Third-Party Punishment of Groups

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts

in the Graduate School of The Ohio State University

By

Clare Marie Sorenson, B.A., B.S.

Graduate Program in Psychology

The Ohio State University

2010

Dissertation Committee:

Michael L. DeKay, Advisor

Hal Arkes

Michael Edwards

Abstract

These two studies were designed to test the dynamics associated with the punishment of an entire group when only a subset of the group violates a social norm. This punishment usually provides no direct benefits to the punisher, classifying it as third-party punishment. In Experiment 1, participants read a hypothetical scenario about a classroom in which one student took more cookies than allowed. The number of cookies was varied as well as the punishment target (individual, a group with knowledge of the offender, and a group without knowledge of the offender). Basic results showed high levels of punishment for most conditions. Participants did punish individuals and groups with knowledge more than groups without knowledge. They also punished them more severely.

In Experiment 2, participants were allowed to punish hypothetical players of a public goods game. Target (individual versus group) was again manipulated as well as magnitude (4 levels). I found higher frequency of punishment and higher severity for individuals versus groups.

These results seem to indicate that while group punishment might occur at a lower frequency than individual punishment and be lower in severity, it does still occur. I observed much higher rates of punishment in Experiment 1 versus Experiment 2 that may be accounted for by entitativity (how well one member of

a group represents the group as a whole) and the social norm being violated (stealing cookies is more clearly a violation than not contributing in a public goods game). Further research will seek to investigate the effects of entitativity and various social norms.

Dedication

Dedicated to my family: my dad, my mom, my sister, and my brother; my best friend and encouragement, Michelle; and Dr. Valerie Staton, who encouraged me to go to grad school to begin with.

Acknowledgments

I would like to acknowledge the help of my advisor, Dr. Michael DeKay, and the help and support of my colleague Seth Miller. I would also like to acknowledge the undergraduates who assisted me in the data collection and for help from several faculty members (such as Dr. Michael Edwards) and other graduate students with my analyses.

Vita

August 1989 ………………………. Our Lady of Perpetual Help School

August 1998 ………………………. Bishop Ready High School

August 2002 ………………………. B.A. Psychology, B.S. Mathematics, Ohio Dominican University

August 2007 to present……………. Graduate Teaching associate, Department of Psychology, The Ohio State University

Field of Study

Major Field: Quantitative Psychology

Table of Contents

List of Tables

List of Figures

Literature Review

Cooperation has come to be a hallmark of the distinction between human beings and most other animals (Shinada, Yamagishi, & Ohmura, 2004; Shinada & Yamagishi, 2007). Cooperation is economically and evolutionarily justified if it involves one's genetic relatives, reciprocity, or reputation building (Gintis, 2000; Hamilton, 1964; Trivers, 1971). Closely associated with cooperation is punishment of those who "free-ride" on the benefits of the cooperation of others without contributing themselves (Fehr & Gachter, 2002; Gintis, 2000; Gintis, Bowles, Boyd, & Fehr, 2003; Kiyonari, Tanida, & Yamagishi, 2000; Kurzban, DeScioli, & O'Brien, 2007; Seymour, Singer, & Dolan, 2007; Shinada et al., 2004; Turillo, Folger, Lavelle, Umphress, & Gee, 2002).

Punishment is also economically rational for the same reasons as cooperation. However, there are situations within human society where the cooperator or the punisher does not seem to gain any direct benefits from the act of punishing (Fehr & Gachter, 2002; Gintis, 2000; Gintis et al., 2003; Kiyonari et al., 2000; Kurzban et al., 2007; Seymour et al., 2007; Shinada et al., 2004; Turillo et al., 2002). Direct benefits might include monetary or material compensation from the violator. This is usually the case in second-party punishment, where one party who is the 'victim' of a norm violation directly punishes the offending party. Third-party punishment (that carried out by a party who is not directly harmed by the action or inaction of the free-rider) is a phenomenon that has important repercussions in society. The criminal justice system and democratic governments rely to a great extent on third-party punishment (Gintis et al., 2003). For this reason, the examination of third-party punishment and its evolution are critical to the understanding of the organization of human society.

Theories of Cooperation

Punishment of any kind, but specifically third-party punishment, is considered second-order cooperation (Fehr & Gachter, 2002; Shinada et al., 2004). Third-party punishment (TPP) exists only if there is cooperation, as it is not sensible to punish if there is no cooperation since there would be no act that would merit punishment if there was no cooperation norm (Shinada et al., 2004). Many instances of cooperation, and hence punishment, can be explained through three reasons: kin selection, direct reciprocity, and reputation building (Gintis, 2000; Hamilton, 1964; Trivers, 1971). Kin selection entails cooperating with others who are genetically related to oneself. In helping them, one ultimately helps to ensure the survival of one's genes. Direct reciprocity involves a mentality of "if I scratch your back, you scratch mine." Cooperating has mutually beneficial results, but this mutual benefit might not occur all at one time, but over a period of time (help now for the promise of help later).

Finally, reputation building involves an audience. There must be others outside of the direct cooperation who can observe the "good act" of the cooperator. This might ensure the cooperator future benefits, including material or social goods. Another important element is the repeatability of the act. One's current cooperation increases the likelihood of future cooperation with the same person or people. All of these scenarios are economically defensible; there is a return for the effort or material invested. However, many people exhibit cooperation in anonymous, non-repeated situations (Fehr & Gachter, 2002; Gintis, 2000; Kiyonari et al., 2000; O'Gorman, Wilson, & Miller, 2005; Shinada et al., 2004).

Gintis et al. (2003) put forward the theory of strong reciprocity to explain the existence of cooperation and punishment in anonymous, non-repeated, non-kin situations. Strong reciprocity is defined as the inclination to cooperate and to punish free-riders even at a cost to oneself without any obvious benefits. This theory does not deny the existence of kin selection, reputation building, and direct reciprocity – all of these surely do exist – but seeks to explain this one seemingly odd phenomenon (Gintis et al., 2003). In order to illustrate the existence of strong reciprocity, Gintis et al. used the economic ultimatum game which involves two players. One, the giver, is given a certain amount of money of which he can offer a part to the other player. The other player, the receiver, has two options: to accept the offer, receiving their portion while the first player receives the residual, or to reject the offer, in which case both players receive nothing. A player who abides strictly by economic rules should offer only a very small amount (if they are the giver) or should accept any amount offered, as it is greater than zero (if they are the receiver). In spite of this, the most common offer is 50% of the money and most offers less than 25% are rejected by the receiver (Gintis et al., 2003).

Over repeated experiments performed by different researchers, it has been found that on average, only one quarter of the participants play in the economically rational way (Fehr & Gachter, 2002). This scenario appears to involve a mixture of cooperation and second-party punishment. The first player seeks to elicit the cooperation of the second player by offering a nontrivial portion of the total money. If the offer is not adequate, the second player is willing to sacrifice a payoff greater than zero in order to punish the perceived selfishness of the other player. The fact that the first player often offers half the total seems to indicate that they expect the second player to reject a lower (and yet "economically fair") offer.

The theory of strong reciprocity seems to describe the existence of TPP, as it is a by-product of our desire to cooperate. Part of cooperating might include punishing those who do not cooperate, even if one is not directly affected by the lack of cooperation. However, the theory does not really address why humans have an inclination to cooperate, and hence, punish (especially as third parties). Turillo et al. (2002) offer the theory of the existence of "deontic emotions," those of duty and obligation. Cooperation and TPP are performed because it is considered moral to do so, which requires a commitment to ethical standards without self-interest. Cooperation or punishment may involve sacrifice, but most importantly, they must exclude the possibility of material or interpersonal gain.

Turillo et al. (2002) designed a series of experiments to eliminate all possible reasons for punishing except for deontic emotions. Through the use of strict anonymity (both to other group members and to the recipient of the punishment), reputation building or message sending was eliminated. In spite of this, almost three-quarters (73%) of participants chose the option that punished the unfair player who divided a $20 payout unequally ($18 for self and $2 for the other player) at a cost of $1 to themselves. However, the ability to reward fair behavior and punish unfair behavior were confounded. The researchers then separated these factors, finding that punishment fell to an average of 10% and rewarding fairness rose to an average of 83% with regard to frequency of the chosen option (punishment or reward). This would seem to indicate that people prefer to reward the fair player rather than punish the unfair player. However, if subjects are given the option to punish while still giving some reward to a fair player, they will do so (Turillo et al., 2002).

