Third-party sanctioning and compensation behavior: Findings from the ultimatum game

Alex K. Chavez[a,*], Cristina Bicchieri[b]

[a] Department of Psychology, University of Michigan, Ann Arbor, MI, USA.

[b] Departments of Philosophy and Psychology, University of Pennsylvania, Philadelphia, PA, USA.

[*] Corresponding author. Fax: 781-623-0407.

Email-addresses: achavez@sas.upenn.edu (A. K. Chavez), cb36@sas.upenn.edu (C. Bicchieri).

Abstract. We measured the beliefs and behavior of third parties who were given the opportunity to add to or deduct from the payoffs of individuals who engaged in an economic bargaining game under different social contexts. Third parties rewarded bargaining outcomes that were equal and compensated victims of unfair bargaining outcomes rather than punishing perpetrators, but were willing to punish when compensation was not an available option. Beliefs of whether unequal bargaining outcomes were fair differed based on the normative context, but actual punishment, compensation, and rewarding behavior did not. This paper makes a contribution to the literature on informal mechanisms of social norm enforcement by comparing negative sanctions, positive sanctions, and compensation behavior by third parties.

1. Introduction

Without legal enforcement, social norms rely on informal sanctions to produce norm-conforming behavior. Empirical studies have reported evidence of negative sanctions (e.g., punishment) both by parties directly involved in the norm-governed interaction (de Quervain, et al., 2004), and by uninvolved third parties (Kahneman, Knetsch, & Thaler, 1986; Fehr & Fischbacher, 2004; Kurzban, DeScioli, & O'Brien, 2007). In these studies, the punished transgression refers to a violation of social norms of fairness or reciprocity, but usually no independent evidence is presented about the players' consensus about the norm's existence and relevance to the experimental situation. In the definition that we adopt, a social norm (Bicchieri 2006, p.11) is a behavioral rule for which it must hold for sufficiently many people that: 1. they know that such an approved behavioral rule exists; 2. they prefer to comply with the rule if they believe a) that others will also comply (empirical expectations) and b) that others believe they ought to comply and might sanction non-compliance (normative expectations). We explore the possibility that individuals may employ compensation and rewards – in addition to sanctions – as mechanisms for upholding social norms.

Notice this definition allows a social norm to exist while not always being followed. Only if sufficiently many people have the appropriate empirical and normative expectations about others' behavior and beliefs will a social norm reliably be followed. Consensus and compliance may thus differ. While observed behavior remains a crucial measure of compliance with elicited norms, recent experimental

work has introduced the explicit use of questionnaires to assess normative consensus (Bicchieri & Chavez, 2010; Bicchieri, Xiao, & Muldoon, 2011; Krupka & Weber, 2013; Rauhut &Winter, 2010; Reuben & Riedl, 2013). In this paper, we therefore measure both the beliefs and behavior of third parties who tradeoff sanctioning, compensating, and rewarding in response to the violation of or compliance with a social norm.

An example of informal sanctioning is the punishment of individuals who divide a good unfairly, when the understanding of what constitutes fair division depends on context. Under an *equality* context, a fair division is one in which goods are allocated equally amongst all parties. Under an *equity* context, a fair division is one that divides goods according to each party's share due to merit or acquired right. Assessing normative consensus should show different judgments of what counts as fair in equity versus equality conditions and this differential assessment should be reflected in behavior. So we would expect third-party punishment levels to differ under contexts that invoke rules either of equity or equality. In particular, we would expect uneven divisions to be punished less under equity contexts in which one party is perceived as being entitled to a greater share of the good.

Although costly punishment is more commonly studied, it is not the sole mechanism for upholding a social norm. Costly rewarding also plays a role in supporting pro-social behavior. When both punishments and rewards were available to the players, second-party rewarding was frequent and promoted cooperation (Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009; Andreoni, Harbaugh, &

Vesterlund, 2003). Third-party rewarding was also common when there was the possibility of reputation formation (Milinski, Semmann, & Krambeck, 2002; Seinen & Shram, 2006). Few experimental studies, however, have compared third party punishment and rewarding in a non-repeated, reputation free game (cf. Almenberg, Dreber, Apicella, & Rand, 2011). Whereas positive and negative sanctions promote norms in that they increase their long-term expected realization, another mechanism for upholding social norms is compensation of the victim of a violation. Compensation is not directly norm promoting. It rather *honors* norms, as it points to the fact that the victim *ought* to have received a fair share or had a right to be treated fairly. If indeed compensation has such an important signaling function, we would expect all three regulatory mechanisms – compensation, punishment, and rewarding – to be employed in upholding social norms.

Our hypothesis is thus that third parties will employ both compensation and positive sanctions in addition to negative sanctions when all three regulatory mechanisms are available. Furthermore, it is worthwhile to investigate whether the presence of all three mechanisms leads individuals to discount one in favor of another. For example, the opportunity to compensate victims might focus third parties on feelings of compassion, which has been found to decrease punitive sentiment (Condon & DeSteno, 2011).

