

NORM ENFORCEMENT: ANGER, INDIGNATION, OR RECIPROCITY?

Jeffrey P. Carpenter
Middlebury College

Peter Hans Matthews
Middlebury College

Abstract

The enforcement of social norms often requires that unaffected third parties sanction offenders. Given the renewed interest of economists in norms, the literature on third-party punishment is surprisingly thin. In this paper, we report the results of an experiment designed to replicate the anger-based punishment of directly affected second parties and evaluate two distinct explanations for third-party punishment: indignation and group reciprocity. We find evidence in favor of both, with the caveat that the incidence of indignation-driven sanctions is perhaps smaller than earlier studies have hinted. Furthermore, our results suggest that second parties use sanctions to promote conformism while third parties intervene primarily to promote efficiency. (JEL: C79, C91, C92, D64, H41)

1. Introduction

A recent article in *The New York Times* (Carey 2008) told the story of Deshan Fishel, a Georgian woman who witnessed a hit-and-run accident and then chased and detained the driver until police arrived, at considerable personal risk. She would later note that “all I could think about was that little kid, getting hit, and this person, getting away with it ... [the situation] just really upset me.” The broader theme of the article, entitled “Citizen Enforcers Take Aim”, was *moralistic punishment*, in particular, the enforcement of social norms by outraged but otherwise not directly affected third parties. Media interest in moralistic punishment owes much to the widespread belief that some part of the current financial crisis in the United States and Europe can be attributed to anti-social behavior in financial markets, and to the reluctance of ordinary people to bail out such behavior. Policymakers cannot afford to ignore this basic impulse, or to misunderstand its causes.

The editor in charge of this paper was Orazio Attanasio.

Acknowledgments: We thank Marco Castillo, Jeremy Clark, Carolyn Craven, Herb Gintis, Corinna Noelke, Louis Putterman, David Sloan Wilson, the editor, and four thoughtful referees for their comments on earlier drafts, as well as seminar and conference participants at the European University Institute, the Economic Science Association, the Canadian Economics Association, and Middlebury College. We also thank Okomboli Ong’ong’a for research assistance, and Middlebury College and the National Science Foundation (SES-CAREER 0092953) for financial support.

E-mail: jpc@middlebury.edu (Carpenter); pmatthew@middlebury.edu (Matthews)

The related academic literature might be traced to the experimental work of psychologists and sociologists in the early 1970s on *bystander intervention*, much of it inspired by another real-world example, the infamous murder of Kitty Genovese in Queens in 1964.¹ The contributions of Latane and Darley (1970), Borofsky, Stollack, and Meese (1971) and Shotland and Straw (1976), for example, reassured readers that the impulse to intervene was, if not universal, not exceptional. In more pedestrian terms, there would be little reason to describe crimes or other forms of anti-social behavior as committed “in broad daylight” if the increased likelihood of observation did not also mean an increased likelihood of intervention.

While most economists now accept the notion that the existence and at least limited local enforcement of norms is characteristic of all human societies (Henrich et al. 2001), most empirical studies are experimental and all but a handful of these are concerned with second-party punishment (SPP).² Consider, for example, experiments using the voluntary contribution mechanism or VCM, the focus of this paper. Fehr and Gächter (2000) allowed subjects to punish one another, at some cost to themselves, and found that free riders were often sanctioned, and that punishment, anticipated or otherwise, was associated with an increase in mean contributions. Masclet et al. (2003) replicated these results and then considered a second treatment in which punishment was nonmonetary, and concluded that even these sanctions were effective, albeit less so. Although punishment in the Masclet et al. protocol was costless, in Carpenter, Daniere, and Takahashi’s (2004) field experiment in southeast Asian urban slums, it was costly, and still even poor participants were prepared to reduce their earnings to show their disapproval of free riders. In related work, Bochet, Page, and Putterman (2006) have demonstrated that communication before contributions are made is an imperfect substitute for financial sanctions. Walker and Halloran (2004) and Gächter and Herrmann (2005) have discerned evidence of norm enforcement even in one-shot experiments, which demonstrates that not all SPP is instrumental, used either to increase one’s own payoff or the future payoffs of group members.³

Fehr and Fischbacher (2004) have observed, however, that the direct effects of norm violations are often circumscribed in the field, which leads them to conclude that most norms would not survive if second parties alone imposed sanctions, a view echoed in the more recent work of Kurzban, DeScioli, and O’Brien (2007). In their view, enforcement often requires the intervention of bystanders who are not directly affected or *third-party punishment* (TPP). The experimental literature on TPP is both

1. The Genovese example is more complicated than this, however. As one referee notes, some have interpreted the outcome as a consequence of the diffusion of responsibility, a manifestation of the *volunteer’s dilemma*.

2. One of the exceptions is Stutzer and Lalive’s (2004) paper, which documents the effect of work norms, and the attendant social pressures on those without work, on the duration of jobless spells and on the self-reported *happiness* of unemployed Swiss workers.

3. A number of other studies, like Bochet, Page, and Putterman (2006), have drawn attention to the punishment that occurs in the last round of multi-round VCM experiments, in which instrumental explanations are not available. A similar argument could be made about perfect strangers treatments. For a fuller discussion, see Falk, Fehr, and Fischbacher (2005).

thinner, however, and much newer.⁴ In fact, to motivate their own contribution, Fehr and Fischbacher (2004) cite just two other papers: Turillo et al. (2002) and an earlier version (Carpenter and Matthews 2002) of this one.

Fehr and Fischbacher (2004) examine both the extent and possible causes of TPP in one-shot dictator and prisoner's dilemma games, and find that a substantial number of third parties sanction violations of distributive or cooperative norms but that, consistent with Carpenter and Matthews (2002), TPP is weaker than SPP. In fact, the level of TPP observed in their experiment was insufficient to render antisocial behavior unprofitable though, as the authors themselves note, this could be an artifact of the design, in which there is only one third party.

