2020-15

Reputational versus Beckerian Sanctions

Claude Fluet
Murat C. Mungan

Décembre / December 2020

Centre de recherche sur les risques les enjeux économiques et les politiques publiques

www.crrep.ca

ABSTRACT

Punishment causes reputational losses in addition to more tangible losses. Lowering the probability of punishment reduces these reputational losses by diluting the informational value of verdicts. These considerations better align the positive as well as normative implications of law enforcement models with intuition and empirics: Crime is more responsive to the certainty rather than the severity of punishment even absent risk-seeking offenders (positive), which causes extreme Beckerian punishments to be inefficient when sanctions are socially costly to impose (normative). Moreover, in some cases optimal enforcement is 'anti-Beckerian': Punishment is symbolic and detection costs are incurred solely to provide reputational incentives.

Keywords: Optimal Deterrence, Law Enforcement, Beckerian Sanctions, Reputational Sanctions.

Claude Fluet : Université Laval, CRREP, CRED

Murat C. Mungan : George Mason University, Antonin Scalia Law School

1 Introduction

The law enforcement literature, building on Becker (1968), has extensively studied how the severity and likelihood of punishment, e.g. via fines or prison terms, affect incentives to commit crime; see the survey in Polinsky and Shavell (2007). Reputational incentives have also been studied, albeit less extensively, e.g., Rasmusen (1996) and Bénabou and Tirole (2011). Existing work has noted, for instance, how reputational sanctions interact with both the frequency of wrongdoing and legal sanctions in place. However, the literature has not analyzed in detail the interactions between the likelihood of apprehending offenders and the reputational sanctions that emerge. We analyze these interactions.

Our approach allows us to revisit two puzzling issues, one normative and one positive. The normative issue that arises in many economic models of law enforcement is that optimal schemes involve extreme punishment due to Becker's (1968) (in)famous maximal sanction result. The positive issue is that when offenders are more responsive to the certainty rather than the severity of punishment, as is widely believed, then they must hold a preference for risk, contrary to typical behavior in other contexts. A number of articles have addressed one or the other of these issues. However, a theory that fails to address *both*, i.e. overturn both results, creates tension with at least one of two commonly held intuitions, namely that punishment ought not to be extreme and that even people with no preference for risk are more responsive to the certainty rather than the severity of punishment. We are unaware of any studies that accomplish this task. We do so by formalizing the interactions between reputational concerns and the probability of punishment.

Simply stated, the Beckerian maximal sanction result suggests that any enforcement scheme which imposes a sanction with a certain probability can be improved upon by reducing the probability of detection and increasing the sanction to keep deterrence unchanged. The new scheme leads to savings in the form of reduced enforcement expenditures. Therefore, without an upper bound on feasible sanctions, there is technically no optimal sanction, since one can always continue reducing the probability of catching offenders while increasing the sanction. Thus, it is conventional to simply assume an upper bound to the severity of punishment. The Beckerian result in this setting is that the optimal sanction is the highest one possible, regardless of how large the maximum feasible sanction is.

This result is robust to many considerations, including, most importantly, the presence of punishment costs, e.g. collection costs when the sanction is a fine or social losses associated with non-monetary sanctions, as with imprisonment or suspension of a license. Given its counter-intuitive nature, scholars have sought to explain why the low probability-high sanction prescription may not be appropriate. Some previous work has been successful in identifying settings in which the result does not hold.[1] Here, we show that when legal sanctions are costly to impose and adversely affect a person's reputation, the Beckerian result practically never holds.

The rationale is a simple one. Low probabilities of catching offenders reduce the informativeness of law enforcement and thereby diminish the reputational incentives to obey the law. Specifically, while a conviction is bad news about the agent involved, a 'clean record' becomes less meaningful when offenders are seldom detected. Because a smaller probability of detection reduces the reputation differential between a clean record and a conviction, it causes a more than proportional reduction in the reputational opportunity cost of committing an offense. Therefore, to preserve the same deterrence, the formal sanction needs to be increased more than proportionally, which in turn leads to an increase in punishment costs. Thus, the Beckerian result does not hold.

The same dynamics that overturn the Beckerian normative result also invalidate the risk-preference implications of the common view among criminal scholars, expressed as early as 1764 in Beccaria's *Dei delitti e delle pene*, that deterrence is more responsive to the certainty rather than the severity of punishment. Some empirical studies measuring elasticities of crime report findings consistent with this view (e.g. Lee and McCrary 2016). In the expected utility framework, however, this implies that offenders must be risk-seekers. Becker (1968) notes that interior solutions for optimal punishment can be justified only with risk-seeking offenders, since optimal sanctions would otherwise be maximal. The idea that offenders are risk-seeking is disturbing because it would imply that offenders are categorically different from non-offenders who exhibit risk-averse behavior in a variety of contexts (Block and Lind 1975, Neilson and Winter 1997, Mungan and Klick 2016). Moreover, some empirical studies provide evidence that criminals and non-

[1] These include settings where offenders can incur costs to avoid punishment (Malik 1990), when wrongful convictions generate direct social costs (Miceli 1991), the jury is less willing to convict when punishment is very severe (Andreoni 1991), where potential offenders are risk averse (Polinsky and Shavell 1979), or where there is a trade-off between deterrence and the provision of productive information to third parties (Deffains and Fluet 2020).

criminals respond similarly to changes in enforcement schemes (Bar-Ilan and Sacerdote 2004).

