The certainty versus the severity of punishment, repeat offenders, and stigmatization

Murat C. Mungan

Antonin Scalia Law School, George Mason University, United States

HIGHLIGHTS

- A two-period law enforcement model is considered.
- p and s respectively denote the certainty and the severity of punishment.
- When repeat offenses are possible, deterrence can be more responsive to p than s.
- Results hold regardless of whether convictions lead to stigmatization.
- Results follow even if people are slightly risk-averse.

ARTICLE INFO

ABSTRACT

There is a widely held presumption among criminologists that the certainty of punishment (p) is a greater deterrent than the severity of punishment (s). This presumption is at odds with recent experimental work as well as the implications of simple law enforcement models. This article shows that when offenses may be committed repeatedly, p may have a greater deterrent effect than s, even when each individual offender is more responsive to s than p. This resolves the discrepancy between experimental results and the common belief held among criminologists.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

A belief shared by many criminologists is that the certainty of punishment (p) affects deterrence more than the severity of punishment (s),[1] and there is some empirical evidence supporting this belief.[2] From a theoretical perspective, this belief is more consistent with offenders having a preference for risk, rather than a preference for avoiding risk.[3] Moreover, there are some experimental studies that provide evidence that offenders are deterred more by s than p, which is contrary to this belief.[4] Thus, there is a

---

[1] See, e.g., Engel and Nagin (2015) calling this the 'certainty effect in criminological parlance'. See also, Mendes (2004, p. 60), reviewing common views held by scholars regarding the differential effects of p versus s.

[2] See Engel and Nagin (2015) for a list of articles providing such evidence.

[3] Becker (1968) formalizes this point. See, Mungan and Klick (2014) for a review of the literature showing how criminals can be risk-averse and still act consistent with this belief.

[4] See, e.g. Friesen (2012) and Engel and Nagin (2015).

seeming discrepancy between commonly held beliefs by criminologists and the empirical studies that support them on the one hand, and, the belief that most individuals act in a risk-averse manner and some experimental results on the other hand.

A possibility is that this discrepancy is the result of a subtle distinction between the questions being asked. When a person thinks about whether p or s is the greater deterrent, she is inquiring about the responsiveness of the aggregate offense rate to these two variables, given all of the enforcement mechanisms in place. On the other hand, when a researcher asks whether a given person is more responsive to these two measures, he generally has to measure the relevant elasticities by focusing either on two (or more) hypothetical scenarios in a lab, or by comparing the responsiveness of two similar individuals who face different punishment schemes. In both instances, one runs the risk (or enjoys the benefit) of eliminating the effects of mechanisms that generate additional effects associated with increases in either p or s.

This article focuses on escalating punishment schemes, where *known* repeat offenders are punished more severely than first time offenders, as an example of this type of mechanism. Under these schemes, an increase in p not only increases the expected punishment for all offenders, but, it also increases the proportion of offenders whose previous offenses are discovered (i.e. the proportion of known repeat offenders). Some repeat offenders are deterred from committing crime when they face the escalated punishment for repeat offenders, but they are not deterred when they face the sanction for first time offenders. Therefore, increasing p has an additional deterrence effect: it moves more offenders into the 'known repeat offender' category and causes some of these offenders to be deterred. On the other hand, increasing the severity of punishment does not have a similar effect, and, therefore, when offenders are risk-neutral, the aggregate crime rate is strictly more responsive to increases in p than to s.

These observations provide an explanation for why offense rates may be more responsive to p than s, whereas individuals may be more responsive to s than p when considering each criminal opportunity in isolation. Given the prevalence of escalating punishment schemes in enforcing non-criminal offenses that range from civil infractions to administrative violations, it is plausible that the mechanism outlined above may explain an important part of the seeming discrepancy previously described. However, the explanation provided thus far is incomplete for an important class of criminal offenses, namely those which cause stigmatization.

When convictions lead to stigma, the total (i.e. formal plus informal) sanctions for first time offenders may actually be greater than the total sanction for repeat offenders, because most informal sanctions are imposed following the first conviction. Thus, stigmatization can lead to criminogenic effects: people with records are more likely to commit crime than people without records. However, in these cases there is a second, countervailing, effect: increases in p have a greater direct effect on deterrence than increases in s,[5] because both expected formal and informal sanctions are affected by changes in p, whereas changes in s only affect the expected formal sanction. The analysis in the next section reveals that the greater direct deterrence effect of p generally dominates whenever the informal sanction is large enough to cause criminogenic effects.

