In: Psychology of Punishment
Editors: N. M. Palmetti et al., pp.
ISBN 978-1-61324-115-8

Chapter 4

THIRD PARTY REWARD AND PUNISHMENT: GROUP SIZE, EFFICIENCY AND PUBLIC GOODS

Johan Almenberg[1], Anna Dreber[2,3], Coren L. Apicella[6] and David G. Rand[3,4,5,*]

[1]Ministry of Finance, Stockholm Sweden[#]
[2]Stockholm School of Economics, Stockholm Sweden
[3]Program for Evolutionary Dynamics
[4]Department of Psychology
[5]Berkman Center for Internet and Society, Harvard University, Cambridge MA USA
[6]Department of Health Care Policy, Harvard Medical School, Boston MA USA

ABSTRACT

Costly third party punishment has been utilized as a tool for studying the enforcement of social norms. Experiments on this topic typically involve a third party observer who can pay to decrease the payoff of a

[*] Corresponding author: drand@fas.harvard.edu
[#] The views expressed in this chapter are those of the authors and do not represent those of the Swedish Ministry of Finance.

player who has behaved selfishly (or generously) toward another. We investigate whether third parties are also willing to engage in costly rewarding, and whether third party responses are sensitive to the number of players affected by the selfish or generous action. Using the 'dictator game', where one player (the dictator) divides a sum of money between herself and a recipient, we allowed dictators to be selfish, fair, or generous. Unlike in other experiments, third parties then had the choice to either punish or reward the dictator. Across all variations, responses followed a consistent and intuitive pattern: selfish behavior was punished while generous behavior was rewarded. Not only were third parties willing to engage in costly rewarding, but rewards were in fact at least as common as punishments. Furthermore, third party response was more pronounced when the dictator transfer had the non-rivalrous character of a public good, in the sense that both (a) the number of recipients increased and (b) the dictator's transfer was multiplied by a constant factor, so that the larger number of recipients did not reduce potential payoff of each recipient. Third party response did not change significantly when either of these manipulations was performed alone, suggesting a particular sensitivity to situations involving a public good.

INTRODUCTION

Shared beliefs of what constitutes appropriate behavior greatly affect human decision making in many social domains, ranging from dress codes and marriage practices to personal conflicts and public policy. Norms appear to be a human universal, with specific norms differing across cultures (Brown, 1991; Hauser, 2006). Depending on the norm, the same action can be classified as either good or bad (e.g., (Ohtsuki and Iwasa, 2006)). It has been proposed that salient norms can be identified through examining the use of costly punishment by bystanders, or third parties, in experimental games (Fehr and Fischbacher, 2004). Not only will subjects punish those who directly harm them (second party punishment, see (Fehr and Gächter, 2000; Fehr and Gächter, 2002)), but they will also engage in costly punishment as third parties. Third party punishment, or "moralistic punishment" (Kurzban et al., 2007), involves situations where I punish you after you defect against somebody else. In the type of non-repeated anonymous interactions that have been explored, third parties can never benefit from choosing to reciprocate (i.e. punishing). Nonetheless, many people are willing to punish others at a cost to themselves in these scenarios (e.g., (Fehr and Fischbacher, 2004)). These third party punishment decisions therefore suggest a commitment to sanctioning

norm violators, and shed light on norms that subjects find important enough to enforce.

While costly punishment has received the lion's share of attention, costly rewarding also plays an important role in human prosociality. Allowing subjects to reward other group members in a public goods game (second party reward) is as effective as second party punishment in promoting cooperation in all but the final period of fixed length public goods games (Sefton et al., 2007; Sutter et al., 2006), and in indefinitely repeated public goods games (Rand et al., 2009a), while neither reward nor punishment promote cooperation in a single one-shot public goods game (Walker and Halloran, 2004). Rand et al. 2009a also find that when both reward and punishment are available in an indefinitely repeated public goods game, a group's probability to reward high contributors is positively correlated with contributions and payoff, while no such correlations exists for the probability to punish low contributors; although surprisingly, the opportunity for rewarding does not decrease the average frequency of punishment. In a two-player proposer game where the responder can either reward, punish, do both, or do neither, the average proposal is highest when both reward and punishment are possible, and is higher for reward only than for punishment only (Andreoni et al., 2003). Taken together, these studies clearly demonstrate that subjects have a taste for second party rewarding, and that reward can be an important force for promoting cooperation.

