Appalachian
STATE UNIVERSITY
BOONE, NORTH CAROLINA

Observed Punishment Spillover Effects: A Laboratory Investigation Of Behavior In A Social Dilemma

By: **David L. Dickinson**, E. Glenn Dutcher, & Cortney S. Rodet

Abstract

Punishment has been shown to be an effective reinforcement mechanism. Intentional or not, punishment will likely generate spillover effects that extend beyond one's immediate decision environment, and these spillovers are not as well understood. We seek to understand these secondary spillover effects in a controlled lab setting using a standard social dilemma: the voluntary contributions mechanism. We find that spillovers occur when others observe punishment outside their own social dilemma. However, the direction of the spillover effect depends crucially on personal punishment history and whether one is personally exempt from punishment or not.

Dickinson, D.L., Glenn Dutcher, E. & Rodet, C.S. Exp Econ (2015) 18: 136. https://doi.org/10.1007/s10683-014-9399-7. Publisher version of record available at: https://link.springer.com/article/10.1007/s10683-014-9399-7

Observed punishment spillover effects: a laboratory investigation of behavior in a social dilemma

David L. Dickinson · E. Glenn Dutcher · Cortney S. Rodet

Abstract Punishment has been shown to be an effective reinforcement mechanism. Intentional or not, punishment will likely generate spillover effects that extend beyond one's immediate decision environment, and these spillovers are not as well understood. We seek to understand these secondary spillover effects in a controlled lab setting using a standard social dilemma: the voluntary contributions mechanism. We find that spillovers occur when others observe punishment outside their own social dilemma. However, the direction of the spillover effect depends crucially on personal punishment history and whether one is personally exempt from punishment or not.

Keywords Punishment · Punishment spillovers · Experiential punishment · VCM · Social dilemma · Experiment

1 Introduction

If punishment alters behavior beyond those directly punished, then the typical model popularized by Becker (1968), which highlights only pecuniary costs and

Electronic supplementary material The online version of this article (doi:10.1007/s10683-014-9399-7) contains supplementary material, which is available to authorized users.

D. L. Dickinson
Department of Economics, Appalachian State University, Boone, NC 28608-2051, USA

E. Glenn Dutcher (✉)
Department of Economics, Finance & Marketing, University of Central Missouri, Warrensburg, MO 64093, USA
e-mail: glenn.dutcher@gmail.com

C. S. Rodet
Economic Science Institute, Chapman University, Orange, CA 92886, USA

benefits, may be incomplete. For instance, punishment could be more effective than traditionally modeled if the observers of punishment subsequently reduce their own undesirable behavior. However, the opposite is also possible: the behavioral improvements of the punished individual could be offset by an adverse reaction from those observing the punishment. It is the goal of this paper to strengthen our understanding of punishment mechanisms by using experimental methods to explore how the observation of others being punished affects behavior when all aspects of punishment are exogenous.

Punishment is intended to discourage undesirable behavior or actions,[1] with effectiveness shown to depend on its frequency, intensity and immediacy (Anderson and Stafford 2003; Johnston 1972). Thus, punishment usually conveys information that adherence to a behavioral standard is socially beneficial and supports a collective goal (Sunstein 1996; Kahan 1998; Cooter 1998; Tyran and Feld 2006; Xiao 2013). Punishment therefore provides information on the acceptability of a behavioral standard, as well as providing an incentive to adhere to it. Of course, the effectiveness of punishment depends greatly on the group members' regard for the established standard. If a boss institutes a particular rule that she sees as integral to the success of the firm, but that the employees find useless, punishment for breaking this rule may not improve behavior and could even have the opposite effect (Trevino 1992).[2] Consider, for example, laws against marijuana use. Through the imposition of exogenously imposed rules, punishment of other users may signal acceptability of the behavior for marijuana users. Thus, even though punishment is meant as a means of social norm enforcement, whose norm is being enforced may alter how others respond to observing punishment. If a new national law is imposed locally (i.e., an exogenously imposed new norm), local citizens who do not observe anyone being punished for violating this law will likely learn that the social norm is to obey the law. On the other hand, if they observe others being punished for breaking the law, they may learn that this behavior is socially acceptable if their relevant social group holds different norms of behavior. This could partially explain the mixed evidence on the effect of decriminalization of cannabis use; some reports actually find that increased marijuana punishments lead to an increase in its use.[3]

In other words, if the probability or severity of punishment is endogenously determined by one's social group, observation of punishment conveys information on internal group preferences and thus establishes an internal or localized social norm. This stands in contrast to situations where the punishment is entirely exogenous or set by a fixed rule. Only in such situations of exogenous punishment is

[1] See Andreoni et al. (2003), Fehr and Fischbacher (2004), Ostrom et al. (1992), and Xiao and Houser (2011) for examples.

[2] The same argument can be made for exogenously imposed laws on groups such as gangs or members of the mafia where they likely do not believe the behavioral standard being enforced is important for their group.

[3] A report on the current knowledge was made to the Judiciary Committee of the Connecticut General Assembly in 1997 where they state in Part 2 of section 4D "However, the increase in marijuana use was even greater in other states and the largest proportionate increase occurred in those states with the most severe penalties." The report can be accessed via http://www.cga.ct.gov/lrc/drugpolicy/drugpolicyrpt8.htm#Sec4D.

there the possibility that punishment may or may not signal acceptability of a behavior, depending on whether one's local group holds different norms than those setting the punishment rules. Our paper is novel in that we experimentally investigate a situation where punishment can be observed but not received, and we do so using entirely exogenous punishment rules so that strategic concerns are eliminated.