In sum, in the non-criminal context, the presence of escalating punishment schemes provides a novel explanation for the certainty effect, even when individual offenders are more responsive to the severity than the certainty of punishment. On the other hand, in the criminal setting, the certainty of punishment leads to greater direct deterrence effects than s, and this differential effect

is generally more than enough to off-set any criminogenic effects that may result from stigmatization. These results reconcile the seemingly contradictory presumptions and empirical results, and because they do not rely on any preference for risk, they also add to the literature reconciling the certainty effect with risk-aversion (e.g. Neilson and Winter (1997) and Block and Lind (1975)).

The next section uses a standard two-period law enforcement model to formalize the points made above. Section 3 concludes.

2. Model and analysis

I consider a standard two-period law enforcement model (see, e.g. Miceli (2013) or Mungan (2014)), where there is a continuum of individuals, who have criminal benefits, b. The distribution of these benefits among individuals is described by the cumulative distribution function F, where $F(0) = 0$, $\lim_{b \to \infty} F(b) = 1$, and $F'(b) = f(b) > 0$ for all $b \geq 0$. Individuals are risk-neutral, but since results are derived through strict inequality conditions, conclusions extend to cases where they may have a slight preference for risk-avoidance.

All individuals are first time offenders in period 1, and, they receive a formal sanction (s) if they are caught subsequent to committing an offense. Moreover, a first conviction results in an informal sanction of σ, and people who are sanctioned in the first period enter the second period with an offense record. If a person with a record commits an offense and is caught, he is punished as a repeat offender, which implies a sanction of πs,[6] where $\pi > 1$ is an inflation factor[7] reflecting escalating formal sanctions. Offenders who do not have records in the second period are considered first time offenders. The probability of punishment, in both periods, is p. Thus, an offender's problem can be represented by the decision tree depicted in Fig. 1.

Two preliminary observations are that individuals with $\frac{b}{p} \geq \max\{\pi s, s + \sigma\}$ commit offenses in both periods,[8] and individuals with $\frac{b}{p} < \min\{\pi s, s + \sigma\}$ never commit offenses. The behavior of individuals with $\frac{b}{p} \in [\min\{\pi s, s + \sigma\}, \max\{\pi s, s + \sigma\})$ depends on whether total sanctions are escalating or declining.

(Weakly) Escalating total sanctions $\pi s \geq s + \sigma$:

A potential offender with $b \in [p(s + \sigma), p\pi s)$ knows that if he enters the second period as a repeat offender he will refrain from committing an offense, since $b < p\pi s$. If, however, he enters the second period without a record, he will commit the offense since $b > p(s + \sigma)$. Thus, his total expected net benefits from committing the offense in the first period is $(2 - p)(b - p(s + \sigma))$. On the other hand, not committing the offense in the first period guarantees a clean record in the second period, and, therefore, a total-net-expected benefit of $b - p(s + \sigma)$ from committing the crime in the second period. Thus, all individuals with $b \in [p(s + \sigma), p\pi s)$ commit offenses in the first period, since $(2 - p)(b - p(s + \sigma)) > b - p(s + \sigma)$.

Thus, when there are escalating total sanctions, it follows that people with $b \geq p\pi s$ commit crime in both periods, people with $b < p(s + \sigma)$ do not commit crimes in either period, and people with $b \in [p(s + \sigma), p\pi s)$ commit crime in the first period and commit

---

[5] See Mungan and Klick (2016) making this point and reviewing the existing literature containing similar observations.

[6] I assume this proportional relationship to compare the deterrent effects of equal percentage changes in the sanctions imposed on first time offenders as well as repeat offenders and the certainty of punishment. Otherwise, there is ambiguity – yet to be clarified in the criminology literature – in what an increase in the sanction means. Due to data limitations, many empirical studies rely on blunt proxies, including average sanctions or past sanctions imposed on individuals (see Friesen (2012, p. 400)). Thus, future empirical studies would need to more specifically define the sanctions they are measuring to enable a precise comparison between the s and p elasticities of deterrence.

[7] I assume $\pi > 1$ to focus on the prevalent case where repeat offenders are punished more severely. Results extend to cases where $\pi < 1$ in a range of cases identified in footnote 10.

[8] I assume that indifferent individuals commit offenses.