Economics Letters 129 (2015) 116–120

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Overdeterrence of repeat offenders when penalties for first-time offenders are restricted

Daniel Müller [a], Patrick W. Schmitz [b,c,*]

[a] Department of Economics, University of Bonn, Adenauerallee 24, 53113 Bonn, Germany
[b] Department of Economics, University of Cologne, Albertus-Magnus-Platz, 50923 Köln, Germany
[c] CEPR, London, UK

HIGHLIGHTS

- Suppose penalties for first-time offenders are restricted.
- It is then typically optimal for the lawmaker to overdeter repeat offenders.
- Now suppose the restriction on penalties for first-time offenders is relaxed.
- Should overdeterrence of repeat offenders now be reduced?
- If the restriction was strong, then overdeterrence should actually be amplified!

ARTICLE INFO

Article history:
Received 5 November 2014
Received in revised form
26 January 2015
Accepted 6 February 2015
Available online 14 February 2015

JEL classification:
D82
H23
K14
K42
L51

Keywords:
Limited liability
Incentives
Repeat offenders
Penalties
Law enforcement

ABSTRACT

When penalties for first-time offenders are restricted, it is typically optimal for the lawmaker to overdeter repeat offenders. First-time offenders are then deterred not only by the (restricted) fine for a first offense, but also by the prospect of a large fine for a subsequent offense. Now suppose the restriction on penalties for first-time offenders is relaxed; i.e., larger fines for a first offense become enforceable. Should overdeterrence of repeat offenders now be reduced? We show that this is the case only if the original restriction was not very strong. Otherwise, overdeterrence of repeat offenders should actually be further amplified.

1. Introduction

The law often sanctions repeat offenders more severely than first-time offenders.[1] The literature has provided various justifications for the fact that the sanction imposed on an offender depends on whether he was convicted previously.[2] Some authors have argued that a record of prior offenses provides information about the offender's characteristics (e.g., a higher-than-average propensity to commit crimes).[3] Yet, making sanctions depend on offense history may be advantageous even when individuals are ex-ante identical such that there are no characteristics to be learned about. As emphasized by Shavell (2004, p. 529), when

* Corresponding author at: Department of Economics, University of Cologne, Albertus-Magnus-Platz, 50923 Köln, Germany. Tel.: +49 221 470 5609; fax: +49 221 470 5077.
E-mail addresses: daniel.mueller@uni-bonn.de (D. Müller), patrick.schmitz@uni-koeln.de (P.W. Schmitz).

[1] For example, with regard to civil penalties in USA, Shavell (2004, Chapter 22) points out that for certain violations of the Occupational Safety and Health Act there is a maximum fine of $7000 for a first offense, while a repeat offender may be fined $70,000.

[2] See Miceli (2013) for a recent literature review.

[3] See e.g. Rubinstein (1980), Polinsky and Rubinfeld (1991), and Chu et al. (2000).

http://dx.doi.org/10.1016/j.econlet.2015.02.010

"detection of a violation implies not only an immediate sanction, but also a higher sanction for a future violation, an individual will be deterred more from committing a violation presently". In this paper, we follow Shavell's (2004) insight and further explore how penalties for repeat offenders should be designed when we take their effect on the deterrence of first-time offenders into account.

Specifically, suppose that there is an exogenously given restriction on the penalties for first-time offenders; i.e., there is an upper limit l which a fine for a first-time offender must not exceed, while there is no (binding) restriction on the fine that a repeat offender has to pay.[4] In each of two periods, a potential offender engages in an activity that may cause a harm h. When a harm is caused, the offender is convicted to pay a fine. If the harm is smaller than the maximum fine l, then by setting the fine equal to the harm h in both periods the negative externality of the activity is internalized and the first-best (i.e., socially optimal) activity level is implemented. Yet, if $l < h$, then first-time offenders in the first period are underdeterred, given that the fine for repeat offenders is set equal to the harm h. As a consequence, in general it will be optimal for the lawmaker to set the fine for repeat offenders larger than h. While in the second period overdeterrence of repeat offenders is ex-post inefficient, the advantage of such a policy is that the large fine for a second offense has a spillover effect on the first period.[5] Individuals in the first period are deterred not only by the (restricted) fine that they have to pay when they cause harm as a first-time offender, but also by the prospect of having to pay a large fine as a repeat offender in the second period.

Let us now explore what happens when the restriction l that society has put on the admissible fines for first-time offenders is relaxed. At first sight, one might guess that the lawmaker should reduce the ex-post inefficient overdeterrence of repeat offenders, because the deterrence of first-time offenders can now be improved by a larger penalty for first offenses. Yet, it turns out that this is the case only if initially the upper limit l was not very restrictive. If l was very small, then an increase in l will actually prompt the lawmaker to further *increase* the fine for repeat offenders; i.e., overdeterrence of second offenses will be further aggravated.

