Erte Xiao

Carnegie Mellon University

Profit-Seeking Punishment Corrupts Norm Obedience

The full version of this paper can be found here

This paper is part of the Decision Making for a Social World webconference

Organized by the International Cognition and Culture Institute and
the Philosophy, Politics and Economics Program at the University of Pennsylvania

The paper is followed by a **discussion**

I. Introduction

Societies with corrupt law enforcement often display pervasive norm violations. For example, Transparency International Global Corruption Barometer 2007 reports that "… *while more than four in five respondents in Cameroon, Ghana, India and Nigeria consider the police to be corrupt, fewer than two in five in Denmark, Germany, Finland, Singapore, Sweden and Switzerland report the same…"* In this paper, we provide a new perspective on the *causal* relationship between legal institutions that embed corrupting temptations (e.g., profitable punishment) and prevalent norm disobedience within the societies such institutions govern. Previous research has focused on the *failure* of corrupt law enforcers to punish norm violators (see, Abbink, 2006). In contrast, we offer both theory and evidence from a laboratory experiment that even if punishment is prevalent, if people know enforcers can earn pecuniary benefits by punishing then people no longer view punishment as signaling a norm violation. The result is that punishment loses its ability to influence behavior.

Punishing someone often means depriving that person of some resources he/she owns (e.g., money or labor). An enforcer can profit by capturing these resources, and this may motivate an enforcement authority to increase the number or severity of punishments. For example, Ukraine traffic police were known to be highly corrupt, in that they imposed on-the-spot fines and frequently demanded bribes. Indeed, so pervasive was this corruption that in 2005, the traffic police were disbanded by Ukrainian President Viktor Yushchenko[1]. Red light cameras can also be used for generating revenue through punishment. For example, Motorists.org reports that in 2008 six cities in the US shortened yellow light times for profit[2].

In principle, profitable punishment could motivate prosecutors to put more effort into catching offenders. Likewise, increasing the severity of punishment could reduce the frequency of violations. On the other hand, profit-seeking punishment may lead to corruption in the sense that punishers may choose to further their own revenue interests rather than pursuing the goal of norm conformity. This means that punishment no longer communicates that a norm has been violated.

If profit-seeking punishment has a detrimental effect on punishment's ability to express norms, this might help to explain why, in generally corrupt societies, norm violation is prevalent. One reason as we suggest here is that in corrupt societies legal authorities like the judiciary and the police are often considered the most corrupt officials, despite the fact that people rely on them to curb corruption (e.g., Hunt, 2006). Indeed, Transparency International (2007) reports survey data from 63,199 people in 60 countries on their perceived level of corruption in key institutions. About half of the respondents report that the police, legal system/judiciary are extremely corrupt, ranking these among the top five most corrupt institutions.

One direct consequence of a corrupt legal system is that more people will violate regulations if they believe they can bribe the authorities in order to avoid punishment. On the other hand, as we point out in this paper, if people know that punishers can benefit from imposing punishment, then people may no longer view punishment as expressing disapproval

[1] http://news.bbc.co.uk/2/hi/europe/4694199.stm. Thanks to Klaus Abbink for drawing my attention to this example.

[2] See http://www.motorists.org/blog/6-cities-that-were-caught-shortening-yellow-light-times-for-profit/ for details.

for norm violations. Previous studies have suggested that punishment promotes compliance not only by changing incentives but by communicating social norms (see, e.g., Kahan, 1998, Cooter, 1998; Sunstein, 1996; Tyran and Feld, 2006; Masclet et al. 2003; Galbiati and Vertova, 2008; Xiao and Houser, forthcoming). We hypothesize that profitable punishment leads to corruption in punishers and as a result, neither those who observe the punishment nor the punishment recipients view punishment as expressing social norms and punishment becomes less effective in preventing norm violations.

To test the hypothesis, we design a lab experiment based on a sender-receiver game used to study cheating behavior (see, Gneezy, 2005). We revise the sender-receiver game so that it can test the norm expression function of punishment. Our result provides direct evidence that profitable punishment has a detrimental effect on the norm communication function of punishment. Moreover, we find that a sender's attempt to lie is constrained by the threat of punishment when punishment is not profitable for the enforcer, but not when enforcers can profit from the punishment.

II. Experiment Design

Our experiment is based on a sender-receiver game. Gneezy (2005) was the first to design experiments using this game to study the nature of people's aversions to lying in a cheap talk environment. Our experiment consists of three treatments: non-profitable punishment treatment (NPP), profitable punishment treatment (PP), and no-punishment treatment (NP).

II. A. Non-profitable punishment treatment (NPP)

In this treatment three subjects form a group. One person acts as a sender, one person acts as a receiver, and one person acts as an enforcer. The receiver chooses between two options: A and B. Each option specifies the payoffs to the sender and receiver if that option is chosen. Only the sender knows the payoffs to each option. The enforcer only knows whether the sent message is true.

Before the receiver chooses an option, the sender sends her one of the following messages:

Message A: Option A earns you more money than Option B.

Message B: Option B earns you more money than Option A.

