# Norm Enforcement: Anger, Indignation or Reciprocity?

by

Jeffrey P. Carpenter
Peter Hans Matthews

March 2005

MIDDLEBURY COLLEGE ECONOMICS DISCUSSION PAPER NO. 05-03

DEPARTMENT OF ECONOMICS
MIDDLEBURY COLLEGE
MIDDLEBURY, VERMONT 05753

http://www.middlebury.edu/~econ

# Norm Enforcement: Anger, Indignation or Reciprocity?*

Jeffrey P. Carpenter† Peter Hans Matthews‡

March 2, 2005

**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, however. In this paper, we report on the results of an experiment designed to evaluate two distinct explanations for this phenomenon, *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.

**Keywords**: experiment, voluntary contribution mechanism, norm, third party punishment, reciprocity, indignation

**JEL Classification Codes**: C79, C91, C92, D64, H41

---

*This is a much revised version of an earlier paper entitled "Social Reciprocity." We thank Marco Castillo, Jeremy Clark, David Colander, Carolyn Craven, Herb Ginitis, Corinna Noelke, Louis Putterman and David Sloan Wilson 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.

†Department of Economics, Middlebury College & IZA; jpc@middlebury.edu.

‡Department of Economics, Middlebury College; pmatthew@middlebury.edu.

# 1 Introduction

Four decades have passed since the infamous murder of Kitty Genovese in Queens, New York, in 1964, but for those concerned about life in urban America, her name still resonates. It is not the crime itself that continues to shock us, but rather the indifference of *all* those who heard her cries. Most of us want to believe such widespread indifference to the violation of basic norms remains the exception and not the rule. This belief is reflected in our rhetorical conventions. There would be little reason to describe some crimes as committed "in broad daylight," for example, if the increased likelihood of observation did not also mean an increased likelihood of intervention. In this context, the experimental work of psychologists and sociologists on "bystander intervention," much of it inspired by the Genovese case, provided some reassurance: Latane and Darley (1970), Borofsky, Stollack and Meese (1971) and Shotland and Straw (1976), for example, would all conclude that the impulse to intervene was, if not universal, not exceptional.

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). One manifestation of the increasingly sociological perspective on exchange is a now substantial literature on norm enforcement. With few exceptions, however, the data are experimental, and all but a handful of these studies are concerned with *second party punishment* or SPP.[1] 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, Noussair, Tucker and Villeval (2003) replicated these results and then considered a second treatment in which punishment was non-monetary, and concluded that even these sanctions were effective, albeit less so. A potential problem with the Masclet *et al* protocol - the punishment points were (also) costless to send - was remedied in Carpenter, Daniere and Takahashi's (2004) field experiment in southeast Asian urban slums, which

[1] 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 self-reported "happiness" of unemployed Swiss workers.

found that even poor contributors were prepared to reduce their earnings to show their disapproval of free riders. In related work, Bochet, Page and Putterman (2003) have demonstrated that communication before contributions are made is an imperfect substitute for financial sanctions. Walker and Halloran (2004) and Gächter and Herrman (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.

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. In their view, enforcement often requires the intervention of unaffected bystanders or *third party punishment* (TPP), the sort of action that could have benefitted Ms. Genovese. The experimental literature on TPP is both thinner, however, and much newer.[2] In fact, to motivate their own contribution to the literature, Fehr and Fischbacher (2004) cite just two other papers: Turillo, Folger, Lavelle, Umphress and Gee (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.

This paper extends the norm enforcement literature in several directions. First, we consider TPP in the context of VCMs, a framework of obvious interest to economists (e.g., as models of the provision of public goods or team production).[3] In particular, each of our experimental sessions comprised two

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

[3]In a survey of student attitudes about team production, for example, we found that almost half of all respondents claimed that they would sanction "shirkers" on other teams, a result with important implications for the evolultion of "factory culture." For more details, see Carpenter, Matthews and Ong'ong'a (2004).

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 Herrman (2005). Three other treatments allowed for SPP and some form of TPP, with important 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. 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, 48, emphasis added). 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.

