CRED UNIVERSITÉ PARIS II PANTHÉON-ASSAS

CENTRE DE RECHERCHES EN ÉCONOMIE ET DROIT
Paris Center for Law and Economics

CRED WORKING PAPER *N^o^* 2020-7

Social Norms and Legal Design

February 2020

Bruno Deffains* Claude Fluet†

*Department of Economics and CRED, University Paris II Panthéon Assas, email: bruno.deffains@paris2.fr
†Université Laval, CRÉFiR, CRED. Email: claude.fluet@fsa.ulaval.ca

Social Norms and Legal Design

Bruno Deffains

Université Panthéon Assas, CRED, and Institut Universitaire de France

Bruno.Deffains@u-paris2.fr

Claude Fluet

Université Laval, CRÉFiR, and CRED

Claude.Fluet@fsa.ulaval.ca

Abstract. We consider legal obligations against a background of social norms, e.g., societal norms, professional codes of conduct or business standards. Violations of the law trigger reputational sanctions insofar as they signal non-adherence to underlying norms, raising the issue of the design of offences. We show that the law generally ought to follow social norms or be stricter than them. When society is only concerned with the trade-off between deterrence and enforcement costs, legal standards defining offences should align with underlying norms so long as the latter are not too deficient. When providing productive information to third parties is also a concern, legal standards should either align with underlying norms with fines that trade off deterrence against the provision of information; or legal standards should be more demanding and enforced with purely symbolic sanctions, e.g., public reprimands. Our analysis has implications for general law enforcement and regulatory policies. (JEL: D8, K4, Z13)

We thank Robert Cooter, Jennifer Reinganum, and Kathryn Spier for very helpful comments on an earlier version. We are particularly grateful to two anonymous referees and the Editor, Andrew Daughety. Claude Fluet acknowledges financial support from SSHRC Canada (grant 435-2013-1671).

1 Introduction

We study the design of legal obligations and their enforcement given preexisting social norms broadly interpreted, e.g., moral or societal norms, professional codes of conduct or business standards. Posner (2000, p. 4) describes the relationship between norms and legal rules as an interplay between background and foreground: " The law is always imposed against a background stream of nonlegal regulation — enforced by gossip, disapproval, ostracism [. . .] The desirability of a proposed legal rule, then, does not depend only on the existence of a collective action problem on the one hand, and competently operated legal institutions on the other hand. It also depends on the way nonlegal systems always already address that collective action problem and the extent to which legal intervention would interfere with those nonlegal systems." We take norms as exogenously given and assume that they are not influenced by the law.

Our contribution is to analyze the role of norms through a simple extension of the economic model of public enforcement of law, as reviewed for instance in Polinsky and Shavell (2000, 2007). In the standard model, which follows Becker (1968), misconduct is deterred solely through the threat of legal penalties. The focus is the trade-off between deterrence and public enforcement costs. We add background norms to this model. To some extent, agents refrain from misconduct because they adhere to norms. Moreover, in so far as it is publicized, non compliance with norms triggers social disapproval or costly reputational sanctions because of what it reveals about an agent's character or disposition. In our analysis, information is not freely available in society at large. Publicity about one's actions arises only through the public enforcement of law, e.g., investigation and prosecution. A by-product of enforcing laws is therefore to allow informal, nonlegal sanctions to operate. We study the implications for legal design and enforcement policies.

First, informal sanctions may complement legal sanctions and are therefore useful in reducing enforcement costs. This raises the question of the design of offences, by which we mean the extent to which legal standards of conduct should differ from background norms of behavior. We find that, for any given level of enforcement as defined by legal sanctions and the probability of detection, reputational sanctions are maximized when legal standards defining offences align with background norms. However, the latter need not be efficient from an utilitarian point of view. Norms of behavior, whose breach would trigger informal sanctions, may fall short (or may exceed) welfare maximizing behavior. Nevertheless, in a society concerned with deterrence and enforcement costs, legal standards should align with background norms provided the latter are not too deficient. When norms are sufficiently inadequate, they become irrelevant and the optimal legal standards will be much more demanding. Convictions then entail no reputational sanctions, which brings us back to the standard Beckerian framework.