The enforcer sees the sent message and decides whether to impose a payoff-cut to the sender. If a payoff-cut is imposed, the sender's payoff, which is decided later by the option chosen by the receiver, is cut by 50%. The receiver sees both the message and the enforcer's payoff-cut decision. Then she chooses an option and thus decides the payoffs for both the sender and receiver.

It is well-established that decisions can be influenced by relative payoff comparisons (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Dawes et al. 2007). We designed our experiment to minimize such confounds. In particular, enforcers are blind to each option's payoff. One of the four payoffs in Option A and Option B is randomly selected as the enforcer's payoff. Thus, the enforcer's payoff is independent of the payoff of either the sender or the receiver. Moreover, the enforcer does not know her randomly assigned payoff until the end of the experiment. That is, the enforcer does not know her payoff when making the payoff-cut

decision in each round. Neither the sender nor the receiver knows the enforcer's payoff. Likewise, the enforcer is kept ignorant of the sender's and receiver's payoffs as well. All these conditions are common knowledge.

We had subjects play three sender-receiver games with different payoff structures. Subjects were not informed of the outcome of each game until the end of the experiment. Thus, the design did not allow learning. In each game, each participant was randomly and anonymously paired with two other participants. The payoffs of the options in each game are listed in Table 1. The payoffs in each game imply different incentives to lie. For example, in Game 1, the sender can earn $10 more, and the receiver earns only $1 less, if the sender succeeds in deceiving the receiver so that the receiver chooses option A. In contrast, in Game 2, successful deception earns the sender only $2 more, and also leaves the receiver earning $2 less.

Table 1. The 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

We define deception based on Sutter (2009): any decision that the sender expects to lead the receiver to choose the low-payoff option. Since we define deception in terms of the intention to lie, it is necessary to obtain data on senders' beliefs regarding whether receivers follow their messages. For each game, we elicited sender's beliefs regarding: 1) which option receivers chose and; 2) whether the enforcers imposed the punishment. We did this after all three games concluded and before anyone was informed of the outcome of any game. One question was randomly chosen and subjects earned an additional $2 for giving the correct answer. All survey questions are in Appendix B.

II. B. Profitable punishment treatment (PP)

This treatment is exactly the same as the Non-profitable punishment treatment except that the enforcer's randomly assigned payoff is increased by 50% if and only if the enforcer imposes a 50% payoff cut to the sender. Both the sender and the receiver are aware of this. Again, the enforcer does not know her own payoff, or the payoffs of the sender or receiver when making the payoff cut decision. This feature of the experiment minimizes the possibility that the enforcer's decision might be affected by concerns related to efficiency or inequality.

II. C. No-punishment treatment. (NP)

The setup of the treatment is identical to the Non-profitable punishment treatment with the exception that the "enforcer" does not have the option to punish. The senders' or receivers' decisions might be affected by the knowledge that someone can observe the sender's decisions.

In view of this, and to minimize all other possible treatment differences other than punishment opportunities, we keep this treatment the same as the previous one in that an enforcer remains involved in the game but she/he can only observe the sender's behavior. As in the other two treatments, the payoff of the enforcers is also randomly assigned. All of these are common knowledge among the game's players. This treatment provides information regarding how senders and receivers behave when there is no punishment available to communicate norms.

II. D. Procedures

The experiment was conducted at P.E.E.L lab using z-tree (Fischbacher, 2007). Subjects were randomly and anonymously assigned a role and the role was fixed in all the three rounds. One round was randomly chosen as the payoff round. Each subject was paid according to the outcome in that round. Subjects were paid privately. Each subject participated in exactly one treatment.

III. Hypothesis

The reason we choose a sender-receiver game is that the norm of truth-telling is clear in this setting. More importantly, a receiver's desire to follow the true message allows us to draw a clean inference regarding the receiver's perception of punishment from his/her behavior: whether to follow the sender's message if the sender is punished. As we discuss below, when we introduce third-party punishment in this game, norm-communicating punishment and norm-irrelevant punishment lead to starkly different predictions on the behavior of both senders and receivers. Consequently, our data provide direct evidence on how incentives for profit-seeking punishment affect norm communication.

We assume it is common knowledge that the enforcer's utility function consists of two parts: 1) the monetary payoff from the implementation of the duty; and 2) the disutility of how poorly she fulfills the duty (e.g. punish the honest message in this case). We show in Appendix C, when punishment is non-profitable, the optimal strategy for the enforcer is to punish the sender if and only if the sender sends a false message. We define norm-communicating punishment as punishment that credibly and perfectly signals that a norm violation occurred. Thus, punishment is norm-communicating when it does not include a pecuniary benefit.

When punishment is profitable, the enforcer will punish a truthful sender if the pecuniary value of doing so is large enough to overcome the dissatisfaction of performing her duty poorly. When a sender sends a false message, punishment increases the monetary payoff and also satisfies the enforcer's duty. Thus, the enforcer is always better off punishing the sender. We define "norm-irrelevant" punishment as punishment that does not convey any information regarding whether a norm violation occurred. Thus, when punishment is profitable and senders and receivers believe the pecuniary value of punishing an honest person is large enough to overcome the dissatisfaction of performing her duty poorly, punishment is norm-irrelevant.