These observations have prompted scholars to seek rationales for why individuals with no preference for risk may nevertheless be less responsive to the severity than the certainty of punishment.[2] Block and Lind (1975), for instance, show that people can be more responsive to the certainty rather than the severity of imprisonment while being risk-averse over monetary outcomes. Mungan and Klick (2014) show that individuals without a preference for risk can have the same type of responses even when sanctions are monetary. This is because they may have to forfeit their criminal gains upon conviction or may be caught red-handed, causing the probability of detection to further reduce expected gains from crime. These studies, however, focus exclusively on behavioral responses and do not derive optimal punishments.

The incorporation of reputational considerations yields similar implications with respect to behavioral responses. As noted in our discussion of non-Beckerian sanctions, unlike the severity of punishment, the probability of detection directly impacts reputational incentives by improving the informational content of punishment. By contrast, expected formal sanctions are symmetrically affected by the certainty and the severity of punishment. The implication, when reputational effects are present, is that risk-neutral offenders are deterred more by the certainty rather than the severity of punishment. By a continuity argument, this result extends to some risk-averse offenders. Therefore, reputational concerns reconcile the model of law enforcement with conventional wisdom without resorting to unusual preferences for risk among offenders.

To summarize, reputational concerns imply results that are more consistent with intuitive notions of justice (i.e. extreme sanctions are not optimal) and with empirical observations (i.e. people with no preference for risk are more responsive to the certainty rather than the severity of punishment). This stands in stark contrast with prior scholarship. The most relevant comparison is Polinsky and Shavell (1999) who consider a set-up similar to ours but wherein reputational considerations are not present. They ask both the normative and positive questions we have described. Like us, they consider costly punishment, but they reach the opposite of our conclusions

[2] The literature on this topic is broad and includes more arguments than we are able to review here. See, e.g., Friesen (2012), Grogger (1991), Mungan and Klick (2015), and Mungan (2017, 2019).

on both accounts: when potential offenders are risk-averse, the Beckerian result holds *and* people are more responsive to the *severity* rather than the *certainty* of punishment.

Our discussion thus far highlights how reputational considerations improves the normative and empirical fit of the economic model of law enforcement. Additionally, our analysis reveals that 'anti-Beckerian' sanction schemes can be optimal in some cases. Optimal punishment may be purely symbolic and the law enforced primarily to provide potential offenders with reputational incentives. Thus, our analysis provides a rationale for shaming penalties, public reprimands or the fact that legal sanctions are often small compared to reputational consequences (Garvey 1998, Karpoff 2012). We also use our model to discuss the interactions between reputational sanctions and the offense rate, which is at the heart of the literature on the interaction between laws and norms (Bénabou and Tirole 2006, 2011; Mazyaki and van der Weele 2019). Finally, we explain how the relative size of reputational versus formal sanctions affects the elasticity gap between the certainty and the severity of punishment.

The next section develops a simple Beckerian law enforcement model that incorporates reputational sanctions. In section 3, we provide concluding remarks.

2 Model

We consider risk neutral potential offenders who have differing private benefits b from engaging in an action which causes external harm h. Benefits are described by the cumulative distribution function $F(b)$ with density $f(b)$ and support $[0, \bar{b}]$, where $\bar{b}$ is the highest private gain obtainable from the action.[3] The government imposes a sanction $s \in [0, \bar{s}]$ to deter the commission of the action, which generates a social loss of ks per person punished with $k > 0$. The sanction may not exceed its upper bound $\bar{s}$, but we will also refer to the case where there is no upper bound when doing so eases derivation of results. The government employs enforcement resources $c(p)$ to detect offenders with a probability of p, where $c(p)$ is increasing and convex with $c(0) = c'(0) = 0$ and $c'(1) = \infty$.

In addition to formal penalties, a conviction triggers informal sanctions

[3] A similar analysis follows when benefits are unbounded from above, but the description of the model becomes more cumbersome. If $f(\bar{b}) = 0$, the density is continuous over $[0, \infty)$ and the interpretation is then essentially the same as with an unbounded support.

by third parties denoted $\varsigma(\hat{b}, p)$ where $\hat{b}$ is the equilibrium threshold benefit such that only people with $b > \hat{b}$ commit the offense. The dependency of ς on the equilibrium threshold as well as the detection probability reflects the fact that reputational sanctions imposed by third parties depend on what kind of information they can infer from a conviction.

Before further explaining ς we derive the best responses of individuals to the enforcement policies in place and to the informal sanction expected from a conviction. For this purpose, it suffices to note that the expected net-benefit of a person engaging in the action is $b - p[s + \varsigma(\hat{b}, p)]$. Therefore, the best response when the informal sanction is expected to be $\varsigma(\hat{b}, p)$ is to commit the action if the private gain is above

$$b^r(\hat{b}, p, s) \equiv \min\{p[s + \varsigma(\hat{b}, p)], \bar{b}\} \tag{1}$$

Thus, a Bayesian equilibrium with consistent beliefs is characterized by

$$b^r(\hat{b}, p, s) = \hat{b} \tag{2}$$

Next, we explain how reputational sanctions emerge.

