strategy (Balliet & Van Lange, 2013). That is, by punishing in-group members who violate fairness norms more strictly, respondents invest in maintaining the in-group preference that is typically expected. Because this preference is not expected from out-group members, the need to punish for a marginally unfair offer would not function to maintain the social system and thus is less urgent. This speculation is consistent with research showing that the degree of both actual and expected punishment is influenced by perceptions of normative cooperation obligations among in-group members (Bernhard et al., 2006; Goette, Huffman, & Meier, 2006; Shinada et al., 2004; Valenzuela & Srivastava, 2012; Yamagishi, Mifune, Liu, & Pauling, 2008). Together, our studies provide new evidence of intergroup bias in the ultimatum game and suggest that the costly punishment of in-group members may represent an important strategy for promoting in-group favoritism in reciprocity contexts.

More broadly, the pattern of costly in-group punishment observed in the present research appears to complement the black sheep effect. Both are strategies for upholding group values and maintaining cohesion, but they may differ in implementation depending on the context (performance evaluation vs. social exchange) and severity of the norm violation (major vs. marginal). A theoretical model that incorporates these two strategies may help explain a wider range of responses to norm-deviant behavior (see Ellemers & Jetten, 2013) and clarify how actions that appear incongruent with pro-in-group attitudes may actually serve a group's interests and promote in-group favoritism.

Acknowledgments

The authors wish to thank members of the Social Perception Amherst Lab for their assistance in data collection and members of the NYU Social Neuroscience Laboratory for their assistance in data collection and assistance in data collection and feedback on earlier versions of this manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was in part supported through faculty funding provided by Amherst College to Saaid A. Mendoza.

Notes

1. Our findings may appear inconsistent with those of Kubota, Li, Bar-David, Banaji, and Phelps (2013). However, their participants primarily considered "unfair" offers and, as in our research, White participants did not differ in their responses to Black and White proposers (anti-Black bias was only observed among non-White participants). Hence, our results are not inconsistent.
2. Although not a central study goal, an additional analysis demonstrated that the expectation of larger offers was associated with higher rejection rates, OR = .77, Wald = 8.88, $p = .003$, regardless of group membership.

Supplemental Material

The online tables and figures are available at http://spp.sagepub.com/supplemental.

References

Abrams, D., Marques, J. M., Brown, N., & Henson, M. (2000). Pro-norm and anti-norm deviance within and between groups. *Journal of Personality and Social Psychology*, *78*, 906–912.

Allport, G. W. (1954). *The nature of prejudice*. New York, NY: Addison-Wesley.

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*, 594–615.

Balliet, D., & Van Lange, P. A. M. (2013). Trust, punishment, and cooperation across 18 societies: A meta analysis. *Perspectives on Psychological Science*, *8*, 363–379.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *44*, 912–915.

Biernat, M., Vescio, T. K., & Billings, L. S. (1999). Black sheep and expectancy violation: Integrating two models of social judgment. *European Journal of Social Psychology*, *29*, 523–542.

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior*, *10*, 95–121.

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

Branscombe, N. R., Wann, D. L., Noel, J. G., & Coleman, J. (1993). In-group or out-group extremity: Importance of the threatened social identity. *Personality and Social Psychology Bulletin*, *19*, 381–388.

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

Brewer, M. B. (2007). The importance of being *we*: Human nature and intergroup relations. *American Psychologist*, *62*, 728–738.

Camerer, C. F., & Thaler, R. (1995). Anomalies: Dictators, ultimatums, and manners. *Journal of Economic Perspectives*, *9*, 209–219.

Castano, E., Paladino, M.-P., Coull, A., & Yzerbyt, V. Y. (2002). Protecting the ingroup stereotype: Ingroup identification and the management of deviant ingroup members. *British Journal of Social Psychology*, *41*, 365–385.

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, *11*, 319–323.

Eidelman, S., & Biernat, M. (2003). Derogating black sheep: Individual or group protection? *Journal of Experimental Social Psychology*, *39*, 602–609.

Ellemers, N., & Jetten, J. (2013). The many ways to be marginal in a group. *Personality and Social Psychology Review*, *17*, 3–21.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *TRENDS in Cognitive Sciences*, *8*, 185–190.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Frable, D. E. S. (1997). Gender, racial, ethnic, sexual, and class identities. *Annual Review of Psychology*, *48*, 139–162.