On the other hand, the *group reciprocity hypothesis* asserts that members of different groups will sometimes exchange gifts of norm enforcement with one another. 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* (Heijden, Nelissen and Potters, 1999). 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*.

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 unaffected bystanders, 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 percent, a dramatic result inasmuch as the dominant strategy is to contribute nothing.[4] 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,

[4]Ledyard (1995) reviews the standard VCM literature. The same is also true, however, for VCMs with SPP (Carpenter, 2004).

the experiment.

First, participants read the lengthy instrunctions at their own pace and were required to answer three control questions correctly before being allowed to continue.[5] Second, inspired by the discussion in Manski (2002), each participant was asked to enumerate some of their beliefs before deciding 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. 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 coopoeration (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 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 unaffected 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 2 EMUs.

In the SPP treatment, participants could only punish 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. In the simultaneous TPP treatment, each participant could punish any of the

[5]For purposes of illustration, we include the instructions for the two-way simultaneous treatment in the appendix.

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.

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 non-random assignment to treatment.

Under the indignation hypothesis, contributions will be higher in the one-way TPP treatment than in the SPP treatment, and under the group reciprocity hypothesis, contributions will be higher still in the simultaneous TPP and sequential TPP treatments. Further, because simple reciprocity is easier to achieve, contributions should be higher in the sequential TPP than in the simultaneous TPP. In sum, if both hypotheses are correct, the predicted rank order of contributions is:

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

Controlling for the norm-specific level of free-riding, we expect there to be more expenditure on punishment in the TPP treatments than in the SPP treatments. Specifically, if indignation motivates norm enforcers, there will be at least some TPP in the one-way treatment and, if SPP and TPP are not perfect substitutes, more punishment in total. When group reciprocity is possible, there should be even more TPP. And because of the difficulties of complex reciprocity, the predicted order of punishment expenditures is the same:

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

# 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 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.

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 SPP, in other words, enforces conformism. This is not true of TPP, however, which seems to be directed entirely at those who fall short of the contribution norm.

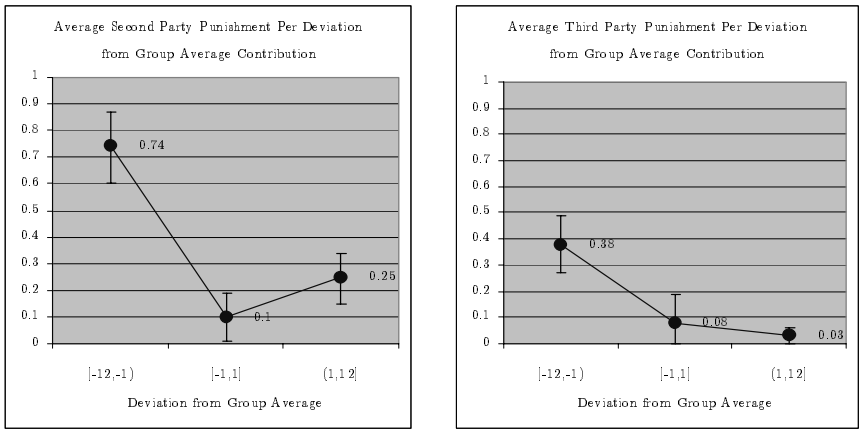


Figure 1: Norm Enforcement (second and third party punishment)

The overall incidence of indignation also appears to be low relative to that reported in Fehr and Fischbacher (2004). In our experiment, only 10 percent of participants punished outside their group in the one-way TPP treatment, compared to the approximately 60 percent 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 to engage in TPP. Nevertheless, bystander intervention triples when group reciprocity is possible.

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. 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.[6] The order of the estimated treatment effects is as predicted: smallest in the standard VCM, then mutual monitoring, one-way TPP, two-way simultaneous and finally two-way sequential. Furthermore, all but the mutual monitoring coefficient are significant, relative to the VCM benchmark, at the 10 percent level or better and (not shown) the difference between the point estimates of the two-way treatments is also statistically significant.