By comparing the choices of senders and receivers among our three treatments, we are able to draw inferences regarding people's perceptions of the norm-communicating function of punishment in the two punishment environments. (see Appendix C. for details).

Hypothesis1: Enforcer's decisions.
Our first hypothesis is that there is more (less) norm-communicating (norm-irrelevant) punishment in NPP treatment than in the PP treatment.
$freq^{PP}$*(norm-communicating punishment)* < $freq^{NPP}$*(norm-communicating punishment)*
$freq^{NPP}$*(norm-irrelevant punishment)* < $freq^{PP}$*(norm-irrelevant punishment)*

Hypothesis2: Receiver's decisions.
Our second hypothesis is that receivers are less likely to interpret punishment as expressing norms in the PP treatment than in the NPP treatment. We can test this hypothesis by comparing, between treatments, the frequency with which receivers follow senders' messages.

$freq^{NPP}$*(follow message |sender is punished)* < $freq^{PP}$*(follow message |sender is punished)* ≤ $freq^{NP}$*(follow message)*

Hypothesis3: Sender's decisions.
Our third hypothesis is that senders in the PP treatment are less likely to expect punishment to communicate norm violations than those in the NPP treatment. We denote that the sender believes with probability l_s that the receiver will believe the message is true and will follow the message. Since a sender's decision depends on the value of l_s (for details see Appendix C), we divide the sender's decisions as follows[3]: 1) T_F : send a true message and expect the receiver to follow (l_s *>1/2*); 2) T_NF: send a true message and expect the receiver not to follow (l_s *<1/2*); 3) F_F send a false message and expect the receiver to follow (l_s *>1/2*); and 4) F_NF: send a false message and expect the receiver not to follow (l_s *<1/2*). As defined above, lies include both T_NF and F_F. If punishment is norm-communicating, then senders send true messages and expect receivers to follow the messages (T_F). If punishment is norm-irrelevant, then the sender's profit-maximizing strategy becomes the same as that in the NP treatment. This means that the sender sends a false message if she believes the receiver is relatively more likely to follow the message (i.e., l_s *>1/2) (F_F)*; otherwise, she sends a true message (T_NF). Therefore, we can test Hypothesis 3 by comparing, between treatments, the frequency of deception and truth-telling.

$freq^{NP}$*(T_F)* ≤ $freq^{PP}$*(T_F)* < $freq^{NPP}$*(T_F)*
$freq^{NPP}$*(F_F+T_NF)* <$freq^{PP}$*(F_F+T_NF)* ≤ $freq^{NP}$*(F_F+T_NF)*

[3] We obtained data on senders' expectations regarding whether the enforcer would impose punishment and whether the receiver would follow the message. In principle, these data can inform the differences in senders' expectations regarding the norm communication functions of PP and NPP. While sample sizes are small, the evidence from the expectations data seems consistent with our findings drawn from the behavioral data. For example, among those who sent at least one false and one true message, compared with the NPP treatment, significantly fewer senders in the PP treatment expect enforcers to punish if and only if the message is false (2out of 16 vs. 14 out of 21, Z-test, $p<0.01$). Since we are interested in the effect of the norm communication function of punishment on behavior, the current paper focuses on the senders' behavioral data. Expectations data are available on request.

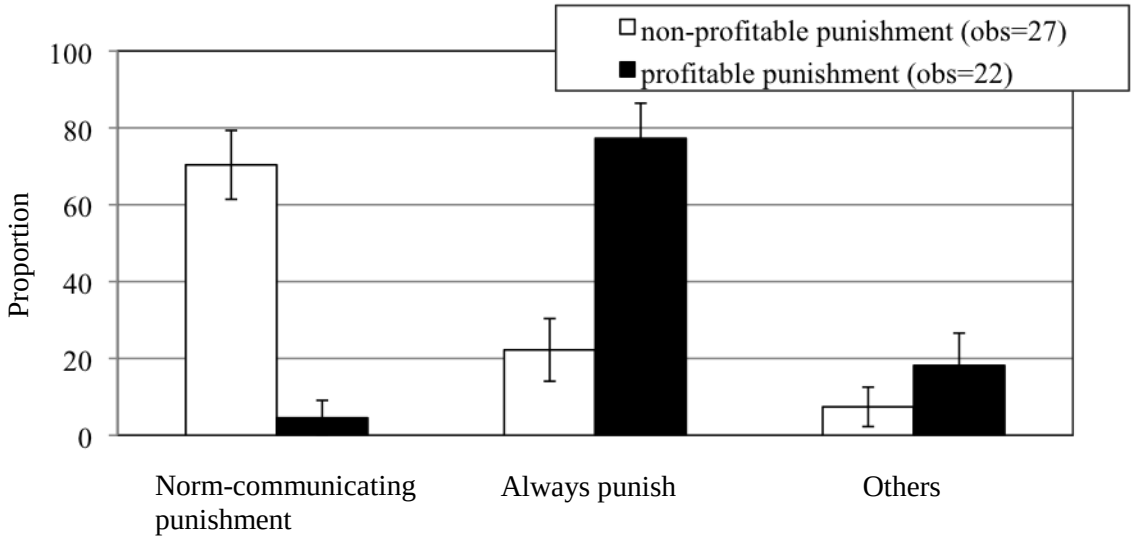
IV. Results

We obtained observations on 252 subjects: 29 groups in the non-profitable punishment treatment; 30 groups in the profitable punishment treatment; and 25 groups in the non-punishment treatment.