Since then, a small number of published and unpublished experimental studies have further enhanced our understanding of TPP. Henrich et al. (2006), for example, find that while there is substantial variation across the small societies from which their subjects are drawn, there is some evidence of TPP in each, and that, at the population level, willingness to punish covaries with a measure of altruism. Marlowe et al. (2008) conclude that individuals in larger, more complex societies engage in more TPP than their small society counterparts, and Ottone et al. (2008) determine that TPP is cost sensitive.

This paper extends the norm enforcement literature in several directions. First, we consider TPP in the context of VCMs, a framework of special interest to economists. Consider, for example, the case of team production. Our design captures the dilemma faced by the members of two teams, who must decide not just whether to sanction free riders on their own team, but the other team as well. We conjecture that the enforcement of norms across teams is important to the development of corporate culture. In a survey of student attitudes about team production, for example, Carpenter, Matthews, and Ong'ong'a (2004) found that almost half of all respondents claimed they would sanction "shirkers" on other teams.

There is no reason that the interpretation of groups should be limited to work teams, however. To the extent that communities can be understood as networks of connected neighborhoods, each of which confronts its own local social dilemmas, the establishment of community values will reflect the transmission and enforcement of norms both within and across neighborhoods.

Our design also reflects a belief that at least some norms are contextual, and that individual behavior is often seen in group relative terms: our response to a free rider or, for that matter, someone who contributes their entire endowment, for example, often depends on whether such behavior is isolated.

Each of our experimental sessions comprised two parallel, one shot, VCMs. In our baseline treatment, there were no opportunities to punish either within or across groups. In the second, SPP treatment, participants could only sanction members of their own group, similar to Walker and Halloran (2004) and Gächter and Herrmann (2005). Three other treatments allowed for SPP and some form of TPP, with important

4. Kahneman, Knetsch, and Thaler's (1986) prescient paper, which included a brief discussion of TPP in dictator games, is the notable exception.

differences. In the third, one-way TPP, members of one group could punish members of the other, but not *vice versa*. The fourth and fifth treatments, two-way sequential TPP and two-way simultaneous TPP, allowed all subjects to punish both within and across groups but in the former, one foursome's TPP decisions were revealed to the other before the latter made their decisions, while in the latter, the sanctions were made, and revealed, at the same time.

Second, we avoid a possible demand effect present in earlier TPP experiments. As is typically the case in this literature, our monitors have no material incentive to punish. They can always earn more by free riding on the punishment doled out by others. That said, if all that third parties are allowed to do is punish—so that participation in the experiment is equivalent to norm enforcement—there is reason to be concerned that more will be spent on TPP than otherwise would be. Under our protocol, there are no isolated third parties: participants were first and foremost contributors to their own VCM. To the extent that their contribution decisions influenced how much was later available to spend on sanctions, there is reason to believe that any unearned income effect was also attenuated.

Third, and perhaps most important, our choice of treatments facilitates a richer discussion of the possible causes of TPP. On one hand, we believe that the desire to punish non-cooperators in the other foursome is at least in part a manifestation of what Elster (1998) calls the “action tendencies” of specific emotions. He observes, for example, that “if I believe that another has violated my interest, I may feel *anger*; if I believe that in doing so he has also violated a norm, I feel *indignation*” (Elster 1998, p. 48, emphasis added).⁵ In the context of our experiment, we conjecture that anger drives SPP but indignation motivates TPP. In particular, our *indignation hypothesis* asserts that when third parties punish, it is the violation of the norm itself that prompts them to do so, a proposition that does not preclude the existence of an inverse relationship between the likelihood, or level, of punishment and social distance.⁶

On the other hand, the *group reciprocity hypothesis* asserts that members of different groups will sometimes exchange gifts of norm enforcement with one another. We are surprised that this simple extension of the gift exchange paradigm—because individuals are (also) members of well-defined groups, it seems reasonable to suppose that the affiliations of donor and recipient are relevant—has received little attention in the literature. If the exchange is sequential, so that individuals are able to condition their gifts on those offered to their group, then the reciprocity is “simple” (van der Heijden et al. 2001). But if the exchange is simultaneous, individuals must condition on the expectation of gifts (Sugden 1984), in which case reciprocity is said to be “complex”.⁷

5. As one reviewer noted, one could interpret Elster (1998) to mean that indignation is never felt apart from anger. If so, our use of the terms differs a little from his. This said, we conjecture that the difference stems from the fact that when he defines the terms, it is in the context of SPP.

6. The identification of anger as motivation for SPP is itself not novel: see, for example, Fehr and Gächter (2002) or de Quervain et al. (2004). For earlier work on the importance of *negative emotions* and punishment, see also Pillutla and Murnighan (1996) and Bosman and van Winden (2002).

7. As one reviewer notes, the surprise, perhaps, is that reciprocity or “trade in sanctions” exists at all, since it costs the same to punish within and across groups.

It follows, then, that in the one-way treatment, TPP is the result of indignation alone, while in the two-way treatments, both indignation and the exchange of enforcement gifts across groups are responsible. The difference between TPP in the one- and two-way treatments is thus a measure of the differential effect of either simple or complex group reciprocity.

2. Design Details and Predictions

We adopted a one-shot framework because we wanted to eliminate some common instrumental explanations for punishment. In a repeated VCM, contributors may choose to punish members of their own group because they believe that punishment will increase how much their targets contribute in the future, thereby increasing their own future payoffs or, in the case of altruists, because they simply want to benefit other contributors. A similar logic applies to punishment outside one's group. As bystanders who are not directly affected, the contributors in one group may punish the free riders in another when engaged in indirect reciprocity (Alexander 1987) or, as altruists, to benefit other contributors. Punishment cannot be instrumental, however, when there are no future rounds.

