Journal of Regulatory Economics; 24:1 91–110, 2003

Punishment in a Regulatory Setting: Experimental Evidence from the VCM*

LISA R. ANDERSON
College of William and Mary
Department of Economics, Williamsburg, VA 23187
E-mail: lrande@wm.edu

SARAH L. STAFFORD
College of William and Mary
Department of Economics, Williamsburg, VA 23187
E-mail: slstaf@wm.edu

Abstract
To analyze the effectiveness of punishment in inducing regulatory compliance, we modify a standard public goods experiment to include a financial penalty for free riding. The design allows us to vary both punishment probability and severity. We introduce the punishment mechanism in both a one-time and a repeated treatment and find that compliance (contributing to the public good) is increasing in expected punishment cost in both treatments. We also find that punishment severity has a larger effect on behavior than punishment probability. In the repeated treatment, we find that past punishment has a negative rather than positive effect on compliance.

1. Introduction

> Public enforcement of law . . . is a subject of obvious importance. Enforcement policy affects, for example, the amount of pollution that firms generate, the extent of compliance with the income tax code, and the incidence of theft, robbery, and other crimes. (Polinsky and Shavell 2000, p. 45)

Although hundreds of papers have been written on the economics of enforcement since the 1968 publication of Gary Becker's seminal paper on the economics of crime and

* We thank Susan Laury and two anonymous referees for their valuable comments. Financial support from the College of William and Mary and the National Science Foundation (SES-0094800) is gratefully acknowledged.

punishment, many fundamental issues are still the subject of considerable debate. One such issue is the deterrent effect of increases in the probability of punishment compared to increases in the severity of punishment. In Becker's model, assuming risk neutrality, punishment probability and punishment severity are perfect substitutes. If potential offenders are risk averse, however, punishment severity should have a larger effect than punishment probability; the opposite is true if potential offenders are risk preferers. Obviously, the relative deterrent effect of punishment probability and punishment severity is important for the literature on optimal enforcement mechanisms. However, there is a lack of conclusive empirical evidence on this issue. Two of the principal contributors to the economic enforcement literature, A. Mitchell Polinsky and Steven Shavell, state in a recent survey article, ''empirical work on law enforcement is strongly needed to better measure the deterrent effect of sanctions, especially to separate the influence of the magnitude of sanctions from their probability of application'' (2000, p. 73).

One problem inhibiting traditional empirical analyzes of the deterrent effect of punishment probability and severity is the difficulty in obtaining appropriate data. In most cases, individual data cannot be collected directly; much data is available only at the aggregate level and must be calculated or estimated. For many offenses the probability of punishment is computed as the ratio of the number of convictions to the number of reported offenses. Because reporting offenses is costly to the victim, offenses are often underreported and thus there may be systematic measurement error in this variable that can bias estimation.[1] In cases where there is no easily identifiable victim, such as environmental regulation and income tax compliance, only detected offenses are reported and the number of undetected offenses must be indirectly estimated. Punishment is also difficult to measure. If the punishment includes incarceration, it is not obvious how to measure it, as the cost to the offender will vary depending on the individual's characteristics and the likelihood of parole. Monetary punishments are also difficult to measure as fines may be reduced from their original level through appeals or other agreements. Additionally it is difficult to judge the effect of changes in punishment on actions, as the number of offenses committed generally can not be measured directly. Because of these measurement problems, most traditional empirical analyzes can only estimate the sign of the effects of punishment probability and severity on compliance. Any estimates of the elasticity of offenses with respect to punishment probability or severity, or even the relative efficacy of these two instruments, are questionable.[2]

Experimental economics provides an alternative to traditional empirical analyses. In a laboratory setting it is possible to explicitly control and accurately measure the conditions under which subjects make decisions. Using an experimental approach, we can vary both punishment probability and punishment severity while accurately measuring probability,

1 Measurement error could operate in more than one direction, depending on whether all offenses are underreported to the same extent, or whether victims are more likely to report specific types of crimes. Not only will the deterrent effect of punishment probability be subject to bias, but the estimated effects of other explanatory variables will be biased as well. See Erlich (1996) for a more detailed discussion.

2 Erlich (1996) provides a thorough exposition of these problems.

severity, and individual actions. In this paper we present an experiment designed to analyze the deterrent effects of punishment probability and punishment severity in a regulatory setting.

Our design models a regulatory regime in which compliance is equivalent to contributing to a public good. The regulation is enforced via a punishment mechanism that sanctions free riding with varying probability and severity. While this model clearly is not applicable to all types of illegal behavior, it is a reasonable model for many regulatory regimes. For example, installing pollution control equipment is analogous to contributing to a public good. Antitrust regulation also can been viewed as a public good. Companies that violate price-fixing laws earn higher profits as a result of their actions, but the total benefit to the companies involved is smaller than the costs borne by the consumers of their goods.

In our experiments subjects respond more to punishment severity than probability. Although traditional empirical studies that have tried to estimate the effect of punishment probability relative to punishment severity find the opposite result, our results are consistent with other experimental findings (see Block and Gerety 1995). Given the obvious importance of this issue in the development of optimal enforcement strategies, our results suggest that additional empirical and experimental analysis is needed. The remainder of the paper is organized as follows. Section 2 describes the related empirical and experimental literature. Section 3 discusses the Voluntary Contribution Mechanism and explains the punishment mechanism we use in the experiment. Section 4 describes the experimental procedures, section 5 presents the results of the experiment, and section 6 concludes.

2. Related Literature

The deterrent effect of punishment has been relatively well documented empirically. See, for example, the surveys by Cameron (1988) on crime, Heyes (2000) on environmental regulations, and Andreoni et al. (1998) on tax compliance. For the most part, however, existing empirical studies can only accurately estimate the existence and direction of the effect of increasing punishment probability and/or severity. In general, estimates of the deterrent effect of punishment probability and severity either in absolute or relative terms are suspect due to the measurement problems mentioned in the introduction. In many regulatory regimes the probability of punishment depends on individual characteristics. For example, the probability of a tax audit depends on the tax payer's income, exemptions, and deductions. Most environmental enforcement is similarly targeted based on the likelihood that a given plant will be in violation. The endogeneity of punishment probability in these situations has made direct empirical analysis of the deterrent effects of probability and severity even more difficult.