The previous literature has identified two effects that may result from the observance of punishment in social dilemmas, and these effects lead to divergent predictions: social learning, and what we have termed 'experiential learning'. A standard public goods provision setting should cause money maximizers to free-ride, or condition their cooperation on the cooperation of others, but complete free riding is not typically observed. It may be that free-riding is initially viewed as socially unacceptable, and people have an aversion to violating social norms.[4] Thus, adding the element of observed punishment to the social dilemma may engage "social learning", which teaches one through a 'punishment signal' that some consider it acceptable to free ride. Seeing others who are willing to repeatedly incur punishment by not contributing to the public good could serve as a very strong signal that low contributions are socially acceptable. Since social norms are learned by observing the others' actions, social learning predicts that observance of punishment will lead to lower levels of cooperation because the benefit of higher levels of cooperation is reduced (Bikhchandani et al. 1992).[5]

The second effect identified in the literature, which contrasts with the social learning effect described above, is that some subjects react to observing punishment as if they experienced it personally, leading to a deterrence effect where pro-social behavior is generally increased.[6] Previous studies on first-person punishment find that subjects often increase their contributions to the public good following punishment (e.g. Anderson and Stafford, 2003). Therefore, an "experiential punishment" effect in our setting where punishment is merely observed implies that if subjects internalize punishment as if it happened to them, they will become more cooperative. In other words, this experiential punishment effect predicts that observing others being punished will lead one to refrain from the punishable behavior.

Due to the lack of control in naturally occurring data, we utilize a laboratory environment where subjects played a linear public goods game in groups of three. In one key treatment, subjects were not punished, but observed if someone else was punished. In naturally occurring settings, it is often the case that the probability of

4 Note that the aversion to local social norm violation has been shown to lead to third-party punishment in Fehr and Fischbacher (2004). This behavior is seen in other settings as well. For instance, Cooper and Glenn Dutcher (2011) show that proposer behavior in an ultimatum game is consistent with the proposer learning which offers are socially acceptable (i.e., socially unacceptable proposals are rejected and the proposers thus learn the social norms. As the saying goes, "When in Rome, do as the Romans do.").

5 There is some confusing usage of the term "social learning" applied to this setting. We are referring to the usage "localized conformity" in Bikhchandani et al. (1992), and not the usage in Arvey et al. (1985), Trevino (1992) and references therein.

6 This internalization of observation of punishment has been labeled vicarious punishment in psychology (Malouff et al. 2009).

punishment is sufficiently low that people do not think they will get caught (Erickson et al. 2013). That is, they will behave as if a small probability of punishment is the same as a zero probability. Our treatment is in line with the use of such mental shortcuts, or decision heuristics (Harsanyi and John 1986), but it also allows us to understand how social norms are transmitted in an environment free of personal punishment. We focus on comparisons of subjects' contributions to the public good when they did and did not observe punishment. We introduce a second key treatment in which subjects both observe punishment and can themselves be punished. Varying the order of these two treatments allows us to manipulate one's history with punishment institutions. Experiential punishment may be more likely when personal punishment history exists, because it can be argued that in order for subjects to react to observing others' punishment, they must first know how they would react if punished themselves.

The social dilemma we utilize is the standard voluntary contribution mechanism (VCM). Our main contribution is that we identify both a positive and a negative spillover effect of observed punishment and we pinpoint when each effect can be expected. Specifically, if a subject was exempt from personal punishment, then she responds to observed punishment by reducing VCM contributions. On the other hand, if a subject was not exempt from personal punishment, then she responds to observed punishment by increasing VCM contributions. Both of these effects are evident only if the subjects had not previously been exposed to any punishment mechanism (i.e., Game 2 of our 3 game design) and both effects decrease over time.

2 Brief review of related literature

The study of punishment effects in economics has its traditions in the law and economics literature. The Becker (1968) or Tullock (1974) approach to crime rests on the rational utility maximization model where deterrence is a function of the probability of being caught, and prosecuted, and the relative size of the payoff from the crime.

Other research exposed the limits of this approach. For instance, Myers (1982) shows that the relative means and variances of legal and illegal income affect the supply of criminal behavior. Likewise, increased public deterrence efforts may substitute for private protection, leading to an increase in criminal behavior (Skogh and Stuart 1982). Spatial spillovers also raise doubts regarding net crime deterrence. If certain regions increase deterrence efforts, therefore becoming crime exporters, surrounding regions may experience higher or lower crime depending on the community's reaction function (Cameron 1988). Industrial organization models of crime networks raise the possibility that the closure of incumbent 'firms' due to deterrence efforts may raise the level of crime, at least in the short run, if competing firms jockey for positions in the markets for illicit goods and services (Buchanan and James 1973). It is also possible that criminal monopolies treat increased efforts as increases in fixed costs of operation, which do not affect criminal behavior on the margin (Holahan 1973). There are also behavioral questions regarding adaptive behavior. As Harsanyi (1986) points out, decreasing the certainty of criminal success from .9 to .8 may be irrelevant if individuals' use heuristics and treat both of these outcomes as if they implied certain success.

Note that the aforementioned literature focuses on the certainty of detection (and prosecution) and the relative size of the payoff. Our focus is on the behavioral response to observing punishment when these variables are known, exogenously determined, and held fixed, which has yet to be studied as closely as laboratory experiments allow. Furthermore, an efficiency argument, supported by behavioral economics insights, has been made regarding how to focus resources in a system meant to deter anti-social or criminal behavior (Harell and Segal 1999). If would-be transgressors prefer more (less) uncertainty regarding detection and less (more) uncertainty regarding punishment, the deterrence system ought to be structured to do the opposite in order to deter crime. Our study contributes to this understanding by analyzing the effects of observable deterrence efforts on cooperative behavior, where we isolate in-group cooperation concerns by making observed punishment exogenous to a subject's VCM group. Another benefit of the experimental methodology is to avoid numerous econometric challenges, like simultaneity and bias, encountered when analyzing reported crime data.