Secondly, convictions or even mere prosecution may provide socially valuable information to third parties. By this we mean information that is useful for productive or allocative decisions shaping future relationships between agents, e.g., trading decisions and job market interactions. A utilitarian policy should take this into account. The consequence is a trade-off between the deterrence of a particular misconduct and the provision of information benefiting future interactions. We find that, generally speaking, optimal legal standards again align with background norms so long as they are not too deficient. However, the optimal enforcement policy is then typically characterized by less than maximal legal sanctions, in contrast to the Beckerian prescription, and by greater effort in detection or investigation. Indeed, the optimal legal sanction may be purely symbolic, e.g., a public reprimand or denunciation. Optimal legal standards are then more demanding than background norms. Moreover, when background norms are very demanding, the optimal zero fine policy may yield overdeterrence compared to the first-best utilitarian level.

Our paper relates to several strands of research. The so-called Laws and Norms literature has looked at closely related issues, more or less formally and from a variety of angles. McAdams (1997), Posner (2000), Shavell (2002), and McAdams and Rasmusen (2007) present general discussions of the relation between laws and norms, whether desirable or undesirable norms. The focus is often the complementarity or substitutability between laws and norms (Ellickson, 1991; Cooter and Porat, 2001; Zasu, 2007). Of particular interest for our purpose is the literature on stigmatization and reputational sanctions (Rasmusen, 1996; Kahan and Posner, 1999; Harel and Klement, 2007; Iaccobuci, 2014; Mungan, 2016a). Some of this literature also bears on legal design. Mungan (2016b) analyzes the stigma dilution effect of over-criminalization, i.e., criminalizing acts that cause little harm dilutes the stigma attached to more serious crimes. In a tort context with imperfectly functioning liability rules, Deffains and Fluet (2013) show that the negligence rule does better than strict liability in harnessing reputational concerns. Also of particular interest for our purpose is the small literature suggesting that the legal system should aim not only at deterrence but also at conveying socially valuable information by screening out types (Rasmusen, 1996; Posner and Rasmusen, 1999; Iaccobuci, 2007).[1]

More generally, our framework relates to the analytical literature on "social preferences". One's actions may reveal unobservable predispositions in situations where some predispositions are socially valued, hence social pressure may influence behavior through the individuals' image concerns (e.g., Bernheim, 1994; Bénabou and Tirole, 2006). This approach has been used, in particular, to study the crowding in or crowding out effects of material rewards. In Bénabou and Tirole (2011), the effect of legal sanctions depends on how they interact with signaling and social norms interpreted as a distribution of intrinsic motivations. They also consider the expressive role of law (roughly defined as in Cooter, 1998) when individuals are uncertain about the distribution of motivations. By contrast, we formalize norms as explicit standards of conduct, with individuals differing in their dispositions to abide by the norms. Our laws or legal standards may be interpreted as having expressive content, but only in the sense that they may be more or less noisy in conveying non-compliance with background norms. Relatedly, Daughety and Reinganum (2010) compare "publicity" and "privacy", when individuals derive utility from others' perception of their type and can engage in actions that generate externalities. While a regime of publicity distorts private choices, it may be socially useful in curtailing negative externalities or enhancing positive externalities. Nevertheless, publicity is undesirable if it leads to too much distortion of private choices by comparison with the social benefits. In our analysis, publicity is costly because it relies on the public enforcement of law. It also depends on the design of offences. When background norms are excessive, we find that it may be optimal to design noisy offences in order to mitigate overdeterrence.

In the standard model of public enforcement of law (i.e., without normative and reputational motivations), strengthening the legal standard of conduct, while keeping the expected sanction constant, never reduces the deterrence of misbehavior. If the standard is binding (i.e., everyone complies with the law), a more demanding standard increases deterrence. If the standard is not binding (i.e., not everyone complies), a more demanding standard has no effect on deterrence. By contrast, in the situation we consider, tighter laws may backfire and result in lower deterrence. The reason is that law-breaking becomes a less meaningful signal of non compliance with background norms. This is a main driver of our results. Acemoglu and Jackson (2017) also study a situation where tighter laws may increase misbehavior. In their analysis, there are no reputational concerns and laws are in part privately enforced through whistle-blowing. Tighter laws may increase misbehavior because of the complementarities between law-breaking behavior, resulting in less whistle-blowing.