The theory of deontic emotions is interesting and should spark further research, but it is challenged by a theory proposed by Kiyonari et al. (2000). The theory of the social exchange heuristic (SEH) explains cooperation in terms of a cognitive change. Self-interest is abandoned in favor of mutual benefit. For example, in the prisoner's dilemma, defection is the economically rational choice (referred to in economic game theory as the dominant action). However, if the SEH is activated, the PD turns into an assurance game where there is no dominant choice. In Kiyonari et al.'s experiment, if both players defect, then both benefit (each get $2); if both players cooperate, both benefit (each get $1), but if they act in opposite ways, then one will benefit and the other will not (the defector receives $3 and the cooperator receives $0). Here, it is clear that defecting is always the dominant choice, as the player will either receive $2 (if the other player also defects) or $3 (if the other player cooperates) rather than receiving either $1 (if the other player cooperates) or $0 (if the other player defects).

This is a strong test of cooperation.

Kiyonari et al. (2000) found that when there is a realistic sense of exchange, which triggers the SEH, cooperation increased, at a cost to the individual players. Realistic sense of exchange is defined as when 'participants' decisions produce significant outcomes for themselves and their partners' (p 417). Cooperation also increased when one player was aware that the other player would cooperate, but only when using real money rather than points. The use of money may be crucial to the realistic sense of exchange. In all of the cases, cooperation is seen as irrational according to strict economic theory because of the structure of the incentives. Therefore, if subjects are being irrational, there must be some advantage to cooperation apart from the strict economic results. Kiyonari et al. have a theory about what motivates the SEH, through either cognitive means (mental transformation of the incentive structure) or subjective means (inequity aversion). To date, these have not been tested.

Punishment

Up to this point, the theories explained have focused mostly on cooperation, with punishment included as a second-order form of cooperation. Punishment has been studied as its own phenomenon, which will now be addressed. The study of punishment often involves the public goods game. It is usually played by four people, although it is easily expanded. Each player is given $X and are given the opportunity to invest in a community fund. All money that is contributed to this fund will be increased by some fixed amount, a percentage of the contribution.

For example, every player is given $10. If every player contributes their entire endowment, then there will be $40 in the community fund. Now, everybody will receive some percentage of that, say 50%. In this case, each person will get $20. There is a benefit to cooperating. But what happens when one member does not cooperate? Imagine that three players contribute their full $10 and one player contributes nothing. There is $30 in the fund and each player receives 50%, which is $15. The player that did not contribute anything to the public good still has their $10 and receives the $15 from the fund, so they now have $25. This person is the free-rider. The remaining three players only receive $15 from the public fund. These two examples are the extreme cases that can occur; most players usually contribute some percentage of their endowment.

Fehr and Gachter (2002) allowed players to punish those who were free-riders at a cost of 1:3; for every one money unit spent to punish, the receiver loses 3 money units. The public goods game was played ten times. For each round, every participant was allowed to punish. An astounding 84.3% punished at least once, with more than a third punishing more than five times. This was amazing because the punisher did not receive direct benefit from punishing and actually reduced their absolute winnings by punishing. Most of the punishment was done by cooperators to free-riders. When the free-rider contributed 14-20 money units less than the average contribution, the group spent almost 10 money units to punish. The free-riding cost the free rider 30 money units, which when combined with their original 14-20 money units they kept for themselves, means that free-riding gave them a net loss. Punishment increased cooperation over time: by the last round, 38.9% contributed all their endowment and 77.8% invested more than three-quarters of their endowment.

A follow-up questionnaire assessed the level of negative emotions in two situations: one in which there were three high contributors and one free rider and one in which there were three medium contributors and one free rider. There was significantly more anger in the first case, but the second case still elicited higher than expected amounts of anger. The questionnaire also addressed expectations of anger. The subjects actually expected more anger than was actually felt (Fehr & Gachter, 2002).

Emotional Bases for Punishment

Anger and guilt are proposed to be driving forces behind punishment (Hopfensitz & Reuben, 2009; Price, Cosmides, & Tooby, 2002). Price et al. (2002) proposed that punitive sentiment drives punishment of free-riders in order to increase cooperation. They defined punitive sentiment as a desire to harm a certain person(s). Punitive sentiment could induce punishment for two reasons: to increase cooperation so that the public good is funded or to decrease the advantage of the free-rider. Using a set of separate predictions for each of the two reasons, Price et al. found that punitive sentiment was positively correlated with the risk of being at a fitness disadvantage by the free-riding behavior of others. Willingness to participate in cooperation was significantly correlated with punishment of free-riders, while self interest in the group goal was significantly correlated with rewarding cooperators. This led the researchers to conclude that punitive sentiment evolved specifically to eliminate differences in fitness, not to increase cooperation directly (Price et al., 2002).

Anger, which could be defined as a punitive sentiment, has been positively associated with an increase in the frequency and the amount of punishment (Hopfensitz & Reuben, 2009). Although there is strong evidence for the existence of punishment as a way to decrease fitness inequalities, Carpenter, Matthews, and Ong'ong'a (2004) found that subjects who are motivated by social norms rather than by payoff differences punish more. Fifty-six percent of the sample punished solely for normative reasons, while 86% was at least partially motivated by normative reasons. This would support the theory of reciprocity rather than that of fitness differences. Further research is needed to examine the relative influences of reciprocity and fitness differences.

Guilt and shame are emotions that are proposed to be felt by the free-rider who is punished (Hopfensitz & Reuben, 2009). Often, in the real world those who are punished have the ability to retaliate, yet most do not. For example, if a person is fired, they seldom take any action to harm their former employer. This lack of retaliation may be due to feelings of guilt and shame that have been associated with reductions in retaliation (Hopfensitz & Reuben, 2009). In the context of a social dilemma game which allowed cooperation, free-riding, punishment, and retaliation, it was shown that punishment of a free-rider cost .149 points to the punisher for every point reduction for the free-rider, but it cost the retaliating free-rider .763 points per point to retaliate. Retaliation was more costly than punishment in this situation. Still, it could be viewed as economically rational, as it reduces the payoff of the person retaliated against by more than it costs the retaliator, thus reducing relative payoff differences.

The lack of retaliation was shown to involve an interaction between guilt and anger (Hopfensitz & Reuben, 2009). At high levels of guilt, retaliation did not change as a function of anger. At low levels of guilt, retaliation was higher for those who experienced high levels of anger versus those who experienced low levels of anger. Overall, punishment did increase levels of cooperation, although it did not lead to more earnings versus a baseline condition. Somewhat surprising, perhaps, was the finding that retaliation did not halt punishment: more than half (56.6%) of those who experienced retaliation went on to punish this retaliation (Hopfensitz & Reuben, 2009). This is a little surprising because one of the reasons suggested for retaliation is to reduce future punishment.

The increase in cooperation when punishment contingencies exist can be ascribed to two different effects (Shinada & Yamagishi, 2007). First, punishment decreases the payoff to the free-rider, leading to a net loss for free-riding. Second, knowing that punishment contingencies are in effect reassures conditional cooperators that others will cooperate. Conditional cooperators are those who cooperate only when others cooperate. This second effect is considered indirect, but Shinada and Yamagishi (2007) found that it was of similar magnitude to the direct effect. Allowing for both the indirect and the direct effects increased payoffs to cooperators from 18.89 yen with just the direct effect to 108.15 yen, after including the cost of punishment (Shinada & Yamagishi, 2007). This indirect effect is a powerful, yet lesser investigated explanation for the existence and success of punishment in many, though not all, situations.

Third-Party Punishment

The studies above examined second-party punishment, which can be economically rational- it costs the players to punish, but punishment can increase cooperation, resulting in net benefits to the group generally and to the punishers specifically. Third-party punishment (TPP) is less clear-cut. TPP is crucial because not all norm violations directly target a person (Fehr & Fischbacher, 2004). For example, if Joe speeds while driving, who is the direct target of his norm violation? Certainly, if he caused an accident, there would be a specific victim, but what about all the times he speeds and does not cause an accident? The behavior is undesirable, and so needs to be corrected. Without a specific target, second-party punishment will not work, as there is no direct victim to inflict punishment. This necessitates the need for punishment by third parties. Fehr and Fischbacher (2004) give the example of one free-rider in a large cooperating group; the size of the cost of the free-rider is very small and does not negatively affect one person to any significant degree. However, punishment is still necessary to correct that person and to deter others from engaging in free-riding behavior.

People do exhibit TPP in cases when there appears to be no obvious gain to themselves (Fehr & Fischbacher, 2004). When given an opportunity to punish "unfair" behavior, participants have been observed to punish the norm violator (about 60% punish). In addition, norm violators expected more punishment than was given (Fehr & Fischbacher, 2004), which raises the issue of why they still choose to deviate. Perhaps the free-rider still expects to have a net surplus, even after punishment. Still, the punishment by third parties is significantly less than punishment by second parties, although both groups will punish severely (Fehr & Fischbacher, 2004).