In addition to evidence regarding second party rewarding, numerous experiments demonstrate a willingness to engage in third party rewarding when subjects can form reputations. In the presence of reputation, evolutionary game theoretic models show that cooperation can spread through indirect reciprocity, where my actions toward you depend on your previous actions towards others (Nowak and Sigmund, 2005; Ohtsuki and Iwasa, 2006). Consistent with these theoretical models, experiments show that under various reputation systems, subjects will frequently reward others, and in particular will preferentially reward those with a past history of being cooperative (Milinski et al., 2001; Seinen and Schram, 2006; Semmann et al., 2005; Wedekind and Milinski, 2000). A very elegant experiment shows that this tendency for third party rewarding can be harnessed to avert the tragedy of the commons in a repeated public goods game (Milinski et al., 2002). After each round, each of the 6 public goods game group members is a donor for another randomly selected member to be a recipient, but direct reciprocity is excluded. The donor has full knowledge of the recipient's past behavior, in both the public goods game and the indirect reciprocity game, and then chooses

whether or not to incur a cost to confer a benefit on the recipient. This setting is a mix of second party rewarding for past actions toward the group (which includes the rewarder) and third party rewarding for past actions towards others in the indirect reciprocity game. Here rewarding is common and leads to stable high levels of contribution in the public goods game. In another experiment (Semmann et al., 2005) the donor is either a member of the recipient's or of another public goods group. Rewarding occurs at the same level in either treatment and induces the same level of contribution to the public good.

There is also a synergistic interaction between this rewarding setup and the opportunity to self-select into an institution with costly punishment (Rockenbach and Milinski, 2006). Groups of eight subjects each played 20 periods of the public goods game. In treatment 'PUNandIR", before each period , each player can choose between joining a group in which the public goods game is followed by both costly punishing and an indirect reciprocity game, and a group in which the public goods game is followed solely by an indirect reciprocity game. In treatment 'PUN', before each period, each player can choose between joining a group in which the public goods game is followed by costly punishing and a group in which the public good game is not combined with any other option. In both treatments the punishment groups achieve higher contributions than the no-punishment groups. Interestingly, the subjects in the PUNandIR treatment prefer the punishment opportunity group despite the presence of an alternative group offering only the reciprocity option, and the combination of indirect reciprocity and punishment results in the highest contributions and the highest efficiency. Additionally, the availability of indirect reciprocity changes how subjects chose to punish. In the indirect PUNandIR treatment, fewer punishments occur, but those that do are more focused on heavy free-riders. It is interesting to note two important differences between these results and those of the experiment with second party reward and punishment discussed above (Rand et al., 2009a): adding reward to punishment resulted in no increase in contributions or payoffs in the second party setup, and no decrease in punishment use; whereas the opposite is true in the Rockenbach and Milinski setup. The source of these differences merits further study, but may lie in the endogenous choice of punishment institution used here – in this setup, punishment works as a sorting tool in addition to its role in actually sanctioning low contributors.

In a related study, subjects have a choice each round between a standard public goods game and a setting with reward and punishment opportunities (Gurerk et al., 2006). Identities are shuffled from round to round, and reward

has a 1:1 technology while punishment has a 3:1 technology. Almost all subjects eventually switch to the reward+punishment institution, and achieve much higher contributions than the game without targeted interaction. The frequency of reward use decreases over time, however. Both the lack of persistent identities and the 1:1 reward technology may contribute to the instability of rewarding in this experimental setup.

As surveyed here, a sizeable amount of evidence exists for the importance of rewarding in human cooperation. Yet the reward-based analog to third party punishment, where I reward you in an anonymous one-shot interaction because you have cooperated with somebody else, remains largely unexplored (a notable exception is (Kahneman et al., 1986)). In such a design, where reputation formation is prohibited, it is never in one's self-interest to reward. Thus choosing to pay to administer a reward gives insight into salient norms in a similar fashion to third party punishment. These third party rewards offer a potential alternative mechanism for promoting norm compliance that avoids the destructive consequences of punishment. In this chapter, we ask whether third parties display a similar willingness to reward as they do to punish.

We also explore factors that effect third party intervention. It has been shown that third party sanctioning is more common in larger and more complex societies than in small-scale societies (Henrich et al., 2010; Marlowe and Berbesque, 2008). Meanwhile, larger and more complex societies give rise to many situations where one individual can either extract a benefit at a cost to many others or incur a cost in order to benefit many others. Yet, to the best of our knowledge, how third party response varies according to the number of individuals affected, and the potentially efficiency gains, has not been investigated. We investigate this issue in the current chapter. We are particularly interested in how third parties respond to money transfers (or lack thereof) with a public good character.

METHODS

In our setup we use a modified dictator game. In the standard version of the dictator game, the dictator is an individual with an endowment to be allocated between herself and a recipient. The recipient does not have an endowment, and has no say over what allocation the dictator chooses.