To test our hypothesis, we focused on a version of the Ultimatum Game (Guth, Schmittberger, & Schwarze, 1982) in which *Proposers* proposed a division of a sum of $10 to *Responders*, who accepted or rejected the offer. In the case of a

rejection, both parties received nothing. Previous studies found that both Proposer behavior and second-party punishment were sensitive to different fairness contexts. When the roles of Proposer and Responder were assigned randomly, Proposers offered an average of 45% of the sum (for a review, see Camerer, 2003, Chapter 2). But when the role of Proposer was earned by higher scorers on a general knowledge quiz, Proposers offered an average of only 35%. Moreover, overall rejection rates were the same whether roles were assigned randomly or based on quiz performance (Hoffman, McCabe, Keith, & Smith, 1994), although Proposers offered moderately less in the latter condition, apparently feeling entitled to a larger share under an equity context. Similarly, Falk, Fehr, and Fischbacher (2003) created different offer contexts by allowing the Proposer to 1) choose between offering 20% and 50%, and 2) choose between offering 20% and 0%. They found that in the first context, Proposers offered 20% only 31% of the time, but in the second context they always offered 20%. Moreover, in the first context, Responders who were offered 20% rejected the offer 44% of the time, whereas in the second context, Responders who were offered 20% rejected only 9% of the time. In the first context, the presence of a 50/50 option prompts a norm of fairness, whereas in the second case an offer of 20% may just be perceived as generous (see Bicchieri, 2006, Chapter 3). Thus, second-party punishment is clearly sensitive to context, and whether a given context elicits a specific norm.

Is third-party punishment, compensation, or rewarding behavior sensitive to fairness context? To answer this question, we conducted a two-stage study in which

participants engaged in an Ultimatum Game under an equity or equality context in Stage 1, and in Stage 2, third parties decided to add to or deduct from the payoffs of participants from Stage 1 based on their bargaining outcome. To determine whether third parties were willing to trade off compensation and punishment, we created a separate experimental condition in which third parties only had the option of deducting from the payoffs of Stage 1 participants. This design allowed us to assess whether third parties preferred to compensate victims or punish perpetrators of unfair bargaining outcomes, and to assess whether third parties would reward fair bargaining outcomes. We discuss data only from Stage 2 below.

2. Materials and Methods

2.1. Participants

198 college-age participants took part in Stage 2 of our study, for which advertisements specified that participants would earn $5.00 in addition to an amount that would depend on decisions made during the experiment. Of these 198 participants, we excluded 2 due to data entry problems, leaving a total sample size of 196.

2.2. Procedure

In Stage 1, participants were asked to read an article and were then given a written quiz on the contents of the article. They then played a single Ultimatum Game, in which a Proposer was randomly and anonymously paired with a Responder, and offered a division of $10 in increments of $1 to the Responder. That is, the Proposer

could choose an offer of $10 for the Proposer and $0 for the Responder, $9 for the Proposer and $1 for the Responder, $8 for the Proposer and $2 for the Responder, etc., so that the two amounts summed to $10 – hereafter, these divisions are denoted ($10,$0), ($9,$1), ($8,$2), and so on, where the first amount is the Proposer's payoff and the second amount is the Responder's payoff. The Responder then chose to accept or reject the offer; in the case of an acceptance, the Proposer and Responder got the amounts specified by the offer, but in the case of a rejection, both parties got nothing. In the *equity condition*, the roles of Proposer and Responder were assigned based on quiz performance, where the instructions emphasized the entitlement of the Proposer: "[T]hose of you who scored in the top half on the quiz have earned the right to be a Proposer," adapting the procedure of Hoffman et al. (1994). In the *equality condition*, roles were assigned randomly. Participants were told that their final payments might be adjusted up or down based on decisions made in other experiments.

In Stage 2, participants read a detailed description of the Stage 1 task described above, including the instructions that Stage 1 participants received in either the equity condition or the equality condition. Participants in Stage 2 were then given an opportunity to adjust the payoffs of one pair of participants from Stage 1 whose photocopied bargaining sheet they saw. The only bargaining outcomes we used were ones in which the Proposer offered ($5,$5), ($7,$3), or ($8,$2), and the Responder accepted. Approximately one week after Stage 1 was completed, an

experimenter paid the participants whose bargaining sheets were used in Stage 2, based on the third parties' average adjustments.

In the *all-adjustments condition*, participants were given $2.50 that they could use at a ratio of 1:4 to add to or deduct from the actual payoff of a Proposer from Stage 1, and another $2.50 that they could use at a ratio of 1:4 to add to or deduct from the payoff the Responder who was paired with the Proposer. For the uneven bargaining outcomes of ($7,$3) and ($8,$2), this condition allowed participants to punish Proposers and/or compensate Responders in various degrees and in any combination. For the even bargaining outcome of ($5,$5), it allowed participants to reward Proposers or exhibit the unlikely spiteful behavior of deducting from the payoff of either Proposers or Responders. The *deduct-only condition* was the same as the all-adjustments condition, except that participants could adjust the Proposer's and Responder's payoffs only by deducting from them, which forced participants to punish Proposers if they wished to enact any sanctions. Participants were informed that they would have to announce their decision to the room full of other third parties at the end of the experiment, and that their decisions were therefore not anonymous. Though the lack of anonymity could create incentives to perform in front of others (third parties and experimenters), we introduced it to make salient the normative expectations associated with norm violation and compliance. For example, if punishment of norm transgressors is socially expected, we would expect a sizable number of third parties to punish unfair Proposers or possibly reward fair ones. If compensation is another mechanism by

which individuals uphold social norms, we would also expect to observe compensating behavior. The public announcement should therefore lend particular salience to those mechanisms that drive norm compliance, and possibly uncover a collective preference for some mechanisms over others.