Most VCM experiments report initial contribution levels close to 50%, a dramatic result inasmuch as the dominant strategy is to contribute nothing.⁸ There is some concern, however, that the common choice of splitting one's token endowment equally is more a reflection of participant confusion than cooperation. In their VCM experiment, for example, Houser and Kurzban (2002) found that the mean contribution was close to half the endowment, despite the fact that players knew that the other members of their group were robots who "chose" their contributions independently. They estimated, in fact, that more than half the tokens contributed could be attributed to confusion. Because confusion is a particular concern in one-shot experiments, a number of measures were taken to ensure that our participants understood, and considered carefully, the experiment.

First, participants read the lengthy instructions at their own pace and were required to answer three control questions correctly before being allowed to continue.⁹ The instructions elucidated the information that players would receive based on the treatment to which they were assigned. For example, from the beginning participants in the one-way TPP treatment knew only one group would be able to engage in TPP. However, to prevent asymmetries among the players at the contribution stage, players did not know whether they were in the group that could punish outside their group until the beginning of the punishment stage. Second, inspired by the discussion in Manski (2002), each participant was asked to enumerate some of their beliefs before deciding

8. Ledyard (1995) reviews the standard VCM literature. The same is also true, however, for VCMs with SPP (Carpenter 2007).

9. For purposes of illustration, we include the instructions for the TPP simultaneous treatment in the Appendix. The instructions made it clear that the experiment would be one-shot and to limit the effects of "confusion", subjects were required to complete a brief quiz before the experiment.

how much to contribute. In particular, each was first asked to estimate how much, on average, others would contribute and then how much others would spend to punish someone who did not contribute anything.¹⁰ In addition to encouraging the participants to think about what might happen in the experiment, the first set of beliefs allowed us to examine the extent to which conditional cooperation (Fischbacher, Gächter, and Fehr 2001) motivated our participants, while the second allowed us to test whether differences in contributions were due to differences in anticipated punishment.

The experiment was conducted over a network of personal computers in a large lab at Middlebury College, which ensured single-blind anonymity by allowing considerable space between the participants while they made their choices. The experimental parameters were as follows: there were 25 sessions (five sessions per treatment) with two four-person groups; each participant was endowed with 25 experimental monetary units or EMUs; and the marginal per capita return on contributions to the public good was 0.5. Because the contribution decisions of each group benefitted only the members of that group—in other words, free riding in one group had no effect on the gross earnings of the other—the members of the other group were bystanders. After participants made their contribution choices, they were given feedback about the group total contribution, the contribution choices of the other participants and their gross payoff. In the punishment treatments the participants were then able to *reduce* the earnings of a subset of the other players. The size of the subset depended on the treatment and each EMU spent out of a participants gross earnings from the first stage reduced the final earnings of the target by two EMUs.

In the SPP treatment, participants could only punish any or all of the other three players in their group. In the one-way TPP treatment, one group could punish only within their group, but members of the other group could punish players in both.¹¹ In the simultaneous TPP treatment, each participant could punish any of the other participants. Finally, in the sequential TPP treatment each participant could again punish any other participant, but one group made their punishment choices before the other group and the second-moving group was told how much each member of the first-moving group had spend on TPP but not who they punished.¹²

There was one other difference between our protocol and the standard VCM experiment. After the experiment was finished, each participant responded to a six question survey. We collected demographics (sex, whether the participant was an economics major, number of economics classes taken, grade point average, and math and verbal SAT scores) to control for any potential nonrandom assignment to treatment.

Our predictions about either contribution or punishment levels are not derived as the equilibrium properties of a specific model, because to be tractable, such a model

10. The elicitation of beliefs was therefore unbalanced: subjects were not also asked, for example, how much they would punish someone who contributed their entire endowment. We leave this robustness check for future research.

11. It is important to remember that there were opportunities for second party punishment in this and all treatments labeled TPP.

12. The second movers in the sequential treatment also knew how much third-party punishment the first movers had imposed before deciding how much to punish within their own group.

would require more behavioral structure than seems warranted. Instead, we start with the conjecture that punishment per unit of norm violation in the TPP treatments will exceed that in the SPP treatment. The simplest explanation is that the number of observers of each norm violator increases from three in the SPP treatment to seven in the TPP treatments, a size effect for which there is some prior evidence (Carpenter 2007, for example) in the experimental literature. To refine this explanation, we further note that if TPP is indeed the action tendency of an emotion or, consistent with the neuroeconomic research of de Quervain et al. (2004), associated with activation of the dorsal striatum, it is improbable that second parties will attempt to reduce their punishment of a norm violator by as much as the introduction of third parties increases it. In fact, our prior is that, as a first-order approximation, second parties will make no adjustment at all. For similar reasons, we also do not expect an individual who does punish third parties to economize (much) on his own resources to do so.

While it is not inevitable that punishment in either two-way treatment should exceed that in the one-way treatment—one reviewer wondered what would happen, for example, in a simultaneous TPP session if no one expected outsiders to exchange gifts of norm enforcement—it was our prior that in practice, at least some of our subjects would both expect, and offer, such gifts, and that when reciprocity was simple, and not complex, there would be still more punishment. Our predicted order of conditional punishment levels per unit of norm violation was therefore:

$$SPP < One\text{-}way\ TPP < Simultaneous\ TPP < Sequential\ TPP.$$

The predicted order of contributions reflects the costs of norm violation. If, for example, a free rider can expect to escape punishment altogether in the VCM, be punished to some extent in the SPP treatment, and punished more still in the TPP treatments, the temptation to free ride is diminished. And while a similar claim about the decision to increase a contribution from, say, seven EMUs to eight is harder to make—and harder still as contributions approach the value of the endowment—it is our sense that the basic intuition is modified, but not overturned, in which case the order of contributions should be:

$$VCM < SPP < One\text{-}way\ TPP < Simultaneous\ TPP < Sequential\ TPP.$$

3. Descriptive Statistics

Our subject pool was large (200 participants) and well paid (average earnings were \$21 in sessions that seldom lasted more than 40 minutes) for a one-shot experiment. The behavior of our subjects is summarized in Table 1.