Numerous studies show that dictators frequently give a non-zero share to the other player (for an overview, see (Camerer, 2003)). The dictator game has been used to examine other-regarding preferences, such as altruism. As in

(Fehr and Fischbacher, 2004), we introduce a third party that can pay a cost to affect the payoff of the dictator. Across conditions, we let the dictator transfer none, half or all of the endowment, and we allow the third party to either punish or reward the dictator.

In our version of the game, the dictator is at all times a single individual endowed with $10. We vary the number of recipients, from 1 to 2 to 8 individuals. The dictator is presented with three options: give the entire $10 endowment to the recipient(s), give $5, or give $0. Third parties are endowed with $3 that can be spent to punish or reward the dictator, or can be kept by the third party. The response technology is 3:1, so that one dollar spent by the third party augments or reduces the dictator's payoff by three dollars.

We explore third party behavior in three settings. In Setting 1, the dictator's transfer is multiplied by *n*, the number of recipients, and then split equally between them. Each recipient thus gets the full dollar amount that the dictator chooses to transfer. In this setting, the money transferred by the dictator is like a public good for the recipients: each recipient experiences a benefit as a result of the dictator's action, and the size of this benefit is independent of the number of recipients.

In addition to asking whether third parties engage in rewarding, the main focus of our experiment is to explore how third party responses in Setting 1 change as the number of recipients varies. When the dictator gives something away, the multiplicative effect means that more money is "created" for the group as a whole, and the total group payoff is increased compared to when the dictator gives nothing.

In other words, increasing the number of recipients introduces the possibility of "waste", which is not present in the standard, zero-sum dictator game. However, in principle waste and the number of recipients are two separate effects. In order to examine these effects in isolation from each other, we run the experiment in two additional settings.

In Setting 2, a single recipient receives a multiple *z* of the amount transferred by the dictator. We let *z* take on the values 1, 2, and 8. For each dollar kept by the dictator, the other player foregoes *z* dollars. For the same values of *n*, in Setting 1, and *z*, in Setting 2, the same amount is foregone by the group when the dictator keeps some or all of the endowment. Although increasing *z* creates a larger potential total group payoff, it also creates the potential for greater payoff inequity. When the dictator transfers money to the recipient, he creates additional wealth through the multiplier but at the same time creates an outcome where he is receiving a smaller payoff than the recipient.

Previous work has shown that people care about both inequity and efficiency (Andreoni, 2007; Engelmann and Strobel, 2004; Fehr and Schmidt, 1999). This setting allows us to compare third party responses between treatments, holding the potential total group payoff constant.

In Setting 3, the number of recipients n takes on the values 1, 2, and 8, as in Setting 1. Here, however, the dictator transfer is not multiplied as in Setting 1, but is simply split equally between the recipients. Now, for one dollar transferred by the dictator, each of the n recipients gets $1/n$ dollars. This allows us to explore the possibility that third party responses are driven only by the number of individuals affected, regardless of the effect on total group payoff.

A total of 275 subjects from the Boston area participated voluntarily in this modified dictator game with a third party. The study was approved by the Harvard institutional review board. Written consent was obtained from all subjects before participating in the study. The subjects ranged in age from 18 to 70, were both students and non-students, and did not know the identity of any of the subjects that they were matched with. Subjects were not allowed to participate in more than one session of the experiment. All sessions were run in May 2007.

Each subject was paid a show up fee of $5. All subjects were informed about the extensive form of the game, the endowments of each player, and that they would all receive a show-up fee. We used neutral language such as “add” or “subtract” money instead of “reward” or “punish”. All subjects interacted anonymously, participated in a single treatment, and were only given one role (dictator, recipient, or third party) within that treatment. They were not informed about the other treatments. We elicited third party responses using the strategy method, as was done in previous work on third party punishment (Fehr and Fischbacher, 2004): Third parties were asked to specify their response to each option available to the dictator.

Once the third parties had indicated their preferences, the dictator's actual choice was revealed. The advantage of this method is that it provides information about responses to outcomes that may occur very infrequently. Previous work has shown that using the strategy method may have a quantitative but not qualitative effect on decisions (e.g., (Falk et al., 2005)).

We expect this effect on decisions to be orthogonal to the variation in conditions, allowing us to compare how third party behavior differs between treatments. In addition, the ratio of third parties to dictators was approximately 10:1. For each dictator, only one randomly selected third party's response was actually carried out. Third parties that were not matched with a dictator kept

their $3 endowment, regardless of the choices they made. Third parties were fully informed of this design feature.

Throughout the data analysis we use a standard ordinary least squares estimator (OLS) with robust standard errors clustered by subject.