2.1 Reputational Sanctions

We adapt the honor-stigma model of Bénabou and Tirole (2006, 2011) to the stochastic law enforcement context. Earlier versions of this approach were used by Rasmusen (1996) and generalized in Mungan (2016). In this framework, third parties adjust their interactions with others based on their inferences regarding their criminal propensities (captured by b), because people's propensities are correlated with a characteristic that is valued by third parties. For example, in some contexts, a high b represents a high opportunity cost of not causing harm and reflects inefficiency or defective organization. In other contexts, b may consist of material gains minus an intrinsic inclination not to cause harm and thus a high b reflects low prosocial motivations. We refer to the characteristics valued by third parties as the individual's quality which is denoted by $q(b)$ with $q'(b) < 0$. The average quality is denoted q_m.[4]

Third parties do not directly observe individuals' behavior, but they observe whether an individual was convicted. Given an equilibrium threshold

[4] That is, $q_m \equiv \int_0^{\bar{b}} q(b) f(b)\, db$.

of $\hat{b}$, a third party infers that the average quality of a person who has not been convicted of an offense is

$$\nu(\hat{b}, p) = \frac{F(\hat{b})\eta(\hat{b}) + (1-p)(1-F(\hat{b}))\lambda(\hat{b})}{1-p(1-F(\hat{b}))} \tag{3}$$

where

$$\eta(\hat{b}) = \frac{\int_0^{\hat{b}} q(b)f(b)\,db}{F(\hat{b})} \tag{4}$$

is the average quality of non-offenders and

$$\lambda(\hat{b}) = \frac{\int_{\hat{b}}^{\bar{b}} q(b)f(b)\,db}{1-F(\hat{b})} \tag{5}$$

is the average quality of offenders. Out of $1-p(1-F(\hat{b}))$ people who are not convicted, $F(\hat{b})$ did not commit the offense and $(1-p)(1-F(\hat{b}))$ are offenders who escaped detection, hence the posterior expected quality in (3). By contrast, because offending is necessary for a conviction, all convicts are offenders. Thus, the average quality of convicts equals $\lambda(\hat{b})$ as defined in (5).

A b-type individual expects to receive a net-benefit of $b + p(\lambda(\hat{b}) - s) + (1-p)\nu(\hat{b}, p)$ from committing the offense, and a benefit of $\nu(\hat{b}, p)$ otherwise. The best-response previously expressed in (1) is then

$$b^r(\hat{b}, p, s) = \min\{p[s + \nu(\hat{b}, p) - \lambda(\hat{b})], \bar{b}\} \tag{6}$$

and the relative stigma or reputational sanction associated with a conviction is

$$\varsigma(\hat{b}, p) \equiv \nu(\hat{b}, p) - \lambda(\hat{b}) \tag{7}$$

Substituting the expressions for the average qualities $\nu(\hat{b}, p)$ and $\lambda(\hat{b})$,

$$\varsigma(\hat{b}, p) = \frac{F(\hat{b})}{1-p(1-F(\hat{b}))}\Delta(\hat{b}) \tag{8}$$

where

$$\Delta(\hat{b}) \equiv \eta(\hat{b}) - \lambda(\hat{b}) \tag{9}$$

is the difference between the average quality of non-offenders and offenders.

If third parties directly observed individuals' behavior, the reputational loss from committing the offense would equal the average quality differential between non-offenders and offenders. Differentials similar to $\Delta(\hat{b})$ have been analyzed in the literature (see Bénabou and Tirole, 2006, 2011, and Adriani and Sonderegger, 2019).[5] As (8) shows, however, the relative stigma from a conviction equals this quality differential discounted by the factor

$$\frac{F(\hat{b})}{1-p(1-F(\hat{b}))}$$

because imperfect detection reduces third parties' abilities to make inferences about people's types. While a conviction reveals that one has committed the offense, no-conviction is a noisy signal about behavior except in the limiting cases of perfect detection or complete deterrence.[6]

Holding the offense threshold constant, the stigma from a conviction is increasing in the probability of detection,

$$\varsigma_p(\hat{b},p) = \frac{F(\hat{b})(1-F(\hat{b}))\Delta(\hat{b})}{[1-p(1-F(\tilde{b}))]^2} > 0, \text{ for all } \hat{b} \in (0,\bar{b}) \tag{10}$$

where ς_p denotes a partial derivative. Moreover, for any positive level of deterrence, a conviction imposes a reputational sanction,

$$\varsigma(\hat{b},p) > 0 \text{ for all } \hat{b} \in (0,\bar{b}] \text{ and all } p \tag{11}$$

In particular, $\varsigma(\hat{b},0)$ is strictly positive if $\hat{b} > 0$ and can be interpreted as the reputational sanction from conviction when the probability of detection is very small. The non-convicted then include practically everyone, irrespective of the proportion of offenders in the population, so third-parties infer that a non-convicted has the average quality q_m. However, a convict is certainly an offender with average quality $\lambda(\hat{b})$. Thus, $\varsigma(\hat{b},0) = q_m - \lambda(\hat{b}) > 0$.

We next characterize the offense threshold that emerges in a Bayesian equilibrium.

[5] $\Delta(0) = q(0) - q_m$ and $\Delta(\bar{b}) = q_m - q(\bar{b})$. When the distribution of qualities $q(b)$ is strictly unimodal with an interior mode, $\Delta(\hat{b})$ is first decreasing then increasing. When the modal quality is $q(0)$, then $\Delta(\hat{b})$ is monotonically increasing, which corresponds to situations where the most likely types are the low benefit-high quality individuals.