Two characteristics of our descriptive statistics stand out. First, the ordering of average contributions is as predicted, consistent with both the indignation and group reciprocity hypotheses, despite the fact that the correlation between participant

TABLE 1. Mean behavior and demographics by treatment (standard deviations).

	Punishment Treatments				
	None	Second Party Punishment	Third Party On Way	Third Party Simultaneous	Third Party Sequential
Contribution	11.08 (5.78)	12.43 (5.46)	12.68 (4.72)	13.93 (4.69)	14.50 (4.58)
Expected Average Contribution	11.48 (6.26)	12.28 (5.90)	11.38 (4.68)	12.03 (5.23)	11.45 (4.72)
Expected Average Punishment (for contributing nothing)	-	3.48 (2.02)	3.63 (1.66)	4.03 (2.27)	4.88 (4.03)
Incidence of Second Party Punishment	-	0.50 (0.51)	0.33 (0.47)	0.30 (0.46)	0.43 (0.50)
Incidence of Third Party Punishment	-	-	0.10 (0.31)	0.30 (0.46)	0.38 (0.49)
Total Expenditure on Second Party punishment	-	2.05 (2.92)	1.40 (2.84)	0.78 (1.39)	1.23 (1.64)
Total Expenditure on Third Party Punishment	-	-	0.20 (0.61)	0.65 (1.41)	1.18 (2.06)
Second Party Expenditure (per offense)/Target EMUs Kept	-	0.26 (0.28)	0.15 (0.09)	0.11 (0.05)	0.13 (0.10)
Third Party Expenditure (per offense)/Target EMUs Kept	-	-	0.07 (0.03)	0.08 (0.04)	0.14 (0.06)
Female	0.30 (0.46)	0.45 (0.50)	0.42 (0.50)	0.35 (0.48)	0.30 (0.46)
Economics Major	0.30 (0.30)	0.30 (0.46)	0.28 (0.45)	0.30 (0.46)	0.10 (0.30)
Number of Economics Classes Completed	1.00 (2.05)	3.00 (4.01)	2.50 (3.65)	1.78 (2.68)	0.55 (0.64)
Grade Point Average	3.42 (0.40)	3.39 (0.33)	3.22 (0.39)	3.26 (0.37)	3.25 (0.34)
VerbalSAT	671.00 (65.66)	688.00 (57.90)	650.00 (81.09)	691.00 (55.20)	666.00 (73.48)
MathSAT	688.00 (69.53)	681.00 (49.38)	650.00 (83.56)	687.00 (63.48)	681.00 (59.35)

Note: Means are calculated over only the active decision makers when that set was smaller than the entire session (e.g. TPP means in the one-way treatment).

expectations of how much others will contribute and treatment is small. Second, and perhaps more important, participant expectations of how much a free rider will be punished correspond to the observed contribution levels in the five treatments. For example, the participants not only contributed the most in the sequential TPP treatment, they anticipated that there would be less toleration for free riding in this treatment, too.

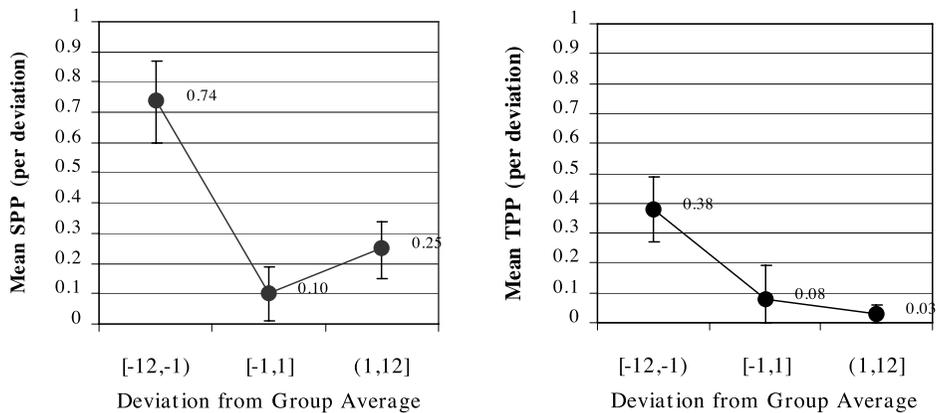


FIGURE 1. Mean second (left) and third (right) party punishment given.

While the focus of later sections is TPP, the data in Table 1 allow for some interesting comparisons between SPP and TPP. As most would expect, participants were, with few exceptions, more likely to engage in, and spend more money on, SPP than TPP, which suggests that anger is a stronger motivation than either indignation or group reciprocity. Furthermore, it seems that SPP and TPP promote different ends. Figure 1 plots the mean number of EMUs spent to punish individual participants as a function of their deviation from the group average contribution. SPP is directed at both those participants who contribute less than the average and, to a lesser extent, those who contribute more. The latter phenomenon, first documented in Fehr and Gächter's (2000) seminal paper, has been replicated several times: see, for example, Cinyabuguma, Page, and Putterman (2006) and Gächter and Herrmann (2005).¹³ The data are consistent with the notion that SPP enforces conformism, but other interpretations are possible, especially with repeated interaction (Ostrom et al. 1992; Cinyabuguma, Page, and Putterman 2006). The same question does not arise with TPP, however, which seems to be directed entirely toward those who fall short of the contribution norm. By itself the SPP–TPP difference is important because it stresses that third parties are not vital just because they provide more eyes to watch norm violators, they are particularly important because the method of third-party scrutiny seems to be inherently different, and more productive.

