Justification and Legitimate Punishment

Xiao, Erte and Tan, Fangfang

Carnegie Mellon University, Max Planck Institute for Tax Law and Public Finance

22 May 2013

Justification and Legitimate Punishment

Erte Xiao
Department of Social and Decision Sciences
Carnegie Mellon University
exiao@andrew.cmu.edu

Fangfang Tan
Department of Public Economics
Max Planck Institute for Tax Law and Public Finance
fangfang.tan@tax.mpg.de

May 22nd, 2013

Abstract

Punishment can lose its legitimacy if the enforcer can profit from delivering punishment. We use a controlled laboratory experiment to examine how justification can combat profit-seeking punishment and promote the legitimacy of punishment. In a one-shot sender-receiver game, an independent third party can punish the sender upon seeing whether the sender has told the truth. Most third parties punish the senders regardless of how the senders behave when they can profit from punishment. However, majority third parties punish the sender if and only if the sender lies when they have to provide explanations for their punishment decisions. Our data also suggests that senders are more likely to perceive punishment as legitimate and behave honestly when they know the enforcer has to justify their punishment decisions. Our findings suggest that justification requirement plays an important role in building efficient punishment institutions.

Key words: third-party punishment, justification, sender-receiver game, experiment

JEL classification: C72, C92, D63, D83

Acknowledgements We gratefully acknowledge the National Science Foundation (SES-0961341) and the National Basic Research Program of China (973 Program) (Serial No. 2012CB955802) for funding that supported this research.

1. Introduction

Justification is widely required in government and other organizations whose decisions can have broader consequences for others. Legislators in the European Union are legally obliged to justify interventions that affect freedom or property rights. In the United States, the Department of Health and Human Services (HSS) began requiring health insurance providers to justify insurance rate increases of 10% or more in September, 2011.[1] Doctors need to offer reasons for prescribing certain costly medications. In large companies and non-profit organizations, human resources must report reasons for hiring and firing employees, and managers need to justify employee evaluations. Justification is particularly important in the court system. In a famously court case, a former Illinois police officer, Drew Peterson, screamed out his innocence right before the verdict, was charged guilty and was sentenced to 38 years in prison for murdering his third wife.[2] In this case, as in many others, the judge needed to offer an explanation along with his sentence.

In view of the pervasive use of justification, it is important to understand how the justification requirement affects decision making. This paper presents novel experimental evidence that requiring justification could promote the legitimacy of third party punishment and curb corrupt punishment behavior.[3]

Whether punishment is perceived to be legitimate determines how effectively it can signal social norms and promote conformity (Lind and Tyler, 1988; Tyler, 2006; Xiao, forthcoming), Compared with the implicated stakeholders, a third party's judgment is less likely to be influenced by negative emotions such as vengeance and anger (Fehr and Fischbacher, 2004;

[1] See http://www.hhs.gov/news/press/2011pres/05/20110519a.html.
[2] See http://abcnews.go.com/US/wireStory/peterson-screams-38-years-murder-18563027.
[3] Legitimacy refers to "the degree of consensus about what (people think) is accepted by others, based on norms or frames about what is valid and appropriate in given situations (Johnson et al., 2006). In this paper, we assume punishment based on norm violations is more legitimate than punishment independent of norm violations.

Cubitt et al. 2011; Tan and Xiao, 2012; 2013)[4]. Punishment decisions in modern societies are thus usually made and implemented by independent third parties such as the court to ensure their legitimacy.

On the other hand, the legitimacy of punishment is sensitive to the nature of the third party mechanism (e.g., Tyran and Feld, 2006; see more papers reviewed in Xiao, forthcoming). For example, when punishment involves depriving violators of resources such as money or labor, those resources can become profit for the enforcers. Profit from punishment could motivate third party enforcers to impose punishment for their own benefit rather than as a way to maintain social order. Such profit-seeking punishment is perhaps most common in corrupt societies. Xiao (2013a) shows that when the third party can profit, many of them impose punishment regardless of how the recipients behave. Consequently, people no longer perceive punishment as legitimate, and punishment fails to signal a norm violation.

Previous research on punishment, however, has not investigated how the pressure to justify imposing punishment may influence the legitimacy of punishment. The effect of justification pressure on judgment and decisions has been discussed extensively in the literature on accountability, although most of those studies are unable to identify the pure effect of justification from other confounding factors such as identifiability.[5] In his seminal work on accountability, Tetlock (1985) argues that people make decisions based on simple heuristics. When they are required to justify their actions but do not know audiences' view, they try to anticipate the objections of potential critics. Xiao (2013b) provide experimental evidence that

[4] One key feature of an independent third party is that his material payoff is independent of the decisions of the implicated stakeholders (see Leibbrandt and Lopez-Perez, 2012). Such categorization distinguishes a third party from centralized punishers randomly chosen or elected among the implicated players (e.g., see Baldassarri and Grossman, 2011).

[5] For instance, Vieider (2011, 2012) finds that when people have to explain their decisions to recipients *face-to-face*, they exert more effort or become less loss-averse. One exception is Xiao (2013b) who excludes identifiability and examines the pure effect of the external pressure from justification. The author finds that it can reduce selfish behavior, even in one-shot environments.

pure pressure of justification can enhance the norm salience by encouraging one to think about what the audience thinks. As a result, the individual becomes more sensitive to any deviation from that expectation. We thus hypothesize that enforcers are more likely to punish consistently with social norms when there is justification requirement. Requiring justification can curb corrupt punishment behavior and promote legitimate punishment, even when the decisions are anonymous and there is no material consequence for poor justification.