Deterrence has been considered a sign of norm internalization where individuals change their preferences in a Pareto improving manner due to the fear or expected cost of punishment (Cooter 1998). Our environment allows us to observe whether individuals internalize the exogenously established norm or not in instances where they are either exempt or nonexempt from punishment. A crucial factor in determining norm internalization is the respect individuals hold for the norm—if it is considered to be illegitimate, internalization will not occur.

Prior research related to our present study of observational punishment has focused on how subjects update an unknown probability of punishment or severity of punishment (or both). Papers in that line of research are very well suited for answering questions relating to endogenously determined norm communication, but not well suited for our purposes. To better illustrate this point, the following model is a typical specification of a subject's utility in public goods games where the social optimum is full contributions to the public good, but the individual payoff-maximization incentives are to contribute nothing.

$$U_{i,P}(x) = \alpha(E - x_i) + \beta \sum_{j=1}^{N} x_j - \mathrm{Pr} \times d \qquad (1)$$

where x_i is the amount contributed to the public good by subject *i, Pr* is the probability of punishment of level *d*. Using this model, Xiao and Houser (2011) examine how the observance of exogenously imposed punishment to fellow group members affects contributions.[7] In their study, they used a repeated linear public goods game where rounds were monitored with probability of 50 %. If a round was

[7] We note that Xiao and Houser (2011) was well designed to answer their research question about how public punishment conveys localized cooperative norms better than private punishment. Our outline here is to merely highlight the differences in their design and our own in that theirs relies partially on endogenous punishment factors while ours does not. Thus, the environments they address and the ones we address are quite different.

monitored, the subject who contributed the lowest amount was punished and the punishment severity was larger the farther the subject was from the average contribution level of his/her other group members. So, subjects can be assumed to be maximizing their utility given uncertainty about *Pr* and *d*. Importantly, because the lowest contributor was punished and the severity of the punishment is a function of one's group members' contribution levels, this key element of punishment is, in effect, endogenous in their design.

In a similar vein, Dickinson (2001) fixes the amount of (private) punishment, but again the punishment is given to the lowest contributors. Since one's optimization is conditional on a rank ordering of others' contributions, behavior can be expected to change as more information is gathered on what others contribute and subjects update beliefs about the actual values of *Pr* and *d* for a given level of contribution. Schnake (1986) examined college students' reaction to a confederate being punished for low output when subjects had no idea that punishment of any sort was possible. Once subjects observed the punishment of the confederate, they were able to update their prior (presumably incorrect) beliefs of *Pr* and/or *d*. In both of these studies (and others like these), observance of punishment leads to an increase in the desirable behavior. Though it is important to understand how subjects update their prior beliefs, our design focuses on whether the behavioral effects of observed punishment are more consistent with experiential punishment or with the social learning we describe above, which is most easily assessed without the added complication of information updating and/or strategic concerns.

Our study diverges from prior studies by examining the behavioral implications of observing punishment that has no direct impact on one's *monetary* payoff. More specifically, assume agent *i* only observes punishment and this observation does not affect monetary payoffs in agent *i's* social dilemma. We can represent this with the following function

$$U_{i,C}(x) = \alpha(E - x_i) + \beta \sum_{j=1}^{N} x_j + C(x_i, P|V) \tag{2}$$

where $C(.)$ is a function accounting for the psychological cost (or benefit) of observing punishment, P is an indicator equal to one if punishment is observed while V represents personal characteristics, which map the agent's reaction to observed punishment into her contributions, x_i.[8] In essence, the personal characteristics, V, determine if the observance of punishment is a psychic cost or benefit, which in turn may lead the agent to either increase or decrease her contributions to the public good. One of our objectives is to uncover some of the specific factors that influence the sign and magnitude of $C(.)$. Behaviorally, if a characteristic in V determines that the sign of $\frac{d^2C(.)}{d_{x_i}dP}$ is negative, this means that an agent should decrease contributions in order to increase utility and vise-a-versa if $\frac{d^2C(.)}{d_{x_i}dP}$ is positive.

[8] Notice that $\frac{dC(.)}{dP} = 0$ when P = 0. See Ku and Salmon (2010) for a similar model.

3 Experimental design

Subjects were recruited through ORSEE (Greiner 2004) and the experiments were programmed in z-Tree (Fischbacher 2007). Experimental sessions lasted about an hour and average subject payoffs were $20.46. In what follows we refer to super-groups and VCM groups to avoid confusion. For example, if a session involved 18 subjects, two super-groups of 9 subjects were randomly formed and remained fixed for the entire experimental session. We will refer to these groups as super-group 1 and super-group 2.[9] Subjects played 3-person VCM games with other subjects within their super-group, and we refer to these 3-person groups as the VCM groups. Two super-groups allow us to randomly match each subject from super-group 1 with a subject from super-group 2 for the purposes of observed punishment treatments while ensuring that these two subjects never play in a VCM game with each other. The two subjects from different super-groups will be referred to as each respective subject's "other group counterpart" (OGC). These OGC pairs remained fixed throughout the experiment.

In each session, there were three 10-period games played by each VCM group: Game A, Game B, and Game C. Before each game, subjects were given written instructions that were read out loud by the experimenter, followed by a short quiz to ensure comprehension of instructions. Afterwards, the first 10-period game (Game A) started. Subjects were given information only about the current treatment but knew others would follow. The same procedure was carried out for all games A,B, and C. Sample instructions are given in the Appendix. Game A is a standard VCM game, which was used to give subjects experience playing the game and allowed learning about the social norms. Subjects allocated 10 tokens to an "individual" or "group" account in whole token increments. Tokens allocated to the individual account yielded $0.025 to that person alone and tokens allocated to the group account returned $0.0125 to all three members of that subject's VCM group. Thus, viewed from the perspective of a single decision round, a money-maximizing agent should place all 10 tokens in the individual account (earning $.25) while the social optimum implies allocation of all 10 tokens to the group account (earning $.375). No punishment exists in Game A, but we maintain a sequential move structure where one super-group makes decisions first. This is done to maintain consistency across Games A, B, and C to ensure that results in Games B and C are not simply a function of sequential decisions across super-groups 1 and 2.