RESULTS

Both third party rewarding and third party punishment occurred frequently in our experiment. As shown in Figure 1, the majority of participants either rewarded, punished, or both.

Figure 1. A majority of third parties spent money in order to reward or punish the dictator (65%). These individuals either only punished (9%), only rewarded (20%), or punished and rewarded (36%). About one third of third parties did neither (35%).

Of 184 total third parties, 66 subjects both rewarded and punished at least once (36%), 37 subjects rewarded at least once but never punished (20%), 17 subjects punished at least once but never rewarded (9%), and 64 subjects never did either (35%). Of all third party responses (3 responses per third party), 27% were rewards, 18% were punishments, and 55% were non-responses. Thus, not only did costly rewards occur, but averaged over all decisions, rewards were in fact more frequent than costly punishments.[1]

[1] Our study focused on third party responses rather than dictator actions. We did not inform the dictators about the costly nature of third party actions, and thus cannot draw any conclusions from dictator behavior. However, 47% of the dictators chose to give away

Table 1. Third party response as a function of dictator transfer

		Setting 1		*Setting 2*		*Setting 3*	
	Baseline	*n =2*	*n = 8*	*z =2*	*z = 8*	*n =2*	*n = 8*
Slope coefficient	0.13	0.20	0.25	0.20	0.18	0.16	0.12
P-value	0.004	<0.001	<0.001	<0.001	0.001	<0.001	0.004

Third party responses followed the same intuitive pattern in all treatments: as the amount transferred by the dictator increased, the average third party response also increased. A highly significant positive correlation between dictator transfer (0 to 10) and third party response (-3 to +3) exists in each treatment (Table 1).

Selfish behavior (transferring nothing) was punished on average, and completely generous behavior (transferring everything) was rewarded on average. Transferring half the endowment was on average punished in some treatments and rewarded in others. However, transferring half was never punished more than transferring nothing, and never rewarded more than transferring everything.

Third party responses are increasing in the amount transferred by the dictator across all treatments, but this relationship may differ between treatments. To investigate this question, we perform an additional regression which includes a dummy variable and an interaction term for each treatment. Each treatment's dummy variable takes on the value 1 for data in that treatment, and 0 for all other data. This allows the regression intercept to differ between treatments. Each interaction term is the product of a treatment dummy and the explanatory variable (dictator transfer). It takes on the same value as the explanatory variable in one treatment and 0 in all other treatments. This allows the slope coefficient to differ between treatments. The *p*-values for the dummies and interaction terms give a direct measure of whether differences in the regression between treatments are statistically significant. We are also able to distinguish between an overall level effect and a slope effect when comparing third party responses in the different treatments: a higher intercept indicates that third party responses are higher for all levels of dictator transfer,

everything and 43% chose to keep half. Due to the small number of dictators (3 in each treatment), we cannot perform a substantive analysis of differences between treatments.

whereas a larger slope coefficient indicates that third parties responds more strongly to increases in the dictator transfer.

In Setting 1, each recipient received the full amount transferred by the dictator, regardless of the number of recipients. In this public goods setting, we found that third parties were sensitive to *n*, the number of individuals affected by the dictator's action (Figure 2).