To test our hypothesis, we adopt an experiment based on a sender-receiver game by Gneezy (2005) widely used to study cheating behavior. We extend this game by introducing an independent third party who could punish the sender after observing whether the sender has sent a true or false message to the receiver on which of two options will earn the receiver a higher payoff.

Our experiment is built on the experiment reported in Xiao (2013a). In the baseline non-profitable punishment treatment (NPP), the punishment decision of the third party is totally independent of the decision itself – he only receives a fixed payment for the task. In the profitable punishment treatment (PP), the third party earns extra money if he punishes the sender, regardless of whether the sender has sent a false message to the receiver. Xiao (2013a) reports that people are less likely to view punishment as signaling a norm violation in PP than NPP treatment. In this paper, we introduce a Justification treatment. Compared to PP, the only difference in the Justification treatment is that the third party must explain his decision (whether to punish or not punish the sender).

We define legitimate punishment as occurring if and only if a sender violates a truth-telling norm. We find that justification increases legitimate punishment to a level similar to the NPP treatment. Moreover, compared with the PP treatment, the senders are significantly more likely

to tell the truth in the Justification treatment, and the receivers are more likely to perceive punishment as signaling a norm violation. These findings support our hypothesis and shed light on the underlying mechanisms of the role of justification on legal enforcement and policy making.

The remainder of the paper proceeds as follows. Section 2 describes the experiment design and procedure. Section 3 reports the results and Section 4 offers some concluding remarks.

2. The experiment

2.1 Design

To provide clean evidence to study how justification promotes legitimate third-party punishment decisions, we design our experiment based three one-shot sender-receiver games with the same payoff structures first used in Xiao (2013a) (see Table 1). The instructions are in Appendix A.

Participants in each game, modified based on Gneezy (2005), play one of the three roles, called sender, receiver, or third party. The receiver must choose between two options, A and B, without knowing what the payoffs will be (see Table 1 for the payoffs of each game). Before the receiver makes the choice, the sender sends one of the two messages to the receiver. The sender, who is fully informed about the monetary payoffs, must tell the receiver either "Option A earns you more money than action B" or "Option B earns you more money than action A." As in Gneezy (2005), we designed the payoffs of the two options such that one option yields a higher payoff for the sender than the other option. Receivers do not know this is the case.

The task of the third party in each game is to decide whether or not to punish the sender after observing his message. The punishment reduces sender's payoff by 50%, depending on the choice later by the receiver. Finally, the receiver makes her decision between Options A and B after observing the sender's message and the payoff-cut decisions of the third party. Subjects

were not informed of the outcome of each game during the experiment. Thus, the design did not allow for learning. At the end of the experiment, one game was randomly chosen as the payoff game and each participant will be informed of the result of that game.

In our experiment, in addition to senders' behavior, the third party's decision can be affected by other concerns such as efficiency or inequality between the sender and the receiver.[6] We designed our experiment to minimize such confounds. First, we do not inform the third party of the sender and receiver payoffs in each game. The only information he knows is whether or not the sender has sent a true message. Second, the payoff of the third party is determined by a random number from the four numbers in the payoff table with equal probabilities. Moreover, a third party learns his earnings only at the end of the experiment. Neither the sender nor the receiver knows the earnings of their matched third party throughout the experiment. All these are common knowledge to all players in the game.

Table 1: Payoffs in each game

Game	Option	Sender's payoff	Receiver's payoff
1	A	10	5
	B	0	6
2	A	4	4
	B	6	2
3	A	4	8
	B	8	4

[6] Previous research finds that decision making is affected by inequality concerns, e.g., Fehr and Fischbacher (2004), Dawes et al., (2007) and Leibbrandt and Lopez-Perez (2012).

Our experiment consists of three treatments: non-profitable punishment (NPP), profitable punishment (PP), and Justification (J). Data from the first two treatments have been discussed in Xiao (2013a). In this paper, we focus on the comparison between the Justification treatment and the other two treatments. The sequence of the experiment is exactly identical across treatments except for the decisions of third parties. In the NPP treatment, a third party makes his decision without any justification and earns a fixed amount of money independent of the decision. In the PP treatment, a third party could earn an extra 50% of his random payoff by punishing the sender. This is common knowledge for all subjects.

The Justification treatment differs from the PP treatment only in that after a third party has made the punishment decision, he has to provide an explanation for his decision. The message is revealed to the other two paired subjects only at the end of the experiment if that game is randomly selected to pay out. This means that during the experiment, a receiver could only learn the decision of the third party, but not the message. The difference in third party punishment behavior between this treatment and that in the PP treatment offers us direct evidence to the extent justification decreases illegitimate third-party punishment.

2.2 Procedure

We conducted our experiment at the Pittsburgh Experimental Economics Laboratory using z-tree (Fischbacher, 2007). A total of 318 students from Carnegie Mellon University and Pittsburgh University participate as subjects (one treatment only for each subject). We used 29 groups of three in the NPP treatment, 30 groups in the PP treatment, and 37 groups in the Justification treatment.

At the beginning of an experiment, subjects were randomly and anonymously assigned a role and the role was fixed throughout the experiment. To compare the potential different reactions of the *same* third party facing true versus false messages, we let subjects play three sender–receiver games with different payoff structures. All sessions begin with Game 1, followed by either Game 2 or Game 3 based on a random order to avoid order effects. After the end of one game, we randomly rematched the subjects within a session to minimize learning and effects from repeated interactions. At the end of the experiment, one game was randomly chosen as the payoff round. Each subject was paid privately according to the outcome in that game.