Several explanations for TPP exist. It may be that third parties want to make the distribution equal between their own payoff and the victim's payoff (Fehr & Schmidt, 1999). Another explanation for TPP is that the third party views the situation as unfair and reacts to this perceived lack of equality, but does not act to restore the equality per se, but to punish the unfair actor (Falk & Fischbacher, 2006). It is this last theory that has so far been supported by experimental outcomes, as it appears that both the intentions of the unfair actor and the outcome are important in decisions of reciprocity (Falk & Fischbacher, 2006).

Another possible explanation is that TPP is motivated by reputation-building. Kurzban et al. (2007) had subjects play a trust game as the first step in a TPP experiment with five rounds. The first players made choices between trusting their partner, putting themselves at risk of losing money but also increasing the total amount of money paid to both players. After playing the game, the first players were given the opportunity to punish their partner at a rate of 1:3. If the experimenter was blind to the punishment action, 38% of first players paid to punish. When the experimenter was aware of the action, punishment increased to 47%. This implies that TPP involves some element of public reputation-building or deterrence.

A second experiment used a sequential prisoner's dilemma game. Player 1 could choose to cooperate or defect. Player 2 was then given the choice to cooperate or defect. If player 1 cooperated, but player 2 defected, the money was distributed $0-$30. If player 2 cooperated as well, the money was distributed $25-$25. Following this game, new participants were shown a game where player 2 defected after player 1 cooperated. The new participants were given $10 and could punish at a rate of 1:3. When their actions were anonymous, 42% of the new participants punished player 2. When the experimenter was aware of the punishment, 65% punished, again reinforcing the idea of the element of public reputation-building or deterrence. Participants did not punish significantly more or less when the previous players (players 1 and 2) were aware of the punishment. Kurzban et al. concluded that only the social presence, not the likelihood of future interactions between parties, is enough to increase TPP. This is in contrast to second-party punishment, which is unaffected by anonymity or observation (Kurzban et al., 2007).

Another important element of TPP is group membership. Theoretically, it only makes sense to punish members of one's own group; punishment increases cooperation, which is something that is desired within one's own group. It is not necessarily important whether one knows the other member, just that this other member is part of one's own group. Shinada et al. (2004) found that subjects who were classified as "cooperators" were more likely to punish in-group than out-group members (36.4% vs. 23.6%). Cooperators (a little over half the sample) were defined as those subjects who gave all of their endowment in the second round of the game. Noncooperators gave less than all, with an average gift of about half of their endowment. This was especially true when punishing the least cooperative players (50.1% vs. 22.0%). However, among those classified as "non-cooperators," there was more punishment of out-group (39.9%) than in-group (22.0%) members (Shinada et al., 2004). It would seem that those who are cooperators will punish according to the TPP definition, but non-cooperators punish for self-interested reasons (competition) (Gintis et al., 2003; O'Gorman et al., 2005; Shinada et al., 2004).

A follow-up questionnaire addressed feelings of anger, unfairness, and guilt, which were all positively correlated with punishment of in-group free-riders but not out-group free riders (Shinada et al., 2004). This supports the theory that TPP is focused on in-group members; it seeks to reduce within-group competition in order to be successful against other groups (Gintis et al., 2003; O'Gorman et al., 2005). Simulations show that only a few strong reciprocators are needed to ensure cooperation (this involves the use of TPP) and that it works better than reciprocal altruism in large groups (Gintis et al., 2003).

Current Framework

Although it has been established empirically that people are willing to and do punish free-riders, this has not been extended to situations in which a particular offender cannot be targeted. In the real world, certain situations lend themselves to punishing an entire group for the transgression of a subset of the group. Teachers, the military, and even the government make use of this technique. While this does exist in the real world, it has not yet been established in the laboratory. This set of studies seeks to do that, and to determine if some simple manipulations, such as the magnitude of the transgression and the number of free-riders, affect how willing subjects are to punish an entire group.

Two different scenarios were employed. In Experiment 1, a simple scenario modeling a classroom environment was given to participants in order to establish empirically the existence of group punishment in the laboratory. In Experiment 2, I manipulated the magnitude of the transgression using a modified public goods game to examine participants' sensitivity to the target of punishment (group versus individual) and level of transgression.

Experiment 1

While the punishment of groups for the violation of a subset can be observed in the field, it has yet to be replicated in the laboratory in a controlled environment. Within the framework of a hypothetical classroom setting, participants had the opportunity to punish (or not) an individual student versus the entire class. This framework allows for the manipulation of two factors: whether an individual or group is to be punished (target) and the magnitude of the transgression (magnitude). The target variable has three levels, individual (IND), group with knowledge (GWK), and group without knowledge (GWOK). The magnitude variable has three levels as well. This will be elaborated on in the methods section.

I anticipated that participants would be more likely to punish individuals and groups who know the identity of the violator and to do so more severely than they would punish groups who do not know the identity of the violator. I also predicted that participants would be sensitive to the magnitude of the transgression, being more likely to punish when the magnitude is higher.

Method

Participants were 312 undergraduate students enrolled in an introductory psychology course at The Ohio State University. Each student received course credit for participating in the study. The sample consisted of 143 (49%) females, 149 (51%) males, and 20 unreported. The majority of the sample was white (81%), with 14 (4.8%) African Americans, 14 (4.8%) Asian Americans or Pacific Islanders, 7 (2.4%) Hispanic or Latinos, 5 (1.7%) participants were of mixed heritage, and 16 (5.5%) other.

Design

The design was a 3 (magnitude) x 3 (target) between-subjects factorial design, so there were nine versions of the paper-and-pencil questionnaire. Participants worked individually in groups of up to 20. Participants read a short scenario that involved a classroom setting in which they were asked to take the role of the teacher (see Appendix A for complete scenarios). In the scenario, each child was allowed to take 1 cookie. The magnitude of the violation (number of cookies) was manipulated by having the one child take 2, 3, or 4 cookies. Participants could decide to punish and if so, how much recess time, out of twenty possible minutes, that the child or class (based on target condition) would lose. It was made clear that this would be an inconvenience to the teacher (they would not get a break for however long the detention lasted) so as to make the punishment costly to the teacher, which is in line with previous research.

For the target variable, the first condition, IND, allowed the participant the choice to punish the specific child or not. The second condition, GWK, allowed the participant to choose to punish the entire class, given that the participants were told that the entire class knew who took the extra cookie(s). The third condition, GWOK, gave the participant the option to punish the entire group, given that the participants were told that the group did not know who took the extra cookie(s). The fact that the group did not know the identity of the offender was made explicit to the participants. If a participant chose to punish, he or she could set the time limit for the detention. There were two dependent measures – the binary decision to punish (yes/no) and the amount of punishment (from 0 to 20 minutes).

At the end of the questionnaire, participants were asked to report some demographic information (see Appendix B), given debriefing forms, and dismissed.

Results

Initial Analyses

I first calculated the frequencies and percentages with which participants punished and the magnitude of the punishment.

Table 1. Proportions and percentages of participants choosing to punish in Experiment 1, and average minutes of punishment for those who punished.

	Target			
# Cookies	Individual (IND)	Group with Knowledge (GWK)	Group without Knowledge (GWOK)	Total
2	24/31 77% 7.20	29/39 74% 7.59	20/35 57% 4.95	73/105 70% 6.74
3	28/31 90% 8.40	30/37 81% 8.07	20/32 63% 6.45	78/100 78% 6.00
4	35/39 90% 8.40	23/32 72% 9.04	21/35 60% 6.43	79/106 75% 8.06
Total	87/101 86% 7.65	82/108 76% 8.17	61/102 60% 7.12	230/311 74% 6.91

Cell contents: ratio of punishers to total, percent who punish, and average amount of punishment in minutes for those who punished

Higher numbers for punishment indicate more sever punishment.

Experiment 1: Proportion of Participants (N = 311) Who Punished

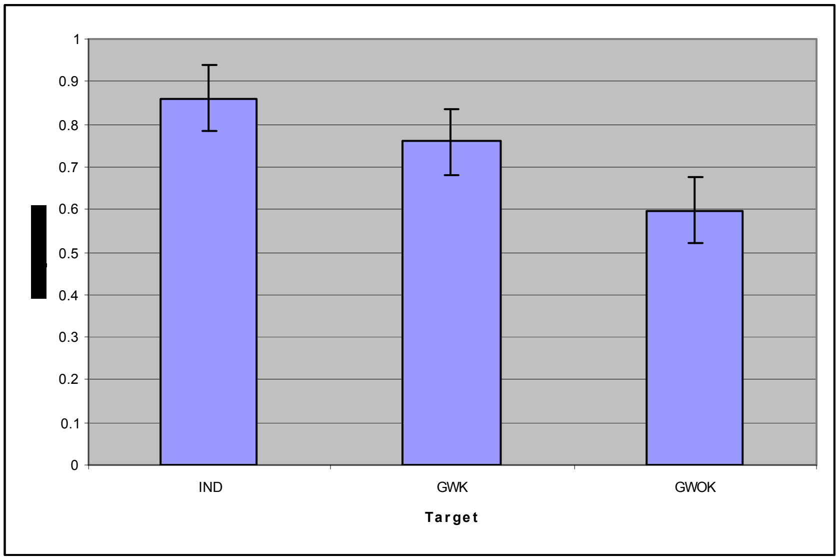

Figure 1. The proportion of participants who chose to punish based on condition. Error bars are ± one standard error.