The overall incidence of indignation also appears to be low relative to that reported in Fehr and Fischbacher (2004). In our experiment, only 10% of participants punished outside their group in the most restrictive, one-way TPP treatment, compared to the approximately 60% who punished selfish dictators in theirs. We would argue, however, that this is still a compelling number, because there was no reason other than indignation

13. The incidence of *perverse punishment* is also similar: in the absence of opportunities for TPP, about 13.4% of punishment expenditure in our experiment was directed toward those who contributed 20 or more EMUs, while Cinyabuguma et al. (2006) report that it is common to find as much as 20% of all punishment directed toward high contributors.

to engage in TPP in this treatment and bystander intervention triples when group reciprocity is possible.

Considering reciprocity more carefully, we can test whether groups differentially engaged in reciprocal punishment gifts by looking at the correlation between TPP in the two groups. Because one group typically elicits higher contributions in a session than the other, there tends to be a negative TPP correlation between groups (i.e. the higher contributing group purchases more TPP than the lower contributing group). This causes a significantly negative baseline correlation between an individual's TPP punishment choice in the simultaneous treatment and the total outgroup punishment expenditure of the other group (Spearman's $\rho = -0.22$, $p < 0.01$). However, because the second-moving group conditions, at least partially, their TPP choices on the expenditures of the first-moving group, this raw correlation is effectively zero in the sequential treatment ($\rho = -0.01$, $p = 0.95$). Further, if we control for the target's contribution and the punisher's demographics, the correlation becomes positive and significant ($p < 0.10$) as one might expect if the sequential treatment fosters reciprocity.

We also list the mean levels of the demographics that we collected in our post-experiment survey at the bottom of Table 1. With the exception of some variation in the number of economics majors or experience in economics classes between treatments, most of the differences are small. Based on these factors, then, we achieved (at least partial) randomization into treatment.

4. Norm Enforcement and Contributions

The first column in Table 2 reports the results for a double-censored Tobit model of individual contributions.¹⁴ Inasmuch as there were few censored observations (four on the left and five on the right), the estimates are all close to the marginal effects conditional on a positive contribution. The order of the estimated treatment effects is as we expected: smallest in the standard VCM, then SPP, one-way TPP, two-way simultaneous and finally two-way sequential. Furthermore, all but the SPP coefficient are significant, relative to the VCM benchmark, at the 10% level or better.

In somewhat different terms, if one starts with the observation that the mean predicted contribution in the baseline VCM treatment was 11.0 EMUs, the increase of slightly more than one (1.481) EMUs in the SPP treatment is less impressive than first seems because the estimated coefficient is not significant at the 10%, or even 20%, level.¹⁵ On the other hand, the increase (relative to the baseline VCM) of more than one and a half EMUs (1.856) in the one-way TPP treatment is significant at the 10%

14. Note that in all the estimates reported in Tables 2 and 3 we control for gender, being an economics major, the number of economics classes one has taken, GPA and SAT scores. None of the demographics appear to robustly predict contributions; however, women and economics majors mete out more TPP.

15. Walker and Halloran (2004) also find that SPP is much less effective in one-shot environments. There is some reason to believe, however, that the results reflect, in some part, our fine-to-fee ratio: Nikofarakis and Normann (2008) find that contributions are sensitive to variations in this price.

TABLE 2. The determinants of contributions.

	Dependent Variable			
	Contribution (All Treatments)		Contribution (Punishment Treatments)	
I(Second Party Punishment)	1.481 [1.227]	0.952 [1.050]		
I(Third Party Punishment (one-way))	1.856 [1.115]*	1.692 [0.850]**	0.852 [0.982]	0.849 [0.951]
I(Third Party Punishment (simultaneous))	2.520 [1.123]**	2.518 [0.899]***	1.661 [1.004]*	1.510 [1.007]
I(Third Party Punishment (sequential))	3.212 [1.118]***	3.423 [0.853]***	2.462 [1.0924]**	2.102 [1.055]**
Expected Average Contribution		0.596 [0.074]***	0.553 [0.086]***	0.537 [0.083]***
Expected Average Punishment				0.318 [0.160]**
Observations	200	200	160	160
F, (p-value)	4.12, (<0.01)	13.75, (<0.01)	9.81, (<0.01)	004.8 (<0.01)
Pseudo R ²	0.03	0.10	0.09	0.09

Notes: I(.) denotes an indicator variable. Double-censored Tobit regressions; [robust standard errors]; Each estimate includes controls for gender, being an economics major, number of economics classes, GPA and SAT scores; *** indicates significant at 1%, ** 5%, * 10%.

level, which suggests that the combination of anticipated anger and indignation can induce more cooperation.

If, in addition, there is some expectation of group reciprocity, individual contributions should increase further still, and this is what our results show: relative to the simple VCM, individual contributions are estimated to be more than two and a half (2.520) EMUs higher in the two-way simultaneous treatment, and almost three and a half (3.212) EMUs higher in the two-way sequential, and both effects are significant at better than the 5% level.

It is not clear, however, how much the prospect of complex group reciprocity stimulates contributions: a one-tailed test of the null hypothesis that the difference between the simultaneous and one-way TPP coefficients is less than or equal to zero can be rejected at the 20%, but not 10%, level. Because punishment based on complex reciprocity is itself predicated on the resolution of a(nother) coordination problem (van der Heijden, Nelissen, and Potters 1999), this is not unexpected.

There is much less doubt about how the prospects for simple group reciprocity influence contributions, however. The difference between the two-way simultaneous and two-way sequential coefficients is significant. We believe, in short, that the contribution levels observed in the last treatment reflect both the effect of expected indignation *and* the possibilities for group reciprocity.