3. Results

In all sessions we ran Game 1 first. This enabled us to test whether our findings were robust to inexperienced subjects. We found that decisions in Game 1 were similar to those in Game 2 and Game 3. Hence, we report data pooling from all three games in the following analysis.

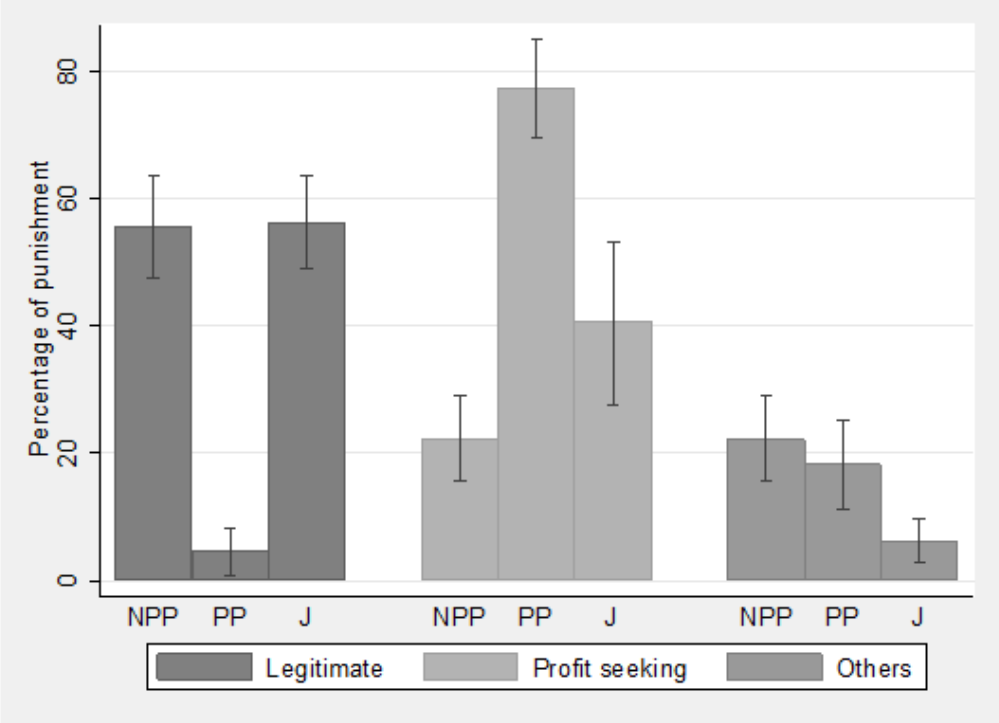
3.1 Third-party decisions

Figure 1 plots the distribution of different types of third parties in all treatments. We sort third parties into one of the following categories based on their punishment behavior in three games: (1) punish if and only if a sender sends a false message (legitimate punishment); (2) always punish regardless of the nature of a message (profit seeking punishment); and (3) exhibit punishment behavior that does not fall into either one of the aforementioned categories (others). We focus on observations from third parties who have seen both true and false messages.

As shown in Figure 1, in the NPP treatment, more than 50% (15 out of 27) of third parties are legitimate. In the PP treatment, nevertheless, only less than 5% (1 out of 22) of third parties

impose legitimate punishment, and the difference is both economically and statistically significant (a two-sided Z-test, $p<0.01$). Most third parties (17 out of 22) always punish the sender regardless of the message sent in the PP treatment, but much fewer do so in the NPP treatment (6 out of 27, a two-sided Z-test, $p<0.01$).

Figure 1. Distributions of third party punishment types by treatment



Notes: The data include only third parties who experienced both false messages and true messages. NPP is short for the Non-Profitable Punishment treatment in which third parties could not monetarily benefit from punishment. PP is short for the Profitable Punishment treatment in which that the third party could benefit from punishment. J is short for the Justification treatment, in which third parties must justify their choices to senders and receivers.

When third parties have to justify their punishment decisions, they are much less likely to seek profitable punishment, even though there are no material consequences of justification. In the Justification treatment, 56.25% (18 out of 32) of third parties punish if and only if a sender sends a false message. This proportion is much higher than in the PP treatment (a two-sided Z-test, $p < 0.01$) and is almost identical to the NPP treatment. Compared with PP treatment,

significantly fewer third parties in Justification treatment always punish regardless of whether the message is true or false (40% vs. 77.27%, a two-sided Z-test, $p<0.01$). Although more third parties in the Justification treatment always punish than in the NPP treatment (40.63% vs. 22.22%), the difference is not statistically significant (a two-sided Z-test, $p=0.12$).

An overview of the justification messages written by the third parties also suggests that the third parties indeed justify their punishment decisions based on norm violations even though all the messages and decisions are anonymous and they have the freedom to write anything they want. Appendix B lists all messages third parties sent to both senders and receivers in the Justification treatment.

In particular, when asked to explain the decision "why did not punish a sender", third parties predominantly wrote, "since he tells the truth" or "since he did not send the wrong message" (32 out of 34). Only one third party wrote "since he lied."

When asked to explain "why punished a sender," over 70% of the time (52 out of 74) third parties wrote that "the sender has lied about the message." About 10% of the answers were the opposite, i.e., the sender was punished because he told the truth (8 out of 74). Third parties explicitly mention that they could profit from punishment only about 5% of the time (5 out of 74). These results suggest that, even though the messages are anonymous, third parties seem to be averse to providing reasons that others might disapprove.