As can be seen, the target of punishment appears to be important in relation to the binary choice to punish. Participants in the IND (individual) condition and the GWK (group with knowledge) condition were more likely to punish than the participants in the GWOK (group without knowledge) condition. There does not appear to be a clear effect of magnitude (number of cookies) on the binary decision to punish or not.

Binary Logistic Models

Two models were run with different sets of orthogonal contrast codes to examine the effect of target, magnitude, and their interactions.

Table 2. Contrast coding for logistic regression for Experiment 1.

Contrast	Individual	Group w/ Knowledge	Group w/out Knowledge
Code Set 1			
Individual vs. Both Groups (IND vs. GWK and GWOK)	2/3	-1/3	-1/3
Group w/ Knowledge vs. Group w/out Knowledge (GWK vs. GWOK)	0	1/2	-1/2
Code Set 2			
Individual and Group w/ Knowledge vs. Group w/out Knowledge (IND and GWK vs. GWOK)	1/3	1/3	-2.3
Individual vs. Group w/ Knowledge (IND vs. GWK)	1/2	-1/2	0

In the models, number of cookies was not significant (b = 0.17, p > .10). The first model contrasted IND vs. GWK and GWOK. It also contrasted GWK vs. GWOK. The first contrast code comparing IND to GWK and GWOK on frequency of punishment showed that the difference in punishment between individual punishment and group punishment was significant (b = 1.08, p < .01). Participants were more likely to punish individuals than the group with knowledge and the group without knowledge. The contrast comparing the GWK and GWOK was also significant (b = 0.75, p < .05). Participants were more likely to punish the group with knowledge than the group without knowledge. None of the interactions between target and magnitude was significant (all *p*s > .10).

The second model contrasted the GWOK to IND and GWK as well as IND to GWK. The contrast for GWOK versus IND and GWK was significant (b

= 1.10, p < .01), but the contrast comparing IND to GWK was not (p > .05). In this second model, participants were more willing to punish individuals and groups with knowledge than groups without knowledge. Individuals and groups with knowledge did not differ in frequency of punishment. None of the interactions between target and magnitude was significant (all *p*s > .10).

Table 3. Binary logistic and Ordinal Logistic Parameter Estimates (N = 312) for Experiment 1

	Binary (punish/not punish)			Ordinal (Amount punishment)		
Variable	b	Std Error	Odds Ratio	b	Std Error	Odds Ratio
Intercept	1.13**	0.14		1.17** -0.90** -3.01**	0.14 0.13 0.25	
Number of Cookies	0.17	0.17	1.18	0.28*	0.13	1.33
Model 1						
Individual vs. Both Groups (IND vs. GWK and GWOK)	1.08**	0.33	2.94	0.72**	0.23	2.05
Group w/ Knowledge vs. Group w/out Knowledge (GWK vs. GWOK)	0.75*	0.30	2.11	1.06**	0.26	2.90
Model 2						
Individual vs. Group w/ Knowledge (IND vs. GWK)	0.70	0.37	2.02	0.19	0.26	1.20
Individual and Group w/ Knowledge vs. Group w/out Knowledge (IND and GWK vs. GWOK)	1.10**	0.27	3.00	1.16**	0.24	3.18

*p < .05, **p < .01
None of the four interactions for either binary or ordinal regressions was significant (p > .20).
Regression performed using all participants (N = 312)

Experiment 1: Average Amount of Punishment Among Punishers (N = 247)

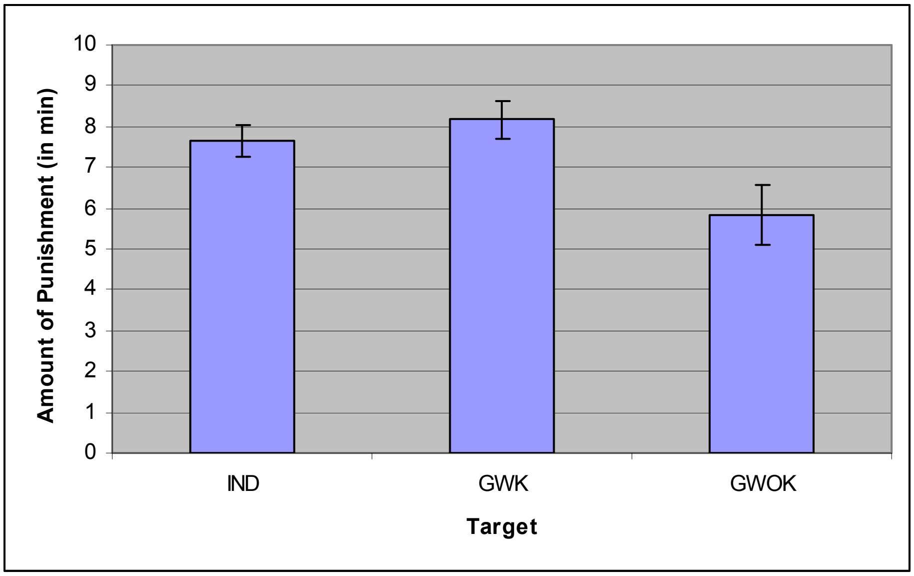

Figure 2. The average amount of punishment (out of 20 possible minutes). Error bars are ± one standard error.

Ordinal Logistic Regression Models

Since I collected data on the amount of punishment, I predicted the level of punishment in an ordinal logistic regression using all of the participants (N = 312). I divided the amount of punishment into four ordinal categories: 0 = no punishment, 1 = punishment between 1 and 5 minutes, 2 = punishment between 6 and 10 minutes, 3 = punishment between 11 and 20 minutes. The last bin is a longer interval because I collapsed two bins due to the lower number of responses in the two highest bins. Number of cookies was found to be significant ($b = 0.28$, $p < .01$). Participants punished the child or the class more when the number of cookies taken was higher.

Again, I used the two sets of orthogonal contrast codes to examine the effect of target (refer to Table 3 above for parameter estimates). In the first model, the code contrasting individual to both groups was significant ($b = 0.72$, $p < .01$). Participants gave higher amounts of punishment to individuals than to either group. The contrast comparing GWK to GWOK was also significant ($b = 1.06$, $p < .01$). Participants punished groups with knowledge more than groups without knowledge. None of the interactions was significant (all *p*s > .10).

The second model contrasted IND with GWK as well as GWOK with IND and GWK. The IND versus GWK contrast was not significant ($b = 0.19$, $p > .10$). Participants did not differ in the amount of punishment they gave to individuals versus groups with knowledge. The GWOK versus IND and GWK comparison was significant ($b = 1.16$, $p < .01$). Participants punished groups with knowledge and individuals more than they punished groups without knowledge. None of the interactions between target and magnitude was significant (all *p*s > .10). Thus, based on level of punishment, participants were more severe in their punishment of individuals and groups with knowledge than groups without knowledge. Individuals and groups with knowledge did not significantly differ in severity of punishment.

It might seem objectionable to include participants who did not choose to punish in the ordinal regression model on amount of punishment. Therefore, a similar ordinal regression using only those who chose to punish ($N = 230$) was also run. The models were exactly the same as the first ordinal logistic models except for restricting analysis to only punishing participants. These models did have a significant main effect for number of cookies ($b = 0.41$, $p < .05$), with more severe punishment for those who took more cookies.

For the first model, IND versus both groups was no longer significant ($b = 0.29$, $p > .10$). Participants who chose to punish did not punish individuals differently from both the group with knowledge and the group without knowledge. GWK versus GWOK was significant ($b = 1.21$, $p < .01$), with participants punishing groups with knowledge more than groups without knowledge. None of the interactions between target and magnitude was significant (all *p*s > .10). In the second model, the contrast for IND versus GWK was not significant ($b = -0.32$, $p > .05$). Participants who chose to punish did not differ in the amount of punishment they gave to individuals versus groups with knowledge. The contrast for IND and GWK versus GWOK ($b = 1.05$, $p < .01$) was significant. Participants who chose to punish gave more punishment to individuals and groups with knowledge than to groups without knowledge. Again, all interactions were not significant (all *p*s > .05). These models give slightly different results than the models incorporating all participants. The sole difference is that the previously significant difference in punishment magnitude between IND and GWK and GWOK became nonsignificant in the model using only participants who chose to punish.