Game B was similar to A except that a punishment mechanism was introduced. Punishment only applied to VCM groups within super-group 1 in Game B. If a VCM group member contributed less than 5 tokens to the group account, there was a 50 % chance of being punished by losing $0.125 (equal to an expected punishment cost of $0.0625, or 2.5 tokens) of her period earnings. This still leaves the strategy structures weakly intact for risk-neutral money-maximizing agents.

[9] We referred to these as Group A and Group B in the experiment instructions. We changed the labels here to Super-group 1 and 2 to avoid confusion in the text when referring to games A, B and C, which are the treatments. In the instructions, treatments were unnamed and simply described by the rules that would apply to each particular block of trials.

In each round of Game B the subjects in super-group 1 made a decision and outcomes (i.e., private payoffs, group payoffs and punishments) were determined. Thereafter, the partner from the unpunished super-group 2 is informed whether her OGC was punished or not and how often they had been punished previously and the costs of such punishment. Those in super-group 2 then made a decision in their respective VCM game without the possibility of being personally punished. We stress that, because VCM groups only include subjects from the same super-group, any behavioral response by a super-group 2 subject cannot be strategic or reciprocal.

Game C is similar to Game B except now subjects from both super-groups 1 and 2 are subject to the same punishment mechanism. Thus, subjects in super-group 2 can observe punishment *and* can be directly punished themselves in Game C. The sequence of decisions was the same. Subjects played all three games in a single session where Game A was always played first. To control for order effects across games, half the sessions were ran in game order ABC and half in order ACB. Counterbalancing the game order for Games B and C is also necessary to generate the differential punishment history needed.

In summary, the main goal of the design was to tightly control what was learned from observing punishment in order to clearly identify the effects without data complications found in other studies (which, to be fair, were designed to answer somewhat different questions). Specifically, the design we implemented precludes conditional cooperation, updating of the probability of punishment and/or strategic learning as a potential cause of behavioral changes due to the observance of OGC punishment.

4 Results

Eight sessions—four with game order ABC and four with game order ACB—were run at a large American university. In total there were 144 subjects. It turns out that the game order matters because our results show that the lack of experience or history with punishment institutions is necessary for punishment spillover effects to display among those players in super-group 2. In our description of results that follow, recall that in Game B, subjects from super-group 2 were exempt from punishment whereas in Game C they were non-exempt and subject to punishment themselves.

4.1 Overview

Figures 1, 2, and 3 give an overview of the results at the more aggregate level. In the related statistical tests of the pooled data, the unit of observation is the individual, though many would consider that independent observations in a VCM context only occur at the session level. For this reason, we intend for the results derived from Figs. 1, 2, and 3 to merely give some indication of the tendencies in the data. We are more careful to control for other relevant factors and non-independence of errors (across periods for a given subject) in the multi-variate analysis that follows. We

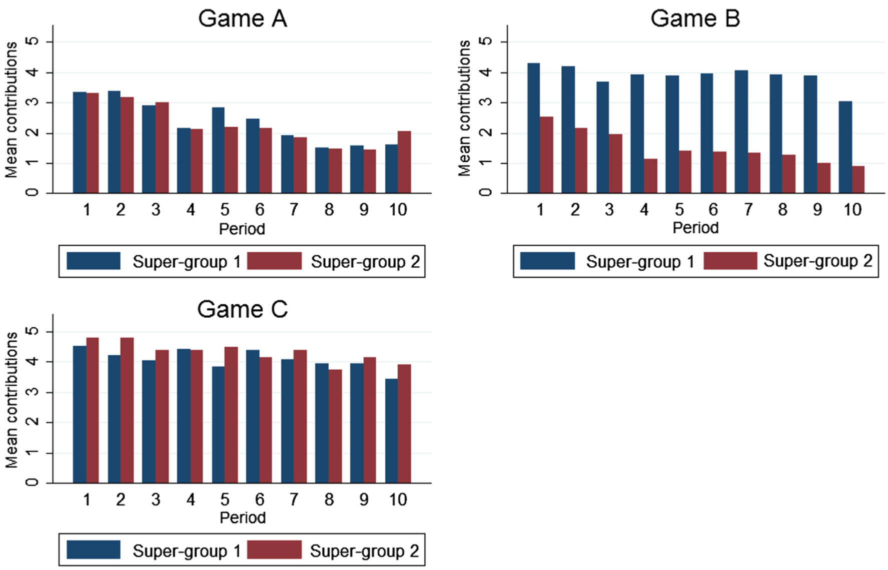

Fig. 1 Mean contributions. Mean contributions in all games by super-group over time

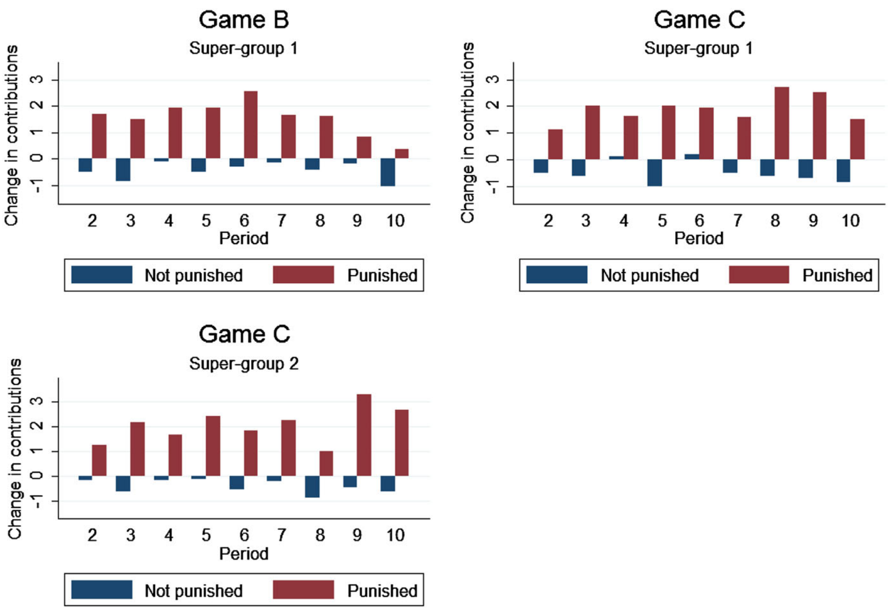

Fig. 2 Mean change in contributions Conditional on personal punishment. The mean change in contributions of both super-groups in both games the period after a subject was or was not personally punished