[*Table 2 about here*]

From another perspective, if one starts with the observation that the mean predicted contribution was 11.0 EMUs in the baseline VCM treatment, the increase of almost one EMU under the mutual monitoring treatment is perhaps less impressive than first seems because the estimated coefficient is not significant at the 10, or even 20, percent level.[7] The increase of more than one-

[6] Full details of the decompositions are available on request from the authors.

[7] Walker and Halloran (2004) also find that SPP is much less effective in one-shot environments.

and-a-half EMUs from the VCM to the one-way TPP treatment, on the other hand, is significant at the 10 percent level, which implies that the combination of anticipated anger and indignation can induce (more) cooperation.

It comes as little surprise, then, that the combination of anger, indignation and group reciprocity produces an effect (relative to the VCM benchmark) that is significant at the 1 percent level: the estimated increase in individual contributions is more than two-and-a-half EMUs in the two-way simultaneous treatment, and almost three-and-a-half in the two-way sequential. It is less clear, however, is how much complex reciprocity adds to this combination: 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 10, but not 20, percent level. Because complex reciprocity is predicated on the resolution of a(nother) coordination problem (Heijden, Nelissen and Potters, 1999), this is not unexpected.

There is much less doubt about the importance of simple reciprocity, however: as mentioned above, the difference between the two-way simultaneous and two-way sequential coefficients is significant. The contribution levels observed in the last treatment, in other words, result from indignation *and* the possibilities for group reciprocity.

The same column also 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 percent level. Participants contributed almost 0.6 EMUs more when the expected mean contribution of other participants increased 1 EMU.[8]

Last, we note that none of the demographic variables are significant at the 10 percent level, from which we conclude that none of the previous results are the consequence of non-randomization into treatment.

In the second two columns of Table 2, we examine some of the mechanisms that could explain the 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

[8]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).

expected TPP, others will increase their own contributions, too. Because the representative participant is conditionally co-operative (recall that the estimated coefficient on expected average contribution in the first column is both positive and significant), 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 of Table 2 considers the same model estimated over a restricted sample, one in which observations from the 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 of indignation-driven TPP. The evidence in favor of group reciprocity is much stronger: contributions in the two-way sequential TPP treatment are significantly higher than under mutual monitoring.

The rationale for the second column, however, is to allow for comparisons with the third, 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 insignificant, as one would expect if contribution decisions reflected differences in exposure to punishment across treatments. Although the coefficient on expected punishment is significant at the 5 percent level, and the treatment coefficients are smaller, so that a channel from norm enforcement to contributions does exist, the reduction in TPP treatment coefficients is modest (7 percent for the simultaneous treatment and 15 percent for the sequential treatment) and the $p$-values do not fall dramatically (from 0.07 to 0.09 for the simultaneous and from 0.01 to 0.03 for the sequential). Further, in an unreported regression, we found that despite the positive correlation between individual expectations about contributions and the punishment of free riders, the estimated coefficient was only significant at the 20 percent level. This suggests to us that participants did indeed increase their contributions in expectation of treatment differences in TPP, but that they did not necessarily anticipate that other participants would react in the same way.

# 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*.[9] 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 two-way simultaneous and two-way 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 punisher's own contribution and the full set of demographic variables. In addition, the second and third columns decompose the marginal effects for the norm in the first.

[*Table 3 about here*]

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

The estimated coefficients on deviation above and below the norm are also significant (at the 1 percent level) and have opposite signs. The latter is important in the context of previous studies that find evidence of "misdirected punishment" (Gächter and Hermann, 2004) or, in other words, punishment of those who also contribute more than the punisher. The implicit focus of all these studies, however, was SPP or punishment within groups. Our results are consistent with these if SPP and TPP are the action tendencies of different emotions that serve different purposes - in the first case, anger that enforces conformism and in the second, indignation that enforces contribution norms.

[9] The results are similar if one clusters errors on the individual instead. Details available on request from the authors.