Table 4. Ordinal Logistic Parameter Estimates (N = 230) for Experiment 1

	Ordinal (Amt punishment)		
Variable	b	Std Error	Odds Ratio
Intercepts	-0.47** -2.76**	0.15 0.27	
# Cookies	0.41*	0.18	1.51
Coding 1			
Individual vs. both groups (IND vs. GWK and GWOK)	0.29	0.28	1.33
Group w/ Knowledge vs. Group w/out Knowledge (GWK vs. GWOK)	1.21**	0.37	3.37
Coding 2			
Individual vs. Group w/ Knowledge (IND vs. GWK)	-0.32	0.31	0.73
Individual and Group w/ Knowledge vs. Group w/out Knowledge (IND and GWK vs. GWOK)	1.05**	0.34	2.87

*p < .05, ** p < .01
None of the interactions was significant (p > .20), Intercept 1 was not significant

Discussion

This first experiment establishes that while group punishment can be replicated in a lab, there are differences between punishing an individual perpetrator, a group with knowledge of the perpetrator, and a group that does not know the identity of the perpetrator. Punishing an individual perpetrator does not differ significantly from punishing a group who knows the identity of the perpetrator, either in the decision to punish or not or in the level of punishment given.

The majority of participants were willing to punish groups without knowledge of the perpetrator, but not to the extent that they would punish individuals or groups with knowledge. This suggests that perhaps a group with knowledge is being punished at equal rates to individuals due to the fact that they are violating a social norm by not revealing the perpetrator. They were, in essence, lying to the teacher. This may be why participants were willing to punish the group with knowledge at the same rate and with the same severity as punishing the individual him or herself. Punishing a group that does not know the perpetrator creates a dilemma: the desire to punish the person who committed the violation is at odds with the desire to avoid punishing innocent people. However, the majority of participants did punish, which is especially surprising when examining the high rates of punishment of the group without knowledge (61%). We will see in the next experiment, however, that the dilemma between punishing the wrongdoer and avoiding punishing the innocent could play out in a slightly different way.

Experiment 2

Method

Participants were 396 undergraduate students enrolled in an introductory psychology course at The Ohio State University. Each student received course credit for participating in the study. The sample consisted of 213 (54%) females, 182 (46%) males, and 1 unreported. The majority of the sample was white (79%), with 15 (4%) African Americans, 33 (8%) Asian Americans or Pacific Islanders, 14 (4%) Hispanic or Latinos, and 4 Native Americans (1%).

Design

Participants worked individually in groups of 1 to 8 on computers. They read a description of a hypothetical public goods game in which there are 5 hypothetical players (X1-X5), each given a $10 endowment. These hypothetical players could choose to contribute any, none, or all of their endowment to the public fund. All money contributed to the public fund would be summed, and each hypothetical player, regardless of their contribution, would receive 40% of the public fund by default. Participants would read a scenario, then make a binary decision regarding punishment (yes or no). If they chose to punish, they could do so by decreasing the default payout from the public fund from 40 for the target (individuals or group). For example, the public fund could contain $30. By default, every player (X1-X5) would receive \$12 (0.4 x \$30). However, a participant could reduce the payout percentage for either individuals or the group (depending on condition) to 20%. Those who are punished would receive \$6 (0.2 x \$30) from the public fund instead of \$12. (For entire scenarios, see Appendix C.)

The experiment was a 4 (contribution disparity) x 2 (target) between-subjects factorial design. Contribution disparity is the difference in the amount donated to the public fund between high and low contributing players. There were 4 levels of manipulation based on the disparity between the higher contributors and the lower contributors, but always maintaining a public good pot of \$30. Contribution disparity was set at equal intervals ranging from \$2.50 to \$10. The identities of the low and high players remained constant (X1, X3, and X5 always contributed an equal higher amount; X2 and X4 always contributed an equal lower amount). The other independent variable, target, had two levels. The participants were permitted to punish the low contributors (X2 and X4) exclusively (individual condition) or the entire group of players (group condition).

Table 5. Contributions of Hypothetical Players in the Public Goods Game

	Contribution Disparity			
Player	Low	Medium-low	Medium-high	High
X1	\$7	\$8	\$9	\$10
X2	\$4.50	\$3	\$1.50	\$0
X3	\$7	\$8	\$9	\$10
X4	\$4.50	\$3	\$1.50	\$0
X5	\$7	\$8	\$9	\$10
Disparity	\$2.50	\$5.00	\$7.50	\$10.00

In the Individual condition, participants were allowed to change the default payout from the public fund from 40% to some lower percentage that would apply only to those hypothetical players who gave less (players X2 and X4). In the Group condition, participants were allowed to change the default payout from the public fund from 40% to some lower percentage that would apply to all the members of the group (players X1-X5). After the initial decision to punish or not, if the participant chose not to change the default they were asked to answer some demographic questions, then dismissed. If a participant chose to change the payout percentage, they were given an opportunity to enter their own payout percentage (constrained to be less than or equal to 40) and then shown the effect of their decision on the payouts to the hypothetical players. If they were not satisfied with the new allocation, they were allowed to enter a new percentage. This was done to allow for people who could not figure out what the payouts would be given a certain percentage. They could see how their first choice altered the payouts and correct accordingly, if they so desired. After reaching a satisfactory payout percentage, they also answered some demographic questions. I had two dependent measures – a binary decision to punish (yes/no) and an amount of punishment (scale of 0 – 40%).

I predicted that participants would be willing to punish both in the Individual and Group conditions, but to a lesser extent in frequency and severity in the Group condition. I also predicted that participants would be somewhat sensitive to the size of the disparity, with more frequent and harsher punishment the larger the disparity became. However, due to past research on quantity insensitivity in between-subjects designs (Bartels & Medin, 2007), I did not expect a large effect for this manipulation.

Results

Initial Analyses

I first calculated the frequencies and percentages of participants who punished and how much punishment was given on average for those who did punish.

Experiment 2: Proportions of Participants (N = 396) Who Punished

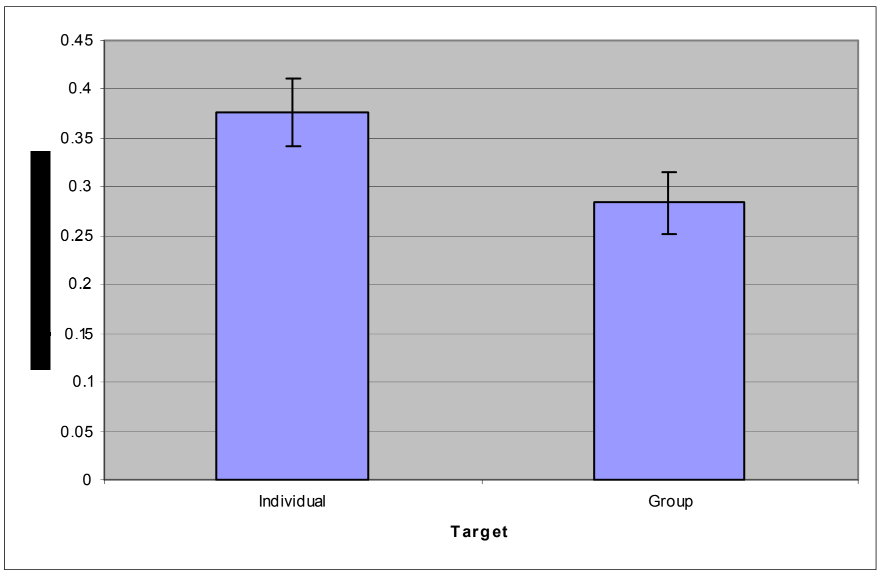

Figure 3. The proportion of participants who chose to punish. Error bars are ± one standard error.

Table 6. Proportion and percentages of punishers and average amount in minutes of punishment for those who chose to punish in Experiment 2.

	Target		
Contribution Disparity	Individuals	Group	Total
Low ($2.50)	11/46 24% 24.72%	12/58 21% 13.33%	23/104 22% 18.78%
Medium-low ($5.00)	21/53 40% 20.48%	6/33 18% 5.83%	27/86 31% 11.37%
Medium-high ($7.25)	24/53 45% 23.29%	19/55 35% 14.63%	43/108 40% 19.47%
High ($10.00)	20/50 40% 25.10%	18/48 38% 16.56%	38/98 39% 21.05%
Totals	76/202 38% 23.20%	55/194 28% 14.02%	131/396 33% 19.34%

Cell contents: ratio of punishers to total, percent who punish, average amount of punishment in minutes for those who chose to punish
Higher percentages indicate greater punishment (calculated by taking 40% - % given by participant).