IV. A Enforcer's behavior

To examine the punishment decisions of the enforcers, we calculate the proportion of enforcers who: 1) punish if and only if the sender sends a false message; 2) always punish regardless of the truthfulness of the message; and 3) exhibit punishment behavior that does not fall into either one of the aforementioned categories. Figure 1 plots the distribution of different types of

Figure 1. Enforcer's behavior by treatment



Note: The data include only those who experienced both false message and true message scenarios.

enforcers in the two punishment treatments. The data includes only those enforcers who have seen both true messages and false messages. As shown in Figure 1, while a majority (19 out of 27) of enforcers imposed norm-communicating punishment in the NPP treatment , only one out of 22 enforcers did so in the PP treatment. The difference is significant between the two treatments (Z-test, $p<0.01$). On the other hand, most enforcers (17 out of 22) always punished the sender regardless of the message sent in the PP treatment, but only 6 out of 27 did so in the NPP treatment (Z-test, $p<0.01$). This result supports Hypothesis 1.

IV. B. Receivers' Choices

Table 2 reports how receivers, overall, follow senders' messages under different conditions in each treatment. As shown in Table 2, in the NPP treatment the enforcer did not punish the sender in 38 cases out of 87 observations. In these cases, receivers followed senders' messages 94.7% of the time (36 out of 38). In the PP treatment, we also found that receivers tend to follow senders' messages most of the time if the sender is not punished. Among the nine observations where punishment was not imposed, receivers followed seven times (77.8% of the messages).

Table 2. Receivers' decisions by treatment

Treatment	Sender	Follow	Not follow
Non-profitable pun (NPP) (# of obs=87)	Punished	7	42
	Not punished	36	2
Profitable pun (PP) (# of obs=90)	Punished	41	40
	Not punished	7	2
No-pun (NP) (# of obs=75)	-	51	24

To test Hypothesis 2, we next focus on treatment effects in cases where punishment is imposed. We compare receivers' tendencies to follow senders' messages among three cases: 1) when senders are punished in the NPP treatment; 2) when senders are punished in the PP treatment; and 3) the NP treatment.

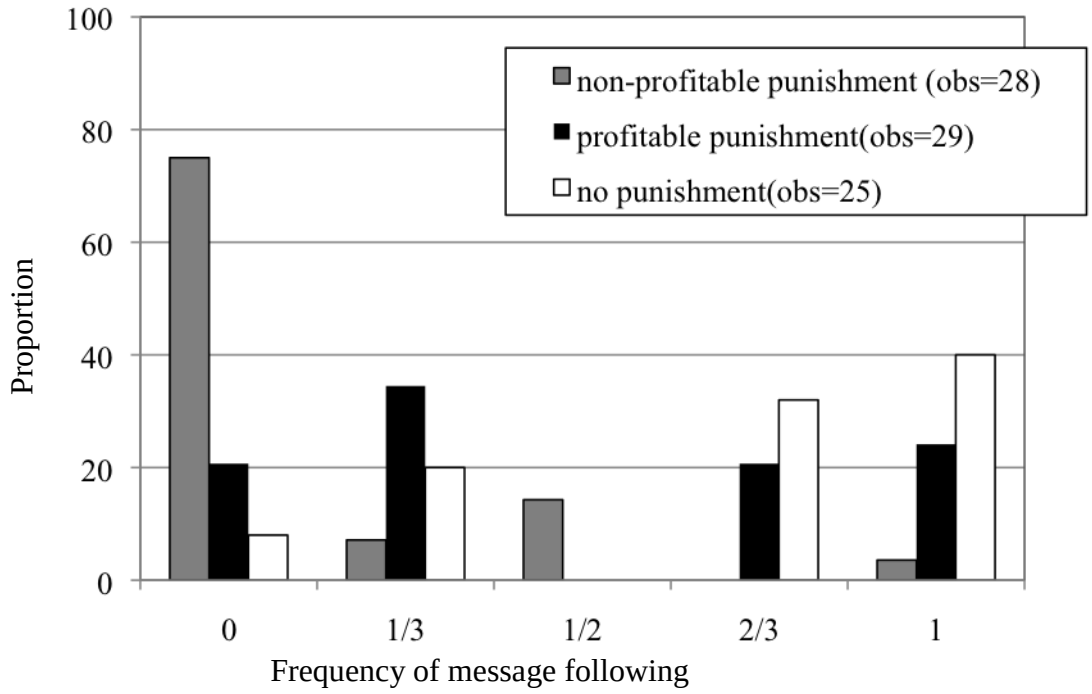
The descriptive data in Table 2 shows that most receivers do not follow messages if the sender is punished in the NPP treatment (42 out of 49). Receivers in the PP treatment, however, seem to make either decision with equal probability (41 decisions follow the message, and 40 decisions do not). Consistent with previous findings (Gneezy, 2005; Sutter, 2009), we find that, in NP treatment receivers tend to follow senders' messages when the message is the only information they have (about 2/3 of decisions follow the sender's message).