Finally, at the end of the study, Stage 2 participants were given a questionnaire. The first question read, "Do you believe the following proposal is fair for both the Proposer and the Responder?" and then listed the 11 possible divisions of the $10.00 sum. These questions elicit players' personal normative beliefs regarding fair divisions in the equity versus equality conditions described in the Stage 1 game. The second question measured second-order normative beliefs (normative expectations) in an incentivized manner:

> "The questions below refer to the first page of this survey, which all participants in this experiment are answering. After we collect all participants' forms, we will randomly select three questions from this page, for which you will earn a $1.00 bonus each, if you guessed correctly. Now for each line below, please guess whether the majority of the participants in this room thought that that particular proposal was fair. Circle your answers,"

The questionnaire then listed the 11 possible divisions again.

The binary responses to these 22 variables comprised the belief data, which were designed to measure third parties' respective first- and second-order normative beliefs about the fairness of each offer. Asking about first-order normative beliefs

does not let us conclude that a social norm is in place, and is perceived as such by the participants. We may all think that a 50/50 division is fair, but this does not necessarily imply that we also think most others hold the same conviction. For a social norm to exist (Bicchieri, 2006, p. 11), there must be collective consensus that a specific behavioral rule is socially prescribed (or proscribed) in a given context. In this case, second-order normative beliefs (normative expectations) about what is fair will be mutually consistent. For example, individuals could believe that ($7,$3) is fair (first-order normative belief), but not expect others to believe it (second-order normative belief), in which case there would be no mutually shared belief that ($7,$3) is socially prescribed.[1] Without agreement in second-order normative beliefs (normative expectations), a behavioral regularity would not be a *social* norm, and consequently there would be no agreement about what behaviors constitute a norm violation and therefore might elicit a sanction, as well as a compensation for the victim.

2.3. Design

This led to a 2 x 3 x 2 between-participants design of adjustment condition (deduct-only or punish/compensate/reward), offer (($5,$5), ($7,$3), or ($8,$2)), and fairness context (equity or equality). The dependent variable was the pair of adjustments in USD that the third party made to the Proposer's and Responder's payoffs. We predicted that the amounts paid to punish the Proposer and compensate the

[1]A typical case in which first and second-order normative beliefs diverge is that of *pluralistic ignorance*, where a majority of group members privately reject a norm, but assume incorrectly that most others accept it and thus keep obeying it (Katz & Allport, 1931)

Responder would decrease with increases in the amount that the Proposer offered to the Responder, as third parties would find fewer reasons to adjust payoffs for divisions which were closer to equality. However, we expected the effect of offer would be moderated by fairness context, as uneven offers would be interpreted as being fairer under an equity context. As a separate hypothesis, we predicted that participants would trade off punishment and compensation in the all-adjustments condition, so that the amounts they deducted from the Proposer's payoff would be lower than in the deduct-only condition. Whenever presented with even offers of ($5,$5) in the all-adjustments condition, we also expected third parties to offer some reward to Proposers for obeying a fairness norm, the reward being possibly greater in a condition of equity, where some deviation from a 50/50 share is justifiable. Because we expected little to no variation in the behavior of third parties for the ($5,$5), deduct-only condition, for which we expected no "punishment" (i.e., zero adjustments, as there was no norm violation), we intentionally created an unbalanced design, assigning only 9 participants to this condition and the remainder roughly evenly to the remaining conditions (see Table 1). As expected, none of the 9 participants in the ($5,$5), deduct-only condition made any adjustments.

MANOVA F-tests used Pillai-Bartlett's statistic. All analyses of variance respected the principle of marginality and used Type-II sums of squares.

Table 1

Sample Size by Adjustment Condition, Offer, and Fairness Context

	Adjustment condition					
	All-adjustments			Deduct-only		
Offer:	($5,$5)	($7,$3)	($8,$2)	($5,$5)	($7,$3)	($8,$2)
Fairness context						
Equity	18	17	19	4	20	19
Equality	18	17	21	5	19	19
Total	36	34	40	9	39	38

3. Results

3.1. Overview

Figure 1 shows bivariate mean payoff adjustments by fairness context, offer, and adjustment condition. As expected, there was a clear main effect of adjustment condition, as third parties on average made positive adjustments to payoffs in the punish/compensate condition, and enacted negative sanctions against Proposers in the deduct-only condition. In addition, the amount that the Proposer offered had a strong effect in the expected directions on the adjustments third parties made to the Proposer's and Responder's payoffs. There appeared to be no effect of fairness context, as evidenced by the visual similarity of the left and right panels of Figure 1.