should highlight, however, that our results are strengthened by the more proper econometric analysis that we will show in Tables 1 and 2, which are used to produce Fig. 4 as well.

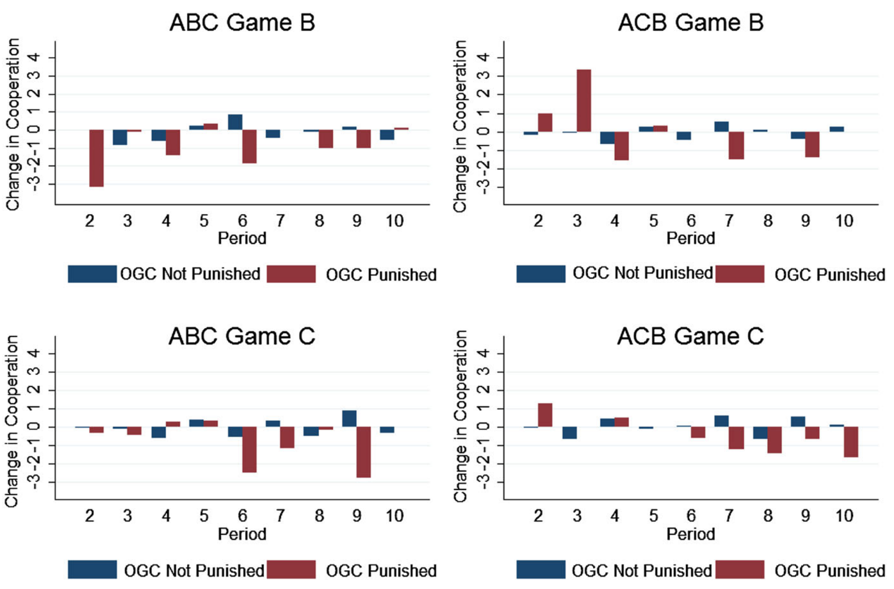

Fig. 3 Mean change in contributions of super-group 2. The mean change in contributions of super-group 2 conditional on game and game order

Table 1 Change in cooperation (Game B)

Variables	(1) ABC	(2) ACB
OGC punished	−1.70** (0.69)	0.73 (0.85)
OGC punished × period	0.17* (0.09)	-0.15 (0.11)
Period	0.04 (0.04)	0.04 (0.04)
Others' contributions previous period	0.06* (0.04)	0.01 (0.04)
Subject's first period contribution	−0.05* (0.03)	−0.02 (0.02)
Constant	−0.32 (0.26)	−0.28 (0.31)
Observations	324	324
Number of subject	36	36

Random effects panel regression of cooperation of super-group 2 subjects in Game B

Robust standard errors are clustered at the individual level and appear in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 1 shows the average contributions each period of those in super-group 1 compared to super-group 2 in Games A, B and C. In Game A, the average contribution of those in super-group 1 (2.37 tokens) is close to the average contributions of those in super-group 2 (2.28; p value = 0.76).[10] The typical decay over time is also seen as both groups contribute more in the first few periods than in

[10] Unless denoted otherwise, the p values are the result of a two-sided *t* test where standard errors are clustered by subject to control for the obvious non-independence of observations over time.

Table 2 Change in cooperation (Game C)

Variables	(1) ABC	(2) ACB
OGC punished	0.40 (0.82)	1.85** (0.93)
OGC punished × period	−0.15 (0.14)	−0.36** (0.18)
Period	0.03 (0.04)	0.02 (0.05)
Personally punished last period	3.23*** (0.65)	2.19*** (0.61)
Others' contributions previous period	0.04* (0.02)	0.01 (0.03)
Subject's first period contribution	−0.05 (0.04)	0.04 (0.03)
Constant	−0.71* (0.40)	−0.70* (0.38)
Observations	324	324
Number of subject	36	36

Random effects panel regression of cooperation of super-group 2 subjects in Game C

Robust standard errors are clustered at the individual level and appear in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

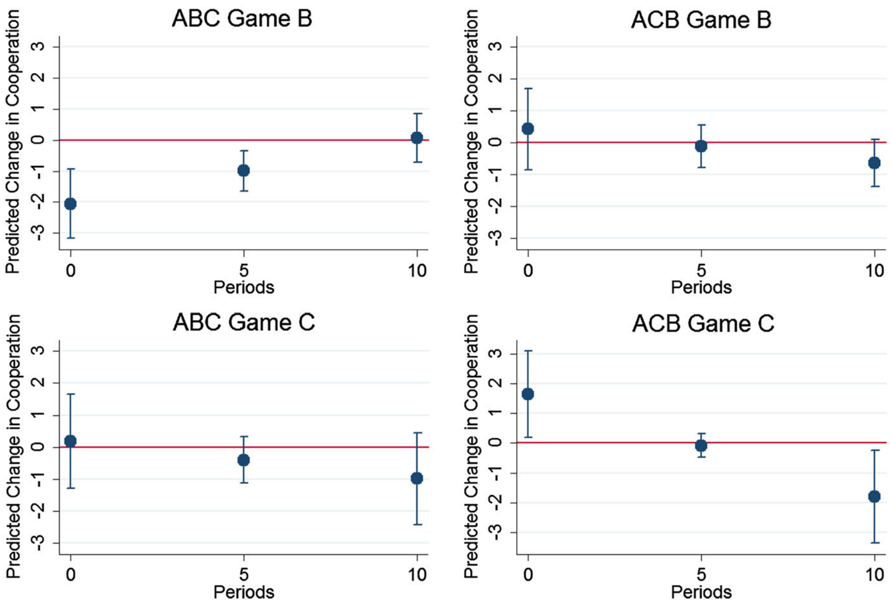

Fig. 4 Predicted change in contributions of super-group 2 resulting from OGC punishment. Predicted outcomes based on random effects panel estimations (see Tables 1, 2) using robust standard errors clustered at the individual level. Bars indicate the 95 % confidence interval