Despite these problems, a few studies have estimated relative elasticities. These studies have generally found that offenders respond more to increases in the probability of punishment than to increases in severity. For example, in a study using aggregate data on crime from the North Carolina police, Trumbull (1989) finds that the elasticity of the probability of arrest (calculated using arrests and reported crimes) is larger than the elasticity of average prison sentence length. Grogger (1991) uses individual crime data and finds a similar result, that probability has a larger quantitative effect than severity, which he measures in terms of sentence delivered. He also finds that the deterrent effect of probability depends on the type of crime: increases in probability have a larger effect on felonies than on non-felonies and on non-property than on property crimes. The deterrent effect of severity is stronger for non-property crimes than for property, but is the same for felonies verses non-felonies. In a study of income tax compliance in Switzerland, Pommerehne and Weck-Hannemann (1996) find that noncompliance is negatively related to the probability of audit, while the penalty has no significant effect on noncompliance.

Given the difficulties in assessing the effect of punishment probability and severity, some researchers have turned to laboratory experiments. A number of papers use an experimental design that simulates an income tax setting, where subjects decide how much income to claim with random auditing and specified fine and audit rates. See, for example, Alm et al. (1992), Beck et al. (1991) and Becker et al. (1987). The general finding from these studies is that more income is reported as fine and audit rates increase. These studies differ from ours in two important ways: While fine and audit rates vary and are considered separately in many of these studies, their relative effects are not the focus of attention. In addition, these tax compliance studies are not conducted in a public goods framework.[3]

Block and Gerety (1995) also test the effectiveness of a probabilistic punishment scheme in a non-public goods framework. Specifically, subjects participate in a sealed bid auction and, in some cases, are given the opportunity to collusively set prices. The focus of this study is a comparison of behavior between a standard experimental subject pool of undergraduate students and prison inmates. They find that students are more deterred by the severity of the punishment than the probability of punishment and that the opposite is true for prisoners, who appear to be risk loving.

Cardenas et al. (2000) study the effect of punishment in a field experiment that conforms to a basic public goods model. Subjects are residents of a Colombian village and their decision variable is how much time to spend harvesting firewood. The amount of firewood harvested is inversely related to water quality in the village, hence, firewood extraction generates a public bad. Like our experiment, their punishment mechanism is probabilistic and proportional to the degree of free riding. However, they consider only one probability/ severity combination and thus cannot isolate the independent effects of the two. Their main finding is that subjects are more cooperative without regulation.

Tyran and Feld (2001) test the effect of increasing the severity of punishment in the context of a standard public goods experiment. Unlike our experiment, punishment is certain (with non-zero severity levels) and the amount of the punishment does not depend on the degree of free riding. They find that severe punishment leads to near-complete contribution to the public good, but mild punishment has a small effect on contribution levels when the punishment mechanism is exogenously imposed. When the punishment level is endogenously determined by a referendum, even mild punishment has a significant effect on contributions to the public good.

---

3 Alm et al. (1992) have one treatment with a public goods framework. In that treatment the fine and audit levels do not change, so they are not able to compare the relative importance of the two.

A number of studies look exclusively at the use of endogenous punishment mechanisms. For example, Ostrom et al. (1992) incorporate the potential for an internal punishment mechanism in a common pool resource experiment. Specifically, punishment is imposed by the participants in the experiment rather than by the experimenter. Fehr and Gächter (2000) examine a similar internal punishment mechanism in a public goods experiment. Our study focuses on the effect of mechanisms available to regulatory agencies, which are all externally determined. Hence, the literature on the effect of internal mechanisms is not reviewed here.

3. The Voluntary Contributions Mechanism and Punishment for Free Riding

The institutional framework most commonly used for studying public good provision is the voluntary contributions mechanism (VCM). In the VCM each person in a group of size N is given a monetary endowment and offered the opportunity to contribute to a group account (i.e., the public good). The sum of all money allocated to the group account, G, is transformed by some factor, w, and each member of the group receives the amount Gw. Importantly, every member of the group receives Gw from the group account, regardless of whether or not they contribute to it. Hence, the group account is a pure public good. Money kept for private consumption is transformed by some factor, v, and converted to cash earnings. In the standard public goods experiment $w < v = 1$, so the dominant-strategy Nash equilibrium is for everyone to free ride on others' contributions. This decision problem is a dilemma (specifically, the prisoner's dilemma) when the social benefit of money allocated to the group account outweighs the value of money kept for private consumption (i.e., $Nw > v$).

Hundreds of VCM experiments have been conducted in the past twenty years with a focus on the effect of different treatment variables on the level of contributions to the group account. This list of variables is long and includes anonymity, repetition, the size of the group, communication and provision points. Some of the effects are quite intuitive. For example, provision points and communication increase contributions to the group account (see Isaac et al. 1989 and Isaac and Walker 1988). Other effects, such as group size, are harder to predict a priori.[4] Ledyard (1995) and Anderson (2001) summarize major findings from VCM experiments, which we build upon by incorporating a probabilistic financial penalty for free riding into the standard VCM framework to disentangle the effects of punishment probability and punishment severity.[5]

Subjects in the experiment receive earnings from two different sources: the group

4 It is reasonable to expect more free riding in a larger group, since detection is harder, but the social benefit of contributions to the group account increases as the group grows in size. Davis and Holt (1993) present experimental evidence that the effect of group size depends on other factors, such as the conversion factor for money allocated to the group account.

5 Because the possibility of a financial penalty introduces uncertainty about the value of money allocated to a person's private account, we assume risk neutrality for the discussion that follows.

account and their private account. Subjects must decide how to allocate their initial endowment between the group account and their private account. Consider first the baseline case with no chance for punishment. Money allocated to the group account is doubled and shared by the ten members of the group (i.e., $w = 2/10$). Hence, each individual receives \$0.20 for every dollar allocated to the group account. Money allocated to the private account is not doubled or shared (i.e., $v = 1$). The ratio of earnings from group account contributions (w) to earnings from private account contributions (v) is termed the marginal per capita return (MPCR) to the group account, and in the example here is equal to 0.20. As long as the MPCR is less than 1, the dominant strategy, with no chance of punishment, is complete free riding. Analogously, in a regulatory scheme with no punishment and the assumption of profit-maximization, we would expect complete noncompliance.