[6] When $\hat{b}$ is very close to the upper bound $\bar{b}$, third parties infer a quality very close to the average q_m from no-conviction and know that only the worst types would be convicted. Thus, $\varsigma(\bar{b},p) = \Delta(\bar{b})$.

2.2 Equilibrium

Let

$$\varphi(\hat{b}, p, s) \equiv \hat{b} - b^r(\hat{b}, p, s) \tag{12}$$

so that the equilibrium condition (2) can be rewritten as

$$\varphi(\hat{b}, p, s) = 0 \tag{13}$$

When $p = 0$, the equilibrium threshold is $\hat{b} = 0$ because $b^r(\hat{b}, 0, s) = 0$ for all $\hat{b}$ and s. When $p > 0$ and $s > 0$, the existence of an equilibrium with $\hat{b} > 0$ follows from the fact that $\varphi(0, p, s) < 0 \leq \varphi(\bar{b}, p, s)$. Uniqueness is ensured if $\varphi_{\hat{b}} = 1 - p\varsigma_{\hat{b}} > 0$ for all interior $\hat{b}$. The latter holds, in particular, when $|q'(b)|$ is never too large. Thus, to simplify the analysis by obviating the need to distinguish between possible multiple equilibria, we impose the following.

Assumption 1 $1 - p\varsigma_{\hat{b}}(\hat{b}, p) > 0$ for all $\hat{b} \in (0, \bar{b})$.

Denote by $\hat{b}(p, s)$ the equilibrium offense threshold as a function of the enforcement policy. Note that $\hat{b}(p, 0) = 0$. An interior threshold therefore requires both $p > 0$ and $s > 0$.[7] For interior thresholds,

$$\hat{b}_s(p, s) = \frac{p}{1 - p\varsigma_{\hat{b}}(\hat{b}, p)} > 0 \tag{14}$$

$$\hat{b}_p(p, s) = \frac{s + \varsigma(\hat{b}, p) + p\varsigma_p(\hat{b}, p)}{1 - p\varsigma_{\hat{b}}(\hat{b}, s), p)} > 0 \tag{15}$$

The signs follow from assumption 1 and the observation that $\varsigma_p(\hat{b}, p) > 0$ at any positive threshold. We now investigate optimal policies.

2.3 Welfare Maximization and Cost Minimization

The government's objective is to maximize welfare which consists of the net-benefits from offenses minus enforcement and punishment costs, which we express as

$$W(p, s) \equiv D(\hat{b}(p, s)) - C(p, s) \tag{16}$$

where

$$D(\hat{b}(p, s)) \equiv \int_{\hat{b}(p,s)}^{\bar{b}} (b - h) f(b)\, db \tag{17}$$

[7] From (8), $\varsigma(0, p) = 0$ so that $\hat{b} = 0$ is the unique equilibrium when the formal sanction is nil.

are net-benefits and

$$C(p,s) \equiv k[1 - F(\hat{b}(p,s)]ps + c(p) \tag{18}$$

are the sum of punishment and enforcement costs.[8]

Before characterizing optimal policies, we identify the cost minimizing combination of instruments given any interior target level of deterrence.

Proposition 1 *Consider any targeted equilibrium threshold $\hat{b} \in (0, \bar{b})$. When the sanction is unbounded, the cost minimizing enforcement scheme consists of a positive probability of detection and a finite sanction. Thus, when there is a sufficiently large maximal sanction, the cost minimizing sanction is non-maximal, i.e. $s < \bar{s}$.*

Proof: Given $\hat{b} > 0$, p and s are positive and solve $p[s + \varsigma(\hat{b}, p)] = \hat{b}$. Therefore

$$\begin{aligned} C(p,s) &= k[1 - F(\hat{b})]ps + c(p) \\ &= k[1 - F(\hat{b})][\hat{b} - p\varsigma(\hat{b}, p)] + c(p) \equiv \widetilde{C}(p) \end{aligned}$$

where

$$\widetilde{C}'(p) = -k[1 - F(\hat{b})][\varsigma(\hat{b}, p) + p\varsigma_p(\hat{b}, p)] + c'(p)$$

with

$$\widetilde{C}'(0) = -k[1 - F(\hat{b})]\varsigma(\hat{b}, 0) + c'(0) = -k[1 - F(\hat{b})]\varsigma(\hat{b}, 0) < 0$$

where the sign follows from (11).■

The result derives from the relative impacts of the probability of detection versus the severity of formal sanctions. When detection is decreased and the severity of the sanction increased so as to keep deterrence constant, there are two countervailing effects. First, marginally lowering detection reduces detection expenditures by $c'(p)$. Secondly, expected punishment costs increase. The increase is proportional to

$$\varsigma(\hat{b}, p) + p\varsigma_p(\hat{b}, p)$$

This expression is bounded from below because $\varsigma(\hat{b}, 0) > 0$, while the reduction in detection costs is negligible for small probabilities of detection.

[8] In defining welfare, reputational effects have been omitted because they cancel out and would only add a constant, i.e. $[1 - p(1 - F(\hat{b}))]\nu(\hat{b}, p) + [1 - F(\hat{b})]\lambda(\hat{b}) = q_m$ for all $\hat{b}$.

Thus, minimizing enforcement costs always entails a positive probability of detection, and therefore a finite sanction. The same scheme continues to be cost-minimizing under a maximum allowable sanction, provided the latter is sufficiently large to be non-binding.

We next consider the implications for the welfare maximizing policy.