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *The American Economic Review*, *96*, 212–216.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, *3*, 367–388.

Güth, W., & Tietz, R. (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, *11*, 417–449.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *The American Economic Review*, *91*, 73–78.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . 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.

Hutchison, P., Abrams, D., Guiterrez, R., & Viki, G. T. (2008). Getting rid of the bad ones: The relationship between group identification, deviant derogation, and identity maintenance. *Journal of Experimental Social Psychology*, *44*, 874–881.

Jetten, J., Spears, R., & Manstead, A. S. R. (1997). Strength of identification and intergroup differentiation: The influence of group norms. *European Journal of Social Psychology*, *27*, 603–609.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*, 829–832.

Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The price of racial bias: Intergroup negotiations in the Ultimatum Game. *Psychological Science*, *24*, 2498–2504.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22.

Marques, J. M., Abrams, D., Paez, D., & Martinez-Taboada, C. (1998). The role of categorization and in-group norms in judgments of groups and their members. *Journal of Personality and Social Psychology*, *75*, 976–988.

Marques, J. M., Abrams, D., & Serodio, R. G. (2001). Being better by being right: Subjective group dynamics and derogation of in-group deviants when generic norms are undermined. *Journal of Personality and Social Psychology*, *81*, 436–447.

Marques, J. M., & Paez, D. (1994). The black sheep effect: Social categorization, rejection of ingroup deviants, and perception of group variability. *European Review of Social Psychology*, *5*, 37–68.

Marques, J. M., & Yzerbyt, V. Y. (1988). The black sheep effect: Judgmental extremity towards ingroup members in inter- and intra-group situations. *European Journal of Social Psychology*, *18*, 287–292.

Marques, J. M., Yzerbyt, V. Y., & Leyens, J.-P. (1988). The black sheep effect: Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, *18*, 1–6.

Plant, E. A., Devine, P. G., & Brazy, P. B. (2003). The bogus pipeline and motivations to respond without prejudice: Revisiting the fading and faking of racial prejudice. *Group Processes and Intergroup Relations*, *6*, 187–200.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision making in the Ultimatum Game. *Science*, *300*, 1755–1758.

SAS Institute Inc. (2008). *SAS/STAT® 9.2 User's Guide*. Cary, NC: Author.

Sellers, R. M., Rowley, S. A. J., Chavous, T. M., Shelton, J. N., & Smith, M. A. (1997). Multidimensional inventory of Black identity: A preliminary investigation of reliability and construct validity. *Journal of Personality and Social Psychology*, *73*, 805–815.

Shinada, M., Yamagishi, T., & Ohmura, Y. (2004). False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. *Evolution and Human Behavior*, *25*, 379–393.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization in intergroup behavior. *European Journal of Social Psychology*, *1*, 149–178.

Tajfel, H., & Turner, J. (1979). An integrative theory of integrative conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Tyler, T. R., & Blader, S. (2000). *Cooperation in groups: Procedural justice, social identity, and behavioral engagement*. Philadelphia, PA: Taylor & Francis.

Valenzuela, A., & Srivastava, J. (2012). Role of information asymmetry and situational salience in reducing intergroup bias: The case of ultimatum games. *Personality and Social Psychology Bulletin*, *38*, 1671–1683.

van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Experimental Brain Research*, *169*, 564–568.

Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism. *Advances in Group Processes*, *16*, 161–197.

Yamagishi, T., Mifune, N., Liu, J. H., & Pauling, J. (2008). Exchanges of group-based favours: Ingroup bias in the prisoner's dilemma game with minimal groups in Japan and New Zealand. *Asian Journal of Social Psychology*, *11*, 196–207.

Author Biographies

Saaid A. Mendoza is a visiting assistant professor of psychology at Amherst College, where he directs the Social Perception Amherst Lab. His research examines how stereotyping processes interact with goals

and norms to influence perceptions and behaviors within intergroup contexts.

Sean P. Lane is a postdoctorate fellow at the University of Missouri, Columbia. His research focuses on affect coregulation in close relationships and between different life domains.

David M. Amodio is an associate professor of psychology and neural science at New York University (NYU) where he directs the NYU Social Neuroscience Laboratory. His research investigates the psychological and neural mechanisms of intergroup bias and self-regulation.