In treatments where punishment occurs, group account earnings continue to be determined as explained above (i.e., $w = 2/10$). However, private account earnings may be reduced if the subject is found to be free riding. In our design, the financial penalty assessed is a function of the amount allocated to the private account, that is the extent of free riding.[6] Specifically, the penalty is equal to the amount allocated to the private account multiplied by s, the punishment severity. Thus, a subject's expected private account earnings are equal to the amount allocated to the private account multiplied by $(1 - ps)$ where p is the probability of being punished (i.e., $v = 1 - ps$). This punishment mechanism is most analogous to a regulatory regime in which noncompliance is discovered through an audit or inspection (for example, income tax audits, health and safety inspections, parking enforcement). In the experiment, punishment severity ranges from 1 to 4 and punishment probability ranges from 10 to 90%.

We term the product of the punishment probability and the punishment severity the "punishment factor." The punishment factor can be interpreted as the amount someone should expect to pay in penalties for each dollar allocated to the private account. We define the expected marginal per capita return (EMPCR) to be the value of a token allocated to the group account divided by the expected value of a token allocated to the private account. Note that the EMPCR is positively related to the punishment factor: As the punishment factor increases, the expected return from the private account decreases, resulting in a higher EMPCR, holding the return from the public account fixed at 0.20. We consider nineteen combinations of punishment severity and probability, some of which have the same punishment factor. For example, probability $\times$ severity combinations of 40% $\times$ 1, 10% $\times$ 4 and 20% $\times$ 2 all yield a punishment factor of 0.4.

Table 1 reports the expected value of one dollar allocated to a private account for each probability $\times$ severity combination. Notice that the expected return from \$1 in the private account (listed in the cells in table 1) is greater than the certain return of \$1 in the group account ($\equiv$ \$0.20) for eleven of the nineteen cases. Hence, free riding is a dominant strategy in these cases. In two cases the expected return of \$1 allocated to one's private

6 We made the size of the penalty contingent on the extent of free riding so that the marginal punishment level would be constant. While in practice marginal punishments are rarely constant, many regulatory penalties do take into account the level or extent of violation.

Table 1. Expected Return* from $1 in Private Account

Punishment Severity	Punishment Probability 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	$0.90 (0.1)	$0.80 (0.2)	$0.70 (0.3)	$0.60 (0.4)	$0.50 (0.5)		$0.30 (0.7)	$0.20 (0.8)	$0.10 (0.9)
2	$0.80 (0.2)	$0.60 (0.4)	$0.40 (0.6)			− $0.20 (1.2)		− $0.60 (1.6)	
3		$0.40 (0.6)	$0.10 (0.9)						
4	$0.60 (0.4)	$0.20 (0.8)	− $0.20 (1.2)		− $1.00 (2.0)				

Note. *Punishment factor (= probability * severity) in parentheses.

account is $0.20, hence, in theory, risk neutral subjects are indifferent between investing in the private account and the group account. In the remaining six cases the expected return from $1 allocated to the private account is less than the certain $0.20 earned by allocating $1 to the group account, making full contribution to the group account the dominant strategy. In three of those six cases, allocating $1 to the private account has a negative expected value, since the expected punishment exceeds $1.

We developed two different treatments for this experiment, each designed to capture the salient features of an existing regulatory regime. The first treatment, the ''One-Time'' treatment, models regulations for activities that occur once or very infrequently (for example, environmental regulations that govern the disposal of hazardous materials from the demolition of old chemical plants or the closure of landfills). In this treatment, subjects make decisions for a series of twenty different punishment scenarios. During the decision-making period, subjects receive no information about other's decisions. This design follows that of Goeree et al. (2002) and can be viewed as a one-shot game. In this design, subjects' decisions do not depend on past results and should not reflect strategic attempts to influence others' behavior.

The second treatment, the ''Repeated'' treatment, models regulations that govern repeated or continuous activities such as regulations on air and water pollution. This treatment is comparable to the majority of VCM experiments where subjects make repeated decisions and receive information about others' decisions between rounds. In this treatment, subjects also make decisions for a series of twenty different punishment scenarios, but the decisions are made one at a time with information about group

7 Of course subjects also receive information on who else has been audited. Although this is also true in some regulatory regimes (parking, for example), it is obviously not true in all repeated regimes. However, in designing the experiment one main goal was that students feel confident in the randomness of the punishment mechanism. Thus, as explained in more detail in section 4, we made the selection of subjects to be punished in front of all participants and could therefore not limit information about who was selected.

contributions and who has been selected for punishment provided between rounds. One of the key features of this treatment is that subjects receive information on whether they have been audited or inspected in the past and thus decisions may depend on past experience as well as on strategic actions to influence others' behavior.[7] In the next section, we discuss in more detail how these two treatments were conducted.

4. Procedures

For each session of the experiment, ten subjects were recruited from undergraduate courses at the College of William and Mary. Upon arriving for the experiment, subjects were given an identification letter and seated in cubicles that blocked their view of others' decision sheets and prevented them from communicating. We used the identification letters to randomly choose among the ten subjects in the punishment phase of the experiment.

Before the subjects made any decisions, they were given a copy of the instructions (see Appendix A) which we read aloud. Each subject started with a \$15 account balance that could not be invested in the group account but could be used to offset punishment losses. Subjects were given a decision sheet with twenty scenarios that consisted of the nineteen probability $\times$ severity combinations described in table 1 and a baseline case with a punishment probability of zero. The punishment was described as an ''earnings adjustment'' in the instructions.[8] In each scenario subjects were told that they controled one tenth of the \$50 (= \$5) in a group account and could move as much of that \$5 as they wished to their private account.[9] As noted above, money left in the group account was doubled and shared equally by the ten members of the group, but money allocated to the private account was neither doubled nor shared. Subjects were told to think carefully about each allocation decision because only one would be chosen for payment by throwing a 20-sided die at the end of the experiment.