In the other three columns of Table 2, we examine some of the mechanisms that could explain the treatment differences in contributions that we observe. There is the obvious direct effect of the TPP treatments on contributions, but there are also

two plausible indirect effects. First, participant i might contribute more because she expected more punishment to be meted out under TPP. Second, to extend this idea one step further, participant i might also anticipate that in response to expected TPP, others will increase their own contributions, too. If the representative participant is conditionally co-operative, it seems reasonable to suppose that she would react to this belief by expecting the average contribution level to rise, and therefore to contribute more herself.

The second column evinces the importance of conditional or expectations-based cooperation in our experiment. The estimated coefficient on the expected average contribution is both substantial and significant at the 1% level. Participants contributed almost 0.6 EMUs more when the expected mean contribution of other participants increased 1 EMU.¹⁶ However, it is important to note that while conditional cooperation seems to affect cooperation, it does not necessarily work through the hypothesized channel. Although expected contributions and expected punishment are positively correlated, the inclusion of the contribution expectations in column 2 does not mediate the treatment effects significantly. It could easily be that this, rather complicated, sequence of logic is not intuitive.

The third column of Table 2 considers the same model estimated over a restricted sample, one in which observations from the simple VCM have been omitted. In this *all punishment treatments* model, the benchmark becomes SPP, and consistent with the first column, there is limited, but far from decisive, evidence that the prospect of indignation-driven TPP influences individual contributions. There is much better evidence that the prospect of group reciprocity affects individual behavior: contributions in the sequential TPP treatment are significantly higher than in the SPP.

The rationale for the third column, however, is to allow for comparisons with the fourth, in which another variable, the expected average punishment, has been added to the restricted sample. Adding the expected punishment for contributing nothing allows us to test whether participant expectations about TPP are behind the results in the first column. In particular, we were interested to see whether or not the estimated treatment coefficients became smaller in size and/or statistically less significant, as one would expect if contribution decisions reflected differences in exposure to punishment across treatments. In column four we see that the coefficient on expected punishment is indeed significant ($p < 0.05$), and its inclusion reduces the treatment coefficients, so that a channel from expected norm enforcement to contributions does exist. As a more formal test of the mediation of the direct treatment effects, we conducted a Hausman test which indicated that the coefficients on the treatment indicators do fall significantly when the punishment expectations are included ($\chi^2 = 17.91$, $p < 0.01$). This suggests to us that participants did increase their contributions in expectation of treatment differences in TPP.

16. Conditional on a positive contribution, the marginal effect is 0.598. This estimate is remarkably close to the 0.625 reported in Fehr and Fischbacher (2003). The relationship between contributions and beliefs was perhaps first explored in Croson (2007).

5. Norm Enforcement Mechanisms

If norm enforcement in our experiment is the result of both indignation and reciprocity, what, exactly, is the norm that is enforced? It is this question that motivates Table 3, which reports the estimates for three double-censored random effects tobit models of individual i 's expenditure on TPP of individual j , each of which embodies a different *situational norm*. The norms are situational in the sense that when no one else has contributed, for example, a decision not to contribute is not perceived as a violation. Each model includes both the simultaneous and the sequential TPP treatment indicators (the omitted category is one-way TPP) and separate measures of j 's deviation above and below the contribution norm, as well as the full set of demographic variables.¹⁷ In unreported regressions we also examined including the contribution of the punisher, but the coefficient was always small and insignificant indicating that our emphasis on situational or relative norms is well-placed. Lastly, the second and third columns decompose the marginal effects for the norm in the first.

The first column measures norm deviation relative to the mean contribution of the target group members. Both treatment coefficients are positive and significant at the 10% level or better—that is, participants in the simultaneous and sequential TPP treatments spent more on punishment than those in the one-way—consistent with our group reciprocity hypothesis. Furthermore, the sequential coefficient is substantially larger than the simultaneous—the null that the two are equal can be rejected at the 10% level—which implies that individuals punish more when reciprocity is easier to achieve.

The estimated slopes on deviations above and below the norm are also significant and indicate that, as seen in Figure 1, punishment falls substantially for those who contribute more than the norm. The latter is important in the context of recent research on “misdirected punishment”, defined in Gächter and Herrmann (2005) as punishment of those who contribute more than the punisher, but understood more broadly here to mean punishment of those who contribute more than some norm. The implicit focus of all these studies, however, was SPP or punishment within groups. Our results are consistent with this literature if SPP and TPP are treated as the action tendencies of two different emotions that serve two different purposes: anger/conformism in the case of SPP and indignation/contribution in the case of TPP.¹⁸

Inasmuch as a substantial number of the observations of punishment are left-censored, it becomes useful, for purposes of interpretation, to decompose the Tobit coefficients. To this end, the second and third columns report the marginal effects on the

17. We follow Carpenter and Matthews (2009) and use a spline specification for deviations from the norm so that the estimated function is continuous at the knot.

18. Because the focus of the paper is on TPP, we omit SPP from the analysis summarized in Table 3. If we include it as a baseline, however, the character of our results is unaffected. The addition of SPP confirms what we see in Table 1: depending on the treatment and contribution norm, SPP produces between 1 and 2.5 more sanctions. Consistent with other research, we also find that SPP is used significantly less often when TPP is available. When SPP is regressed on treatment indicators and other controls, point estimates of the treatment effects vary from -1.06 to -1.94 . There does not seem to be significant variation in the size of this effect across treatments in which TPP is available, however.

TABLE 3. The determinants of third-party punishment. (Dependent variable is the amount of punishment given by one player to another player in the other group).