We next report data at the individual level. Supporting Hypothesis 2, we find that the frequency of receivers to follow the senders' messages is highest in the NP treatment and lowest in the NPP treatment. (68% vs. 49% vs. 13%, Jonckheere test, $p<0.01$; 49% vs. 13%, Mann-Whitney test, one-tail $p<0.01$) [4]. We classify receivers according to the fraction of times they follow the sender's decisions, and plot the subsequent distribution in Figure 2. As shown, most receivers (75%) never follow the sender's message when the sender is punished in the NPP

[4] To test Hypothesis 2 and 3, we first run Jonckheere tests to establish the hypothesized trend and then use Mann-Whitney or Z-tests to provide evidence on specific differences between the NPP and PP treatments.

treatment. In contrast, this percentage is only 21% in the PP treatment and 8% in NP treatment (Jonckheere test, $p<0.01$; 75% vs. 21%, Z-test, one-tail $p<0.01$). On the other hand, while only one out of 28 (4%) receivers in the NPP treatment always follows the sender's message when the sender is punished, about 24% receivers do so in the PP treatment and 40% receivers follow the sender's message in the NP treatment (Jonckheere test, $p<0.01$; 4% vs. 24%, Z-test, one-tail $p=0.01$). These results support Hypothesis 2.

Figure 2. Distributions of receivers' frequencies of following the message when the sender is punished in the two punishment treatments, and in the no-punishment treatment



Note: The data excludes one receiver in each of the punishment treatment who did not see any punished message.

IV. C. Sender's decisions

We report the descriptive data of sender's decisions and beliefs in Table 3. As shown, the fraction of lies is higher in the PP and NP treatments than in the NPP treatment. The fraction of T_F is higher in NPP treatment than in the PP and NP treatments.

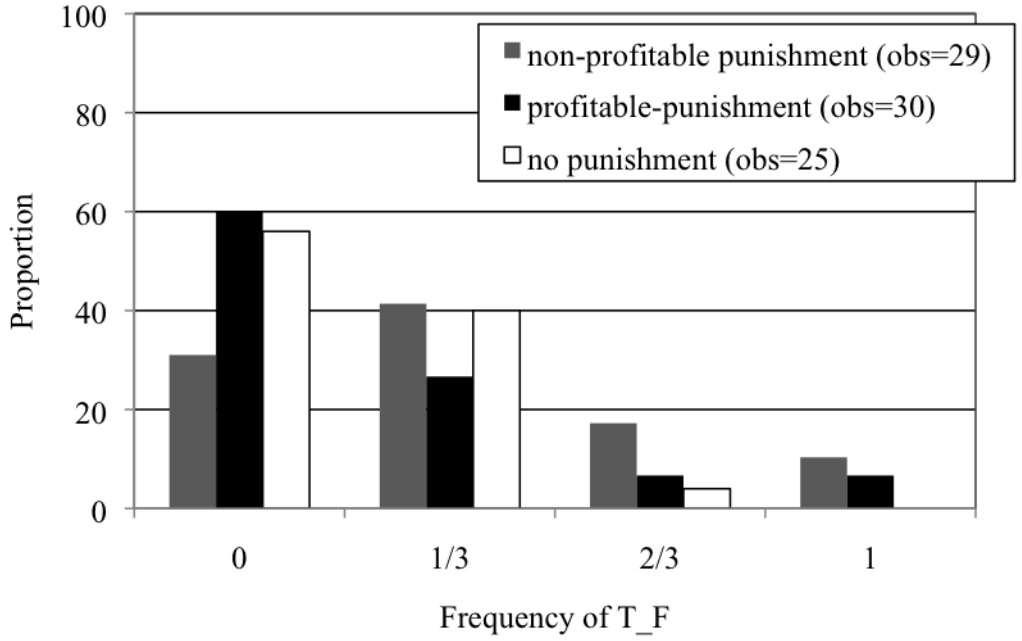
We next analyze our data at the individual level. Each sender played the three games. For each sender, we calculate the percentage of lies (T_NF or F_F) and T_F. Supporting Hypothesis 3, we find that in the NPP treatment, on average, senders lie less frequently (34%) compared to those in PP treatment (56%) and NP treatment (60%) (Jonckheere test, $p<0.01$; 56% vs. 34%, Mann-Whitney test, one-tail $p=0.02$). On the other hand, the average percentage of T_F is higher in NPP treatment (36%) than that in PP treatment (20%) and NP treatment (16%) (Jonckheere test, $p<0.01$; 20% vs. 36%, Mann-Whitney test, one-tail $p=0.02$).