The maximum possible payoff in the experiment was \$29. This payoff was only achievable if (1) a subject moved his entire \$5 to his private account; (2) he was not chosen to be punished; and (3) all of the other nine subjects left their entire \$5 in the group account. Under this scenario, the other nine members of the group would each receive \$24.

8 We used the term ''earnings adjustment'' to avoid any ethical considerations students might associate with the word punishment. Since noncompliance penalties are often associated with unethical behavior, our results are potentially biased against our punishment mechanism. Specifically, if there is some non-monetary cost associated with being punished, contributions to the public good might be higher than an analysis based solely on financial gain would suggest.

9 The \$5 could be split between the two accounts in any penny increment. Our design is a variation of the negatively framed version of the VCM used by Andreoni (1995). Andreoni changes the description of the allocation process from a focus on the positive externality generated by putting money in the group account to a focus on the negative externality associated with moving money to the private account. He finds that contributions to the group account are lower in the negatively framed version of the experiment. This framing is appropriate for our application since violation of laws and regulations generally exerts a negative externality on society.

The socially optimal outcome was for all ten subjects to allocate their entire \$5 to the group account, guaranteeing each person a total of \$25 (= \$10 from the group account plus the \$15 starting balance). Actual earnings for the experiment averaged \$22.91.

After the instructions were read, subjects made hypothetical allocation decisions and calculated earnings under three different scenarios so we could verify that they understood how payoffs were calculated. Next we started the decision-making period which followed one of two formats depending on the treatment. In the One-Time treatment, subjects were given 20 minutes to make the twenty allocation decisions with no additional information provided. In contrast, in the Repeated treatment subjects made one allocation decision at a time. Between their decisions, we totaled and announced the amount in the group account and randomly selected subjects to be punished.[10]

After subjects made all twenty allocation decisions, we chose one of the ten subjects at random to throw a 20-sided die to determine which scenario would be used for payoffs. Next we announced the total amount allocated to the group account for the chosen scenario, the doubled amount, and each person's share (1/10) of the group account. At this point in the One-Time treatment we selected the appropriate number of subjects to be punished for the chosen scenario. (In the Repeated treatment, subjects were selected for punishment after each allocation decision.) The final task was for subjects to calculate their earnings which consisted of (1) the \$15 initial balance, (2) money allocated to their private account and (3) their share of the group account. This amount was adjusted for subjects who were chosen to be punished. The specific adjustment depended on the severity of the punishment. If the severity was 1, they simply subtracted the amount allocated to their private account. If the severity was 2, they subtracted two times the amount allocated to their private account and so forth.

To control for order effects, we created three lists, each with a different order for the scenarios described in table 1. On all lists, the ''no punishment'' decision was the first scenario. As subjects moved from one scenario to the next on the list, either the probability or severity changed, but never both at the same time. We conducted one session of each of the three ordered lists under both the One-Time and the Repeated treatments, although subjects in the One-Time treatment were not constrained to make their decision in the order presented. We ran 6 sessions of the experiment with a total of 60 subjects participating. On average the One-Time sessions lasted 45 minutes and the Repeated sessions lasted 1 hour.

5. Results

When free riding is the Nash equilibrium (i.e., the punishment factor is less than 0.8), our results are consistent with the general experimental finding that group contributions are

10 For both treatments, subjects were selected to be punished by drawing Scrabble© tiles from a cup without replacement. In the Repeated treatment, Scrabble© tiles were drawn at the end of every period. In the One-Time treatment, Scrabble© tiles were drawn after all decisions had been made.

Table 2. Summary of Group Contributions in the One-Time and Repeated Treatments

Punishment Factor (Probability × Severity)	Number of Observations	Average Public Contribution	$5 to Private Account (%)	$5 Split between Accounts (%)	$5 to Group Account (%)
One-time treatment					
Less than 0.8	360	$2.87	13*	69*	18*
Exactly 0.8	60	$4.15	3	50	47
More than 0.8	180	$4.37*	1	48*	51*
Repeated treatment					
Less than 0.8	360	$2.84	21*	54*	25*
Exactly 0.8	60	$4.31	5	43	52
More than 0.8	180	$4.63*	3	27*	70*

Note. *Differences across treatments are statistically significant at the 95% level using a two-tailed *t*-test.

higher than economic theory predicts.[11] When the punishment factor is greater than 0.8, the average contribution increases to approximately 90% of the maximum possible contribution. Table 2 reports the average group contribution for three cases: when free riding is optimal (punishment factor < 0.8), when all actions are optimal (punishment factor $= 0.8$), and when full contribution is optimal (punishment factor > 0.8).[12] Figure 1 shows the average group contribution by treatment for all punishment factors.[13] In general, group contributions increase as the punishment factor increases (i.e., as the EMPCR increases). Overall, our results suggest that the punishment mechanism encourages compliance (i.e., group contributions) in both treatments, even when the chance of punishment does not affect the Nash prediction.

Table 2 also reports statistical tests of differences between the two treatments.[14] When the punishment factor is greater than 0.8, subjects in the Repeated treatment contribute significantly more on average than subjects in the One-Time treatment. When the punishment factor is equal to and less than 0.8, average contributions between the two treatments are not significantly different.

To analyze the relative effects of punishment probability and severity, we ran a random-effects tobit regression on group contribution controlling for the punishment probability, the punishment severity, the punishment factor, whether it was early or late in the decision

11 See, for example, the summary provided in Ledyard (1995). Interestingly, a number of empirical studies have also found ''overcompliance'' with existing regulations. For example, Harrington (1988) provides a number of environmental examples.

12 Appendix B provides a more complete breakdown of the average contribution by punishment probability and severity.

13 It is important to note that this figure is not constructed to show a time trend. Reading from left to right, average contributions are not reported in the order in which scenarios were listed on decision sheets.

14 Our experiment is not designed to pinpoint the effect of each information-related variable on group contributions, since the informational environment differs in two dimensions moving from one treatment to the other. However, we do test for differences that result from the package of changes considered.