Proposition 2 *Let* $\hat{b} = \hat{b}(p, s)$ *where* s *and* p *maximize welfare. When* $\hat{b} \in (0, \bar{b})$ *and the maximum sanction is sufficiently large,*
(i) the optimal sanction trades off the benefits of deterrence against punishment costs,

$$(h + kps - \hat{b})\frac{\partial F(\hat{b}(p, s))}{\partial s} = kp[1 - F(\hat{b})] \tag{19}$$

(ii) the optimal probability of detection trades off detection costs against the savings in punishment costs from reputational sanctions,

$$k[1 - F(\hat{b})][\varsigma(\hat{b}, p) + p\varsigma_p(\hat{b}, p)] = c'(p) \tag{20}$$

An interior equilibrium threshold requires that both p and s are positive. Thus, when the maximum allowable sanction is sufficiently large, the optimal policy satisfies the first-order condition with respect to s, yielding (19). The immediate implication of this is that Beckerian maximal sanctions are sub-optimal. In (19), $h + kps - \hat{b}$ is the net social loss from the marginal offender, so the left-hand side captures the benefits from greater deterrence through a marginal increase in the sanction. The right-hand side is the extra deadweight punishment cost from increasing the sanction imposed on undeterred individuals.

With respect to detection effort, the first-order condition is

$$(h + kps - \hat{b})\frac{\partial F(\hat{b}(p, s))}{\partial p} = ks[1 - F(\hat{b})] + c'(p) \tag{21}$$

The right-hand side is the sum of the extra punishment and detection costs from an increase in the probability of detection. Combining (19) and (21) yields

$$\underbrace{\frac{kp(1 - F(\hat{b}))}{\hat{b}_s}}_{\substack{\textit{marg. cost of}\\ \textit{punishment}\\ \textit{to increase } \hat{b}\\ \textit{via sanctions}}} = \underbrace{\frac{ks(1 - F(\hat{b}))}{\hat{b}_p}}_{\substack{\textit{marg. cost of}\\ \textit{punishment}\\ \textit{to increase } \hat{b}\\ \textit{via detection}}} + \underbrace{\frac{c'(p)}{\hat{b}_p}}_{\substack{\textit{marg. detection}\\ \textit{costs to}\\ \textit{increase } \hat{b}\\ \textit{via detection}}} \tag{22}$$

The marginal costs of achieving deterrence through sanctions or detection are equalized, as would be expected. Substituting for $\hat{b}_s$ and $\hat{b}_p$ from their explicit formulation, (22) is equivalent to the condition (20) in the proposition.

That condition can itself be further simplified. From (8) and (10),

$$(1 - F)(\varsigma + p\varsigma_p) = \varsigma_p \tag{23}$$

thus yielding the following result.

Corollary 1 *Let s and p maximize welfare with $s < \overline{s}$. Then $k\varsigma_p(\hat{b}(p,s),p) = c'(p)$.*

The social benefits from a marginal increase in reputational sanctions, due to greater detection of offenders, equals $k\varsigma_p$. In a utilitarian policy, reputational sanctions are valued on the basis of punishment costs.

2.4 Discussion and Extensions

We relate our results to the literature and briefly discuss some implications and extensions.

Severity versus certainty of punishment. As noted in the introduction, there is a widely held presumption that people are more responsive to the certainty rather than the severity of punishment. We explain here how the incorporation of reputational concerns aligns the enforcement model with this presumption, and relate the gap between the $p-$ and $s-$elasiticities of crime to the size of reputational sanctions.

The certainty and severity elasticities of crime at any deterrence threshold $\hat{b}$ is given by

$$\frac{\partial(1 - F(\hat{b}))}{\partial x}\frac{x}{1 - F(\hat{b})}, \qquad \text{for } x \in \{p, s\}$$

Therefore, the ratio between the $p-$ and $s-$elasticities of crime is

$$\frac{p\hat{b}_p}{s\hat{b}_s}$$

and we can express the percentage difference between the p-elasticity and s-elasticity as

$$\delta \equiv \frac{p\hat{b}_p - s\hat{b}_s}{s\hat{b}_s} \tag{24}$$

such that potential offenders are more responsive to the certainty rather than the severity of punishment if $\delta > 0$. Substituting from (14) and (15),

$$\delta = \frac{\varsigma + p\varsigma_p}{s} > 0 \tag{25}$$

where the inequality follows from (10) and (11), which implies the following.

Corollary 2 *Crime is more responsive to the certainty rather than the severity of punishment.*

The intuition behind this result is one that we have already highlighted: while the severity and certainty of punishment symmetrically affect formal sanctions, an increase in the certainty of punishment increases the frequency with which both formal as well as reputational sanctions are imposed (this is Rasmusen's "double deterrent effect", 1996, p. 532) and it has the additional effect of increasing the stigma differential.

We note that corollary 2 is derived under the assumption that potential offenders maximize their expected net-benefits, and hence are assumed to be risk-neutral. The result extends through continuity to cases where they are slightly risk-averse. Thus, individuals may be more responsive to the certainty rather than the severity of punishment, even when they are risk-averse.