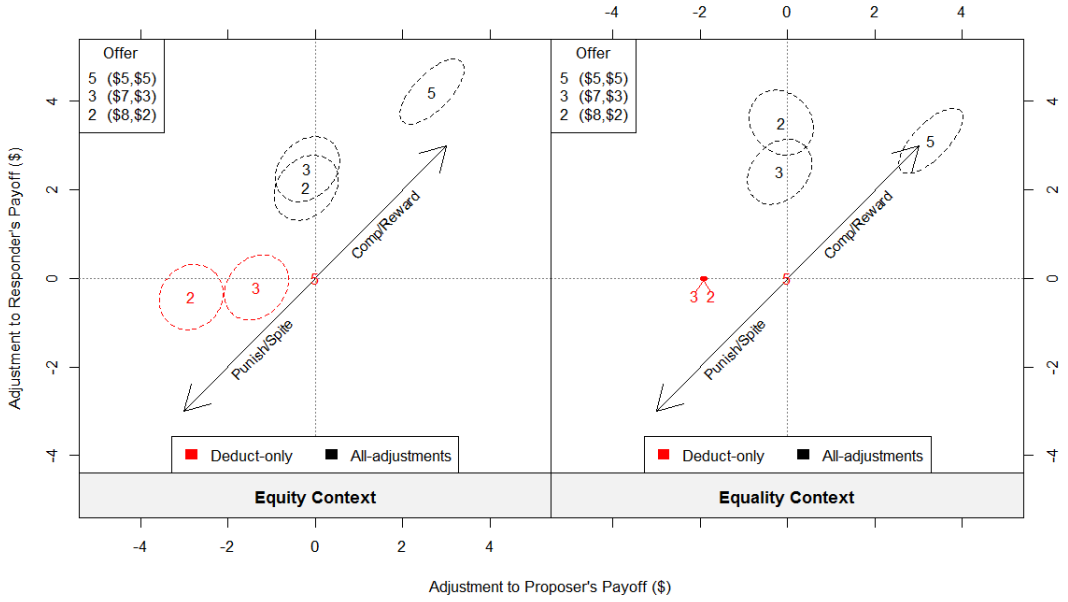

Figure 1. Ellipses show direction of covariation but are omitted for adjustments of $0.00.

3.2. Adjustment behavior

A 3 x 2 x 2 MANOVA of adjustments made to the Proposer's and Responder's payoffs confirmed the presence of main effects of offer ($F(4, 370) = 11.0$, $p < .0001$) and adjustment condition ($F(2, 184) = 46.1$, $p < .0001$)). There were no significant effects of fairness context ($F(2, 184) = 0.4$, $p = .67$), Offer x Fairness Context ($F(4, 370) = 1.7$, $p = .15$), Offer x Adjustment Condition ($F(4, 370) = 0.67$, $p = .61$), Fairness Context x Adjustment Condition ($F(2, 184) = 0.1$, $p = .94$), or the three-way interaction ($F(4, 370) = 1.0$, $p = 0.44$). Although the MANOVA detected no effect of fairness context, because we had a specific, directional hypothesis that adjustments

to the Proposer would be higher and adjustments to the Responder would be lower in the equity context, we reanalyzed the data using more powerful statistical methods. However, an exact, multivariate permutation test revealed no significant effects of fairness context ($p = .52$), fairness context by offer ($p = .25$), fairness context by adjustment condition ($p = .82$), or fairness context by offer by adjustment condition ($p = .28$).[2] Therefore, we collapsed over fairness context in subsequent analyses.

The effect of adjustment condition was driven by more negative adjustments in the deduct-only condition than in the all-adjustments condition, for both the Proposer's payoff (M = +$0.85 for deduct-only and M = -$1.79 for all-adjustments) and Responder's payoff M = +$2.96 for deduct-only and M = -$0.14 for all-adjustments). The effect of offer amount was driven by higher adjustments for ($5,$5) offers for both the Proposer's payoff (M = $2.33 for ($5,$5), M = -$0.95 for ($7,$3), and M = $1.26 for ($8,$2)) and the Responder's payoff (M = $2.87 for ($5,$5), M = $1.08 for ($7,$3), and M = $1.35 for ($8,$2)).

That adjustments to the Proposer's payoff were higher when the Proposer offered ($5,$5) was consistent with our prediction that third parties would reward fair outcomes. However, we did not expect adjustments to the Responder's payoff to also be positive for ($5,$5) offers, as Responders did nothing to generate the even

[2] The permutation test of Strasser & Weber (1999) implemented in the R *coin* package and described by Hothorn, Hornik, van de Wiel, & Zeileis (2006) maps a multivariate linear statistic, T, into a univariate test statistic by standardizing T and taking the maximum of absolute values over T. For the present application, T is the sum of bivariate adjustments in the equality and equity conditions, where the adjustments to the Responder's payoff is pre-multiplied by -1. Standardizing this vector and taking the higher of the absolute value of adjustments to the Proposer's and Responder's payoff yields the univariate statistic. Permuting the fairness context labels then yields the null distribution from which the proportion of test statistics greater than the observed statistic determines the p-value. This method is exact up to arbitrary Monte Carlo precision.

outcome. This unexpected finding may be due to third parties not wanting to create inequality by rewarding only Proposers. Indeed, creating an inequality in this case would defeat the purpose of positively sanctioning norm-abiding behavior, as in this case the desired outcome is equality.

As expected, there was no evidence of spite (negative adjustments to the Proposer for even offers, or any negative adjustments to the Responder).

3.3. Inequality Aversion and Half as a Focal Point

Did third parties attempt to equalize the payoffs of Proposers and Responders? Figure 2 shows mean payoffs before and after third party adjustments. An ANOVA of the difference between the Proposer's and Responder's *adjusted* payoffs revealed significant effects of offer amount ($F(2, 185) = 28.7$, $p < .0001$) and adjustment condition ($F(1, 185) = 5.5$, $p = .02$), and no other main effects or interactions. The Proposer's adjusted payoff was \$3.40 higher than the Responder's for offers of (\$8,\$2), $p < .0001$, \$1.97 higher for offers of (\$7,\$3) , $p < .0001$, but not significantly different for offers of (\$5,\$5), $p = .14$. In the all-adjustments condition, the Proposer's adjusted payoff was \$1.28 higher than the Responder's ($p < .0001$), and \$2.81 higher in the deduct-only condition ($p < .0001$). These remaining payoff inequalities should be judged relative to the original inequalities, which were \$0.00 for offers of (\$5,\$5), \$4.00 for (\$7,\$3), and \$6.00 for (\$8,\$2). On average, third parties reduced the amount of inequality between the Proposer's and Responder's payoffs, but did not equalize them, as indicated by the sizable and statistically significant remaining differences.