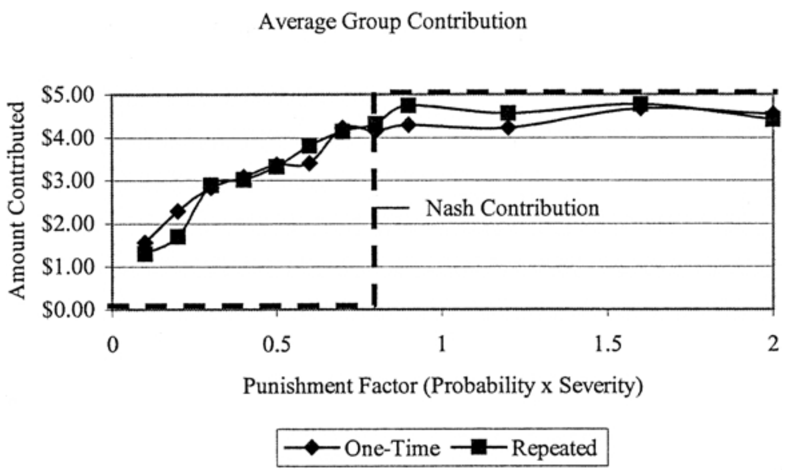

Figure 1. Average group contribution by punishment factor.

making order, and the subject's gender.[15] Period 1 decisions are not included in this regression because there was no punishment in period 1.[16] The estimated coefficients, standard errors, and marginal effects for both treatments are presented in table 3. For both treatments, the coefficients on Punishment Probability and Punishment Severity are positive and significant, as expected, but the coefficient on Probability × Severity is not significant. Thus, subjects do not consider probability and severity to be perfect substitutes. Interestingly, for both treatments the marginal effect of punishment severity is quantitatively larger (by about one third) than the marginal effect of punishment probability, indicating that subjects react more to increases in punishment severity than they do to increases in punishment probability.[17] As noted in section 2, traditional empirical studies find the opposite result, i.e., an increase in punishment probability has a larger effect than a similar increase in punishment severity. However, these results are consistent with the experimental findings of Block and Gerety (1995) reported above. Namely, their student subjects react more strongly to increases in punishment severity than probability, although prisoners react more strongly to increases in punishment probability.

There are two qualitative differences between the results for the two treatments. The first is that the constant in the One-Time treatment is positive and significant while the constant in the Repeated treatment is negative and significant. The lower overall

15 The tobit regression includes an upper limit of $5 and a lower limit of $0. The random-effects model controls for unobserved subject differences by allowing for a subject-specific error component in addition to the customary error term (see Maddala 1987 for a description of this model).

16 The average first period (no punishment) contribution to the group account was $1.87 in the One-Time treatment and $2.37 in the Repeated treatment. Results do not change qualitatively if we include these results in the econometric analysis. However, our model does not converge if we include period one results and a female dummy variable in the regression.

17 To facilitate comparisons between punishment probability and severity, Punishment Probability is equal to the probability of being punished times 10.

Table 3. Results of Random Effects Tobit Regression of Group Contribution

	One-Time Treatment ($N = 570$)			Repeated Treatment ($N = 570$)		
Coefficient	Coefficient	Standard Error	Marginal Effect	Coefficient	Standard Error	Marginal Effect
Constant	0.49*	0.27	—	− 1.31**	0.60	—
Punishment probability	0.58**	0.05	0.37	0.89**	0.10	0.40
Punishment severity	0.79**	0.10	0.50	1.17**	0.20	0.53
Probability × severity	− 0.04	0.03	− 0.02	− 0.03	0.05	− 0.01
Early	− 0.09	0.12	− 0.05	− 0.69**	0.22	− 0.31
Female	0.28	0.17	0.17	0.63	0.48	0.29
Percentage contributing $0			6.1%			12.6%
Percentage contributing $5			31.1%			42.3%
Log likelihood			− 761.9			− 769.9

Notes. *Significant at the 90% level; **Significant at 95% level.

contribution level is consistent with the general finding in the VCM literature that repetition with the same group of people normally deteriorates group account contributions.[18] The second is that the coefficient on Early in the Repeated treatment is both significant and larger (in absolute terms) compared to the insignificant coefficient on Early in the One-Time treatment. The insignificant coefficient in the One-Time treatment makes sense given that subjects in that treatment make all decisions during the same time period without receiving any feedback. The significant negative coefficient in the Repeated treatment indicates that contributions in the Repeated treatment are approximately $0.31 lower in earlier periods (2 through 10) than in later periods (11 through 20). Interestingly, a general finding of the VCM literature is that contributions to the group account fall across decision periods, the opposite of what we find in the Repeated treatment (see Anderson 2001).[19]

Recall that the Repeated treatment differs from the One-Time treatment in the following ways: (1) subjects play the game repeatedly; (2) subjects receive feedback about group contributions; and (3) subjects receive feedback about who has been punished. To further explore how these features affect contribution levels in the Repeated treatment, we ran a second random-effects tobit on the data from the Repeated sessions with a number of additional variables to try to separate these effects to the extent possible. Table 4 reports the results.[20] The results for Punishment Probability, Punishment Severity, Probability × Severity, Early and Female are essentially the same as the results reported in table 3. The variables Selected Last Period and Punished Last Period indicate whether the subject

18 For example, Andreoni (1988) finds that subjects contribute less in repeated encounters than in one-time encounters.

19 We also do not find any effect of gender on contributions, although many studies report higher contributions from females (see Eckel and Grossman 2003).

20 We did not include decisions from period 1 in this regression because there was no punishment in period 1. Additionally, since we include Group Account Last Period and Punished Last Period as explanatory variables, it is not possible to include period 1 in the regression.