Substituting from (8) and (10), the elasticity gap can be re-expressed as

$$\delta = \left(\frac{1}{1 - p(1 - F)}\right)\frac{\varsigma}{s} \tag{26}$$

The gap depends on the relative importance of reputational concerns and formal sanctions (i.e. ς/s) multiplied by a factor related to the frequency of convictions or detected crime rate (i.e. $p(1-F)$). When the detected crime rate is small, the ratio between reputational and formal sanctions will act as a good approximation of the percentage difference between the certainty and severity elasticities of crime. This implication can potentially be tested by comparing δ's across crimes which involve different ratios of reputational to formal sanctions.

Optimal policy mix. Starting from an interior optimal level of deterrence, it is easily seen that an increase in the external harm h leads to an increase in the welfare maximizing $\hat{b}$. This requires an increase in either the legal sanction or the probability of detection, or both. How does the mix of instruments vary as greater levels of deterrence are sought? We provide a partial characterization.

Corollary 3 *In optimal policies, if $q(\hat{b}) < q_m$ at the initial equilibrium, then $ds/d\hat{b} > 0$ and $dp/d\hat{b} < 0$.*

Proof: The legal sanction and probability of detection solve:

$$p[s + \varsigma(\hat{b}, p)] = \hat{b} \tag{27}$$

$$k\varsigma_p(\hat{b}, p) = c'(p) \tag{28}$$

The first equation ensures that p and s implement $\hat{b}$, the second that enforcement costs are minimized. Totally differentiating with respect to $\hat{b}$,

$$\frac{dp}{d\hat{b}} = \frac{k\varsigma_{p\hat{b}}}{c''(p) - k\varsigma_{pp}} \tag{29}$$

$$\frac{ds}{d\hat{b}} = \frac{(c''(p) - k\varsigma_{pp})(1 - p\varsigma_{\hat{b}}) - (s + \varsigma + p\varsigma_p)k\varsigma_{p\hat{b}}}{p(c''(p) - k\varsigma_{pp})} \tag{30}$$

where $c''(p) - k\varsigma_{pp} > 0$ because of the second-order condition for cost minimization. We next consider what determines the sign of $\varsigma_{p\hat{b}}$. Rewrite (10) as

$$\varsigma_p(\hat{b}, p) = \gamma(\hat{b}, p)Z(\hat{b})$$

where

$$\gamma(\hat{b}, p) \equiv \frac{1}{[1 - p(1 - F(\hat{b}))]^2}$$

$$Z(\hat{b}) \equiv F(\hat{b})(1 - F(\hat{b}))\Delta(\hat{b})$$

Hence,

$$\varsigma_{p\hat{b}}(\hat{b}, p) = \gamma_{\hat{b}}(\hat{b}, p)Z(\hat{b}) + \gamma(\hat{b}, p)Z'(\hat{b}) \tag{31}$$

Now,

$$\begin{aligned} Z(\hat{b}) &= F(\hat{b})(1 - F(\hat{b}))\left(\frac{\int_0^{\hat{b}} q(b)f(b)\,db}{F(\hat{b})} - \frac{\int_{\hat{b}}^{\bar{b}} q(b)f(b)\,db}{1 - F(\hat{b})}\right) \\ &= (1 - F(\hat{b}))\int_0^{\hat{b}} q(b)f(b)\,db - F(\hat{b})\int_0^{\bar{b}} q(b)f(b)\,db \\ &= \int_0^{\hat{b}} q(b)f(b)\,db - F(\hat{b})q_m \end{aligned}$$

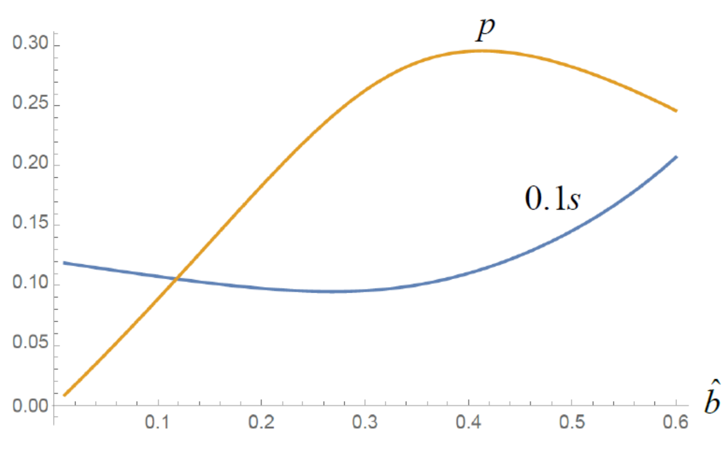
so that

$$Z'(\hat{b}) = f(\hat{b})[q(\hat{b}) - q_m]$$

Therefore, using (31) and noting that $\gamma_{\hat{b}} < 0$, $q(\hat{b}) - q_m < 0$ implies $\varsigma_{p\hat{b}} < 0$. From (29) and (30), this implies $dp/d\hat{b} < 0$ and $ds/d\hat{b} > 0$.■

The condition $q(\hat{b}) < q_m$ means that the marginal agent's quality is below average, hence the undeterred are all below average quality. This is necessarily the case at sufficiently large levels of deterrence. To increase deterrence even further (say, following an exogenous increase in the harmfulness of offenses), the optimal policy is then to invest less in detection and to strengthen the severity of the sanction. This is driven by the fact that the reputational dividend of more detection, $\varsigma_p(\hat{b}, p)$, becomes smaller as $\hat{b}$ gets sufficiently large. Loosely speaking, the optimal policy adjusts by relying more on formal sanctions and less on reputational sanctions, by increasing s and reducing p.