Punishment could range from reducing the default public fund payout from 40% (no punishment) down to 0. I converted this by taking 40% - % given by the participant, so that larger percentages indicate greater punishment. In Table 6, there appears to be a general trend to punish more often as magnitude of disparity increased (players who gave less deviated more from those who gave more).

Also, participants in the Group condition were less likely to punish than those in the Individual condition (28% versus 38%, respectively) and punished less severely (14.02% versus 23.20%).

Experiment 2: Average Amount of Punishment for Punishers (N = 131)

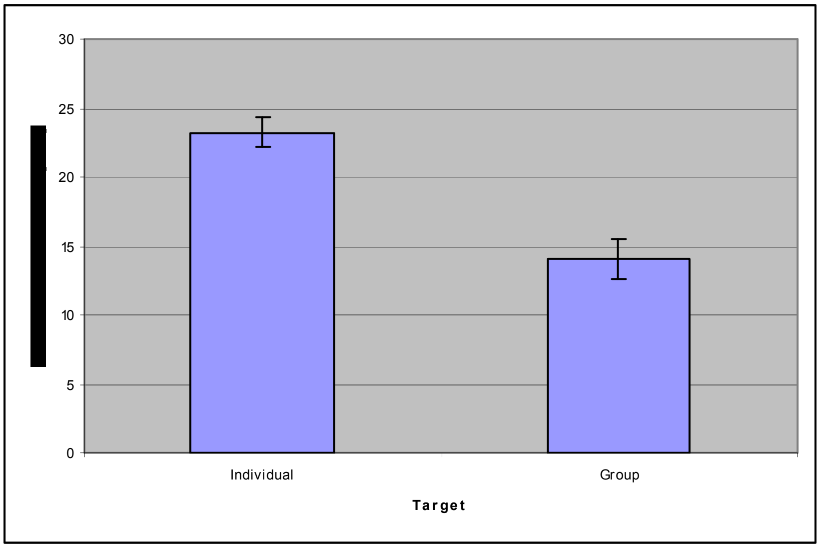

Figure 4. The average amount of punishment among participants who chose to punish (out of a possible 40%). Error bars are ± one standard error.

Binary Logistic Regression Model

As in Experiment 1, I analyzed the binary choices and the ordinal amount of punishment. Analyses were performed based on the binary choice to punish (coded as 0) or not to punish (coded as 1). Target was coded as ½ for Individual and -½ for Group. The independent variable contribution disparity was centered and coded as: 1.5 = \$10, 0.5 = \$7.50, -0.5 = \$5.00 and -1.5 = \$2.50. Confirming what is seen in Table 6, target was found to be significant ($b = 0.43$, $p < .05$). In this case, participants were more likely to punish an individual than a group (odds ratio = 1.54). Contribution disparity was also found to be significant ($b = 0.27$, $p < .01$). The more extreme the disparity between high versus low donors to the public fund, the more willing participants were to punish (odds ratio = 1.32). The interaction of target and contribution disparity was not significant (b = -0.10, p > .10).

Table 7. Binary and ordinal logistic regression estimates for Experiment 2 using all participants (N = 396)

	Binary (punish/not punish)			Ordinal (amount of punishment)		
Variable	b	Std Error	Odds Ratio	b	Std Error	Odds Ratio
Intercepts	-0.74**	0.11				
				-0.82**	0.12	
				-1.23**	0.13	
				-2.09**	0.16	
				-3.40**	0.28	
Target (individual vs. group)	0.43*	0.22	1.54	0.76**	0.23	2.13
Contribution Disparity	0.27**	0.10	1.32	0.29**	0.10	1.34
Target x Contribution Disparity	-0.10	0.20	0.91	-0.07	0.20	0.93

*p < .05, **p < .01
For ordinal, calculated based on all participants

Ordinal Logistic Model

An incremental variable was created from the amount to punish variable with bins that ranged as follows: 0 = no punishment, 1 = (39%, 30%], 2 = (29%, 20%], 3 = (20%, 0%]. These numbers are based on the conversion so that higher numbers reflect greater punishment. An ordinal logistic model using all participants (N = 396) predicting magnitude of punishment on the basis of target (contrast coded as + ½ for individual punishment, - ½ for group punishment), contribution disparity (0 - 3), and their interaction was used to analyze the data (see Table 7 above for parameter estimates). There was a significant effect of target (b =0.76, p < .01). The odds ratio for condition indicated that participants punished individuals 2.15 times more severely than groups. Contribution disparity was also significant (b = 0.29, p < .01). The more that the players who contributed less deviated from those players who contributed more, the more punishment a subject was willing to dole out (odds ratio = 1.34). There was no interaction effect (b = -0.07, p > .10).

In addition to this ordinal logistic model using all participants, another similar ordinal logistic was run using only those participants who chose to punish (N = 131). The bins remained the same. Target remained significant (b = 1.77, p < .01), but contribution disparity became nonsignificant (b = 0.42, p > .05). The interaction remained nonsignificant (b = -0.29, p > .20).

Table 8. Ordinal logistic regression estimates for Experiment 2 using only participants who chose to punish (N = 131)

	Ordinal (amount of punishment)		
Variable	b	Std Error	Odds Ratio
Intercepts	0.96**	0.23	
	-0.94**	0.23	
	-2.61**	0.32	
Target (individual vs. group)	1.77**	0.40	5.85
Contribution Disparity	0.42	0.17	1.53
Target x Contribution Disparity	-0.29	0.33	0.75

*p < .05, ** p < .01
Intercept 1 was *ns*

Discussion

As in Experiment 1, I found that there is a difference in the punishment of individuals versus groups. In keeping with Experiment 1, participants were more likely to punish individuals than groups and to do so more severely. This scenario is more complicated and may reflect a different dimension, such as the strength of the norm violated in this experiment and/or of the dilemma between balancing the desire to punish the low contributors versus avoiding punishing the high contributors. It may be that subjects want to send a message to the perpetrator or the whole group. The theoretical differences between the results found in Experiment 1 versus Experiment 2 will be covered more fully in the general discussion.

General Discussion

Two experiments have supported the fact that punishment of groups can be reproduced in a laboratory setting. In both Experiment 1 (cookie study) and Experiment 2 (public goods study), group punishment was observed to varying degrees. In Experiment 1, punishment occurred at least 50% and up to 95% of the time, with the smallest frequency of punishment for groups without knowledge of the identity of the perpetrator. The punishment rate was quite high, even when punishing the group. Experiment 2 also found punishment of both individuals and groups, but less often. Whereas in Experiment 1, 74% of participants punished on average, in Experiment 2 this was drastically reduced to an average of 33%.

Comparison of Frequency of Punishment for Experiment 1 versus Experiment 2

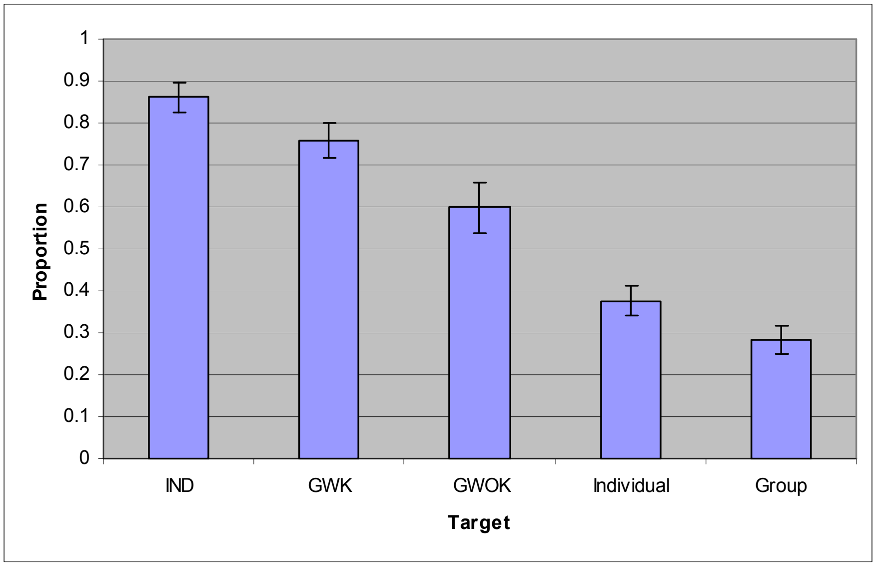

Figure 5. The proportion of participants who chose to punish for Experiment 1 (IND, GWK, and GWOK) and Experiment 2 (Individual and Group). Error bars are ± one standard error.