3.2 Sender Decisions

Senders' choice of whether to lie can show us their perception of the legitimacy of punishment. Sutter (2009) argues that the definition of deception should depend on the intention to deceive, not message sent (i.e., whether the message is the same suggested by a computer). We adopt this

and define deception in this paper as any decision that the sender expects to lead the receiver to choose the low payoff option. At the end of the experiment, after everyone has decided their actions, we elicit senders' beliefs regarding whether the receivers will follow their messages. Thus, lies in our experiment include the cases where a sender sends a false message and expects the receiver to follow the message (F_F) and the cases where a sender sends a true message and expect the receiver not to follow the message (T_NF).

If senders expect punishment to be legitimate and believe it is imposed to enforce the norm of honesty, senders will send true messages and expect receivers to follow them. On the other hand, if senders expect that third party punishment is driven by profit, they will send a false message if they believe the receiver will follow the message, or a true message if they believe the receiver will not follow the message (see Xiao (2013a) for theoretical analysis in details).

We report the descriptive data of sender's decisions and beliefs in Table 3. The right panel reports the messages by the senders conditional on their beliefs. The left part of Table 3 describes the aggregate frequency of lies, which is the sum of F_F and T_NF messages from the right panel. Table 3 shows the frequency of lying is approximately 20% higher in the PP than in the NPP treatment (55.56% vs. 34.48%). In the Justification treatment, however, the frequency of lying drops to almost the same level as in the NPP treatment (34.23% vs. 34.48%). Relative to the PP treatment, the frequency of lies is significantly lower in the Justification treatment (55.56% vs. 34.23%).

To provide statistic evidences for the treatment effects, we calculate the percentage of lies (T_NF or F_F) for each sender. We find that the justification effect is statistically significant. In particular, senders lie significantly less in the Justification treatment than in the PP treatment (38%

vs. 50%, a two-sided Mann-Whitney test, $p < 0.05$), and similarly to the NPP treatment (38% vs. 35.63%, a two-sided Mann-Whitney test, $p = 0.86$).

Table 3. Senders' decisions and beliefs by treatment

Treatment (# of obs.)	Lie		Sender's decision_ belief		
	Freq.	%		Freq.	%
Non-profitable pun (NPP) (87)	30	34.48	T_NF	10	11.49
			F_F	20	22.99
			T_F	31	35.63
			F_NF	26	29.89
Profitable pun (PP) (90)	50	55.56	T_NF	36	40.00
			F_F	14	15.56
			T_F	18	20.00
			F_NF	22	24.44
Justification (J) (111)	38	34.24	T_NF	19	17.12
			F_F	19	17.12
			T_F	39	35.14
			F_NF	34	30.63

Notes: T_NF: a sender sends a true message and expects the receiver not to follow;
F_F: a sender sends a false message and expects the receiver to follow;
T_F: a sender sends a true message and expects the receiver to follow;
F_NF: a sender sends a false message and expects the receiver not to follow.

The results suggest that, senders are more likely to expect punishment to be legitimate and less likely to lie when they know that the enforcers are required to provide explanations for their decisions.

3.3 Receiver decisions

We summarize receivers' decisions in Table 4. The descriptive data show that, if the senders are not punished by the third party, the receivers predominantly follow the senders' choices across all three treatments. We next focus on treatment effects in cases where punishment is imposed.

We can infer receivers' perception of the punishment legitimacy from their decisions on whether to follow sender messages when the third party punishes the sender. If receivers expect punishment to be legitimate, they should not follow the messages when there is a punishment. On the other hand, if they think the punishment decisions of the third parties are profit-oriented and hence norm-irrelevant, they will decide whether to follow the message based on their initial belief about whether senders will send a true or false message. Thus, the difference in the rate of message following when the receiver observes a punishment reflects the difference in the receiver's perception of the legitimacy of punishment. The more likely a receiver views punishment as legitimate, the less likely she will follow sender's message when it is punished (see details in Xiao, 2013a).

In 85.71% (42 out of 49) cases of the NPP treatment, receivers do not follow messages if the sender is punished. By contrast, in the PP treatment, if they observe the third party punishing the sender, then receivers make either decision with equal probability (41 decisions follow the message, and 40 decisions do not). The Justification requirement recovers the receivers' perception of the punishment's legitimacy. When the third party punishes the sender in the Justification treatment, about 72.60% (53 out of 73 cases) of times receivers do not follow the message of the senders, which is a sharp decrease of nearly 25% compared to that in the PP treatment.

To compare these percentages statistically, we average out the percentage of decisions across three games for each receiver upon receiving a punishment message. We find receivers in the Justification treatment are less likely to follow senders' advice than those in the PP treatment (21% vs. 35%, a two-sided Mann–Whitney ranksum test, $p < 0.05$) when they know that the sender is punished. Although the percentage of message following in the Justification treatment is still higher than that in the NPP treatment, the difference is no longer statistically significant (21% vs. 14.04%, a two-sided Mann-Whitney ranksum test, $p=0.11$).