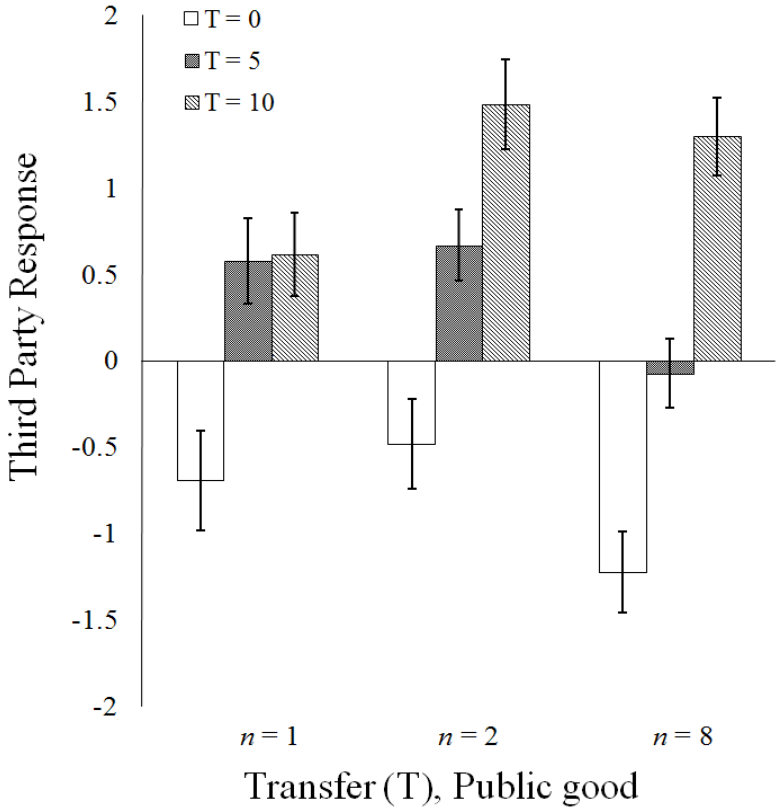

Figure 2. In Setting 1, each recipient got the full amount transferred by the dictator, regardless of the number of recipients. Thus, the dictator's transfer here is a public good. We found a significant increase in the extremeness of third party response when the number of recipients increased from one to eight, but not from one to two. Full generosity was rewarded more, and full selfishness was punished more. This suggests third party responses are at least moderately sensitive to the number of individuals affected when benefits are like a public good to the recipients. Error bars indicate standard error of the mean.

Table 2. Third party response as a function of dictator transfer, with treatment dummies and interaction terms

Treatment		Slope	*P*-value	Intercept	*P*-value	Observations
Baseline setup						
n = 1	*z = 1*	0.13	0.001	-0.49	0.065	*78*
Treatment interactions and treatment dummies						
Setting 1						
n = 2	*z = 2*	0.07	0.253	0.06	0.868	*81*
n = 8	*z = 8*	0.12	0.030	-0.77	0.027	*81*
Setting 2						
n = 1	*z = 2*	0.07	0.24	-0.35	0.364	78
n = 1	*z = 8*	0.05	0.41	-0.01	0.987	78
Setting 3						
n = 2	*z = 1*	0.03	0.60	-0.14	0.676	81
n = 8	*z = 1*	-0.01	0.86	-0.02	0.953	78

Third parties rewarded generosity and punished selfishness significantly more when a dictator had 8 receivers than when a dictator only had 1receiver (OLS regression with dummies and interactions; *p*-value 0.030, indicating a significantly steeper slope for the correlation between dictator transfer and third party response; see Table 2). However, the size of this effect is fairly moderate compared with the increase in possible group payoff and waste. Dictators transferring all $10 were rewarded on average 2.1 times more when $n = 8$ as compared to $n = 1$; and dictators keeping all $10 were punished on average 1.8 times more when $n = 8$ as compared to $n = 1$.

In Setting 1, we also observed a change in the perception of a transfer of $5, which resulted in the dictator and each recipient all receiving $5. When $n = 1$ or $n = 2$, keeping half was on average rewarded (*T*-test, two-tailed; $n = 1$ response not equal to 0, *p*-value 0.0291; $n = 2$ response not equal to 0, *p*-value0.0034). But when $n = 8$, this egalitarian allocation was no longer considered generous, and drew a neutral average response (*T*-test, two-tailed; $n=8$ response not equal to 0, *p*-value 0.71).

This suggests that a considerable fraction of third parties are sensitive to the number of individuals affected and potential group payoff. However, we

note that there is no statistically significant difference between *n*=1 and *n*=2 when considering responses to all three possible dictator transfers (OLS regression with dummies and interactions; *p*-value 0.253; see Table 2). Thus it seems the increase in *n* must be sufficiently large to alter the third party response.

In Setting 2, a single recipient received a multiple *z* of the amount transferred by the dictator. For equal values of *z* in this setting, and *n* in the previous setting, the same total amount of money was foregone by the group for each dollar kept by the dictator. This mimics the possibility of waste in Setting 1, without increasing the number of recipients (fixed at1). If third parties were only concerned about the amount of money foregone by others as a result of the dictator's action, we should observe the same change in responses as in Setting 1. However, this was not the case (see Figure 3).