At low initial levels of deterrence, the comparative statics is less clear cut. For $\hat{b}$ small (because h is itself small), $\varsigma_p(\hat{b}, p)$ is small and increasing in the level of deterrence. The reputational dividend of more detection then increases as more individuals refrain from the harmful act. The optimal policy to achieve greater deterrence is now to increase detection. Whether the sanction should also increase is then in general ambiguous. Figure 1 provides a numerical example showing that, when the initial $\hat{b}$ is sufficiently small, greater deterrence may be optimally achieved by increasing p while reducing s.



Symbolic sanctions. So far, we considered a model wherein everyone is willing to commit the harmful action but for the threat of public enforcement. Now we consider the case where some individuals are intrinsically motivated to avoid causing harm. Let the support of the distribution of private benefits be redefined as $[\underline{b}, \overline{b}]$ where $\underline{b} < 0$. For instance, the net benefit from committing the harmful action is $b = y - v$ where y is the material gain and v is the moral cost, so that b can be negative. Some individuals trade off material gain against their reluctance to cause harm.

Without public enforcement, the fraction of people refraining from the harmful act is now $F(0) > 0$. With a policy that publicizes offenders but imposes no legal sanctions, the equilibrium threshold is $\hat{b} > 0$ solving $p\varsigma(\hat{b}, p) = \hat{b}$. This raises the possibility that the optimal policy could rely solely on reputational incentives without actually imposing sanctions, thereby avoiding punishment costs.

Proposition 3 *When the lower bound of private benefits is* $\underline{b} < 0$*, an optimal policy* $\hat{b}(p, s) < \overline{b}$ *is either as in proposition 2 or it imposes a symbolic sanction* $s = 0$ *with* $p > 0$ *satisfying*

$$(h - \hat{b}(p, 0)) \frac{\partial F(\hat{b}(p, 0))}{\partial p} = c'(p) \tag{32}$$

and with

$$(h - \hat{b}(p, 0)) \left. \frac{\partial F(\hat{b}(p, s))}{\partial s} \right|_{s=0} \leq kp[1 - F(\hat{b}(p, 0))] \tag{33}$$

Proof: The conditions (32) and (33) are necessary for an optimum with $s = 0$. It remains to prove that $p > 0$. From (8) and recalling that $\underline{b} < 0$ implies $F(0) > 0$, we now have $\varsigma(0, p) > 0$ including at $p = 0$. Therefore, using (15),

$$\hat{b}_p(0, 0) = \frac{\varsigma(\hat{b}(0, 0), 0)}{1 - p\varsigma_{\hat{b}}(\hat{b}(0, 0), 0)} > 0$$

where $\hat{b}(0, 0) = 0$. It follows that

$$W_p(0, 0) = hf(0)\hat{b}_p(0, 0) - c'(0) = hf(0)\hat{b}_p(0, 0) > 0$$

The sign follows from the fact that $f(0) > 0$ because $\hat{b} = 0$ is in the interior of the support.■

In the situation described in the proposition, the law is enforced only for the purpose of creating reputational incentives, hence the trade-off between detection costs and deterrence benefits in (32). This is optimal only if adding a formal sanction, which would further increase deterrence, is not worth the punishment costs that would ensue, hence condition (33).

When would such a policy be optimal? The punishment cost parameter would need to be large, since otherwise additional deterrence can be obtained at low cost. However, even when k is large, the imposition of a substantial formal sanction could improve welfare, because at high levels of deterrence both the frequency of offenses and punishment costs will be negligible. Thus, with k large, it could be preferable to deter most individuals from committing the act. The next result provides a sufficient condition that takes these considerations into account.

Corollary 4 *The symbolic sanction policy with $p > 0$ is optimal if k is sufficiently large and*

$$h \leq \int_0^{\bar{b}} \left(\frac{bf(b)}{1 - F(0)} \right) db \tag{34}$$

Proof: Condition (34) implies $h < \bar{b}$ and also implies

$$\int_{\hat{b}}^{\bar{b}} (h - b) f(b)\, db < 0 \text{ for all } \hat{b} \in (0, \bar{b}) \tag{35}$$

We first show that, for any positive p, condition (35) and k sufficiently large imply

$$L(p, s) \equiv W(p, s) - W(p, 0) < 0 \text{ for all } s > 0 \tag{36}$$

Substituting from (16)-(18), and noting that detection costs cancel out,

$$L(p, s) = \int_{\hat{b}(p,0)}^{\hat{b}(p,s)} (h - b) f(b)\, db - psk[1 - F(\hat{b}(p, s))] \tag{37}$$

If $\hat{b}(p, 0) \geq h$, the expression is trivially negative, so consider the case $\hat{b}(p, 0) < h$. From (35), the first term in the right-hand side of (37) is negative if $\hat{b}(p, s)$ is close to $\bar{b}$, so that $L(p, s)$ is then also negative. Therefore, let $\hat{b}(p, s)$ be sufficiently small for the first term in (37) to be positive, implying that $1 - F(\hat{b}(p, s))$ is bounded away from zero. But then a sufficiently large k will make $L(p, s)$ negative, for any positive s. This is also true for arbitrarily small values of s, when

$$k > \lim_{s \to 0} \frac{\int_{\hat{b}(p,0)}^{\hat{b}(p,s)} (h - b) f(b)\, db}{psk[1 - F(\hat{b}(p, s))]} = \frac{(h - \hat{b}(p, 0)) f(\hat{b}(p, 0)) \hat{b}_s(p, 0)}{p[1 - F(\hat{b}(p, 0))]}$$

which is condition (33) with a strict inequality. Finally, by proposition 3, there exists p such that $W(p,0) > W(0,0)$, which concludes the proof.$\blacksquare$

Condition (34) states that the external harm is less than the average private benefit of those who would commit the act under a no-enforcement policy, i.e. individuals with $b > 0$. Thus, welfare would be larger under no deterrence at all than under a policy seeking near complete deterrence. Now, an intermediate policy, with formal sanctions and a nonnegligible fraction of undeterred individuals, cannot be optimal if k is sufficiently large. A symbolic sanction policy, aiming only at publicizing behavior, then improves on no-enforcement by deterring some individuals with private benefits less than the external harm.