Intuitively, when l is very small, then a first-time offender in the second period faces only a very small fine, which provides indirect incentives in the first period not to cause a harm. Now consider an increase in l, such that a first-time offender in the second period can be more severely punished. When the fine for a repeat offender does not go up, then the indirect incentives in the first period are reduced, which the lawmaker may prefer to offset by further increasing the fine for repeat offenders.[6]

2. The model

In each of two consecutive periods, $t = 1, 2$, a risk-neutral individual chooses the level $a_t \in [0, 1]$ of a potentially harmful activity. With probability a_t the individual causes a harm $h > 0$ in period t.[7] For simplicity, assume that whenever a harm is caused, the individual is convicted to pay a fine.[8] The individual's private benefit from pursuing the activity is $b(a) : [0, 1] \to \mathbb{R}$ with $b'(a) > 0$, $b''(a) < 0$, $\lim_{a\to 0} b'(a) = \infty$ and $\lim_{a\to 1} b'(a) = 0$.

Let $y \in \{0, 1\}$ denote the individual's offense history at the beginning of period $t = 2$. If $y = 1$, then the individual is pre-convicted because he caused a harm h in period $t = 1$. If $y = 0$, the individual has a clean slate. In period $t = 0$ the lawmaker commits to a constitution, in particular stipulating the (finite) fine $F_t \geq 0$ to be paid by an individual in period t if he causes a harm. While the fine in period $t = 2$ may condition on the individual's offense history, $F_2 = F_2(y)$, the lawmaker is not allowed to discriminate according to whether a first offense was committed in $t = 1$ or $t = 2$.[9] Thus, $F_1 = F_2(0) \equiv F^0$ and $F_2(1) \equiv F^1$. Moreover, while there is no (binding) restriction regarding the punishment F^1 of a repeat offender, by social convention punishment of a first-time offender must not be overly drastic, $F^0 \leq l$, where $l \geq 0$.

If the activity level in each period were directly enforceable, then the lawmaker would implement the activity levels that maximize the expected social surplus $S(a_1) + S(a_2)$, where

$$S(a_t) = b(a_t) - h \cdot a_t. \tag{1}$$

Thus, the first-best solution is given by $a_1 = a_2 = a^{FB} > 0$, where a^{FB} is implicitly characterized by $b'(a^{FB}) = h$.

3. The analysis

In period $t = 2$, an individual with offense history $y \in \{0, 1\}$ chooses the activity level $a_2(F^y) = \arg\max_{a_2\in[0,1]} U(a_2; F^y)$, where

$$U(a; F) = b(a) - F \cdot a \tag{2}$$

denotes an individual's expected utility from activity level a when facing fine F in case of a harm. The second-period activity level that is optimal for the individual satisfies

$$b'(a_2(F^y)) = F^y \tag{3}$$

with $da_2(F^y)/dF^y = 1/b''(a_2(F^y)) < 0$; i.e., the higher the fine, the lower the individual's optimal activity level. Note that $U(a; h) = S(a)$, hence $a_2(F^y) \geq a^{FB}$ if and only if $F^y \leq h$. Application of the envelope theorem reveals that the individual's expected second-period utility is decreasing in the second-period fine, $dU(a_2(F^y); F^y)/dF^y = -a_2(F^y)$.

In period $t = 1$, the individual chooses his activity level $a_1 \in [0, 1]$ in order to maximize his overall expected utility,

$$\begin{aligned} EU(a_1) = {} & U(a_1; F^0) + a_1 \cdot U(a_2(F^1); F^1) \\ & + (1 - a_1) \cdot U(a_2(F^0); F^0), \end{aligned} \tag{4}$$

which is strictly concave, $d^2EU(a_1)/d(a_1)^2 = b''(a_1) < 0$. In consequence, if $dEU(a_1)/da_1\big|_{a_1=1} = -F^0 + U(a_2(F^1); F^1) - U(a_2(F^0); F^0) > 0$, then the optimal first-period activity level is

[4] There may be various reasons why society does not accept larger penalties for first-time offenders. For instance, Stigler (1970, p. 528) has pointed out that a "first-time offender may have committed the offense almost accidentally" and Polinsky and Shavell (1998, p. 313) argue that "considerations of fairness might constrain the sanction imposed on first-time offenders but not on repeat offenders".

[5] Note that related spillover-of-incentives effects also occur in sequential moral hazard models with limited liability. See e.g. Schmitz (2005) and Ohlendorf and Schmitz (2012), who show how second-period rents may act as carrot and stick for an agent's first-period effort choice, such that optimal contracts exhibit memory even though the periods are technologically independent. Recent papers that exploit related effects include e.g. Kräkel and Schöttner (2010), Tsai and Kung (2011), Chen and Chiu (2013), and Pi (2014).

[6] This argument holds provided that the direct punishment in the first period is still rather small, which is the case for relatively small values of the upper limit l.

[7] It is straightforward to generalize the model to the case in which a may be larger than 1, provided that the probability $p(a)$ with which a harm is caused is strictly convex. When $p(a)$ is strictly concave, in general the problem is no longer well-behaved; yet, one can construct examples with $a \in [0, \infty)$ such that our main insights still hold. Hence, the upper bound on a is not crucial.

[8] Throughout, we suppose that the individual has sufficient wealth to pay the fine.

[9] Qualitatively similar results hold in the case in which a first-time offender may face different fines in the two periods and there is an exogenous restriction on fines in the first period only.