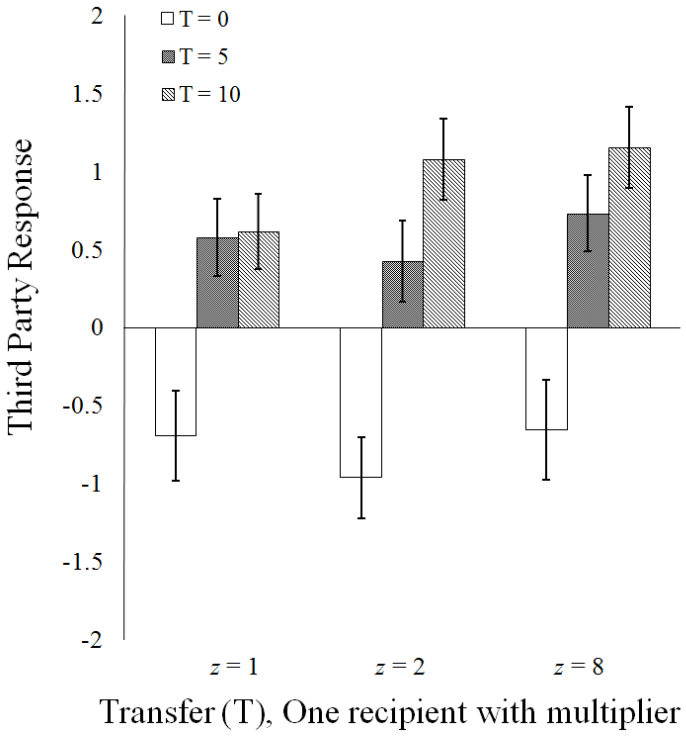

Figure 3. In Setting 2, a single recipient received a multiple z of the amount transferred by the dictator. The amount foregone by other players, as a result of the dictator'sdecision, was the same in settings 1 and 2 for the same values of n and z. Yet when z increased, we found no significant change in third party responses. Efficiency

concerns alone are clearly not driving third party responses. Error bars indicate standard error of the mean.

In fact, we found no significant change in third party responses as z increased (OLS regression with dummies and interactions; z = 2: p-value 0.240, z = 8: p-value 0.411; see Table 2). Yet in Setting 1, where the potential total group payoff was the same as it is here for equal values of n and z, there was a significant change in response going from 1 to 8.

Based on observed third party responses in Setting 2 we conclude that concerns about total group payoff alone do not seem to be driving third party responses. At this point, however, we have not ruled out the possibility that the results in Setting 1 were simply driven by a concern for the number of people affected, regardless of the extent to which they were affected. In our third and final setting, we show that this was also not the case.

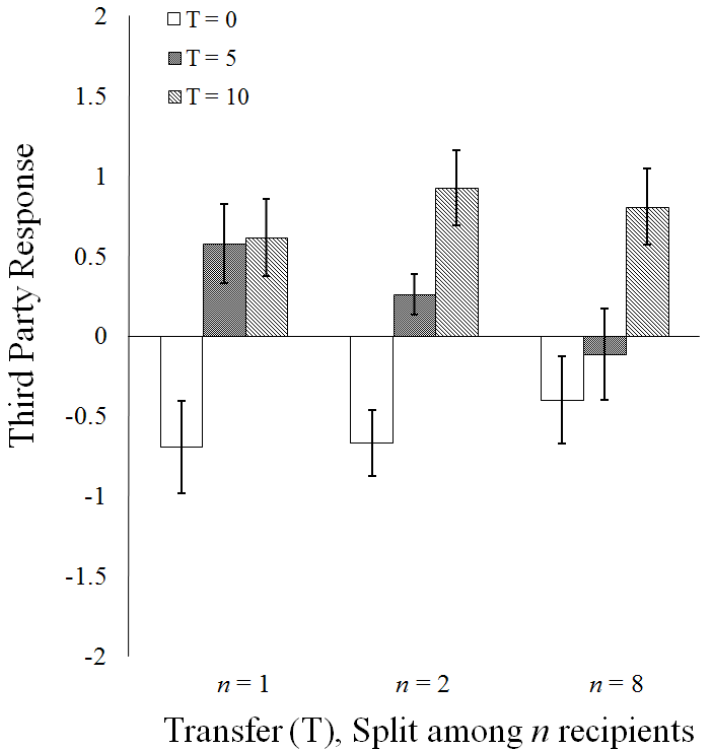

Figure 4. In the third setting, the dictator's transfer T was simply shared by the recipients. For each dollar transferred by the dictator, each one of n recipients got 1/n dollars. In this setting, increasing the number of recipients had no significant effect on

third party responses. Even the response to T = 5 was not significantly different when *n* = 1 versus when *n* = 8 (Mann-whitney test, p-value 0.11). Error bars indicate standard error of the mean.

In Setting 3, we once again varied *n*, the number of individuals affected by the dictator's action. However, unlike in the first setting, here the transfer from the dictator was split equally by the *n* recipients. One more dollar for the dictator now meant 1/*n* dollars less for each recipient. We find no significant change across *n* values in third party responses (Figure 4). The overall difference in third party responses when we varied *n* in this setting was not statistically significant (OLS regression with dummies and interactions; *n* = 2: *p*-value 0.60, *n* = 8: *p*-value 0.86; see Table 2).