The surprise, perhaps, is that while individual characteristics did not seem to influence contribution decisions, the same cannot be said about the decision to punish norm violators. Under this norm, sex and math SAT score were both significant. In fact, the substantial female coefficient is robust across norm specification: women do not contribute more than men, ceteris paribus, but are more committed to norm enforcement.[10]

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 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 8.0 and 11.4 percent 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. Further, the difference between these likelihoods is itself significant.

Those who contribute less than the prospective punisher are significantly more likely to be punished, and *vice versa*. In the one-way treatment, someone who does not contribute to the group project at all is $28 = 25(0.011)$ percent more likely to be punished by someone who has contributed her entire endowment. In the two-way sequential, treatment, however, this figure increases to almost 40 percent.

The sex effect manifests itself as an 8.8 percent differential in the likelihood that a particular norm violation is punished.

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 punish than those in the one-way, but the size of these effects is perhaps smaller than expected. In the two-way simultaneous treatment, for example, 0.23 more EMUs were spent, and in the two-way sequential, 0.31 more

[10] 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.

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 20 fewer EMUs will receive 0.35 more EMUs punishment than someone who has contributed 10 fewer in the one-way treatment, for example, and 0.66 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, both coefficients (above and below) on the extent of norm violation, and the coefficient on sex 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.

# 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 non-conformism 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 (2002), for example, is difficult to reconcile with the different motivations of second and third parties.

# 7 Appendix: Experimental Instructions for the Simultaneous TPP Treatment

You have been asked to participate in an economics experiment. For participating today and being on time you will be paid a show-up fee of $5. You may earn an additional amount of money depending on your decisions in the experiment. All your earnings will be paid to you in cash at the end of the experiment.

During the experiment the 8 participants will be randomly divided into 2 groups of 4. The experiment has two stages.

At the beginning each participant receives a 25 EMU endowment. In Stage One each of you will decide how much of the 25 EMUs to contribute to a group project and how much you want to keep for yourself. You are asked to contribute whole EMU amounts (i.e., a contribution of 5 EMUs is alright, but 3.85 should be rounded up to 4). Your payoff and the payoff of everyone else in your group will be determined by how much each member contributes to the group project and how much each member keeps.

To record your decision, you will type EMUs amounts in two text-input boxes, one for the group project labeled GROUP ALLOCATION and one for yourself labeled PRIVATE ALLOCATION. These boxes will be yellow. Once you have made your decision, there will be a green SUBMIT button that will

record your decision.

After all the participants have made their decisions, each of you will be informed of your gross earnings for the period. Your Gross Earnings will consist of two parts: 1) Earnings from your Private Allocation. You are the only beneficiary of EMUs you keep. More specifically, each EMU you keep increases your earnings by one. 2) Earnings from the Group Project. Each member of the group gets the same payoff from the group project regardless of how much he or she contributed. The payoff from the group project is calculated by multiplying 0.5 times the total EMUs contributed by the members of your group.

Your Earnings can be summarized as follows: $1\times$(EMUs you keep) + $0.5\times$ (Total EMUs contributed by your group)

Let's discuss three examples. Example 1: Say each member of your group contributes 15 of their 25 EMUs. In this case, the group total contribution to the project is $4\times15 = 60$ EMUs. Each group member earns $0.5\times60 = 30$ EMUs from the project. The gross earnings of each member will then be the number of EMUs kept, 25-15 = 10, plus the earnings from the group project, 30 EMUs, for each member. Hence, each member would earn 10+30 = 40 EMUs.

Example 2: Now say everyone in the group contributes 5 EMUs. Here the group total contribution will be 20 and each member will earn $0.5\times20 = 10$ EMUs from the group project. This means that the total earnings of each member of the group will be 20 (the number of EMUs kept) plus 10 (earnings from the group project) which equals 30 EMUs.

Example 3: Finally, say three group members contribute all their EMUs and one contributes none. In this case, the group total contribution to the project is $3\times25 = 75$ EMUs. Each group member earns $0.5\times75 = 37.5$ EMUs from the project. The three members who contributed everything will earn 0+37.5 = 37.5 EMUs and the one member who contributed nothing will earn 25+37.5 = 62.5 EMUs.