Yet, the average severity of the punishment based on the participants who punished was similar in both studies.

Comparison of Percentage Punishment for Experiment 1 versus Experiment 2

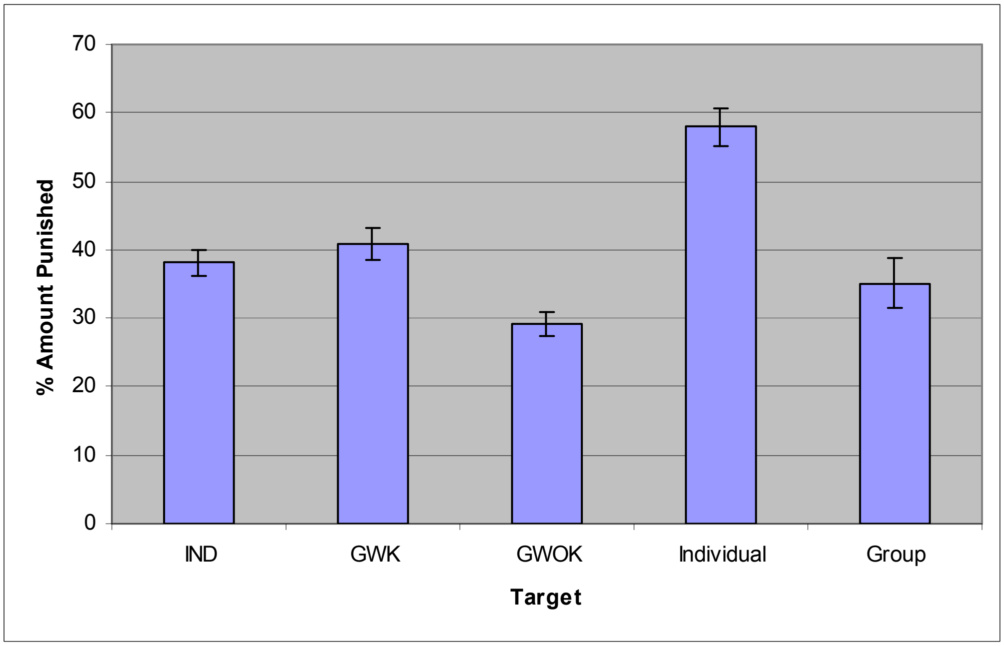

Figure 6. The average amount of punishment for Experiment 1 and Experiment 2 participants standardized onto a percentage scale (IND, GWK, and GWOK are from Experiment 1; Individual and Group are from Experiment 2). Error bars are ± one standard error.

One possible explanation is that in Experiment 1, it was clear that the child who took extra cookies was violating a social norm, taking more than one is allowed to take. In Experiment 2, it is not so clear that the hypothetical free-riders were violating a social norm. Participants could feel that the free-riders were not stealing. Free-riders were given a payout from the public fund, but they did not ask for one or take it by choice. They could also be seen as being strategic – free-riding allows for maximum economic benefits. Another person or organization had already determined that everyone would receive payouts from the public fund; the responsibility of that choice was not with the free-rider but with some other person(s), although the players do choose to free-ride. This may account for why participants in Experiment 2 were punished less frequently than in Experiment 1.

A second explanation could be that in Experiment 1, participants in either group condition (group with knowledge and group without knowledge) saw a scenario in which there was a clearer difference between the guilty and the innocent. For the group with knowledge, the child who took more than allowed was obviously guilty, and the rest of the group was guilty to a lesser degree perhaps, by not telling the teacher the identity of the cookie robber. In the group without knowledge, each child either took extra cookies or did not. Those who did not take extra cookies did not know who did, and were completely innocent. Thus, it would be easier to determine whether to punish and how much to punish.

In Experiment 2, participants in the group decision had a less clear choice to make. Not only did the contributions of free-riders increase but the contributions of the "high" contributors decreased as well, compared to Experiment 1 where the only the behavior of the violator varied. The high contributors are not completely "innocent." Punishing the group without knowledge in Experiment 1 necessarily punishes completely innocent people (at least as far as cookie stealing goes). Punishing the group in Experiment 2 punishes two possible types of guilty players (except in the case of high disparity, where players X1, X3, and X5 each donate their entire endowment of $10 to the public fund). While there is a relative difference between high and low contributors, for all but one contribution disparity level, neither type contributed all of their endowment to the public fund. While this might lead us to believe that Experiment 2 should have had higher punishment rates due to the increased number of "guilty" versus Experiment1, I believe that the social norm being violated (taking cookies versus giving to a public fund) overrides this tendency when examining punishment frequency. However, once the decision to punish has been made, the magnitude of the punishment does not seem to be influenced by the norm violated.

Both experiments indicate that participants will punish groups and that participants are sensitive to the magnitude of a violation when determining the amount of punishment, but not the necessarily when deciding whether to punish. These initial results require future research to clarify the phenomenon of group punishment.

Future Directions

I anticipate extending this paradigm to other domains, using different scenarios such as the military or other economic games. This would be useful in determining how robust group punishment is across different domains. I would also like to have experiments that involve actual money at stake which would be an especially strong test of group punishment.

Additionally, experiments could be designed to examine what impact group punishment has on the behavior and affect of actual members of the group, both violators and non-violators. Punishment could result in non-violators reducing their conformity to the social norm and becoming angry or frustrated. It could also result in violators increasing their conformity to avoid future punishment.

Another interesting direction would be to examine entitativity. Entitativity is the degree to which one member of a group is seen as a legitimate representative of that group (Rydell, Hugenberg, Ray, & Mackie, 2007). Increasing entititativity would be predicted to result in increasing levels of group punishment. Those who violate social norms would be seen as representative of his or her group, which may imply to outsiders that the group espouses or tolerates the violation. Therefore, punishing the group would send a message to the entire group that the violation will not be tolerated and therefore encourage other members of the group to adhere to social norms.

In contrast, when entitativity is low, the group would be less likely to be viewed as endorsing or tolerating the social norm violation. This may also explain why the rates of punishment were higher in Experiment 1 than in Experiment 2. In Experiment 1, the group was a class of schoolchildren. They often work and play together. They could be seen as high in entitativity. They will most likely interact with one another on a regular basis. The use of peer pressure in such a situation could force conformity in the offender. In Experiment 2, the players of the public goods game do not know each other (this is stated to participants explicitly) and will not ever play the public goods game together again. The groups could be seen as low in entitativity, therefore punishing non-offenders may not seem as appropriate as in high entitativity groups.

Limitations

There are some limitations to the two experiments. First, these results would need to be replicated and extended to other situations. The use of the classroom and an economic game was convenient to use to create group punishment conditions in the lab, but there are other domains where group punishment could be explored (e.g. legal or military situations). Field studies could be especially helpful in identifying these areas and generalizing lab findings.

As usual, our sample may not be representative of the population in important respects. Our samples were overwhelmingly white in ethnicity and probably of different socio-economic and cultural backgrounds than the general population. Also, the sample size for punishers for the ordinal regression for Experiment 2 was significantly smaller than the total number of participants, perhaps introducing instability in the regression estimates.

It is traditional to use real money in these types of experiments, especially the public goods game. Since it did not actually cost the participants anything to punish the individual or groups, the proportions of those who punished could be inflated. However, the relative proportions and punishment magnitudes should remain similar between conditions (target and magnitude/contribution disparity), just at lower levels. A person should be indifferent to whether he or she is punishing a group or an individual if the cost is the same to him or her for both, although some people might believe they are getting more for their money when punishing a group or less if punishing innocents is bad. In addition, for many economic games, it has been found that providing actual monetary payment does not significantly alter the results (Camerer & Hogarth, 1999).

Conclusions

People generally desire to punish those who violate social norms. This has an evolutionary advantage, as it increases cooperation, in turn increasing the ability of one's group to survive. This desire to enforce social norms can even extend to punishing those who could be termed "innocent." Although these findings are preliminary, I believe that they require further research to fully explore this phenomenon. Ideally, I would like to find boundary conditions for such punishment, when it is most likely to occur, and how successful group punishment is. The ultimate goal of this line of research would be to determine if and when group punishment increases cooperation and reduces the frequency and severity of norm violations.

References

Bartels, D. M., & Medin, D. L. (2007). Are morally motivated decision makers insensitive to the consequences of their choices? *Psychological Science, 18*, 24-28.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*, 7-42.

Carpenter, J. P., Matthews, P. H., & Ong'ong'a, O. (2004). Why punish? Social reciprocity and the enforcement of prosocial norms. *Journal of Evolutionary Economics, 14,* 407-429.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*, 293-315.

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

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

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

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical*

Biology, 206, 169-179.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24*, 153-172.

Hamilton, W. D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology, 37*, 1-52.