	Definition of the Contribution Norm					
	Deviation from Target Group	Deviation from Target Group Mean (Marginal Effects)		Deviation from Punisher's		Deviation from Punisher's Contribution
		Mean	Pr(0 < TPP)	E(TPP 0 < TPP)	Expectation	
I(Third Party Punishment (simultaneous))	1.518 [0.817]*	0.087 [0.050]*	0.231 [0.127]*	1.599 [0.808]**	1.826 [0.827]**	
I(Third Party Punishment (sequential))	2.206 [0.817]**	0.134 [0.056]**	0.343 [0.131]**	2.439 [0.813]**	2.463 [0.830]**	
Slope Below the Norm	-0.348 [0.076]**	-0.018 [0.004]**	-0.051 [0.010]**	-0.308 [0.05]**	-0.270 [0.055]**	
Slope Above the Norm	-0.171 [0.061]**	-0.009 [0.004]**	-0.025 [0.012]*	-0.174 [0.055]**	-0.219 [0.059]**	
Individual Random Effects	Yes	Yes	Yes	Yes	Yes	
Observations	400	400	400	400	400	
Wald χ^2 , (p-value)	34.23, (<0.01)			35.80, (<0.01)	33.01, (<0.01)	

Notes: Punishment observations from the one-way treatment come only from the groups of four individuals that were allowed to punish outside their group. Observations from the other four players in these sessions were treated as missing. The notation I(.) denotes an indicator variable. Random effects Tobit regressions censored at 0; [standard errors]; Each estimate includes controls for gender, being an economics major, number of economics classes, GPA and SAT scores; ** indicates significant at 1%, * 5%, * 10%.

likelihood that punishment is observed and on punishment expenditures, conditional on the decision to punish, evaluated (for continuous variables) at sample means.

We first observe that conditional on the level of norm violation, participation in the two-way simultaneous and sequential TPP treatments increased the likelihood that sanctions would be imposed 8.6% and 13.4% relative to the one-way default, and that both likelihoods are statistically significant. To reprise one of the themes of this paper, punishment is not the result of indignation alone. In the one-way treatment, someone who does not contribute at all to a group in which the other three members contribute fully (and so the average in the target group is 18.75) is $(0.018 \times 18.75) = 34\%$ more likely to receive TPP. In the sequential treatment, however, this figure increases to 47%.

Conditional on the decision to punish—that is, on the observation of positive punishment—and norm deviation, subjects in the two-way treatments spent significantly more on punishment than those in the one-way, but the size of these effects is perhaps smaller than expected. In the simultaneous treatment, for example, 0.23 more EMUs were spent, and in the sequential, 0.34 more EMUs were. The coefficients on the extent of norm deviation are also significant but small: once the decision to sanction has been made, someone who has contributed 10 EMUs less than the average will receive $(0.051 \times 5) = 0.25$ more EMUs punishment than someone who has contributed five fewer in the one-way treatment, for example, and 0.59 more in the two-way sequential.

Looking at the remaining columns in Table 3, we find that all of our principal conclusions, and indeed most of the incidental ones, are robust with respect to the choice of norm. In the fourth column, for example, it is the deviation of actual from expected contributions that determines norm violation, and the estimates of both treatment coefficients and both coefficients (above and below) on the extent of norm violation are all close in size and significance to those in the first. The same holds true for the estimates in the fifth column, in which the norm is defined in terms of the punisher's own contribution, a particular implementation of the Sugden (1984) norm. In his theoretical model of public goods provision, each individual would prefer to contribute the minimum of all other contributions, in which case individuals would perhaps treat their own contributions as the relevant benchmark. Table 3 cannot tell us, however, which of these, or perhaps other, norms fits the data best, a question that we explore in considerable detail in Carpenter and Matthews (2009).

6. Concluding Remarks

To understand the nature of third-party punishment is to understand how, to invoke a popular phrase, “it takes a village”. Enforcement of prosocial norms often requires the intervention of bystanders who are nevertheless connected to the affected parties in loose networks, the sorts of networks that are common to villages. Indeed, if, as the literature on misdirected punishment hints, it is the desire to punish nonconformism that drives second parties, the enforcement of some norms would become difficult without

third parties. While we do not find as much indignation-driven punishment as, say, Fehr and Fischbacher (2004), a substantial number of our subjects were nevertheless prepared to sanction antisocial behavior even in environments where traditional notions of reciprocity were not possible. When gifts of norm enforcement can be exchanged across groups, however, there was a substantial increase in both contributions and punishment per violation.

Three possible extensions of our work come to mind. First, while our focus has been on punishment, there are some environments in which rewards are more common. Is it the case, for example, that individuals will reward both insiders and outsiders, or that more will be rewarded when reciprocal behavior is possible? As a related matter, it remains to be seen whether our results are robust with respect to the choice of frame: would it make much difference, for example, if the sanctions or rewards were cast in terms of workplace relations? Second, our reliance on student subjects will be a source of concern to some, so that it is important to know whether the same results would obtain with subjects—workers, for example—for whom contribution decisions and norms could be more salient. Third, there remains much to do on the theoretical front. The evolutionary model of group reciprocity in Carpenter and Matthews (2010), for example, is difficult to reconcile with the different motivations of second and third parties.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Appendix: experimental instructions (pdf file)

Please note: Blackwell Publishing are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

- Alexander, Richard S. (1987). *The Biology of Moral Systems*. Aldine de Guyter.
- Bochet, Olivier, Talbot Page, and Louis Putterman (2006). “Communication and Punishment in Voluntary Contribution Experiments.” *Journal of Economic Behavior and Organization*, 60, 11–26.
- Borofsky, Gerald, Gary Stollak, and Lawrence Messe (1971). “Sex Differences in Bystander Reactions to Physical Assault.” *Journal of Experimental Social Psychology*, 7, 313–318.
- Bosman, Ronald, and Frans van Winden (2002). “Emotional Hazard in a Power-to-Take Experiment.” *Economic Journal*, 112, 147–169.
- Carey, Benedict (2008). “Citizen Enforcers Take Aim.” *New York Times*, 7 October.
- Carpenter, Jeffrey (2007). “Punishing Free-Riders: How Group Size Affects SPP and the Provision of Public Goods.” *Games and Economic Behavior*, 60, 31–51.
- Carpenter, Jeffrey, Amrita Daniere, and Lois Takahashi (2004). “Cooperation, Trust, and Social Capital in Southeast Asian Urban Slums.” *Journal of Economic Behavior and Organization*, 55, 533–551.