Our analysis is motivated by the observation that the conduct of individuals or firms partly depends on normative considerations and that reputation matters. Regarding individuals, numerous experimental or field studies show that people may be willing to sacrifice private benefits against socially efficient actions (e.g., Charness and Rabin, 2002; Engelmann and Strobel, 2004; Charness et al., 2016) and that image concerns are important motivators (Masclet et al., 2003; Dana et al., 2006; Andreoni and Bernheim, 2008; Ariely et al., 2010; Funk, 2010, among others). The reputational consequences of judicial decisions is well documented. In the labor market, individuals with a criminal record face signification stigmatization, even for minor offences; see Agan and Starr (2017), Uggen et al. (2014), and the references therein. Karpoff (2012) surveys the important empirical literature on reputational sanctions for corporate misconduct.[2] Concerning the role of norms, Parsons et al. (2018) explain the large differences in the rates of corporate financial misconduct between U.S. cities in terms of differences in "city-level norms", as measured by non-business types of misbehavior. They show that corporate financial misconduct is strongly related to an index of city-level norms.

The paper develops as follows. Section 2 presents the public enforcement model and incorporates social norms and reputational concerns. Section 3 describes the equilibria under different legal regimes and enforcement policies. Section 4 derives the implications for optimal legal design and enforcement. Section 5 discusses extensions and qualifications. Section 6 summarizes our results and concludes. Proofs are in the Appendix.

2 The Model

We use the model of public law enforcement to define legal regimes and enforcement policies.[3] Next, we incorporate social norms and reputational concerns.

Legal regimes and enforcement. Risk-neutral agents encounter situations where they may obtain a private benefit, equivalently avoid a private cost, from an action that imposes an external harm of amount h.[4] The benefit depends on the agent's circumstances. g denotes both the objective circumstance and the associated benefit. The cumulative probability distribution is $Z(g)$ with density $z(g)$ and support $[0, \infty)$.

Engaging in the action is denoted by $e = 1$, not doing so by $e = 0$. Private benefits are a legitimate part of social welfare. Letting $e(g)$ be the action in the circumstance g, social welfare is

$$W = \int_0^\infty (g - h)e(g)z(g)\, dg.$$

The action is socially efficient when $g \geq h$.

A legal regime is defined by a threshold $\hat{g}_L$ such that the action is illegal when $g < \hat{g}_L$; otherwise, it does not constitute an offence. The law is enforced with a probability p of detecting violations. The sanction is a fine f which may not exceed f_m, the maximum allowed by the legislature or administrative guidelines. Under the regime $\hat{g}_L$, an agent's expected utility conditional on circumstance g is $(g - pf)e$ if $g < \hat{g}_L$ and ge if $g \geq \hat{g}_L$, where e is 0 or 1. The agent therefore engages in the action when $g \geq \min(\hat{g}_L, pf)$.

The per capita enforcement cost is $c(p)$, a strictly increasing and convex function with $c'0) = 0$. Taking enforcement cost into account, social welfare is

$$W = \int_{\min(pf, \hat{g}_L)}^\infty (g - h)\, z(g)\, dg - c(p) \tag{1}$$

and is maximized with respect to the legal regime, the fine, and the probability of detection. In the optimal policy, the fine is set at the highest feasible level and the legal threshold satisfies $\hat{g}_L \geq pf_m$; otherwise, costs could be reduced without affecting deterrence. Maximizing (1) with respect to the probability of detection then yields the first-order condition

$$(h - pf_m)z(pf_m)f_m = c'(p). \tag{2}$$

Let p^* denote the solution, so that $g^* := p^* f_m$ is the threshold at which individuals are just indifferent between engaging in the action and not. Condition (2) implies $g^* < h$. Optimal enforcement trades off some underdeterrence against savings in enforcement costs.