Modal adjusted payoffs were ($5,$5) for both ($5,$5) offers and ($7,$3) offers; for ($8,$2) offers, ($5,$5) was the second most frequent outcome. In the all-adjustment condition, third parties appeared to be motivated to bring the Responder's payoff up to $5, the 'rightful' share dictated by a norm of fairness. In the deduct-only condition, third parties adjusted the Proposer's payoffs of $8 or $7 to $5, but no lower. Thus, $5 was a focal point for most third parties, and their primary motivation was to ensure payoffs were close to this focal point, although the difference between the modal and mean adjusted payoffs implied significant heterogeneity. Thus, third parties sought to pursue a variety of sanctioning motives subject to the constraint that they create no additional inequality, and were not averse to allowing inequalities in payoffs to remain.

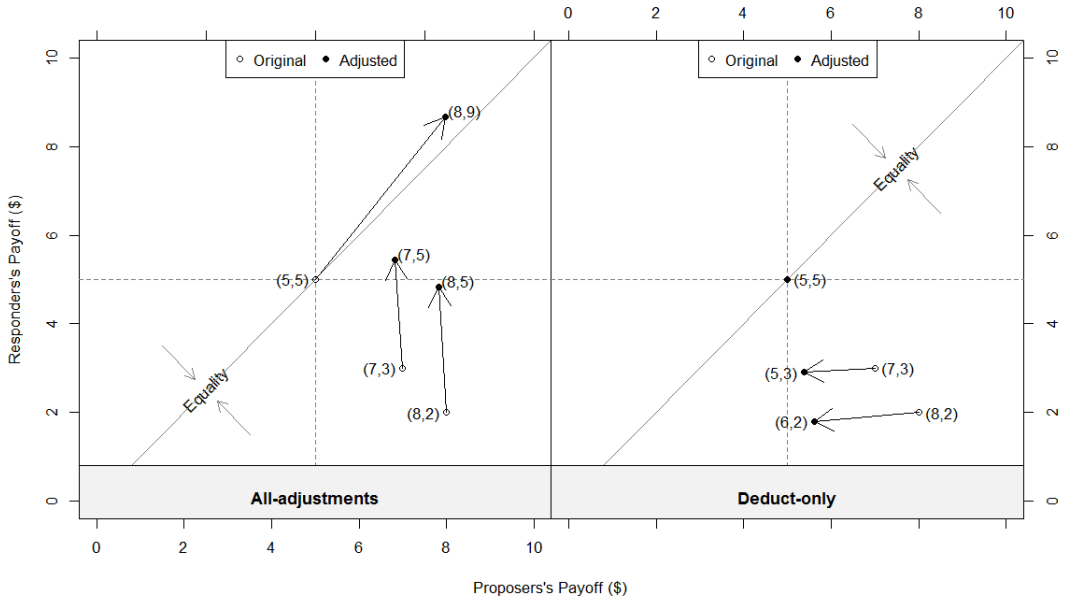

Figure 2. Payoff adjustments by adjustment condition. Black arrows join payoffs before and after third party adjustments. Payoff labels were rounded to the nearest whole number to facilitate visualization. Gray arrows represent the directions of the greatest increases in payoff equality. The 45 degree line corresponds to complete equality.

3.4. Willingness to Spend on Third-Party Sanctioning and Compensation

In addition to considering the pattern of adjusted payoffs, it is also worthwhile to ask whether the cost that third parties incurred to adjust Proposer's and Responder's payoffs differed by condition. Because third parties had to spend \$0.25 for each \$1.00 they added to or deducted from the Proposer's or Responder's payoffs, total expenditures could reach \$5.00. An ANOVA of the amounts spent by third parties revealed significant effects of offer amount ($F(2,185) = 3.7$, $p = .0268$), adjustment condition ($F(1,185) = 21.0$, $p < .0001$), and their two-way interaction ($F(2,185) = 8.7$, $p = .0063$). No other main effects or interactions reached significance.

For the deduct-only condition, expenditures were \$0.00 for offers of (\$5,\$5), \$0.43 for offers of (\$7,\$3), and \$0.65 for offers of (\$8,\$2). For the all-adjustments condition, third parties spent \$1.66 for offers of (\$5,\$5), \$0.88 for (\$7,\$3), and \$1.08 for (\$8,\$2). Of particular note is that the respective expenditures by offer amount in the all-adjustments were all significantly higher than in the deduct-only condition.

3.5. Heterogeneity in Third Party Adjustment Behavior

The fact that, on average, third parties adjusted payoffs only partially to equality could belie significant heterogeneity in motives and behavior. The average movement of adjustments in the all-adjustment condition was ($7,$5) and ($8,$5) for offers of ($7,$3) and ($8,$2), and ($8,$9) for ($5,$5) offers. Yet ($5,$5) was the modal outcome for both ($5,$5) offers and ($7,$3) offers, and it was the second most frequent outcome for ($8,$2) offers. This shows that the adjusted payoff of ($5,$5) was a focal point for most, but not all, players.