Table 4. Receivers' decisions by treatment

Treatment	Sender	Follow	Not follow
Non-profitable pun (NPP) (# of obs=87)	Punished (49)	7	42
	Not punished (38)	36	2
Profitable pun (PP) (# of obs=90)	Punished (81)	41	40
	Not punished (9)	7	2
Justification (J) (# of obs=111)	Punished (73)	20	53
	Not punished (38)	36	2

4. Concluding remarks

We study the pure effect of justification on third party punishment decisions using a controlled laboratory sender-receiver game experiment. The results support the hypothesis that, when punishment is profitable for the enforcers, requiring justification can balance the scale by promoting legitimate punishment, even if there are no reputation or material consequences for poor justification. This level of balance is comparable to the case where a third party does not benefit from punishment. As a result, profitable punishment is more effective in promoting honest behavior when justification is required.

The fact that justification promotes legitimate punishment even in the presence of monetary temptation can be applied to policy makers, who must explain or justify policy decisions. Traditionally, policy explanations are merely perceived as information displayed for the public's benefit. However, the results from our experiment indicate that providing explanations can also change the behavior of decision makers in a direction more consistent with the perceived social norm. Hence, an obligation to constantly provide justifications to the public could be a low-cost way to combat corruption, as a supplement to conventional mechanisms such as punishment.

Our results also suggest the importance of designing appropriate compensation packages for the law enforcers. Our experiment data clearly support the claim that how law enforcers are paid strongly influences the public's perception of the legitimacy of punishment. As discussed in Xiao (forthcoming), the choice between a guaranteed, fixed payment and performance-based payment has a tricky trade-off. On the one hand, some scholars have suggested that paying a high salary reduces the temptation for law enforcers to earn extra money by exerting their power (Abbink, 2006). On the other hand, a fixed, guaranteed payment decreases the incentives to convict guilty defendants. Performance-based payment can incentivize enforcers to work harder to catch violators. However, it also might distort the enforcer's incentive to convict innocent defendants to pursue profit, which could deteriorate the public's perception of the legal system. This challenge opens avenues for future research, such as comparing the costs and benefits of different compensation packages, and evaluating the effect of introducing non-monetary incentives. This paper shows that the requirement of justification can be an effective solution to build into compensation plans to overcome the dilemma of incentives and promote the legitimacy of punishment.

References:

Abbink, K. (2006), Laboratory experiments on corruption, in *International Handbook on the Economics of Corruption*, ed. S. Rose-Ackerman, Cheltenham: Elgar: 418 – 437.

Baldassarri, D. and G. Grossman (2011), Centralized sanctioning and legitimate authority promote cooperation in humans, *Proceedings of the National Academy of Sciences* 108 (27): 11023 – 11027.

Cubitt, R. P., M. Drouvelis, S. Gächter, and R. Kabalin (2011), Moral judgments in social dilemmas: How bad is free riding? *Journal of Public Economics*, 95, 253-264.

Dawes, C. T., J. H., Fowler, T. Johnson, R. McElreath, and O. Smirnov (2007), Egalitarian motives in humans. *Nature* 446, 794–796.

Fehr, E. and U. Fischbacher (2004), Third-party punishment and social norms, *Evolution and Human Behavior* 25(2), 63-87.

Fischbacher, U. (2007), z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics*, 10(2), 171–178.

Gneezy, U. (2005), Deception: the role of consequences. *American Economic Review*, 95, 384 – 394.

Johnson, C., T. Dowd and C. Ridgeway (2006), Legitimacy as a social process, *Annual Review of Sociology*, 32(1), 53-78.

Leibbrandt, A. and R. López-Pérez (2012), An exploration of third and second party punishment in ten simple games, *Journal of Economic Behavior and Organization*, 84, 753– 766.

Lind, A. E. and T. Tyler (1988), *The Social Psychology of Procedural Justice*. By New York: Plenum Press.

Sutter, M. (2009), Deception through telling the truth?! Experimental evidence from individuals and teams, *Economic Journal*, 119, 47-60.

Tan, F. and E. Xiao (2012), Peer punishment with third-party approval in a social dilemma game, *Economics Letters*, 117, 589-591.

Tan, F. and E. Xiao (2013), Third-party punishment: retribution or deterrence? Working paper of the Max Planck Institute for Tax Law and Public Finance.

Tetlock, P. E. (1985), Accountability: A social check on the fundamental attribution error, *Social Psychology Quarterly*, 48(3), 227-236.

Tyler, T. (2006), *Why People Obey the Law*, Princeton: Princeton University Press.

Tyran, J.R. and L. Feld. (2006), Achieving compliance when legal sanctions are non-deterrent, *Scandinavian Journal of Economics*, 108 (1), 135–156.

Vieider, F. M. (2011), Separating real incentives and accountability, *Experimental Economics* 14(4), 507–518.

Vieider, F. M. (2012), Moderate stake variations for risk and uncertainty, gains and losses: methodological implications for comparative studies, *Economics Letters* 117, 718-721.

Xiao, E. (forthcoming). Punishment, social norms and cooperation, Research Handbook on Behavioral Law and Economics, ed. Joshua C. Teitelbaum and Kathryn Zeiler, Edward Elgar.

Xiao, E. (2013a), Profit seeking punishment corrupts norm obedience, *Games and Economic Behavior*, 77, 321 – 344.

Xiao, E. (2013b), Justification and Cooperation, working paper, Carnegie Mellon University

Appendix A

Instructions (Person 1)

General Information

Thank you for coming! You've earned $5 for showing up on time, and the instructions explain how you can make decisions and earn more money. So please read these instructions carefully! There should be no talking at any time during this experiment. If you have a question, please raise your hand, and an experimenter will assist you.

Each participant is in the role of either Person 1, or Person 2, or Person 3. You are in the role of Person 1.