An important observation for what follows is that the legal regime is irrelevant so long as $\hat{g}_L \geq g^*$, because deterrence essentially depends on the expected fine. Whether offences are defined over a wide range of circumstances or more narrowly does not matter. In particular, welfare is the same if the action is illegal *per se* as in strict liability offences (i.e., $\hat{g}_L = \infty$).

Social norms and reputational concerns. Social norms are non legal standards of conduct potentially supported by informal sanctions.[5] A norm is a threshold $\hat{g}_S$. The action is viewed as inappropriate in circumstances $g < \hat{g}_S$, otherwise not. A natural interpretation is that, to some extent, one ought to take into account the potential harm to others. The social norm reflects the socially acceptable trade-off in this respect, e.g., causing harm in order to avoid negligible private costs is blameworthy. It may also well be that the norm is very demanding, some threshold $\hat{g}_S > h$.[6]

There is a continuum of agents. They differ by their disposition to comply with the social norm. A proportion λ, referred to as type G, are "good citizens" who have internalized the norm and always comply. They are willing to "do the right thing" irrespective of their material self-interest. Others, referred to as type B, are the "bad citizens" who have no such motivation. Adherence to the norm is valued by others, either because internalization of the group norm is valued by itself or because it is associated with a more principled character and less opportunistic tendencies, which may be desirable in other situations.[7] Those who are known to be good citizens earn social esteem or status, a direct source of utility, or they earn greater benefits from future social or economic interactions because they are seen as more reliable partners.

The foregoing can be formalized as follows. The utility of a type-t agent is

$$u_t = y - e\delta_t + v(\mu), \quad t = B, G;\ e = 0, 1. \tag{3}$$

The first term, y, is the material payoff as defined in the standard model. The middle term captures the intrinsic disutility of not complying with the social norm, e.g., guilt or loss of self-esteem. For the bad citizen, $\delta_B = 0$ irrespective of circumstances; for the good citizen, δ_G is positive and large if $g < \hat{g}_S$ and is otherwise equal to zero.

The last term in (3) is the reputational payoff to which we now turn. v is an increasing function and μ is the posterior probability that the agent is a good citizen given the information about him. A benchmark is the case where $v(\mu)$ is linear. For instance, let q_B and q_G be the marginal products of the bad and good types in future jobs, with $q_G > q_B$. Given the belief μ, the wage paid is equal to the expected productivity

$$v(\mu) = (1 - \mu)q_B + \mu q_G = q_B + \mu(q_G - q_B).$$

Alternatively, q_B and q_G are the purely hedonic utility of social esteem *per se*. Although the linear specification has been widely used[8], we will also interpret $v(\mu)$ as the value of future interactions with third parties who can take "productive" actions conditional on their information about the agent. $v(\mu)$ is then an increasing strictly convex function reflecting the value of the information (see Section 4).

Both types of agents face the same distribution of potential circumstances and differ only in their adherence to the social norm. Types are private information. With respect to others' beliefs, actions matter to the extent that they signal one's disposition or character. We assume that there is no direct informal enforcement of social norms because society at large does not observe the agents' circumstances nor their behavior. This assumption cuts off the direct social or market pressure on bad types to mimic the good types.[9] However, public enforcers can detect harmful actions and can verify circumstances.

Legal proceedings then constitute public information from which others draw inferences. Those who are detected engaging in the action are prosecuted, which is public information. The outcome is then either C for "convicted", i.e., detected and found guilty of an offence, or D for "detected and dismissed", meaning that the agent was detected for engaging in the action but not found guilty. Otherwise, the information about the agent is N for "no news", meaning that either the agent did not engage in the action or that he did but was not detected. Agents will therefore be labelled as C, D or N. The significance of these labels depend on what they reveal about types, given the legal regime and the enforcement policy.

3 Equilibria

This section takes the legal regime and enforcement policy as given and describes the equilibrium outcome. A perfect Bayesian equilibrium is defined by the agents' strategies and by the beliefs conditional on the information C, D or N. In a first step we characterize the strategies taking the beliefs as given. Next we derive the beliefs as a function of strategies. Finally, we solve for the equilibrium wherein strategies and beliefs are consistent with one another.