Hopfensitz, A., & Reuben, E. (2009). The importance of emotions for the effectiveness of social punishment. *The Economic Journal, 119*, 1534-1559.

Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: Confusion or a heuristic? *Evolution and Human Behavior, 21*, 411-427.

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

O'Gorman, R., Wilson, D. S., & Miller, R. R. (2005). Altruistic punishing and helping differ in sensitivity to relatedness, friendship, and future interactions. *Evolution and Human Behavior, 26*, 375-387.

Price, M. E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior, 23*, 203-231.

Rydell, R.J., Hugenberg, K., Ray, D., & Mackie, D. M. (2007). Implicit theories about groups and stereotyping: The role of group entitativity. *Personality and Social Psychology Bulletin, 33*, 549-558.

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature, 8*, 300-312.

Shinada, M., & Yamagishi, T. (2007). Punishing free riders: Direct and indirect promotion of cooperation. *Evolution and Human Behavior, 28*, 330-339.

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

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35-57.

Turillo, C. J., Folger, R., Lavelle, J. J., Umphress, E. E., & Gee, J. O. (2002). Is virtue its own reward? Self-sacrificial decisions for the sake of fairness. *Organizational Behavior and Human Decision Processes, 89*, 839-865.

Appendix A: Materials for Experiment 1

Individual Punishment

Imagine that you are a teacher at a public middle school. You work very hard to teach your students. You are busy almost all of the time with your class of 30 students. Every day, the class gets cookies for snack. Each child is allowed to take 1 cookie. However, on this day, someone took *(2, 3, or 4)* cookies. You find out that Bobby took the extra cookies. You have two options: to ignore it or to reduce the number of minutes of recess for Bobby. What would you like to do?

1.

☐ Ignore it

☐ Reduce number of recess minutes for Bobby (if yes, go to #2)

2.

You have chosen to reduce the number of recess minutes for Bobby. Normally, recess is 20 minutes long. Bobby (and you) will not be able to have a break for a certain number of minutes, which is up to you. How many minutes would you like to reduce recess?

__________ minutes

Group Punishment w/ Group Knowledge

Imagine that you are a teacher at a public middle school. You work very hard to teach your students. You are busy almost all of the time with your class of 30 students. Every day, the class gets cookies for snack. Each child is allowed to take 1 cookie. However, on this day, someone took *(2, 3, or 4)* cookies. Since you are so busy, you do not have the time or the energy to figure out who took the extra cookies. The other students know who took the extra cookies, but won't tell who. You have two options: to ignore it or to reduce the number of minutes of recess for everyone. What would you like to do?

1.

☐ Ignore it

☐ Reduce number of recess minutes (if yes, go to #2)

2.

You have chosen to reduce the number of recess minutes. Normally, recess is 20 minutes long. The students (and you) will not be able to have a break for a certain number of minutes, which is up to you. How many minutes would you like to reduce recess?

__________ minutes

Group Punishment w/out Group Knowledge

Imagine that you are a teacher at a public middle school. You work very hard to teach your students. You are busy almost all of the time with your class of 30 students. Every day, the class gets cookies for snack. Each child is allowed to take 1 cookie. However, on this day, someone took *(2, 3, or 4)* cookies. Since you are so busy, you do not have the time or the energy to figure out who took the extra cookies. None of the other students know who took the extra cookies. You have two options: to ignore it or to reduce the number of minutes of recess for everyone. What would you like to do?

1.

☐ Ignore it

☐ Reduce number of recess minutes (if yes, go to #2)

2.

You have chosen to reduce the number of recess minutes. Normally, recess is 20 minutes long. The students (and you) will not be able to have a break for a certain number of minutes, which is up to you. How many minutes would you like to reduce recess?

__________ minutes

Appendix B: Demographic Materials

Background Information

Finally, we would like you to answer a few questions about yourself. This information will be very useful in helping us describe the types of people who participated in our study.

1. What is your sex? *Please mark one box.*

    ☐ Female

    ☐ Male

2. What is your age?

    __________ years old

3. What is your race or ethnicity? *Please mark all boxes that apply.*

    ☐ African American

    ☐ Asian American or Pacific Islander

    ☐ Native American or Alaskan

    ☐ Hispanic or Latino

    ☐ White (not of Hispanic origin)

☐ Other (please specify): ___________________________________

4. Is English your first language? *Please mark one box.*

☐ Yes

☐ No

Appendix C: Materials for Experiment 2

Instructions

Thank you for agreeing to participate in this study. You will be reading about a special kind of game and then be presented with a hypothetical scenario after which you will answer a couple questions. You may stop at any time and there are no right or wrong answers. Although the situations are hypothetical, please respond to them as if they were real.

The Public Goods Game

One kind of game used by researchers to examine human behavior and decision making is the public goods game. There are five players, labeled 1 through 5. Each player is given an endowment of $10. They may keep any or all of this money or contribute some or all to the public fund. All the money that is donated to the fund is added together, and then each player receives 40% of this total back from the fund. Each player gets this 40% payout, regardless of whether they contributed to the fund.

If every player contributes an equal share, then everyone comes out ahead – that is, they make more than $10 - and everyone has the same amount of money. For example, if every player contributes all $10 to the fund, there will be $50 in the fund (5 players x $10 each). 40% of $50 is $20, so each player leaves the game with $20.

On the other hand, if some players contribute more than others, then some players could lose money (earn less than $10) and others would gain money. For example, if two players each contribute their full $10 and the other three players contribute nothing, the fund contains $20. Each player receives 40% of this, which is $8. The two players who contributed all of their endowments earn $8, a net loss of $2. The three players who contributed nothing each have their original $10 plus the $8 from the fund, earning them a total of $18.

This game is played by many groups. After each game is over, the players are shuffled and placed into new groups with different players. They then play the game again with this new group.

Your Task

You are a sixth player in the game. You do not participate in the first part, the public goods game. However, you do observe how much each player contributes to the public fund. After each player has made their decision about how to contribute, you will have a decision to make. Based on the contributions, you may choose to change the payout percentage from the default of 40%. You can only reduce the percentage and you may reduce it as far down as 0%. This may be done to reduce payoffs. (Experimental condition: *This new percentage will be applied to all the players – each player will now receive the new percentage of the fund.) (*Control condition*: This new percentage will only be applied to the players you choose – you may choose any or all of the players to receive the new percentage of the fund.)*

You are now ready to observe one of these games and to make some decisions.

Remember that you do not have to change the payoff percentage and that while a specific group will not play the game again, each player within the group will with a new group.

Task

Each player, 1 through 5, is given a $10 endowment. They may contribute any or all of it to the public fund. Any money placed in the fund will be added together. This total will be multiplied by 40% and given to each player, regardless of their contribution level. Each player contributed to the fund as follows:

Player 1: $10

Player 2: $0

Player 3: $10 *These numbers will vary as part of the manipulation.*

Player 4: $0

Player 5: $10

Total in public fund: $30

If the default 40% is applied, then each player receives $12 from the public fund. The total earnings for each player are as follows:

Player 1: $12

Player 2: $22

Player 3: $12

Player 4: $22

Player 5: $12

Group condition:

Would you like to

□ Use the default, keeping the payouts as above

□ Change the payoff percentage for all of the players, changing the payouts

If they select to change, they then see this screen:

You have chosen to change the payoff percentage. What would you like to change it to? Remember that this applies to all players.

__ %

You have chosen to change the payoff percentage to X%. This changes the payoffs to the players to

Player 1: ($30)*(X%)

Player 2: $10 + ($30)*(X%)

Player 3: ($30)*(X%)

Player 4: $10 + ($30)*(X%)

Player 5: ($30)*(X%)

Are you sure you would like to implement this change?

□ Yes

□ No (*If no, go back to screen to select percentage*)

Individual condition:

Would you like to

□ Use the default, keeping the payouts as above

□ Change the payoff percentage for one or more of the players, changing their payoffs

If they select to change, they then see this screen:

You have chosen to change the payoff percentage for one or more of the players. What would you like to change it to? Remember that this applies only to the players you will select on the next screen.

__ %

You have chosen to change the payoff percentage to X%. Which players would you like this new percentage to apply to? *Check all that apply*.

□ Player 1

□ Player 2

□ Player 3

□ Player 4

□ Player 5

You have selected Player A (B, C, etc) to receive the new payoff percentage of X%. The total payoffs to the players are now

Player 1: ($30)*(40 or X%)

Player 2: $10 + ($30)*(40 or X%)

Player 3: ($30)*(40 or X%)

Player 4: $10 + ($30)*(40 or X%)

Player 5: ($30)*(40 or X%)

Are you sure you want to implement this change:

□ Yes

□ No (*If no, go back to screen to select percentage*)