This session consists of **three** rounds. At the beginning of each round, the computer will randomly group one Person 1 with one Person 2 and one Person 3. Thus, your counterpart in each round will change randomly throughout the experiment. No one will ever be informed of the identity of the two counterparts.

Below are the decision tasks in each round

In each round, two possible monetary payments are available to Person 1 and Person 2 in the experiment.
The two payment options are:

Option A: $*W* to Person 1 and $*X* to Person 2;
Option B: $*Y* to Person 1 and $*Z* to Person 2

The payoff structure of Option A and Option B will be different in each round. Only Person 1 will know the exact values of *W, X, Y*, and *Z* in each round. Neither Person 2 nor Person 3 will know those values in any round. The computer will randomly assign *W,X, Y, or Z , with equal chance*, as Person 3's payoff.

- **Person 2's decision:** In each round, **Person 2** will decide to choose either Option A or Option B and thus decide the payoffs of Person 1 and Person 2.

- **Person 1's decision:** In each round, **Person 1** needs to decide which one of the following messages to send to Person 2 before Person 2 decides which option to choose.

  Message A: "Option A will earn you more money than option B.";

  Message B: "Option B will earn you more money than option A."

- **Person 3's decision: Person 3** will NOT know the exact values of *W, X, Y*, and Z **but** will know whether the message sent by the matched Person 1 in each of the three rounds is true. (**Note:** Since the payoff structure of Option A and Option B changes from one round to another, which message is true in each round will also change accordingly.) After Person 1 decides which message to send to Person 2, Person 3 decides whether to assign a payoff-cut to Person 1.

(the Profitable Punishment and the Justification treatments)

If Person 3 assigns the payoff-cut, Person 1's payoff (decided by the option Person 2 chooses) is cut by 50% and Person 3's payoff is increased by 50%.

If Person 3 does not assign the payoff-cut, Person 1's payoff is not reduced by any amount and Person 3's payoff is not increased by any amount. Person 2's payoff will not change no matter what decision Person 3 makes. There is no cost for Person 3 to assign the payoff cut. Person 3 will make his/her decision prior to knowing his/her randomly assigned payoff.

(the justification treatment)

After Person 3 has made the payoff cut decision, he/she must write a message to explain why he/she decided to do so. The explanation must be related to Person 1's behavior. Any message written in a round will be sent to and reviewed by Person 1 and Person 2 at the end of the experiment only if that round is randomly selected as the payoff round (details below).

Important: *Person 3 will forfeit all earnings from the experiment and receive only the show up bonus if he/she did not write a message to explain his/her decision or the explanation is not written in the required format as explained below. Person 3 should not identify him or herself by name or ID number or gender or appearance. Violations will also result in Person 3 receiving only the $5 show-up bonus. In particular:*

- *If Person 3 assigned the payoff cut, the message should be written in the following format:*

"Person 1 should be assigned the payoff cut because Person 1 …"

- *If Person 3 did not assign the payoff cut, the message should be written in the following format:*

"Person 1 should not be assigned the payoff cut because Person 1…"

After Person 1's and Person 3's decisions, Person 2 sees the decisions of both Person 1 and Person 3, and decides whether to choose Option A or Option B. (Note: neither Person 2 nor Person 1 will see Person 3's message at this point).

For example:

Suppose, in one period, Person 1 sent a message and Person 3 decided to impose the payoff-cut. Person 2 then decided to choose Option B. The random payoff assigned to Person 3 turns out to be $X.

Person 1's payoff in that period = Y-0.5*Y

Person 2's payoff in that period= Z

Person 3's payoff in that period =X+0.5*X

Suppose, in one period, Person 1 sent a message and Person 3 decided NOT to impose the payoff-cut. Person 2 then decided to choose Option B. The random payoff assigned to Person 3 turns out to be$X.

Person 1's payoff in that period = Y

Person 2's payoff in that period = Z

Person 3's payoff in that period =X

After Person 2's decision, a new round starts. Each participant will be randomly paired with another two participants. Each round will proceed in the same way.

<u>You will not know the result of each round during the experiment.</u> In the end of the experiment, one round will be randomly chosen to be your payoff round. Every participant will be informed

of the result of that round. Both Person 1 and Person 2 will also see Person 3's message that explains his/her decision. Each participant will be paid accordingly.

To repeat the key parts of this experiment, Person 1 and Person 2 will earn the amounts specified in the option chosen by Person 2. However, Person 2 will never know what amounts were actually offered in the option not chosen (that is, he or she will never know whether Person 1's message was true or not). Moreover, Person 2 will never know the amounts to be paid to Person 1 according to the different options. Also, Person 2 and Person 1 will never know what payoff Person 3 randomly received. Person 1's earnings will also be affected by Person 3's payoff-cut decision. Person 3's earning will be increased if he/she assigns the payoff-cut to Person 1. Person 3 must write a message to explain his/her decision and the explanation must be related to Person 1's behavior. The message will be viewed by Person 1 and Person 2 at the end of the experiment only if that round is randomly chosen to be the payoff round. Person 3 will lose all earnings from the experiment and receive only the show up bonus if he/she did not write a message to explain her/his decision or if the explanation is not written in the required format.

Your ID_____

The next several pages outline the procedure of the experiment and the computer screens when Person 1, Person 2 and Person 3 make their decisions.