CONCLUSION

Many actions, especially in larger and more complex societies, can potentially influence the welfare of more than one other person. In economics, a benefit that is nonrivalrous (and non-excludable) is labeled a public good. Here we examine how third party response varies with the number of individuals affected by a dictator's action when the transfer has the nonrivalrous character of a public good. This gives us insight into how social norms, and their enforcement, may contribute to the provision and regulation of such goods.

We find that not only do third parties use costly punishment when given the opportunity, as has been found previously, but third parties also use costly rewarding. In fact costly rewards are at least as common as costly punishments. This suggests that people are inclined to reward behavior that they deem good or generous at a cost to themselves. The preference for rewarding over punishing might in part be explained by a fear of retaliation. In most real life scenarios, both in early human societies (Marlowe and Berbesque, 2008) and today, punishers are not anonymous and so bear the risk of retaliation. Retaliation has been found to be common among animals (Clutton-Brock and Parker, 1995) and among humans in studies of cooperation games that allow for counter-punishment (Cinyabuguma et al., 2006; Denant-Boemont et al., 2007; Dreber et al., 2008; Nikiforakis, 2008; Wu et al., 2009), and costly punishment is disfavored by natural selection in repeated games (Rand et al., 2009b). Thus, if individuals have limited resources to put toward enforcing cooperation it seems plausible that they may be better served by

rewarding rather than punishing. In line with this, in a public goods game followed by two rounds of targeted interactions, individuals who reward end up better off than those who punish (Kiyonari and Barclay, 2008). Consistent with second party punishment studies run in the West, we find very little third party anti-social punishment (Gächter and Herrmann, 2009, In press; Herrmann et al., 2008; Rand et al., 2010) with only 2 out of 184 third parties choosing to punish dictators who transferred the entire endowment.

When it comes to the number of individuals affected by the dictator in our experiment, previous work has produced related results, but in somewhat different contexts. One study compares groups of size 4, 10, 40 and 100 and shows that group size under some circumstances correlates positively with the group's ability to provide the optimal level of a public good (Isaac et al., 1994). Their results also show that cooperative behavior is influenced by a subtle interaction between group size and the marginal return of an individual's contribution that cannot be explained by either of these two things alone. The marginal return resembles our *z*, the multiplier of the dictator transfer. Another experiment on altruism and group size (Andreoni, 2007) finds that altruism in a modified dictator game is partly a *congestible* good, in the sense that people care both about the total benefit and the average benefit resulting from a monetary gift, with slightly more emphasis put on the former.

Our results, in combination with these previous findings, lead us to believe that the regulation of public goods plays a role in promoting cooperation that is, so to speak, greater than the sum of its parts. In our study, third party response is more accentuated when the number of recipients sharing a public good increases from 1 to 8. This increase contains two components: an increase in the number of individuals that are affected by the outcome, and an increase in the size of the aggregate transfer. In parallel treatments we show that third party response does not react to changes in one of these two components alone.

The experimental setup used in our study and in other third party studies is ultimately a non-repeated, reputation-free version of an indirect reciprocity game. The fundamental assumption of indirect reciprocity is that individuals have reputations which persist across interactions, such that my potentially cooperative behavior toward you depends on what you have done to others (Nowak and Sigmund, 2005; Ohtsuki and Iwasa, 2006). In such situations, it is often in the individual's self-interest to cooperate and reciprocate. Thus, it could be that the propensity for costly intervention by third parties in non-repeated anonymous interactions is a misapplication of moral tendencies that evolved through indirect reciprocity. This possibility merits further study.

Whether third party norm enforcement has played an evolutionary relevant role in the provision of public goods remains an open question. As has been suggested previously, third party norm enforcement may create a group-level incentive for creating a public good (Fehr and Fischbacher, 2004). If this is the case, holding the action itself constant, helping or hurting more people should thus provoke more extreme third party responses. We find some evidence of this, but the effects are small. However, it has been suggested (Fehr and Fischbacher, 2004) that more than one third party may be needed to enforce a norm, and that this condition is probably met frequently in real life. Whether this conjecture holds true still remains to be explored.

ACKNOWLEDGEMENTS

We are grateful for comments from Martin A. Nowak, Magnus Johannesson, Thomas Pfeiffer, David Cesarini, Drew Fudenberg, Manfred Milinski, Hisashi Ohtsuki, Jorge M. Pacheco, Bettina Rockenbach and Arne Traulsen. Johan Almenberg and Anna Dreber thank the Jan Wallander and Tom Hedelius Foundation, Johan Almenberg thanks the Torsten and Ragnar Söderberg Foundations, for financial support, and David Rand is supported by a grant from the John Templeton Foundation.

REFERENCES

Andreoni, J. (2007). Giving Gifts to Groups: How Altruism Depends on the Number of Recipients". *Journal of Public Economics, 91*, 1731-1749.