the last few. Game B shows a much different results. Those in super-group 1 contribute much more (3.88) than those in super-Group 2 (1.51; p value = 0.00). Thus, the potential punishment in Game B, which is only applied to super-group 1, increases cooperation. Finally, we can see that in Game C average contributions from both super-groups are very similar; those in super-group 1 contributed 4.08 while those in super-group 2 contributed 4.32 (p value = 0.42).

Before going to our main results of observed punishment, we highlight in Fig. 2 the direct effect of personal punishment. Figure 2 shows the change in subject contributions from period t − 1 to period t as a function of being personally punished in period t − 1, and it shows that being punished increases contributions. Across all periods (2–10), those in super-group 1 who were punished in round t − 1 contributed 1.87 more the next period (p value = 0.00) in Game B and 2.41 more (p value = 0.00) in Game C than their super-group counterparts who were not punished in period t − 1. Likewise, those in super-group 2 who were punished in round t − 1 contributed 2.24 more across all periods (p value = 0.00) in Game C compared to those who were not punished.

4.2 Game B results

Our primary interest is in the behavioral changes of super-group 2 upon observing OGC punishment. Figure 3 shows the change in cooperation in Games B in the upper half and C in the lower half from period t − 1 to period t for subjects in super-group 2 conditional on the game order and if they observed their OGC punishment. Looking first at the upper panel of Fig. 3 highlighting Game B results, we find that those who did not observe OGC punishment in game order ABC had a slight increase in contributions (0.03 tokens). However, those in Game B who observed OGC punishment in game order ABC decreased their contributions by an average of 0.62 tokens. This difference is statistically significant (p value = 0.079) and the decrease in contributions is also economically substantial given that the average contribution in Game B was 1.51 tokens. This effect is seen to be the most pronounced initially and dies off by period 10. For those in the ACB game order, there appears to be little difference between those who observed OGC punishment (−0.20) and those who did not (−0.25; p value = 0.885) in Game B.

The above simple analysis suggests that in Game B, observation of OGC punishment has a negative effect on cooperation in the ABC condition, but does not affect those in the ACB condition. To put this finding on firmer statistical grounds, we will turn to regression analysis that controls for other co-variates. Figure 4 displays the predicted change in cooperation (measured in tokens) when OGC punishment is observed. These are predicted changes from one period to the next when punishment has been observed relative to when punishment has not been observed. Thus, positive (negative) changes indicate that a subject who observed OGC punishment increased (decreased) her contributions from the previous period compared to one who did not observe OGC punishment. We want to highlight how punishment spillovers change overtime, so three periods (1, 5 and 10) are displayed. These predictions are based on the random effects panel regressions, the results of which are displayed in Table 1.[11] Note that we control for the level of cooperation among an individual's group members in the previous period. The fitted model uses our observed sample to predict cooperation as a function of two key variables in

[11] We estimated the regression models for the ABC and ACB conditions separately to make interpretations simple since the data gathered from these two settings are vastly different. If the regressions are combined and the appropriate triple interaction effects are included, the same result emerges. This also motivated us to separate Game B and Game C analysis.

Fig. 3 observed punishment and the period within the game. The predictions seen below are the average predicted levels of cooperation at different periods assuming punishment has been observed. We will address the predictions first and then move to the regression estimates to support our conclusions.

The upper panels of Fig. 4 show the predicted changes in cooperation from Game B. The left panel shows those from the ABC condition; the right shows those from the ACB condition. The results from the ABC condition indicate a negative effect from observed punishment on cooperation that shrinks as punishment is observed later in the game. We take this to mean that subjects who observed punishment significantly decreased their contribution levels early in the game, but this effect of observed punishment dissipates over time and is no longer significant in the final periods of Game B. The finding that subjects are predicted to decrease their relative contributions in the first round by 2 tokens is notable. The average contribution in the first round was 2.6 tokens, meaning those who observed punishment contributed only 0.6 tokens, or 77 % less than the average subject.

On the other hand, as previously noted, observed punishment had a statistically insignificant impact on the predicted cooperation of subjects in game B from the ACB condition. Recall that these subjects have already played Game C where the punishment rules were simultaneously in place in the OGC's VCM group as well as in their own. We take this to mean that once one has had experience with punishment institutions in a previous treatment, the spillover effects of observing punishment disappear. Of course, history of play in game C implies history both with observing punishment as well as being personally non-exempt from punishment. Because of this, results from game B alone cannot pinpoint which dimension of punishment history is most important. However, we will show in the next section on game C results that observed punishment effects also disappear in game C of the ABC order. Thus, experiencing punishment history of any sort (i.e., past observance of others' punishment or past experience of being non-exempt from punishment) eliminates the estimated spillover effects of observed punishment.

Figure 4 is based on Table 1, which displays the random effect estimations from regressing the change in cooperation (measured in tokens) on key covariates including the observation of OGC punishment and its effect over time.[12] The variable *OGC punished* indicates whether the subject observed OGC punishment in the current period. Other independent variables include the contributions of the other players in the previous period's VCM group as well as the subject's first period contribution, which proxies for an individual's cooperative nature. The results from the ABC condition are in column 1 and column 2 displays the results from the ACB condition. The key result is that observing punishment leads to a negative change in contributions from the previous period in the ABC condition.