- Carpenter, Jeffrey, and Peter Hans Matthews (2002). "Social Reciprocity." Middlebury College Department of Economics Working Paper 2002-29.
- Carpenter, Jeffrey, Peter Hans Matthews, and Okomboli Ong'ong'a (2004). "Why Punish? Social Reciprocity and the Enforcement of Prosocial Norms." *Journal of Evolutionary Economics*, 14, 407–429.
- Carpenter, Jeffrey, and Peter Hans Matthews (2009). "What Norms Trigger Punishment?" *Experimental Economics*, 12, 272–288.
- Carpenter, Jeffrey, and Peter Hans Matthews (2010). "Norm Enforcement: The Role of Third Parties." *Journal of Institutional and Theoretical Economics*, 166, 239–258.
- Cinyabuguma, Matthias, Tablot Page, and Louis Putterman (2006). "Can Second-Order Punishment Deter Perverse Punishment?" *Experimental Economics*, 9, 265–279.
- Croson, Rachel (2007). "Theories of Commitment, Altruism and Reciprocity." *Economic Inquiry*, 45, 199–216.
- de Quervain, Dominique J.-F. et al. (2004). "The Neural Basis of Altruistic Punishment." *Science*, 305, 1254–1258.
- Elster, Jon (1998). "Emotions and Economic Theory." *Journal of Economic Literature*, 36, 47–74.
- Falk, Armin, Ernst Fehr, and Urs Fischbacher (2005). "Driving Forces Behind Informal Sanctions." *Econometrica*, 73, 2017–2030.
- Fehr, Ernst, and Urs Fischbacher (2003). "The Nature of Human Altruism." *Nature*, 425, 785–791.
- Fehr, Ernst, and Urs Fischbacher (2004). "Third Party Punishment and Social Norms." *Evolution and Human Behavior*, 25, 63–87.
- Fehr, Ernst, and Simon Gächter (2000). "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90, 980–994.
- Fischbacher, Urs, Simon Gächter, and Ernst Fehr (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economic Letters*, 71, 397–404.
- Gächter, Simon, and Benedikt Herrmann (2005). "Norms of Cooperation among Urban and Rural Dwellers: Experimental Evidence from Russia." Working Paper, University of Nottingham.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Robert McElreath (2001). "In Search of Homo Economics: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review*, 91, 73–78.
- Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, and Clark Barrett et al. (2006). "Costly Punishment Across Human Societies." *Science*, 312, 1767–1770.
- Houser, Daniel, and Robert Kurzban (2002). "Revisiting Kindness and Confusion in Public Goods Experiments." *American Economic Review*, 92, 1062–1069.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1986). "Fairness and the Assumptions of Economics." *Journal of Business*, 59, s285–s300.
- Kurzban, Robert, Peter DeScioli, and Erin O'Brien (2007). "Audience Effects on Moralistic Punishment." *Evolution and Human Behavior*, 28, 75–84.
- Latane, Bibb, and John Darley (1970). *The Unresponsive Bystander: Why Doesn't He Help?* Appleton-Century-Crofts.
- Ledyard, John (1995). "Public Goods: A Survey of Experimental Research." In *The Handbook of Experimental Economics*, edited by John Kagel and Alvin Roth. Princeton University Press.
- Manski, Charles (2002). "Identification of Decision Rules in Experiments on Simple Games of Proposal and Response." *European Economic Review*, 46, 880–891.
- Marlowe, Frank W. et al. (2008). "More 'Altruistic' Punishment in Larger Societies." *Proceedings of the Royal Society B*, 275, 587–590.
- Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval (2003). "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, 93, 366–380.
- Nikofarakis, Nikos, and Hans-Theor Normann (2008). "A Comparative Statics Analysis of Punishment in Public Goods Games." *Experimental Economics*, 11, 358–369.

- Ostrom, Elinor, James Walker, and Roy Gardner (1992). "Covenants with and without a Sword: Self Governance is Possible." *American Political Science Review*, 86, 404–416.
- Ottone, Stefania, Ferruccio Ponzano, and Luca Zarri (2008). "Moral Sentiments and Material Interests Behind Altruistic Third-Party Punishment." Università degli Studi di Verona, Working Paper.
- Pillutla, Maitan M., and J. Keith Murnighan (1996). "Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers." *Organizational Behavior and Human Decision Processes*, 68, 208–224.
- Shotland, Lance, and Marget Straw (1976). "Bystander Response to an Assault: When a Man Attacks a Woman." *Journal of Personality and Social Psychology*, 34, 990–999.
- Stutzer, Alois, and Rafael Lalive (2004). "The Role of Social Work Norms in Job Searching and Subjective Well-Being." *Journal of the European Economic Association*, 2, 696–719.
- Sugden, Robert (1984). "Reciprocity: The Supply of Public Goods through Voluntary Contributions." *Economic Journal*, 94, 772–787.
- Turillo, Carmelo J., Robert Folger, James J. Lavelle, Elizabeth E. Umphress, and Julie O. Gee (2002). "Is Virtue Its Own Reward? Self-Sacrificial Decisions for the Sake of Fairness." *Organizational Behavior and Human Decision Processes*, 89, 839–865.
- van der Heijden, Eline, Jan Nelissen, Jan Potters, and Harrie Verbon (2001). "Simple and Complex Gift Exchange in the Laboratory." *Economic Inquiry*, 39, 80–97.
- Walker, James, and Matthew Halloran (2004). "Rewards and Sanctions and the Provision of Public Goods in One-Shot Settings." *Experimental Economics*, 7, 235–247.