Table 4. Alternative Specification for Random Effects Tobit Regression of Group Contribution in the Repeated Treatment

	Repeated Treatment (N = 570)		
Coefficient	Coefficient	Standard Error	Marginal Effect
Constant	− 0.87	0.78	—
Punishment probability	0.93**	0.11	0.42
Punishment severity	1.25**	0.20	0.56
Probability × severity	− 0.04	0.06	− 0.02
Early	− 0.73**	0.26	− 0.33
Female	0.41	0.58	0.28
Selected last period	0.74	0.58	0.33
Punished last period	− 1.18*	0.62	− 0.53
Ever selected	− 1.69**	0.62	− 0.76
Ever punished	1.11*	0.61	0.47
Group account last period	0.01	0.01	0.00
Percentage contributing $0			6.1%
Percentage contributing $5			31.1%
Log likelihood			− 791.3

Notes. *Significant at the 90% level; **Significant at 95% level.

was selected in the previous period for punishment and whether there was an effective punish-ment (i.e., whether the subject has moved any money into his private account). The coefficient on Selected Last Period is positive and not significant while the coefficient on Punished Last Period is negative and significant. Additionally, the net effect of the two variables is negative, indicating that getting punished last period actually decreases the subject's contribution to the group account this period by about $0.20. Ever Selected and Ever Punished indicate whether the subject has ever been selected and has ever had an effective punishment. The coefficients on both of these variables are significant. If someone was selected in the past, but was not caught free riding, his group contribution decreases by $0.77. However, if he was caught free riding at some point in the past, the decrease is only $0.29.

In our setup, being punished in a prior period should not affect a subject's current allocation decision, as people are selected randomly. However, if there were any effect of past audits and punishment on behavior, one would expect them to increase contributions to the group account. For example, in several tax experiments, compliance improved significantly in later rounds among participants that had been audited in earlier rounds (Andreoni et al. 1998, p. 843). One possible explanation for our perverse result is a ''lightning doesn't strike twice'' attitude. If a subject was randomly selected for punishment last period, he may think (incorrectly, of course) that it is unlikely that his letter will come up again and may thus be more likely to free ride.[21] A second possible

21 Camerer (1995) refers to this type of belief as the ''gambler's fallacy,'' and provides the example that there are significantly fewer bets on a specific lottery number in the few days after it is chosen as a winner.

explanation is that subjects who have been punished may feel they cannot afford to contribute to the public good. This is consistent with Block and Gerety's (1995) finding that subjects were more likely to risk punishment the lower were their experimental earnings.

Finally, Group Account Last Period is included in the regression to determine whether subjects respond to information about others' contributions. While the coefficient is positive, suggesting that subjects are positively influenced by the contributions of group members, it is not significant and qualitatively it is very small. This suggests that while feedback about punishment can help to increase compliance, information about overall compliance levels is much less important. These results indicate areas for further research, using both experimental data and data from actual regulatory programs.

6. Summary

We modify a standard public goods experiment to analyze the effectiveness of punishment in inducing regulatory compliance. Specifically, we incorporate a financial penalty into a standard voluntary contribution mechanism to examine the effect of changes in both punishment probability and punishment severity on free riding. Two different treatments are developed to simulate different regulatory environments. The One-Time treatment models regulation of activities that occur very infrequently. The Repeated treatment models regulation of continuous activities.

In both treatments, our findings are generally consistent with the existing experimental literature in terms of average contribution levels and violation of one-shot dominant strategies. The punishment mechanism encourages compliance, i.e., group contributions are increasing in the expected cost of punishment (punishment probability $\times$ punishment severity). Contrary to most empirical evidence, our results indicate that punishment severity has a quantitatively larger effect on behavior than punishment probability. Although this difference could be the result of the measurement problems that exist for most traditional empirical analyses, another possible explanation is differences in risk preferences between the subjects in our experiment and the individuals represented in the empirical data. More work needs to be done to determine why this discrepancy exists, but it does suggest that optimal enforcement policies depend on the risk preferences of the regulated universe.

We also find evidence that past punishment has a negative rather than positive effect on contribution levels, perhaps due to a "lightning doesn't strike twice" attitude. The results of this experiment suggest that this is a fruitful area for research. In particular, further study of the relative effect of punishment severity and punishment probability, the effect of different forms of feedback, and the effect of punishment history on behavior appears warranted.

Appendix A: Instructions for Repeated Treatment

This experiment is a study of individual behavior. The instructions are simple. If you follow them carefully and make good decisions, you may earn a considerable amount of money, which will be paid to you privately in cash at the end of the experiment today.

In this experiment you will make a decision in each of 20 rounds. At the end of the experiment, we will throw a 20-sided die to determine which round will be used to determine your earnings. The throw of the 20-sided die guarantees that all 20 rounds are equally likely to be chosen for payment, so you should think carefully about each decision. Your earnings will depend on the decisions that you and the other members of this group make. Below, we describe how your earnings are determined.

Your Decision

We have created a private account for each person participating in this experiment. Your private account currently contains $15 dollars. You and the other nine people completing this experiment also share a group account that currently contains $50. Each of you will be given control over $5 from this group account. In each round of the experiment, you will be choosing what to do with this money. You may choose either (1) to keep the money in the group account or (2) move some or all of the money to your private account.

Once each person has made his or her decision, we will determine the amount remaining in the group account and will then double it. Once we have doubled the group account, it will be evenly distributed to all members of the group. You share equally in the group account regardless of whether you leave your money in the group account or move some to your private account. For example, if there was $10,000 in the group account, we would double it for a total of $20,000. Each group member would then receive $2000.

You may choose to move some of the $5 over which you have control into your private account. We will not double your private account. Additionally, you may not move any of the $15 in your private account to the group account. Your task is to decide how much of the $5 to leave in the group account and how much to move to your private account. You may choose to move some money into your private account and leave some in the group account. Alternatively, you can leave all of the money in the group account or move all of it to your private account.

For some people, earnings will simply equal their share of the group account plus the amount in their private account (the $15 placed in the private account at the beginning of this experiment plus any money moved from the group account to the private account). Recall that group account earnings are determined by multiplying the total dollars left in the group account by 2 and dividing that total amount evenly among the 10 members of your group.