To explore potential heterogeneity in adjustment behavior, we categorized each outcome into one of the nine combinations formed by whether the Proposer's and the Responder's payoffs respectively were increased, decreased, or left unchanged; Table 2 tabulates these categories by offer amount and adjustment condition. There was a substantial number of failures to make any payoff adjustment (see *(n,n)* category), comprising 12-30% of the data in the all-adjustments condition, and 34-100% of the data in the deduct-only condition. Because adjustments had a monetary cost to participants, it is likely that such behavior reflects greater weighting of selfish motives compared to the desire to punish, reward, or compensate. Moreover, for some participants this selfish concern appears to be stronger when outcomes are very unequal, as it costs more to restore equality. Despite the observation that, on average, third parties spent more when faced with ($8,$2), in the all-adjustment condition, 30% chose to do nothing (*n,n*), as opposed to only 12% of third parties who were presented with the ($7,$3) outcome. And whereas 35% of

third partied presented with ($7,$3) compensated the Responder (*n,*+), only 15% of those presented with ($8,$2) did so.

Although the majority of the remaining adjustments reduced the amount of inequality, there was clearly a separate motive to increase total payoffs. 61% of participants increased both the Proposer's and Responder's payoffs for ($5,$5) offers, 18% did so for ($7,$3) offers, and 25% did so for ($8,$2) offers. On the whole, it appears that third parties sought to pursue a variety of sanctioning motives subject to the constraint that they create no *additional* inequality.

Table 2

Counts of Adjusted Payoff by Adjusted Payoff Category, Adjustment Condition, and Offer

		Adjusted Payoff Category[1]								
	Offer	(+,+)	(n,+)	(-,+)	(+,n)	(+,-)	(n,n)	(-,n)	(n,-)	(-,-)
All-Adj.	($5,$5)	22	4	2	1	0	7	0	0	0
	($7,$3)	6	12	12	0	0	4	0	0	0
	($8,$2)	10	6	12	0	0	12	0	0	0
Deduct-Only	($5,$5)						9	0	0	0
	($7,$3)						18	19	1	1
	($8,$2)						13	24	0	1

[1]"+" represents a positive adjustment to the offer, "-" a negative adjustment, and "n" no adjustment. In each pair, the first position represents an adjustment to the payoff

of the Proposer, and the second position represents an adjustment to the payoff of the Responder. For example, for an offer of ($7,$3), an adjusted payoff of ($5,$5) would be classified as "(-,+)" and an adjusted payoff of ($7,$7) would be classified as "(n,+)".

3.6. Third-Party Normative Fairness beliefs

Table 3 shows average third party first and second-order normative beliefs about fair offers by fairness context and offer. 47 participants were excluded due to missing data. Normative beliefs were unimodal, with almost all participants indicating that ($5,$5) was fair and that they believed the majority of other third parties believed it was fair. At the extremes, only 5% of participants considered ($10,$0) to be fair, whereas 15% considered ($0,$10) to be fair.

There were significant differences in beliefs by fairness context. Consistent with our hypotheses, participants in the equity context were more likely than participants in the equality context to view splits favoring the Proposer as being "fair for both the Proposer and the Responder," averaging across offers. On the other hand, participants in the equality context were more likely to view splits favoring the Responder as being "fair for both the Proposer and the Responder." Second-order normative beliefs (whether the participant believed the majority of other participants believed each option was fair) exhibited a similar pattern, characterized by participants in the equity context being more likely to believe that others though splits favoring the Proposer were fair. Similarly, in the equality condition, splits

favoring the Responders were believed to be judged as fair by the majority of participants. It is interesting to note that, though the "entitlements" of Stage 1 were simply due to earning high scores on a general knowledge quiz, and thus had little substantial justification, they strongly influenced third parties' first and second-order beliefs about fair divisions.

Because Table 3 reveals only average fairness beliefs, it potentially masks a variety of important individual-level patterns. Consider, for example, that 5% of participants believed ($10,$0) was "fair for both the Proposer and the Responder" (hereafter "fair for both players") and that 15% believed ($0,$10) was fair for both players. It would seem highly unlikely that these two groups overlapped at all. More generally, we expected that participants who held Proposer-favoring normative beliefs (i.e., ($10,$0), ($9,$1), ($8,$2), etc. are fair for both players) would be largely distinct from participants who held Responder-favoring beliefs (i.e., ($0,$10), ($1,$9), ($2,$8), etc. are fair for both players). In addition, we expected such individual-level patterns to vary by fairness context, with participants in the equity context being more likely to adopt Proposer-favoring normative beliefs, and participants in the equality context being more like to adopt Responder-favoring beliefs.