In stage two you will be shown the allocation decisions made by all the other participants, and they will see your decision. Also at this stage you will be able to reduce the earnings of other participants, if you want to, and the other participants will be able to reduce your earnings. You will be shown how much each member of your group kept and how much they allocated to the group project. You will also be shown how much each member of the other group kept

and how much they contributed to their group project. Your allocation decision will also appear on the screen and will be labeled 'YOU'.

At this point you will decide how much (if at all) you wish to reduce the earnings of the other participants. You reduce someone's earnings by typing the number of EMUs you wish to spend to reduce that person's earnings into the input-text box that appears below that participant's allocation decision.

For each EMU you spend you will reduce the earnings of the other participant by 2 EMUs. You can spend as much of your accumulated earnings as you wish to reduce the earnings of the other participants.

Consider this example: suppose you spend 2 EMUs to reduce the earnings of a participant in the other group, you spend 9 EMUs reducing the earnings of a participant in your group, and you don't spend anything to reduce the earnings of the remaining participants. Your total cost of reductions will be (2+9+0) or 11 EMUs. When you have finished you will click the blue DONE button.

How much a participant's gross earnings are reduced is determined by the total amount spent by all the other participants in this session. If a total of 3 EMUs is spent, then this person's earnings will be reduced by 6 EMUs. If the other participants spend 4 EMUs in total, the person's earnings would be reduced by 8 EMUs, and so on.

Again, for each EMU you spend you will reduce the earnings of the other participant by 2 EMUs. You can spend as much of your accumulated earnings as you wish to reduce the earnings of each of the other participants. When you have finished click the blue DONE button.

Nobody's earnings will be reduced below zero by the other participants. For example, if your gross earnings were 40 EMUs and the other participants spent 50 EMUs to reduce your earnings, your gross earnings would be reduced to zero and not minus 60.

Your NET EARNINGS after the third stage will be calculated as follows: (Gross Earnings from Stage One) - (2×the Number of EMUs spent on reductions directed towards you) - (your expenditure on reductions directed at other participants).

If you have any questions please raise your hand. Otherwise, click the red TAKE QUIZ button when you are done reading. You will then answer a few questions about the experiment so that we make sure that everyone understands.

Pay attention because you will not be allowed to continue until you provide the correct answers.

# 8 References


Alexander, R. The Biology of Moral Systems. New York: Aldine de Guyter, 1987.

Bochet, Olivier; Page, Talbot and Putterman, Louis. "Communication and Punishment in Voluntary Contribution Experiments." Journal of Economic Behavior & Organization, 2003, forthcoming.

Borofsky, Gerald; Stollak, Gary and Messe, Lawrence. "Sex Differences in Bystander Reactions to Physical Assault." Journal of Experimental Social Psychology, 1971, 7, pp. 313-18.

Carpenter, Jeffrey. "Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods." Games and Economic Behavior, 2004, forthcoming.

Carpenter, Jeffrey; Daniere, Amrita and Takahashi, Lois. "Cooperation, Trust, and Social Capital in Southeast Asian Urban Slums." Journal of Economic Behavior & Organization, 2004a, 55(4), pp. 533-51.

Carpenter, Jeffrey and Matthews, Peter. "Social Reciprocity," Middlebury College Department of Economics Working Paper 2002-29. 2002.

Carpenter, Jeffrey; Matthews, Peter and Ong'ong'a, Okomboli. "Why Punish? Social Reciprocity and the Enforcement of Prosocial Norms." Journal of Evolutionary Economics, 2004b, 14(4), pp. 407-29.

Elster, Jon. "Emotions and Economic Theory." Journal of Economic Literature, 1998, 36(March), pp. 47-74.

Fehr, Ernst and Fischbacher, Urs. "The Nature of Human Altruism." Nature, 2003, 425(October), pp. 785-91.

Fehr, Ernst and Fischbacher, Urs. "Third Party Punishment and Social Norms." Evolution and Human Behavior, 2004, 25, pp. 63-87.

Fehr, Ernst and Gaechter, Simon. "Cooperation and Punishment in Public Goods Experiments." American Economic Review, 2000, 90(4), pp. 980-94.