Table 3. Senders' decisions and beliefs by treatment

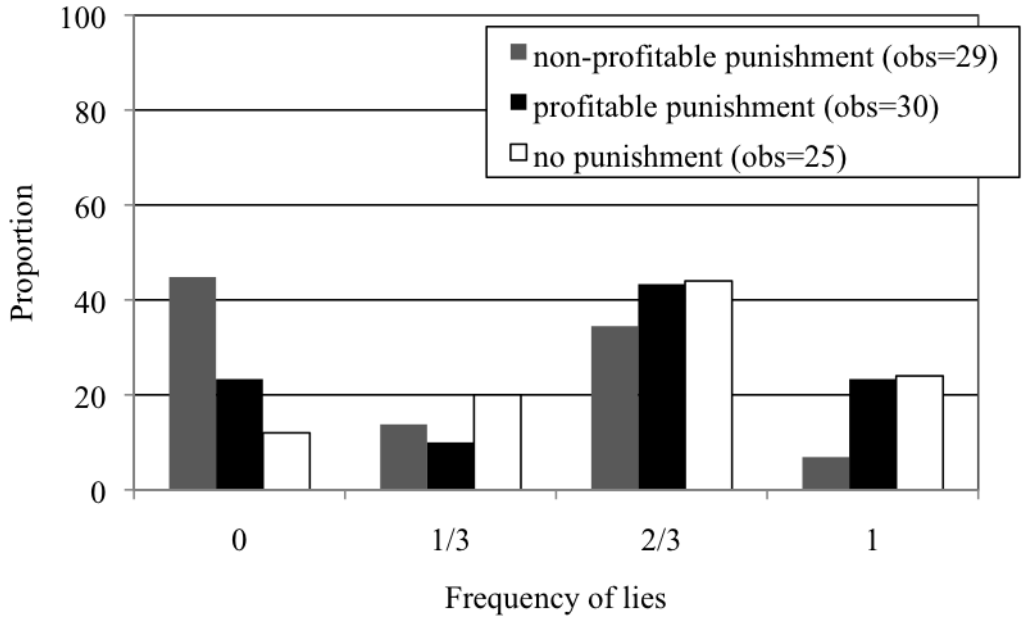
Treatment (# of obs.)	Lie		Sender's decision_ belief		
	Freq.	%		Freq.	%
Non-profitable pun (NPP) (87)	30	34.48	T_NF	10	11.49
			F_F	20	22.99
			T_F	31	35.63
			F_NF	26	29.89
Profitable pun (PP) (90)	50	55.56	T_NF	36	40.00
			F_F	14	15.56
			T_F	18	20.00
			F_NF	22	24.44
No-pun (NP) (75)	45	60	T_NF	35	46.67
			F_F	10	13.33
			T_F	12	16.00
			F_NF	18	24.00

Note: 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.

Figure 3. (A). Distribution of the frequency of T_F by treatment



(B). Distribution of the frequency of lies (T_NF or F_F) by treatment



Note: T_F: a sender sends a true message and expects the receiver to follow.
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.

Figure 3 plots the distribution of sender types according to the frequencies of T_F and lies. First, note from Figure 3(A) that in the NPP treatment, about 69% of senders "send a true message and believe the receiver will follow the message" at least once. This percentage is only

40 % in the PP treatment and 44% in the NP treatment (Jonckheere test, p=0.03; 69% vs. 40%, Z- test, one-tail p=0.02). Figure 3(B) shows that about 77% of senders in the PP treatment and 88% of senders in NP treatment lie at least once and only 55% do so in the NPP treatment (Jonckheere test, p<0.01; 77% vs. 55%, Z-test, one-tail p=0.04). These results support our Hypothesis 3.

V. Discussion

By definition, corrupt societies include persistent and pervasive norm violations. This may be one important reason that corruption is widely perceived to be a major impediment to economic development. To curb corruption, scholars have highlighted the importance of punishment. Unfortunately, in corrupt societies punishment enforcement authorities are typically perceived as also being highly corrupt. Rather than simply reflecting a culture of norm disobedience, however, we argued that legal institutions that embed corrupting temptations (e.g., profitable punishment) may be causally connected to systemic patterns of norm violations in a society. In particular, we designed experiments to provide direct evidence on the detrimental effect of profitable punishment on the norm communication function of punishment, and demonstrated that this led to further norm disobedience.

Our experiments are based on sender-receiver games in which norm-communicating and norm-irrelevant punishment yield starkly different predictions about the decisions of senders and receivers. The results support our hypothesis that punishment can effectively express norm violations to receivers when enforcers do not benefit from the penalty. Its expression function is significantly diminished, however, when the punishment becomes a source of revenue for enforcers[5]. Importantly, deception occurs more often when punishment is profitable for the enforcer than when it is not. In fact, deception occurs just as frequently in the case of profitable punishment as it does when punishment is not possible.

A substantial literature reveals that people care about equality and fairness and often share with others and even unaffected bystanders are willing to incur a cost to punish unfair behavior (e.g. Fehr and Fischbacher, 2004). We nevertheless find that when third-party enforcers can profit by punishing, a high percentage of them (about 80%) take advantage of this option and abuse their authority. Doing so benefits themselves but is costly both to senders as well as receivers. Given that enforcers often benefit from their duty to maintain social order, our paper calls attention to the role of third-party punishment in promoting cooperation when enforcement is profitable.

It is worthwhile to emphasize that in our experiment receivers do not know whether enforcers are "corrupt" because they do not receive any feedback. This suggests that the detrimental effect of profitable punishment can stem simply from the corrupting temptations embedded within enforcement intuitions. In other words, an enforcement institution that enables corrupt behavior can perhaps have a detrimental impact on punishment's effectiveness, regardless of enforcers' actual decisions.