Table 3

Summary of Fairness Beliefs by Fairness Context

Offer	Fairness Context		
	Equity	Equality	All
Is the offer fair for the both Proposer and the Responder?			
($10,$0)	5%	4%	5%
($9,$1)	11%	4%	7%
($8,$2)	23%	5%	14%
($7,$3)	37%	18%	28%
($6,$4)	72%	52%	62%
($5,$5)	93%	100%	97%
($4,$6)	41%	61%	51%
($3,$7)	17%	27%	22%
($2,$8)	9%	23%	16%
($1,$9)	9%	22%	16%
($0,$10)	8%	21%	15%
Do the majority of other participants believe the offer is fair?			
($10,$0)	1%	0%	1%
($9,$1)	7%	0%	3%
($8,$2)	17%	3%	10%
($7,$3)	49%	18%	34%
($6,$4)	80%	65%	73%
($5,$5)	95%	100%	97%
($4,$6)	48%	61%	55%
($3,$7)	21%	23%	22%
($2,$8)	7%	12%	9%
($1,$9)	4%	9%	7%
($0,$10)	4%	8%	6%
N	74	75	149

To investigate participant-level heterogeneity in beliefs, we subjected the 11 first-order fairness beliefs to a latent class analysis. A four-class solution shown in Table 4 was chosen based on successive goodness of fit tests and interpretability. Class 1 had an estimated population share of 42%, and was characterized by participants who believed that roughly even splits – namely ($6,$4), ($5,$5), and ($4,$6) – were fair. Class 2 accounted for an estimated 34% of the population, and was

characterized by the belief that only ($5,$5) was fair. Class 3 had an estimated population share of 16%, and was characterized primarily by the belief that Responder-favoring splits –namely ($0,$10), ($1,$9), ($2,$8), ($3,$7), ($4,$6) – were fair. Finally, Class 4 accounted for 9% of population and was characterized by the belief that Proposer-favoring splits – namely ($9,$1), ($8,$2), ($7,$3), ($6,$4) – were fair.

Table 4

Estimated Latent Class Parameters by Fairness Beliefs (N = 149)

	Latent Class[1]			
Offer	1	2	3	4
($10,$0)			0.2	0.2
($9,$1)			0.3	0.4
($8,$2)			0.3	1.0
($7,$3)			0.4	1.0
($6,$4)	0.9	0.2	0.7	0.8
($5,$5)	1.0	1.0	1.0	0.8
($4,$6)	0.8		1.0	0.2
($3,$7)	0.1		1.0	0.1
($2,$8)			1.0	
($1,$9)			1.0	
($0,$10)			0.9	

Est. Population Share	41.4%	33.7%	16.1%	8.7%

Note: Estimates less than 0.1 are not shown.

[1]Class descriptions: 1) Roughly even-split; 2) Even-split; 3) Responder-favoring; 4) Proposer-favoring.

Table 5 tabulates predicted class membership – that is, the latent class to which each of the 149 participants was most likely to be assigned based on posterior modes – by fairness context. As expected, class membership varied by fairness context ($p = .0002$, Fisher's exact test for independence), with the effect being driven by the complete absence of Class 4 (those who believed Proposer-favoring splits were fair) in the equality condition, and a smaller proportion of Class 3 (those who believed Responder-favoring splits were fair) in the equity condition. Fairness context thus had a strong association with whether participants adopted fairness beliefs that favored the Proposer or the Responder. Although we do not report them here, nearly identical analyses and conclusions for the latent class analysis and the association between predicted class membership and equity context obtained for second-order beliefs.

Finally, a 4 x 2 MANOVA of adjustments made to the Proposer's and Responder's payoffs revealed no effects of predicted class membership ($F(6, 284) = 0.97$, $p = .45$), fairness context ($F(2, 141) = 0.18$, $p = .84$)), or their two-way interaction ($F(4, 284) = 0.09$, $p = .99$)).

Table 5

Cross-Tabulation of Predicted Class Membership by Fairness Context

	Latent Class				
Fairness Context	1	2	3	4	Total
Equity	25	29	7	13	74
Equality	31	27	17	0	75
Total	56	56	24	13	149

4. Discussion

Third party norm-regulating behavior appeared to be driven by at least two distinct motives: 1) to reward equal outcomes without creating inequality and 2) to adjust parties' unequal payoffs to a normative value. When adding to the payoffs of the Proposer or the Responder was an option, third parties increased the payoffs of both parties in equal amounts when the Proposer offered ($5,$5). For uneven offers of ($7,$3) or ($8,$2), however, third parties reduced inequality either by compensating the Responder when compensation was an available option or – when the only option for adjustment was to deduct – by punishing the Proposer. Of note is that adjustments in both conditions give one of the players the "rightful" payoff of $5, in

line with a fairness norm. Finally, this pattern of behavior is not compatible with a theory of inequality aversion, as third parties allowed payoff inequalities to remain and also exhibited rewarding behavior.

First and second-order normative beliefs show that a majority of third parties endorsed (and agreed upon) fairness norms and were sensitive to contextual differences. Fairness as equality (the $5,$5 outcome) was endorsed by the largest majorities of participants in both conditions. In the equity condition, a majority of 72% also believed that ($6,$4) was fair, as opposed to a small majority of 52% in the equality condition. This pattern was reversed for the ($4,$6) outcome; in the equality condition, 61% considered this outcome to be fair, whereas only 41% of third parties thought so in the equity condition. The second-order normative beliefs show the same pattern, suggesting that there is an implicit agreement as to what fairness dictates in each condition.

Note that third parties in the equity condition were judging the fairness of different possible offers under the Stage 1 description provided by the experimenters. It was thus clear to them that the "entitlement" was related to scoring higher on an arbitrary quiz. Though one may doubt that performance on an arbitrary quiz would justify entitlement, third parties believed otherwise. For example, 49% of third parties held the belief that a majority of other parties would find a ($7,$3) division to be fair in the equity condition, whereas only 18% of third parties held such belief in the equality condition.