Fischbacher, Urs; Gaechter, Simon and Fehr, Ernst. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." Economic Letters, 2001, 71, pp. 397-404.

Gaechter, Simon and Herrmann, Benedikt. "Norms of Cooperation among Urban and Rural Dwellers: Experimental Evidence from Russia," University of Nottingham Working Paper. 2005.

Henrich, Joseph; Boyd, Robert; Bowles, Samuel; Camerer, Colin; Fehr, Ernst; Gintis, Herbert and McElreath, Richard. "In Search of Homo Economics: Behavioral Experiments in 15 Small-Scale Societies." American Economic Review, 2001, 91(2), pp. 73-78.

Houser, Daniel and Kurzban, Robert. "Revisiting Kindness and Confusion in Public Goods Experiments." American Economic Review, 2002, 92(4), pp. 1062-69.

Kahneman, Daniel; Knetsch, J. and Thaler, R. "Fairness and the Assumptions of Economics." Journal of Business, 1986, 59(4), pp. s285-s300.

Latane, Bibb and Darley, John. The Unresponsive Bystander: Why Doesn't He Help? New York: Appleton-Century-Crofts, 1970.

Ledyard, John. "Public Goods: A Survey of Experimental Research," J. Kagel and A. Roth, The Handbook of Experimental Economics. Princeton: Princeton University Press, 1995, 111-94.

Manski, Charles. "Identification of Decision Rules in Experiments on Simple Games of Proposal and Response." European Economic Review, 2002, 46(4-5), pp. 880-91.

Masclet, David; Noussair, Charles; Tucker, Steven and Villeval, Marie-Claire. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." American Economic Review, 2003, 93(1), pp. 366-80.

Shotland, Lance and Straw, Margret. "Bystander Response to an Assault: When a Man Attacks a Woman." Journal of Personality and Social Psychology, 1976, 34(5), pp. 990-99.

Stutzer, Alois and Lalive, Rafael. "The Role of Socail Work Norms in Job Searching and Subjective Well-Being." Journal of the European Economic Association, 2004, 2(4), pp. 696-719.

Sugden, Robert. "Reciprocity: The Supply of Public Goods through Voluntary Contributions." Economic Journal, 1984, 94(376), pp. 772-87.

Turillo, C.J.; Folger, R.; Lavelle, J.J.; Umphress, E.E. and Gee, J. O. "Is Virtue Its Own Reward? Self-Sacrificial Decisions for the Sake of Fairness." Organizational Behavior and Human Decision Processes, 2002, 89, pp. 839-65.

van der Heijden, Eline; Nelissen, Jan; Potters, Jan and Verbon, Harrie. "Simple and Complex Gift Exchange in the Laboratory." Economic Inquiry, 2001, 39(2), pp. 280-97.

Walker, James and Halloran, Matthew. "Rewards and Sanctions and the Provision of Public Goods in One-Shot Settings." Experimental Economics, 2004, 7(3), pp. 235-47.

# 9 Tables

Table 1: Mean Behavior & Demographics by Treatment (standard deviations)