[5] Our findings regarding the effect of profitability on the effectiveness of punishment are also consistent with Kuang et al. (2007)'s investigation of the effectiveness of advice. That paper reports that advice from a party with a monetary stake in whether the advice is followed is less effective than the same advice given by a neutral independent party.

This finding speaks to the importance of establishing institutions that signal intolerance for corruption in the legal system. Such institutions have proven effective in natural environments. For example, during the economic growth period in Hong Kong, the ICAC (Independent Commission Against Corruption, http://www.icac.org.hk/en/home/index.html) was established to clean up its endemic corruption in law enforcement and many other government departments. Meeting challenges initially, the ICAC ultimately proved vital in transforming Hong Kong from a graft-ridden city into a metropolis known for its cleanliness and lawfulness.

Our findings help to explain why the existence of organizations such as ICAC is important: their presence enhances the effectiveness of the legal system by restoring the norm communication function of law enforcement. Indeed, such organizations may be a necessary first step towards mitigating pervasive norm violations in severely corrupt societies, and doing so might help set the stage for rapid and peaceful economic expansion.

Acknowledgements. I thank Klaus Abbink, Jonathan Baron, Andreas Blume, Brit Grosskopf, Uri Gneezy, Daniel Houser, George Lowenstein, Roberto Weber, and participants at seminars at Texas University A&M; 2009 International ESA; 2009 North-American ESA; 2010 Behavioral Decision Research in Management Conference; 2010 American Law & Economics Association conference ;CBDR seminar at Carnegie Mellon University for valuable comments. I gratefully acknowledge the Berkman Fund and National Science Foundation (SES-0961341) for funding that supported this research.

References

Abbink, K., Irlenbusch, B. and Renner, E. (2002), "An experimental bribery game", Journal of Law, Economics, and Organization, 18(2), 428-454.

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

Andreoni, J., Harbaugh, W., Vesterlund, L. (2003). "The Carrot or Stick: Reward, Punishment and Cooperation". American Economic Review. 93(3), 893-902.

Ariely, D., Bracha, A., Meier, S. (2009). "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially" American Economic Review, 99(1), 544–555

Bicchieri,C., The Grammar of Society: The Nature and Dynamics of Social Norms, Cambridge University Press, 2006

Benabou, R., Tirole, J. 2003. Intrinsic and Extrinsic Motivation. Review of Economic Studies 70, 489-520.

Blume, Andreas; DeJong, Douglas V.; Kim, Yong Gwan and Sprinkle, Geoffrey B. (2001), "Evolution of Communication with Partial Common Interest." Games and Economic Behavior, 37(1), pp. 79-120

Boles, T. L., Croson, R. T. A. and Murnighan, J. K. (2000) "Deception and retribution in repeated ultimatum bargaining," Organizational Behavior and Human Decision Processes, 83, 235-259.

Bolton, G. E and Ockenfels, A. (2000) "ERC: A Theory of Equity, Reciprocity, and Competition." American Economic Review, 90, pp166-193.

Brandts, J., and G. Charness (2003) "Truth or consequence: an experiment," Management Science, 116–130.

Crawford, Vincent P. and Sobel, Joel. (1982). "Strategic Information Transmission," Econometrica, 50, 1431–1451.

Crawford, Vincent P. (1998) "A Survey of Experiments on Communication via Cheap Talk." Journal of Economic Theory, 78(2), pp. 286–298.

Crawford, Vincent P. (2003). "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." American Economic Review, 93 (1), pp. 133-49.

Charness, G., and Dufwenberg, M. (2006). "Promises and Partnership," Econometrica, 74(6), 1579-1601.

Cooter, Robert. D. (1998). "Expressive Law and Economics", Journal of Legal Studies 27(2), pp.585-608

Croson, R., Boles, T., and Murnighan, J. K. (2003) "Cheap talk in bargaining experiments: lying and threats in ultimatum games," Journal of Economic Behavior and Organization, 51, 143-159.

Dawes CT, Fowler JH, Johnson T, McElreath R, Smirnov O. (2007) "Egalitarian Motives in Humans," Nature 446 (12), p 794-796

Duggan, M. and Levitt, S. (2002). "Winning Isn't Everything: Corruption in Sumo Wrestling." American Economic Review, 92(5),1594-1605.

Dickinson, D. 2001. The Carrot vs. the Stick in Work Team Motivation. Experimental Economics 4, 107-124.

Dreber, A. and M. Johannesson (2008) "Gender Differences in Deception," Economics Letters. 99, 197-99.