Result 1 Super-group 2 subjects in the ABC condition respond to observed punishment in Game B by decreasing their cooperation. This response shrinks over time. This is consistent with a short-term social learning effect. Those from the ACB

[12] In some sessions, subjects had the chance to briefly chat with their OGC. This had no effect on our results and thus we pool the data. These regressions are available upon request.

condition are unaffected by observed punishment in Game B, which indicates that history with punishment eliminates this effect.

4.3 Game C results

We now explore what insights Game C can provide. In particular, we are interested in understanding the behavior of super-group 2 subjects from the ACB condition because their first exposure to punishment occurred in Game C. The lower panel in Fig. 3 displays the changes in contributions conditional on the game order and if the subject observed OGC punishment. For those subjects in the ABC treatment, no difference in the change in cooperation is found between those who observed OGC punishment (0.05) and those who did not in Game C (0.42; p value = 0.404). The same is not true in Game C for subjects in the ACB treatment. Those who observed OGC punishment *decreased* their contributions by 0.73 while those who did not observe OGC punishment *increased* their contributions by 0.25 (p value = 0.028). Once again, this result is not constant over time as the effect seems to be the strongest in the later periods. Thus based on Fig. 3, we would expect an initial positive response to observance of OGC punishment, but a negative sign on the interaction of OGC punishment with period.

We once again turn to regressions for a more careful analysis. The lower panel of Fig. 4 displays similar predicted changes in cooperation for the ABC and ACB conditions using Game C results. The underlying regression results are found in Table 2; however, the underlying models now include an indicator variable controlling for whether the subject was personally punished in the previous period. As already highlighted, personal punishment leads to a significant increase in cooperation as one would expect. This is a clear indication that subjects respond to being personally punished by increasing their contributions, and it is important to control for this effect in the Game C analysis of OGC punishment effects.

In Fig. 4 we see that subjects in Game C in the ACB condition responded to OGC punishment by significantly *increasing* their contributions. Likewise, this effect diminishes with time to the point that the predicted change in cooperative behavior becomes negative. We take this to mean that observed punishment led to a positive spillover effect when these subjects were also exposed to the possibility of personal punishment. This experiential punishment effect diminishes over time and was no longer present when they entered Game B where they were exempt from punishment in their own VCM group. Our results indicate that a necessary condition for experiential punishment to arise is non-exemption from punishment. This result is intuitive and encouraging if sanctions are meant to serve as deterrents for would be violators.

Furthermore, we see that observing OGC punishment had no significant effect on cooperation for those in the ABC condition during Game C. In this case, recall that Game C is the last treatment of the experiment for these subjects and we previously saw that the negative "social learning" effect of observed punishment in Game B diminished over time. Of course, these subjects were no longer exempt from punishment within their own VCM group in game C. But coupled with Result 1 above, we conclude that previous history of play in a punishment treatment diminishes all punishment spillover effects.

Result 2 Super-group 2 subjects in the ACB condition respond to observed punishment in Game C by increasing contributions. This response shrinks over time and becomes negative. This is consistent with a short-term experiential punishment effect. Those from the ABC condition are unaffected by observed punishment in Game C, which indicates that history with punishment eliminates this effect.

The effect highlighted in Result 2 is very intuitive. If an observer is going to process punishment experientially, it seems she has to be potentially subject to a similar outcome as the subject who was initially punished. Thus, observation of punishment serves as a reminder of the rule-sanction connection, and/or makes what is perceived as a punishment probability seem more real. The second interesting finding is that, not only does this effect go away, but it becomes increasingly negative by the last round. One explanation is that subjects are learning to take more risks by observing risky behavior. Remember, in this setting, subjects who contribute fewer than half of their tokens are subject to punishment. Thus, contributions which trend towards zero are riskier choices. We see that this is true. In the first 5 periods of Game C, those who observed punishment contributed an average of 4.63 tokens compared to 3.51 tokens in the last 5 periods.

5 Discussion

In this study we examined the basic behavioral response of observing punishment as it relates to cooperation in a social dilemma. A priori, there were two competing thoughts on the effect the observation of punishment might have. A social learning effect dictates that observing the punishment of someone else, when such punishment is entirely exogenous to one's social group, signals that the behavior was socially acceptable and may increase such behaviors. In our setting, that would imply that the observer would contribute less in the VCM game. On the other hand, an experiential punishment effect implies that observing punishment will make one feel as if personally punished, which would lead to increased contributions in a VCM context.

Our results indicate an adverse social learning effect exists when subjects are exempt from punishment and have no previous exposure to punishment institutions. This effect decreases over time and entirely disappears once subjects are no longer exempt from punishment. In other words, they turn their attention to their own punishment rather than that being observed. On the other hand, we also identify an experiential punishment effect when subjects are non-exempt from punishment and have no previous exposure to punishment institutions. This implies that observing punishment does have a deterrent effect, at least temporarily. While being exempt or non-exempt from personal punishment seems to be the key determinant of whether observed punishment leads to social learning or experiential punishment effects, these effects only manifest when subjects are exposed to a punishment treatment condition for the first time. Observed punished has no significant impact on contributions in game 3 of our 3-game design.

The main contribution of our paper is an experimental treatment where punishment can be observed but not received. Our design also imposed a completely exogenous punishment institution that removes any possibility that punishment reflects the norm preference of one's own group. This is important if one wishes to identify social learning or experiential punishment effects. Previous research did not attempt to distinguish between these two effects, and our results imply that the positive spillover effects of observed punishment found in previous studies can either be attributed to information updating (Schnake 1986) or to a combination of experiential punishment and strategic concerns (i.e., Xiao and Houser 2011).