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

Table 2: The Determinants of Contributions

| | Dependent Variable | | |
|---|---|---|---|
| | Contribution (All Treatments) | Contribution (Punishment Treatments) | |
| Second Party Punishment | 0.952 | | |
| | [0.926] | | |
| Third Party Punishment (one-way) | 1.692 | 0.852 | 0.849 |
| | [0.930]* | [0.891] | [0.877] |
| Third Party Punishment (simultaneous) | 2.518 | 1.616 | 1.510 |
| | [0.918]*** | [0.895]* | [0.882]* |
| Third Party Punishment (sequential) | 3.423 | 2.462 | 2.102 |
| | [0.888]*** | [0.932]*** | [0.929]** |
| Expected Average Contribution | 0.596 | 0.553 | 0.537 |
| | [0.055]*** | [0.067]*** | [0.063]*** |
| Expected Average Punishment | | | 0.318 |
| | | | [0.128]** |
| Female | -0.651 | -0.912 | -0.906 |
| | [0.635] | [0.700] | [0.689] |
| Economics Major | -0.598 | -0.685 | -1.159 |
| | [0.954] | [0.992] | [0.998] |
| Economics Classes Completed | 0.034 | 0.037 | 0.110 |
| | [0.134] | [0.139] | [0.140] |
| Grade Point Average | -1.249 | -0.388 | -0.473 |
| | [0.883] | [1.021] | [1.006] |
| Verbal SAT | 0.006 | 0.004 | 0.004 |
| | [0.005] | [0.006] | [0.006] |
| Math SAT | 0.004 | 0.006 | 0.008 |
| | [0.005] | [0.005] | [0.005] |
| Intercept | 1.847 | -0.187 | -1.982 |
| | [3.725] | [4.213] | [4.209] |
| Observations | 200 | 160 | 160 |
| $Chi^2$, (p-value) | 125, (<0.01) | 83, (<0.01) | 90 (<0.01) |
| Pseudo $R^2$ | 0.10 | 0.09 | 0.09 |

Notes: Double-censored Tobit regressions; [standard errors]; *** indicates significant at 1%, ** 5%, * 10%.

Table 3: The Determinants of Third Party Punishment

| | Definition of the Contribution Norm | | | | |
|---|---|---|---|---|---|
| | Deviation from Reference Group Mean | Deviation from Reference Group Mean (Marginal Effects) | | Deviation from Punisher's Expectation | Deviation from Punisher's Contribution |
| | | Pr(0<TPP) | E(TPP\|0<TPP) | | |
| Third Party Punishment (simultaneous) | 1.537 | 0.080 | 0.227 | 1.703 | 2.270 |
| | [0.813]* | [0.047]* | [0.123]* | [0.821]** | [0.875]*** |
| Third Party Punishment (sequential) | 2.077 | 0.114 | 0.313 | 2.563 | 2.699 |
| | [0.810]*** | [0.052]** | [0.126]** | [0.833]*** | [0.879]*** |
| Deviation Above Norm | -0.483 | -0.022 | -0.067 | -0.194 | -0.144 |
| | [0.188]*** | [0.007]*** | [0.024]*** | [0.082]** | [0.079]* |
| Deviation Below Norm | 0.244 | 0.011 | 0.035 | 0.197 | 0.155 |
| | [0.083]*** | [0.004]** | [0.012]*** | [0.054]*** | [0.051]*** |
| Female | 1.593 | 0.088 | 0.240 | 1.958 | 1.996 |
| | [0.515]*** | [0.033]*** | [0.080]*** | [0.532]*** | [0.561]*** |
| Economics Major | 0.978 | 0.054 | 0.148 | 1.649 | 1.063 |
| | [0.676] | [0.045] | [0.109] | [0.690]** | [0.697] |
| Economics Classes Completed | -0.211 | -0.010 | -0.030 | -0.259 | -0.241 |
| | [0.160] | [0.007] | [0.022] | [0.162] | [0.177] |
| Grade Point Average | -0.142 | -0.006 | -0.020 | 0.450 | 0.421 |
| | [0.788] | [0.037] | [0.112] | [0.809] | [0.837] |
| Verbal SAT | 0.003 | 0.000 | 0.001 | -0.001 | -0.001 |
| | [0.004] | [0.0002] | [0.0006] | [0.004] | [0.004] |
| Math SAT | -0.008 | -0.0004 | -0.001 | -0.003 | -0.008 |
| | [0.004]** | [0.0002]** | [0.0005]** | [0.004] | [0.004]** |
| Intercept | -1.791 | | | -5.113 | -1.931 |
| | [2.863] | | | [2.982]* | [2.943] |
| Individual Random Effects | Yes | Yes | Yes | Yes | Yes |
| Observations | 400 | 400 | 400 | 400 | 400 |
| Wald $chi^2$, (p-value) | 31.76, (<0.01) | | | 33.77, (<0.01) | 27.53, (<0.01) |

Notes: Random Effects Tobit Regressions censored at 0; [standard errors]; *** indicates significant at 1%, ** 5%, * 10%.