Incentives. Denote the beliefs by μ_C, μ_D and μ_N. We first consider the bad citizens' behavior. Abstracting from variables not affected by the agent's action, the expected utility of a type B agent who takes action $e \in \{0, 1\}$ in the circumstance g is

$$(g - pf)e + (1 - pe)v(\mu_N) + pev(\mu_C) \text{ if } g < \hat{g}_L$$

and

$$ge + (1 - pe)v(\mu_N) + pev(\mu_D) \text{ if } g \geq \hat{g}_L.$$

The expected utility includes the expected net material benefit from engaging in the action and the expected reputational utility.

Refraining from the action always yields the utility $v(\mu_N)$. When the action is illegal, i.e., in circumstances $g < \hat{g}_L$, engaging in the action yields the expected utility $g - pf + (1-p)v(\mu_N) + pv(\mu_C)$. An offence is therefore committed if

$$g \geq p[f + v(\mu_N) - v(\mu_C)]. \tag{4}$$

In words, the law is violated if the private benefit exceeds the expected disutility from the fine and from the reputational loss between the "no news" and "convicted" labels.

When the action does not constitute an offence, i.e., in circumstances $g \geq \hat{g}_L$, the expected utility from the action is $g + (1-p)v(\mu_N) + pv(\mu_D)$. The agent therefore engages in the action if

$$g \geq p[v(\mu_N) - v(\mu_D)]. \tag{5}$$

One may refrain from the action, even though it is legal, because mere prosecution entails a reputational loss.

The right-hand side of (4) and (5) defines possible thresholds for the bad citizen to engage in the action. Which of these thresholds is effective in determining the bad citizens' best response depends on the legal standard. As will be seen, $\mu_C \leq \mu_D \leq \mu_N$ at equilibrium. The right-hand side of (4) is therefore larger than that of (5). The possibilities are illustrated in Figure 1.

Under the legal standards $\hat{g}_L^2$ and $\hat{g}_L^3$, condition (5) is always satisfied when the action is legal. With the standard $\hat{g}_L^3$, the bad citizens' threshold is the right-hand side of condition (4). Bad citizens then sometimes do not comply with the law. With the standard $\hat{g}_L^2$, by contrast, condition (4) is never satisfied by illegal acts. The bad citizens' threshold then equals the legal standard $\hat{g}_L^2$, i.e., bad citizens engage in the action only when it is legal. Finally, with the standard $\hat{g}_L^1$, condition (5) is not satisfied in some circumstances where the action is legal. The bad citizens' threshold is then the right-hand side of condition (5). The threat of prosecution then provides deterrence, irrespective of conviction.

<<COMP: Place Fig. 1 about here>>

We will denote by g_B and g_G the bad and the good citizens' thresholds, respectively, for engaging in the action. In circumstances $g < \hat{g}_S$, a good citizen never engages in the action. In circumstances $g \geq \hat{g}_S$, the action does not violate the social norm, so the good citizen then behaves the same as the bad citizen. Therefore, $g_G = \max(\hat{g}_S, g_B)$.

Beliefs. How others behave affects the payoffs from one's actions through the effect on beliefs conditional on the events C, D or N. Beliefs are obtained from Bayes' rule given the frequency of detected acts and convictions among good and bad citizens. Because the good citizens' best-response threshold g_G is a function of the bad citizens' threshold g_B, posterior beliefs can be written as a function of g_B. Recall that λ is the proportion of good citizens.

Lemma 1 *If $g_B \geq \hat{g}_L$, the event C never occurs and $\mu_D \leq \lambda \leq \mu_N$ where*

$$\mu_N = \frac{\lambda(1-p+pZ(\max(\hat{g}_S, g_B)))}{\lambda(1-p+pZ(\max(\hat{g}_S, g_B)) + (1-\lambda)(1-p+pZ(g_B))} \tag{6}$$

$$\mu_D = \frac{\lambda(1-Z(\max(\hat{g}_S, g_B))}{\lambda(1-Z(\max(\hat{g}_S, g_B)) + (1-\lambda)(1-Z(g_B))} \tag{7}$$