Alternatively, some people may have their earnings adjusted for some rounds. In each round, a number of people will be randomly selected to have their earnings adjusted. The number of people whose earning will be adjusted varies across rounds, as shown on the RECORD SHEET at the back of these instructions. We will decide the specific people in the group who will have their earnings adjusted by drawing scrabble tiles with letter A through J (to match the letters on your folders). For example, note that 8 people will have their earnings adjusted for round 4. If that decision is chosen (by the throw of the 20-sided die) to determine earnings for this experiment, we will draw 8 scrabble tiles, without replacement, from this cup.

If you are chosen to have your earnings adjusted, the amount of the adjustment is based on the dollars moved to your private account and the adjustment factor listed in column (E). Specifically, if you are chosen to have your earnings adjusted, you must subtract the dollars moved to your private account (listed in column B) times the adjustment factor (listed in column E) from your net earnings in the experiment. Note that a significant earnings adjustment may require you to return some of the $15 placed in your private account at the beginning of the experiment, but it is mathematically impossible for you to make less than $0 in the experiment.

Procedures

When we begin the experiment, you will make a decision about how to allocate your money in the first round. The procedure to follow for each round is as follows. First, you should record the number of dollars you want left in the group account in column (A) and the number of dollars you want moved to your private account in column (B). Notice that the sum of column (A) and (B) should equal $5 for **each** separate decision-making round. Also record your decision on the piece of paper in your folder labeled with your player letter (the letter on your folder) and Round 1. **Do not discuss your decision with any other participant!** We will collect everyone's decisions and determine how much was allocated to the group account and double it. We will then announce the group account amount which you should record in column (E) on your record sheet. Next we will randomly select the people whose earning will be adjusted by drawing the appropriate number of scrabble tiles. If your letter is selected, please indicate that on your RECORD SHEET by placing an asterisk next to the round number. Once a round is finished, you will make your decision for the next round.

Once all 20 rounds are completed, we will throw a 20-sided die to determine which round will be used to determine your earnings. You should use the form at the bottom of the record sheet to calculate your earnings at the end of the experiment.

Before we begin the decision making period, you will make some hypothetical decisions so we can verify that you understand how earnings are calculated. **These decisions will not affect your actual earnings in any way**.

1. Choose how your $5 will be split between the group account and your private account.
   amount 1a.________ Dollars left in group account
   amount 1b.________ Dollars moved to private account
   Recall that the amounts above must sum to $5.

2. Choose the total amount left in the group account by the other 9 members of your group. Note that in the actual round that determines your earnings, your group members will choose their own group account allocations.
   amount 2._____ Total dollars left in the group account by the other 9 members of your group.
   Recall that this amount must be between $0 and $45 since each of the 9 members of your group has $5 to allocate.

3. Calculate your earnings from the group account.
   _________ + ________ = ________*amount 3a*.
   (*amount 1a*) (*amount 2*)

_________ * 2 = _________ *amount 3b*.
(*amount 3a*)
_________ / 10 = _________ *amount 3c*. = Your earnings from group account.
(*amount 3b*)

4. Calculate your earnings for the hypothetical information you chose above assuming that **earnings are not adjusted**.
_________ + _________ = _________ *amount 4*. = Your total earnings.
(*amount 1b*) (*amount 3c*)

5. Recalculate your earnings for the hypothetical information you chose above assuming that your earnings are adjusted with an **adjustment factor of 1**.
_________ minus _________ * 1 = _________ Your adjusted earnings.
(*amount 4*) (*amount 1b*)

6. Recalculate your earnings for the hypothetical information you chose above assuming that your earnings are adjusted with an **adjustment factor of 3**.
_________ minus _________ * 3 = _________ Your adjusted earnings.
(*amount 4*) (*amount 1b*)

Record Sheet Player Letter:

Round	Dollars left in group account (A)	Dollars moved to private account (B)	Total Amount to Allocate (A) + (B) = (C)	Number of people to have earnings adjusted (D)	Adjustment Factor (E)	Group Account (after doubling) (F)
1			$5	0	0	
2			$5	3	1	
3			$5	3	2	
4			$5	8	2	
5			$5	6	2	
6			$5	2	2	
7			$5	2	4	
8			$5	2	1	
9			$5	2	3	
10			$5	3	3	
11			$5	3	4	
12			$5	5	4	
13			$5	5	1	
14			$5	9	1	
15			$5	4	1	
16			$5	7	1	
17			$5	8	1	
18			$5	1	1	
19			$5	1	2	
20			$5	1	4	

Calculate your earnings:

Step 1: Record your earnings from the group account (Amount in column (F)/10) for the round chosen by the throw of the 20-sided die.		____________
Step 2: Add the amount listed in column (B) for the round chosen by the throw of the 20-sided die. Skip to Step 4 if you were not chosen to have your earnings adjusted.	**plus**	____________
Step 3: If you were chosen to have your earnings adjusted, multiply the amount in column (B) times the adjustment factor in column (E) for the chosen decision.	**minus**	____________
Step 4: Make the appropriate calculations.	**Earnings:**	____________
	plus	$15.
		____________
Total amount to record on receipt:		____________

Appendix B: Individual Behavior Patterns

Tables B-1 and B-2 show average group contribution by punishment probability and punishment severity for the One-Time and Repeated treatments respectively. Notice that for a given punishment probability, average contributions are always increasing with the punishment severity. For a given punishment severity, average contributions increase as punishment probability increases in all but one cases. The patterns displayed in these tables do not hold at the individual level. Of the 60 subjects participating in the experiment, only 12 (7 from the One-Time treatment and 5 from the Repeated treatment) always responded to higher punishment probabilities or severity (holding the other constant) consistently, that is by increasing their group account allocation or holding it constant. The remaining 48 subjects had at least one ‘‘inconsistent decision,’’ which we define as a lower group contribution for a higher punishment factor. Subjects were slightly more likely to respond consistently with respect to punishment severity than with respect to punishment probability: 26 subjects (14 from the One-Time treatment and 12 from the Repeated) had consistent responses with respect to severity while only 15 (9 from the One-Time and 6 from the Repeated) had consistent responses with respect to probability. The higher consistency of subjects in the One-Time treatment compared to the Repeated treatment is not surprising given that subjects in the Repeated treatment may have been influenced by past results and may have engaged in strategic behavior to influence others' decisions. The high number of inconsistent decisions in the One-Time treatment is a puzzling result, given that all decisions were made at the same time.