However, the manner in which roles were assigned to Proposers and Responders did not affect adjustment behavior. Because our analyses showed strong effects of fairness context on first and second-order fairness beliefs, we consider our manipulation of the relevant fairness norm -- equality, in the case of random role assignment, and equity, in the case of performance-based role assignment -- to have been successful. Against the background of previous findings that equity contexts lowered offers (Hoffman et al., 1994), one interpretation of our findings is that the equality norm overshadowed the effects of fairness context. Thus, though third parties recognize the difference between equity and equality, their behavior is much less fine-tuned, showing that an egalitarian principle is far more important than equity in deciding how to adjust players' payoffs.

Another possibility is that normative consensus is not a good predictor of behavior, and that actual behavior better measures the intensity with which normative principles are adhered to. Whereas our questionnaire measures the level of normative consensus, actual behavior measures the intensity with which these principles are adhered to (Rauhut & Winter, 2010). Thus a norm may be present but, as Ostrom (2000) observed, "social norms may lead individuals to behave differently in the same objective situation depending on how strongly they value conformance with (or deviance from) a norm" (p. 144). In our experiment, the public declaration of adjustment should have made second-order normative expectations of fairness as equality particularly salient (and safe to publicly endorse). If third parties want to

conform to the expected majority opinion, they would discount context differences in favor of the focal ($5,$5) allocation.

In addition, third parties were more willing to incur personal costs to modify the payoffs of others when they had the opportunity to sanction or compensate, as opposed to having only the option to negatively sanction. This finding is relevant to theories of justice and tort law in which the compensation of a victim may be traded off against the punishment of a perpetrator, and suggests that on the whole, third parties are averse to punishing and prefer instead to compensate.

Although third party punishment has been extensively studied in laboratory settings, our paper contributes to the literature by comparing a broader set of regulatory mechanisms that can serve to uphold a norm, namely negative sanctions, positive sanctions, and compensation behavior by third parties. Future work could compare brain activation patterns implicated in negative sanctioning (e.g., Buckholtz, et al., 2008; de Quervain, et al., 2004) to those in positive sanctioning or compensation to test whether different mechanisms of regulating social norms activate the same neural pathways. As interest grows in studying the regulation of norms other than fairness – such as trust, reciprocity, and corruption – it will be increasingly important to consider both positive sanctioning and compensation behavior in addition to negative sanctioning to better understand how norms operate in the real world.

References

Almenberg, J., Dreber, A., Apicella, C. L., & Rand, D. G. (2011). Third party reward and punishment: Group size, efficiency and public goods. In N. M. Palmetti, & J. P. Russo (Eds.), *Psychology of Punishment*. Nova Science Publishers.

Andreoni, J., Harbaugh, W. & Vesterlund, L. (2003). The carrot or the stick: Rewards, punishments and cooperation. *The American Economic Review, 93(3)*, 893-902.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms.* Cambridge: Cambridge University Press.

Bicchieri, C. (2008). The fragility of fairness: An experimental investigation on the conditional status of pro-social norms. *Philosophical Issues, 18(1)*, 227-246.

Bicchieri, C. & Chavez, A. K. (2010) Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making, 23(2)*, 161-178.

Bicchieri, C., Xiao, E. & Muldoon, R. (2011). Trustworthiness is a social norm, but trusting is not. *Politics, Philosophy & Economics, 10(2)*, 170-187.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zaid, D. H., Gore, J. C., Jones, O. D., Marois, R. E. (2008). The neural correlates of third-party punishment. *Neuron*, *60(5)*, 930-940.

Camerer, C. (2003). *Behavioral game theory: Experiments on strategic interaction*. Princeton, NJ: Princeton University Press.

Condon, P. & DeSteno, D. (2011). Compassion for one reduces punishment for another. *Journal of Experimental Social Psychology, 47(3)*, 698-701.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science 305(5688)*, 1254-1258.

Falk, A., Fehr E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry, 41(1)*, 20-26.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms.

Evolution and Human Behavior, 25(2), 63-87.

Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research, 21(2)*, 245-253.

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3(4),* 367-388.

Hoffman, E., McCabe, K. A., Keith, S., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7(3)*, 346–380.

Hothorn, T., Hornik, K., van de Wiel, M. A., & Zeileis A. (2006). A lego system for conditional inference. *The American Statistician, 60(3)*, 257-263.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness and the assumptions of economics. *Journal of Business, 59(4)*, S285-S300.

Katz, D., & Allport F. H. (1931). *Student Attitudes*. Syracuse, N.Y.: Craftsman.

Krupka, E. L. & Weber, R. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association, 11(3)*, 495-524.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior, 28(2)*, 75-84.

Milinski, M., Semmann, D., & Krambeck, H. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature, 415(6870)*, 424-426.

Ostrom, E., 2000. Collective action and the evolution of social norms. *Journal of Economic Perspectives, 14(3)*, 137-158.

Rahut, H. & Fabian, W. (2010). A sociological perspective on measuring social norms by means of strategy method experiments. *Social Science Research, 39(6)*, 1181-1194.

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science, 325(5945)*, 1272-1275.

Reuben, E., & Riedl, A. (2013). Enforcement of contribution norms in public good

games with heterogeneous populations. *Games and Economic Behavior, 77(1)*, 122-137.

Seinen, I. & Schram, A. (2006). Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review, 50(3)*, 581-602.

Strasser, H., & Weber, C. (1999). On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics*, *8*, 220-250.