3 Conclusion

Our analysis shows that incorporating reputational penalties significantly mitigates the shortfalls of the standard economic model of law enforcement. It provides rationales for non-maximal optimal sanctions and for the greater aversion of offenders to the certainty rather than the severity of punishment. Moreover, it supplies a new perspective for interpreting alternative methods of punishment, such as shaming or public reprimands. Overall, our findings suggest that economic analyses of crime yield more relevant and interesting insights by taking the issue of stigmatization more seriously.

References

[1] Adriani, F. and S. Sonderegger (2019) "A Theory of Esteem Based Peer Pressure" 115 Games and Economic Behavior 314-335.

[2] Andreoni, J. (1991) "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?" 22 The RAND Journal of Economics 385-395.

[3] Bar-Ilan, A. and B. Sacerdote (2004) "The Response of Criminals and Noncriminals to Fines" 47 Journal of Law and Economics 1-17.

[4] Becker, G. (1968) "Crime and Punishment: An Economic Approach" 76 Journal of Political Economy 169-217.

[5] Bénabou, R. and J. Tirole (2006) "Incentives and Prosocial Behavior" 96 American Economic Review 1652-1678.

[6] Bénabou, R. and J. Tirole (2011) "Laws and Norms" . National Bureau of Economic Research Working Paper 17579.

[7] Block, M. and R. Lind. (1975) "An Economic Analysis of Crimes Punishable by Imprisonment." 4 The Journal of Legal Studies 479-492.

[8] Deffains, B. and C. Fluet (2020) "Social Norms and Legal Design." 36 Journal of Law, Economics, and Organization 139-169.

[9] Friesen, L. (2012) "Certainty of Punishment versus Severity of Punishment: An Experimental Investigation" 79 Southern Economic Journal 399-421.

[10] Grogger, J. (1991) "Certainty vs. Severity of Punishment." 29 Economic Inquiry 297-309.

[11] Karpoff, J.M. (2012) "Does Reputation Work to Discipline Corporate Misconduct?" Oxford Handbook of Corporate Reputation, Oxford University Press, 361-382 (M. L. Barnett and T. G. Pollock, eds).

[12] Lee, D. and J. McCrary (2017) "The Deterrence Effect of Prison: Dynamic Theory and Evidence" in Regression Discontinuity Designs: Theory and Applications. Emerald Publishing Limited, 73-146.

[13] Malik, A. (1990) "Avoidance, Screening, and Optimum Enforcement" 21 Rand Journal of Economics 341-353.

[14] Mazyaki, A. and J. van der Weele (2019) "On Esteem-Based Incentives" 60 International Review of Law and Economics 105848.

[15] Miceli, T. (1991) "Optimal Criminal Procedure: Fairness and Deterrence" 11 International Review of Law and Economics 3-10.

[16] Mungan, M. (2016) "A Generalized Model for Reputational Sanctions and the (Ir)Relevance of the Interactions Between Legal and Reputational Sanctions" 46 International Review of Law and Economics 86-92.

[17] Mungan, M. (2017) "The Certainty versus the Severity of Punishment, Repeat Offenders, and Stigmatization" 150 Economics Letters 126-129.

[18] Mungan, M. (2019) "Salience and the severity versus the certainty of punishment" 57 International Review of Law and Economics 95-100.

[19] Mungan, M., and J. Klick (2014) "Forfeiture of Illegal Gains, Attempts, and Implied Risk Preferences" 43 The Journal of Legal Studies 137-153.

[20] Mungan, M. and J. Klick (2015) "Discounting and Criminals' Implied Risk Preferences" 11 Review of Law and Economics 19-23.

[21] Mungan, M. and J. Klick (2016) "Identifying Criminals' Risk Preferences" 91 Indiana Law Journal 791-821.

[22] Neilson, W. and H. Winter (1997) "On Criminals' Risk Attitudes" 55 Economics Letters 97-102.

[23] Polinsky, M. and S. Shavell. (1979) "The Optimal Tradeoff between the Probability and Magnitude of Fines." 69 The American Economic Review 880-891.

[24] Polinsky, M. and S. Shavell (1999) "On the Disutility and Discounting of Imprisonment and the Theory of Deterrence" 28 The Journal of Legal Studies 1-16.

[25] Polinsky, M., and S. Shavell. (2007) "The Theory of Public Enforcement of Law" Handbook of Law and Economics, North Holland: Elsevier, 403-454 (A. M. Polinsky & S. Shavell, eds.).

[26] Rasmusen, E. (1996) "Stigma and Self-Fulfilling Expectations of Criminality" 39 Journal of Law and Economics 519-543.