Table B-1. Average Group Contribution for One-Time Treatment

Punishment Severity	Punishment Probability								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	\$1.56	\$2.29	\$2.83	\$3.16	\$3.63		\$4.22	\$4.41	\$4.77
2	\$2.28	\$2.76	\$3.27			\$4.23		\$4.67	
3		\$3.53	\$3.81						
4	\$3.30	\$3.88	\$4.21		\$4.54				

Table B-2. Average Group Contribution for Repeated Treatment

Punishment Severity	Punishment Probability								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	\$1.30	\$1.49	\$2.89	\$3.30	\$3.32		\$4.14	\$4.69	\$4.97
2	\$1.95	\$2.76	\$3.68			\$4.49		\$4.77	
3		\$3.92	\$4.51						
4	\$3.00	\$3.93	\$4.63		\$4.41				

References

Alm, J., B. R. Jackson and M. McKee. 1992. ''Estimating the Determinants of Taxpayer Compliance with Experimental Data.'' *National Tax Journal* 45(1): 107–114.

Anderson, L. R. 2001. ''Public Choice as an Experimental Science.'' In *The Elgar Companion to Public Choice*, edited by William Shughart, and Laura Razzolini. Northampton, MA: Edward Elgar Publishing.

Andreoni, J. 1988. ''Why Free Ride? Strategies and Learning in Public Goods Experiments.'' *Journal of Public Economics* 37: 291–304.

Andreoni, J. 1995. ''Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments.'' *Quarterly Journal of Economics* 110: 1–21.

Andreoni, J., B. Erard, and J. Feinstein. 1998. ''Tax Compliance.'' *Journal of Economic Literature* 36(2): 818–860.

Beck, P. J., J. S. Davis, and W.-O. Jung. 1991. ''Experimental Evidence on Taxpayer Reporting under Uncertainty.'' *Accounting Review* 66(3): 535–558.

Becker, G. 1968. ''Crime and Punishment: An Economic Approach.'' *Journal of Political Economy* 76(2): 169–172.

Becker, W., H.-J. Buchner, and S. Sleeking. 1987. ''The Impact of Public Transfer Expenditures on Tax Evasion: An Experimental Approach.'' *Journal of Public Economics* 34(2): 243–52.

Block, M. K., and V. E. Gerety. 1995. ''Some Experimental Evidence on Differences Between Student and Prisoner Reactions to Monetary Penalties and Risk.'' *Journal of Legal Studies* 22: 123–138.

Camerer, C. 1995. ''Individual Decision Making.'' In *The Handbook of Experimental Economics*, edited by John Kagel, and Alvin Roth. Princeton, NJ: Princeton University Press.

Cameron, S. 1988.''The Economics of Crime Deterrence: A Survey of Theory and Evidence.'' *Kyklos* 41: 301–323.

Cardenas, J. C., J. Stranlund, and C. Willis. 2000. ''Local Environmental Control and Institutional Crowding Out.'' *World Development* 28(10): 1719–1733.

Davis, D. D., and Charles A. Holt. 1993. *Experimental Economics*. Princeton, NJ: Princeton University Press.

Eckel, C., and P. Grossman. 2003. ''Differences in the Economic Decisions of Men and Women: Experimental Evidence.'' Forthcoming in the *Handbook of Experimental Economic Results*, edited by Charles Plott and Vernon Smith. New York, NY: Elsevier Press.

Ehrlich, I. 1996. ''Crime, Punishment, and the Market for Offenses.'' *The Journal of Economic Perspectives* 10: 43–67.

Fehr, E., and S. Gächter. 2000. ''Cooperation and Punishment in Public Goods Experiments.'' *American Economic Review* 90(4): 980–994.

Goeree, J., C. Holt, and S. Laury. 2002.''Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior.'' *Journal of Public Economics* 83(2): 255–276

Grogger, J. 1991. ''Certainty vs. Severity of Punishment.'' *Economic Inquiry* 29(2): 297–309.

Harrington, W. 1988. ''Enforcement Leverage When Penalties are Restricted.'' *Journal of Public Economics* 37: 29–53.

Heyes, A. 2000.''Implementing Environmental Regulation: Enforcement and Compliance.'' *Journal of Regulatory Economics* 17(2): 107–129.

Isaac, R. M., D. Schmidtz, and J. M. Walker. 1989. ''The Assurance Problem in a Laboratory Market.'' *Public Choice* 62: 217–236.

Isaac, R. M. and J. M. Walker. 1988. ''Communication and Free-Riding Behavior: The Voluntary Contributions Mechanism.'' *Economic Inquiry* 26: 585–608.

Ledyard, J. O. 1995. ''Public Goods: A Survey of Experimental Research.'' In *The Handbook of Experimental Economics*, edited by John Kagel and Alvin Roth. Princeton, NJ: Princeton University Press.

Maddala, G. S. 1987. ''Limited Dependent Variable Models Using Panel Data.'' *Journal of Human Resources* 22: 307–338

Ostrom, E., J. Walker, and R. Gardner. 1992. ''Covenants With and Without a Sword: Self-Governance is Possible.'' *American Political Science Review* 86: 404–417.

Polinsky, A. M., and S. Shavell. 2000. ''The Economic Theory of the Public Enforcement of Law.'' *Journal of Economic Literature* 38: 45–76.

Pommerehne, W. W., and Hannelore Weck-Hannemann. 1996. ''Tax Rates, Tax Administration and Income Tax Evasion in Switzerland.'' *Public Choice* 88(1–2): 161–170.

Trumbull, W. N. 1989. ''Estimations of the Economic Model of Crime Using Aggregate and Individual Level Data.'' *Southern Economic Journal* 56(2): 423–439.

Tyran, J.-R., and L. P. Feld. 2001. ''Why People Obey the Law: Experimental Evidence from the Provision of Public Goods.'' Working paper, Department of Economics, University of St. Gallen.