Ellingsen, T., Johannesson, M. (2004) “Promises, threats and fairness,” Economic Journal,114, 397-420.
Ellingsen, T., Johannesson, M. Lilja, J. and Zetterqvist, H. (2009) “Trust and truth,” Economic Journal,119, 252-276.
Elster, J. Alchemies of the Mind: Rationality and the Emotions, Cambridge, 1999.
Erat, S. and Gneezy, U. (2009). “White Lies”, Rady School of Management, UCSD working paper.
Falk, A. and Kosfeld, M. (2006). “The Hidden Cost of Control” American Economic Review, 96 (5), 1611-1630
Fehr, E. and Falk, A. (2002). Psychological Foundation of Incentives. European Economic Review 46, 687 – 724
Fehr, E. and Fischbacher, U., (2004), "Third Party Punishment and Social Norms," Evolution and Human Behavior, 25, 63-87.
Fehr, E., Gächter,S. (2000) “Cooperation and Punishment in Public Goods Experiments”. American Economic Review 90 (4), 980-994
Fehr, E. and List, J. (2004). “The Hidden Costs and Rewards of Incentives”. Journal of the European Economic Association, 2(5), 743-771.
Fehr, E., Rockenbach, B. (2003).” Detrimental Effects of Sanctions on Human Altruism”. Nature, 422, 137-140
Fehr, E.and Schmidt, K. (1999) “A Theory of Fairness, Competition, and Cooperation.” Quarterly Journal of Economics, 114, pp817-868
Frey, B. and Oberholzer-Gee, F. (1997) “The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding –Out”. American Economic Review. 87, 746-755.
Fischbacher, Urs (2007): z-Tree: Zurich Toolbox for Ready-made Economic Experiments, Experimental Economics 10(2), 171-178.
Fuster, A. and Meier, S. (2009). “Another Hidden Cost of Incentives: The Detrimental Effect on Norm Enforcement”. Management Science, doi 10.1287/mnsc.1090.1081
Galbiati, Roberto and Vertova, Pietro. (2008). “Obligations and cooperative behaviour in public good games”, Games and Economic Behavior. doi:10.1016/j.geb.2007.09.004
Gneezy, U. (2005), “Deception: The Role of Consequences,” American Economic Review 95, 384-94
Gneezy, U., Rustichini, A. (2000). “A Fine Is A Price”. Journal of Legal Studies. 29(1), 1-17.
Herrmann, B. C. Thöni, C. and S. Gächter, S. (2008), “Antisocial Punishment Across Societies” Science, Vol. 319. no. 5868, pp. 1362 – 1367.
Houser,D., Xiao,E., McCabe, K. and Smith,V. (2008). “When punishment fails: Research on sanctions, intentions and non-cooperation”, Games and Economic Behavior, 62 (2), 509-532.
Hunt, J., “Why Are Some Public Officials More Corrupt Than Others?” In Susan Rose-Ackerman ed. International Handbook on the Economics of Corruption Northampton, M.A.: Edward Elgar, 2006
Hurkens, S. and Kartik, N. (forthcoming). “Would I lie to you? On social preferences and lying aversion”, Experimental Economics.
Kahan, D. M. (1998) “Social Meaning and the Economic Analysis of Crime. Journal of Legal Studies 27(2), pp. 661-672.
Lundquist, T., Ellingsen, T., Gribbe E.and Johannesson,M. (2009) “The aversion to lying,” Journal of Economic Behavior and Organization, 70, 81-92.

Masclet, D. Noussair, C., Tucker, S. and Villeval,M. (2003) "Monetary and Non-monetary Punishment in the Voluntary Contributions Mechanism", American Economic Review, 93, pp366–380.
Mazar, N., O. Amir and D. Ariely (2008) "The dishonesty of honest people: A theory of self-concept maintenance," Journal of Marketing Research, 45, 633-644.
Rode, J. (2010). "Truth and trust in communication. an experimental study of behaviour under asymmetric information", Games and Economic Behavior, 68 (1), pp. 325–338.
Ostrom, E., Walker, J., Gardner, R. (1992). "Covenants With and Without a Sword: Self-Governance is Possible". American Political Science Review. 86, 404 - 417.
Sánchez-Pagés, S. and Vorsatz, M. (2007). "An experimental study of truth-telling in a sender-receiver game", Games and Economic Behavior, vol. 61, pp. 86–112.
Sánchez-Pagés,, S. and Vorsatz, M. (2009). "Enjoy the silence: an experiment on truth-telling", Experimental Economics, 12, pp. 220–241
Sefton, M., Shupp R. and Walker, J. (2007). "The Effect of Rewards and Sanctions in Provision of Public Goods," Economic Inquiry, Vol. 45, pp. 671-690.
Sunstein, C., R. (1996). "On the Expressive Function of Law". University of Pennsylvania Law Review, 144. pp2021-2031.
Sutter, M. (2009). "Deception through telling the truth?! Experimental evidence from individuals and teams". Economic Journal 119, 47-60
Tyran, J and Feld, L. "Achieving Compliance when Legal Sanctions are Non-Deterrent", Scandinavian Journal of Economics, 2006, 108(1), pp135-156
Transparency Internal (2007). "Report on the Transparency International Global Corruption Barometer 2007". (http://www.transparency.org/content/download/27256/410704/file/GCB_2007_report_en_02-12-2007.pdf)
Yamagishi, T. (1986). "The provision of a Sanctioning System as a Public Good". Journal of Personality and Social Psychology. 51(1), 110-116
Xiao, E. and Houser, E. (forthcoming). "Punish in Public", Journal of Public Economics