Andreoni, J., Harbaugh, W. T., and Vesterlund, L. (2003). The Carrot or the Stick: Rewards, Punishments and Cooperation. *American Economic Review, 93*, 893-902.

Brown, D. E. (1991). *Human Universals*. New York: McGraw-Hill.

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

Cinyabuguma, M., Page, T., and Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics, 9*, 265-279.

Clutton-Brock, T. H., and Parker, G. A. (1995). Punishment in animal societies. *Nature, 373*, 209-216.

Denant-Boemont, L., Masclet, D., and Noussair, C. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory, 33*, 145-167.

Dreber, A., Rand, D. G., Fudenberg, D., and Nowak, M. A. (2008). Winners don't punish. *Nature, 452*, 348-351.

Engelmann, D., and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review, 94*, 857-869.

Falk, A., Fehr, E., and Fischbacher, U. (2005). Driving Forces Behind Informal Sanctions. *Econometrica, 73*, 2017-2030.

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

Fehr, E., and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*, 980-994.

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

Fehr, E., and Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics, 114*, 817-868.

Gächter, S., and Herrmann, B. (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 791-806.

Gächter, S., and Herrmann, B. (In press). The Limits of Self-Governance when Cooperators Get Punished: Experimental Evidence from Urban and Rural Russia. *European Economic Review*.

Gurerk, O., Irlenbusch, B., and Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science, 312*, 108-111.

Hauser, M. (2006). *Moral Minds*. New York: HarperCollins.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., et al. (2010). Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. *Science, 327*, 1480-1484.

Herrmann, B., Thoni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science, 319*, 1362-1367.

Isaac, R. M., Walker, J. M., and Williams, A. W. (1994). Group size and the voluntary provision of public goods : Experimental evidence utilizing large groups. *Journal of Public Economics, 54*, 1-36.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the Assumptions of Economics. *The Journal of Business, 59*, S285-S300.

Kiyonari, T., and Barclay, P. (2008). Free-riding may be thwarted by second-order rewards rather than punishments. *Journal of Personality and Social Psychology, in press*.

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

Marlowe, F. W., and Berbesque, J. C. (2008). More 'altruistic' punishment in larger societies. *Proc Biol Sci, 275*, 587-590.

Milinski, M., Semmann, D., Bakker, T. C. M., and Krambeck, H.-J. r. (2001). Cooperation through indirect reciprocity: image scoring or standing strategy? *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*, 2495-2501.

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

Nikiforakis, N. (2008). Punishment and Counter-punishment in Public Goods Games: Can we still govern ourselves? *Journal of Public Economics, 92*, 91-112.

Nowak, M. A., and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*, 1291-1298.

Ohtsuki, H., and Iwasa, Y. (2006). The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J Theor Biol, 239*, 435-444.

Rand, D. G., Armao IV, J. J., Nakamaru, M., and Ohtsuki, H. (2010). Anti-social punishment can prevent the co-evolution of punishment and cooperation. *Journal of theoretical biology, 265*, 624-632.

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., and Nowak, M. A. (2009a). Positive Interactions Promote Public Cooperation. *Science, 325*, 1272-1275.

Rand, D. G., Ohtsuki, H., and Nowak, M. A. (2009b). Direct reciprocity with costly punishment: Generous tit-for-tat prevails *J Theor Biol, 256*, 45-57.

Rockenbach, B., and Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature, 444*, 718-723.

Sefton, M., Schupp, R., and Walker, J. M. (2007). The Effect of Rewards and Sanctions in Provision of Public Goods. *Economic Inquiry, 45*, 671 - 690.

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

Semmann, D., Krambeck, H.-J., and Milinski, M. (2005). Reputation is valuable within and outside one’s own social group. *Behavioral Ecology and Sociobiology, 57*, 611-616.

Sutter, M., Haigner, S., and Kocher, M. G. (2006). Choosing the Stick or the Carrot? Endogenous Institutional Choice in Social Dilemma Situations. *CEPR Discussion Paper No. 5497*.

Walker, J. M., and Halloran, M. (2004). Rewards and Sanctions and the Provision of Public Goods in One-Shot Settings. *Experimental Economics, 7*, 235-247.

Wedekind, C., and Milinski, M. (2000). Cooperation Through Image Scoring in Humans. *Science, 288*, 850-852.

Wu, J.-J., Zhang, B.-Y., Zhou, Z.-X., He, Q.-Q., Zheng, X.-D., Cressman, R., et al. (2009). Costly punishment does not always increase cooperation. *Proceedings of the National Academy of Sciences, 106*, 17448-17451.