Appendix B

We asked the subjects in the role of third parties write a message to explain their decisions. We pool all the answers (three by each third party) from the three games and put them into three categories: reasons about why they punish the sender, why they do not punish and other unidentifiable reasons. Despite that we discourage empty messages, there are three occasions that subjects did not provide any explanations. Therefore, the total number of messages is 108.

Reasons why the sender should be punished (74):
Reason 1: Person 1 lied to person 2 (52):

- Person 1 should be assigned the payoff cut because they sent the Message A which is not true.
- Person 1 should be assigned the payoff cut because Person 1 lied about which message was true.
- Person 1 should receive the payoff cut because Person 1 was not honest in their answer.
- Person 1 should be assigned the payoff cut because person 1 chose the not true option and that should disqualify them from earning more than person 2.
- Person 1 should be assigned the payoff cut because Person 1 did not act truthfully. Person 1 gave Person 2 incorrect information.
- Person 1 should receive the payoff cut because Person 1 was not honest in their answer.
- Person 1 should be assigned the pay cut because person 1 chose the not true option and that detract from the overall group's earnings.
- Person 1 should not be assigned payoff cut because person 1 has sent message A which is not true.
- Person 1 should be assigned the payoff cut because Person 1 sent an untrue message.
- Person 1 should be assigned the payoff cut because person 1 sent a message which is untrue.
- Person 1 should be assigned the payoff cut because Person 1 sent a message to Person 2 that was not true.
- Person 1 should be assigned the payoff cut because Person 1 chose the message that is not true even though person 1 knew the message was not true.
- Person 1 should be assigned the payoff cut because Person 1 sent a false message to Person 2. This is deceitful, and there are negative consequences for such actions.
- Person 1 should be assigned the payoff cut because Person 1 sent an untrue message.
- Person 1 should be assigned the payoff cut because Person 1 sent an untrue message.
- Person 1 should be assigned the payoff cut because person 1 could be lying to person 2 in order to get more money for themselves.
- Person 1 should be assigned the payoff cut because person 1 lied.
- Person1 should be assigned the payoff cut because Person 1 did not tell the truth.
- Person 1 should be assigned the payoff cut because Person 1 sent out the untrue message.
- Person 1 should be assigned the payoff cut because Person 1, while presumably acting in his/her own self-interest, has given person 2 false information.
- Person 1 should be assigned the payoff cut because Person 1 sent Person 2 an untrue message that cause Person 2 to believe they would receive more money when they

wouldn't.

- Person one should be assigned the payoff cut because Person 1 sent an untrue message to Person 2 that would cause person two to believe they would receive more money when they would not.
- Person 1 should be assigned the payoff cut because Person 1 has given Person 2 false information.
- Person 1 should be assigned the payoff cut because Person 1 lied about which message was true.
- Person 1 should be assigned the payoff cut because Person 1 sent Person 2 a message that was untrue which may cause person 2 to believe they would receive more money when they wouldn't.
- Person 1 should be assigned the payoff cut because Person 1 sent a false message to Person 2.
- Person 1 should be assigned the payoff cut because Person 1 didn't tell the truth.
- Person 1 should be assigned the payoff-cut because they considered their own fiscal well-being over Person 2 by lying about the better option, so I considered my own gain over theirs.
- Person 1 should be assigned the payoff cut because Person 1 sent Message A which is not true.
- Person 1 should be assigned the payoff cut because Person 1 sent a message to Person 2 that was not true instead of sending the message that was true.
- Person 1 should be assigned the payoff cut because Person 1 lied.
- Person 1 should be assigned the payoff cut because Person 1 lied to Person 2.
- Person 1 should be assigned the payoff cut because Person 1 lied to Person 2.
- Person 1 should be assigned the payoff cut because Person 1 sent Message A to Person 2, which was a false statement.
- Person 1 should be assigned the payoff cut because person 1 was not truthful.
- Person 1 should be assigned the payoff cut because Person 1 lied.
- Person 1 should be assigned the payoff cut because Person 1 lied to Person 2.
- Person 1 should be assigned the payoff cut because Person 1 did not send the true message.
- Person 1 should be assigned the payoff cut because Person one was dishonest.
- Person 1 should be assigned the payoff cut because Person 1 did not send a true message.
- Person 1 should be assigned the payoff cut because Person 1 lied to Person 2.
- Person 1 should be assigned the payoff cut because Person 1 did not send a true message.
- Person 1 should be assigned the payoff cut because Person 1 was dishonest.
- Person 1 should be assigned the payoff cut because Person 1 did not send a true message.
- Person 1 should be assigned the payoff cut because Person 1 was not truthful.
- Person 1 should be assigned the payoff cut because Person 1 sent Message A to Person 2, which was a false statement.
- Person 1 should be assigned the payoff cut because person 1 lied to person 2 about which option was better.

- Person 1 should be assigned the payoff cut because person 1 lied.
- Person 1 should be assigned the payoff cut because Person 1 might not be telling the truth.
- Person 1 should be assigned the payoff cut because Person 1 lied to Person 2 in order to earn more money for himself.
- Person 1 should be assigned the payoff cut because Person 1 sent Message A to Person 2, which was a false statement.
- Person 1 should be assigned the payoff cut because person 1 lied.