Only by removing observed punishment from one's own social dilemma group and by including environments where one is exempt and non-exempt from personal punishment are we able to identify a clean experiential punishment effect. Even so, we reiterate that this experiential punishment effect requires one to be non-exempt from personal punishment and new to punishment institutions in general—yet the effect dissipates over time. After exposure to a punishment treatment, whether or not one is personally exempt from punishment, we find no significant remaining spillover effect of punishment. This last result implies that beneficial long-term effects of observed punishment in the real world may actually rely on the interplay of punishment, information, and/or strategic concerns as others have documented. To this end, a useful extension to the current paper would include endogenous punishment mechanisms; specifically in Game B where personal punishment is not possible.

Finally, we note that the external punishment mechanism we used is similar to prior work on regulation in social dilemmas. Consistent with one of our results, the primary finding of these studies is that external punishment leads to lower levels of cooperation. Cardenas et al. (2000) examine regulation in a common pool resource game where they attribute the lower levels of cooperation to a crowding-out effect. Similarly, Anderson and Stafford (2003) utilize a VCM game where exogenous punishment was also used. They find that past punishment negatively affects cooperation of the observers, which they link to some version of the gambler's fallacy. Though our design points to different explanations for a negative punishment spillover effect, our results, taken in conjunction with these, gives policy and law makers an additional reason to carefully consider the implementation of regulations.

The paper closest to our own in design is Xiao and Houser (2011), which also examines the effects of the observation of punishment. Our observed punishment result differs from theirs in our Game B, but is similar to theirs in our Game C environment. We have previously noted the design differences between their study and ours, but both studies identify a positive observed punishment spillover effect under certain conditions. In their paper, the subjects themselves had some role in shaping punishment, while in ours they do not. Because the beneficial effect of observed punishment is more enduring in their environment than in ours, it would appear that having some punishment parameters be endogenous to the group's behaviors is a wise idea.

References

Anderson, L., & Stafford, S. (2003). Punishment in a regulatory setting: Experimental evidence from the VCM. *Journal of Regulatory Economics, 24*(1), 91–110.

Andreoni, J., Harbaugh, W., & Vesterlund, L. (2003). The carrot or the stick: Rewards, punishments, and cooperation. *The American Economic Review, 93*(3), 893–902.

Arvey, R., & Jones, A. P. 1985. The use of discipline in organizational settings: A framework for future research. In B. Staw & L.L Cummings (Eds.), Research in organizational behavior (vol. 7, pp. 367–408). Greenwich, CT: JAI Press.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*, 169–217.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascade. *Journal of Political Economy, 100*(5), 992–1026.

Buchanan, J. M. 1973. A defense of organized crime? In S. Rottenberg (ed.), *The economics of crime and punishment*. Washington, DC: American Enterprise Institute

Cameron, S. (1988). The economics of crime deterrence: A survey of theory and evidence. *Kyklos, 41*, 301–323.

Cardenas, J. C., Stranlund, J., & Willis, C. (2000). Local environmental control and institutional crowding-out. *World Development, 28*(10), 1719–1733.

Cooper, D., & Glenn Dutcher, E. (2011). The dynamics of responder behavior in ultimatum games: A meta-study. *Experimental Economics, 14*(4), 519–546.

Cooter, Robert. (1998). Expressive law and economics. *Journal of Legal Studies, 27*, 585–608.

Dickinson, D. (2001). The carrot vs. the stick in work team motivation. *Experimental Economics, 4*(1), 107–124.

Erickson, P., van der Maas, M., & Hathaway, A. (2013). Revisiting deterrence: Legal knowledge, use context and arrest perception for cannabis. *Czech Sociological Review, 49*(3), 427–449.

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

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

Greiner, B. (2004). The online recruitment system ORSEE 2.0—A guide for the organization of experiments in economics. Working Paper Series in Economics 10, University of Cologne.

Harell, A., & Segal, U. (1999). Criminal law and behavioral law and economics: Observations on the neglected role of uncertainty in deterring crime. *American Law and Economics Review, 1*, 276–312.

Harsanyi, J. C. 1986. Practical Certainty and the Acceptance of Empirical Statements. In *Recent Developments in the Foundations of Utility and Risk Theory*. Dordrecht: Springer.

Holahan, J. (1973). The economics of control of the illegal supply of heroin. *Public Finance Quarterly, 1*, 467–477.

Johnston, J. M. (1972). Punishment of human behavior. *American Psychologist, 27*(11), 1033–1054.

Kahan, D. M. (1998). Social meaning and the economic analysis of crime. *Journal of Legal Studies, 27*(2), 661–672.

Ku, H., & Salmon, T. (2010). The incentive effects of inequality: An experimental investigation. *Southern Economic Journal, 79*(1), 46–70.

Malouff, J., Thorsteinsson, E., Schutte, N., & Rooke, S. (2009). Effects of vicarious punishment: A meta-analysis. *Journal of General Psychology, 136*, 271–286.

Myers, S. L. (1982). Crime in urban areas: Some new evidence. *Journal of Urban Economics, 14*, 148–158.

Ostrom, E., Walker, J. M., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review, 86*(2), 404–417.

Schnake, M. (1986). Vicarious punishment in a work setting. *Journal of Applied Psychology, 71*, 323–345.

Skogh, G., & Stuart, C. (1982). An economic analysis of crime rates, punishment and the social consequences of crime. *Public Choice, 38*, 171–180.

Sunstein, C. (1996). On the expressive function of law. *University of Pennsylvania Law Review, 144*, 2021–2031.

Trevino, L. K. (1992). The social effects of punishment in organizations: A justice perspective. *The Academy of Management Review, 17*(4), 647–676.

Tullock, G. (1974). Does punishment deter crime? *The Public Interest, 14*, 103–111.

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

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

Xiao, E., & Houser, D. (2011). Punish in public. *Journal of Public Economics, 95*(August), 1006–1017.