Reason 2: The sender sent a true message (8):

- Person 1 should be assigned the payoff cut because Person 1 choose message A and it was true.
- Person 1 should be assigned the payoff cut because Person 1 sent a true message to Person 2 when a false message could have benefitted Person 1 more than a true message.
- Person 1 should be assigned the payoff cut because Person 1 choose message A which is true.
- Person 1 should be assigned the payoff cut because Person 1 sent a true message to Person 2 when a false message could have benefitted Person 1 more than a true message.
- Person 1 should be assigned the payoff-cut because Person 1 probably chose to send the message that would result in them making more money. Because of this, it is okay for them to receive a payoff-cut since he/she will already make enough money compared to Person 2, and perhaps Person 3, as well.
- Person 1 should be assigned the payoff cut because person 1 was honest.
- Person 1 should be assigned the payoff cut because Person 1 sent the correct statement, which he probably believes that Person 2 will perceive as false and hence earn himself a higher pay in the end.
- Person 1 should be assigned the payoff cut because Person 1 sent a true message to Person 2, even though he probably thinks that Person 2 will not believe him. Therefore, by sending the true message, Person 1 believes that Person 2 will choose the opposite message and hence earn Person 1 more money.

Reason 3: The third party could earn more (5):

- Person 1 should be assigned the payoff cut because person 1 chooses message b where I can be assigned a higher payoff.
- Person 3 has the opportunity to gain an additional 50% on their earnings, making it the logical option for person 3 to choose.
- Person 1 should be assigned payoff cut because person 1's payoff will be decreased by 50% which make mine's payoff increased by 50%.
- Person 1 should be assigned the payoff cut because Person 1 has shown that he/she does not mind acting in the best interest of others; thus he/she will not object to Person 3's best interest.
- Person 1 should be assigned a payoff cut because Person 1 will give Person 3 more

money this way.

Reason 4: Others/unidentifiable reasons (9):

- Person 1 should be assigned the payoff cut because person 1 stands to lose 50% of their earnings.
- Person 1 should receive the payoff cut because Person 1 was not acting in their best economic interest.
- Person 1 should be assigned the payoff cut because Person 1 chose an option that lost themselves money.
- Person 1 should be assigned the payoff cut because they sent Message A.
- Person 1 should be assigned the payoff cut because Person 1 sent Message A.
- Person 1 should be assigned the payoff cut because Person 1 was not crafty enough with his/her decision.
- Person 1 should be assigned the payoff cut because Person 1 sent Message A to Person 2.
- Person 1 should be assigned the payoff cut because Person 1 should make the least amount of money.
- Person 1 should be assigned the payoff cut because Person 1 sent Message A.

Explanations why the sender should not be punished (34):

Reason 1: The sender sent a truthful message (32):

- Person 1 should not be assigned the payoff cut because Person 1 sent a truthful message to Person 2, in order for Person 2 to base a decision on.
- Person 1 should not be assigned the payoff cut because Person 1 sent a truthful message to Person 2.
- Person 1 should not be assigned the payoff cut because Person 1 sent Message A which was true.
- Person 1 should not be assigned the payoff cut because Person 1 did not lie about which option is greater.
- Person 1 should not be assigned the payoff cut because Person 1 sent Message B which was true.
- Person 1 should not be assigned the payoff cut because Person 1 sent the message that was true.
- Person 1 should not be assigned the payoff cut because person 1 decided to send the message which is true.
- Person 1 should not be assigned the payoff cut because Person 1 was honest and sent the truthful message to Person 2. Therefore, he/she should be assigned the amount they deserve, without a payoff-cut.
- Person 1 should not be assigned the payoff cut because Person 1 sent the true message.
- Person 1 should not be assigned the payoff cut because Person 1 sent the truthful message.
- Person 1 should not be assigned the payoff cut because Person 1 did NOT lie.
- Person 1 should not be assigned the payoff cut because Person 1 sent out the true

message.

- Person 1 should not be assigned the payoff cut because Person 1 did NOT lie.
- Person 1 should not be assigned the payoff cut because Person 1 sent an accurate and truthful message to Person 2
- Person 1 should not be assigned the payoff cut because Person 1 told the truth.
- Person 1 should not be assigned the payoff cut because Person 1 told the truth.
- Person 1 should not be assigned the payoff because Person 1 told the truth to person 2
- Person 1 should not be assigned the payoff cut because Person 1 told the truth.
- Person 1 should not be assigned the payoff cut because Person 1 sent the true message.
- Person 1 should not be assigned the payoff cut because he/she is correct
- Person 1 should not be assigned the payoff cut because Person 1 should be telling the truth about which option to pick.
- Person 1 should not be assigned the payoff cut because Person 1 is correct.
- Person 1 should not be assigned the pay-off cut, because they did what was best for everyone in the group and didn't lie about which option was better.
- Person 1 will not be assigned the payoff cut because Person 1 is correct.
- Person 1 should not be assigned the payoff cut because Person 1 told the truth.
- Person 1 should not be assigned the payoff cut because Person 1 told the truth.
- Person 1 should not be assigned the payoff cut because Person 1 sent the true message.
- Person 1 should not be assigned the payoff cut because Person 1 sent the true message to person 2 and in fact told person 2 which one would make him more money.
- Person 1 should not be assigned the payoff cut because Person 1 told the truth.
- Person 1 should not be assigned the payoff cut because Person 1 told the truth.
- Person 1 should not be assigned the payoff cut because Person 1 didn't lie.
- Person 1 should not be assigned the payoff cut because Person 1 send a message to Person 2 that was true and that would result in Person 2 making more money.

Reason 2: The sender has told the truth (1)

- Person 1 should not be assigned the payoff cut because Person 1 chose Message A which is untrue.

Reason 3: Unclear of the game (1)

- I did not assign the payoff cut because I don’t even get what’s going on in this experiment.