If $g_B < \hat{g}_L$, then $\mu_C \leq \mu_D \leq \lambda \leq \mu_N$ where μ_N is as defined in (6) and

$$\mu_C = \frac{\lambda \max[0, Z(\hat{g}_L) - Z(\max(\hat{g}_S, g_B))]}{\lambda \max[0, Z(\hat{g}_L) - Z(\max(\hat{g}_S, g_B))] + (1-\lambda)[Z(\hat{g}_L) - Z(g_B)]} \tag{8}$$

$$\mu_D = \frac{\lambda(1-Z(\max(\hat{g}_S, \hat{g}_L))}{\lambda(1-Z(\max(\hat{g}_S, \hat{g}_L)) + (1-\lambda)(1-Z(\hat{g}_L))} \tag{9}$$

The numerator of (6) is the population of good citizens labeled as "no news". This is the sum of the fraction $1-p$ who are not monitored and of the fraction p who are monitored and do not engage in the action. The denominator is the total population labeled as "no news". Therefore, μ_N is the posterior probability that one is a good citizen given "no news". A similar reasoning applies for the other cases.

For the time being, suppose that (5) always holds when the act is legal, i.e., bad citizens are motivated by the threat of conviction. Their threshold is then either the right-hand side of (4) or it is equal to $\hat{g}_L$. Thus,

$$g_B = \min[\hat{g}_L, p(f + \Delta)]$$

where $\Delta := v(\mu_N) - v(\mu_C)$ is the reputational loss or stigma from a conviction. We use Lemma 1 to determine how the stigma varies with the bad citizens' behavior and with the legal standard. There are two cases to consider.

Case 1: $\hat{g}_L \leq \hat{g}_S$

Good citizens then always comply with the law. A conviction therefore reveals that one is for sure a bad citizen, hence $\mu_C = 0$. The belief conditional on "no news" depends on the proportion of bad types who comply with the law, so we can write the posterior as $\mu_N(g_B)$. As more of the bad types abstain from the action (i.e., g_B increases), "no news" becomes less indicative that one is a good citizen, so $\mu_N(g_B)$ is decreasing. Figure 2a depicts the resulting stigma curve $\Delta = v(\mu_N(g_B)) - v(0))$. The relevant portion is for values of the bad citizens' threshold satisfying $g_B \leq \hat{g}_L$. If the legal standard is increased, a greater portion of the curve becomes relevant. When $\hat{g}_L = \hat{g}_S$, the legal standard is equal to the social norm. In the particular case where bad citizens always comply with the law, i.e., $g_B = \hat{g}_L$, both good and bad citizens behave the same. The event "no news" is then uninformative, so the posterior probability conditional on "no news" equals the prior λ that an individual is a good citizen.[10]

<<COMP: Place Fig. 2 about here>>

Case 2: $\hat{g}_L > \hat{g}_S$

Both good and bad citizens are now at times convicted, therefore $\mu_C > 0$. From Lemma 1, μ_C depends on g_B and $\hat{g}_L$, which we write as $\mu_C(g_B, \hat{g}_L)$.

This is increasing in g_B because convictions become less unfavorable news as the proportion of bad citizens behaving like good citizens increases. As before, μ_N is decreasing in g_B. So long as violating the law is more likely for bad citizens, the stigma $\Delta = v(\mu_N(g_B)) - v(\mu_C(g_B, \hat{g}_L))$ is therefore positive and decreasing in g_B. When $g_B \geq \hat{g}_S$, everyone behaves the same. The events "convicted" and "no news" are then uninformative and the stigma from a conviction is zero. Figure 2b depicts the stigma curve. The curve shifts downwards (at values of g_B below $\hat{g}_S$) when the legal standard is increased because $\mu_C(g_B, \hat{g}_L)$ is increasing in $\hat{g}_L$. A greater proportion of good citizens is then convicted, hence a conviction is less unfavorable news.

Equilibrium. An equilibrium is a solution (g_B, Δ) to the system of equations

$$g_B = \min[\hat{g}_L, p(f + \Delta)], \tag{10}$$

$$\Delta = v(\mu_N(g_B)) - v(\mu_C(g_B, \hat{g}_L)), \tag{11}$$

provided the solution satisfies

$$p[v(\mu_N(g_B)) - v(\mu_D(\hat{g}_L))] \leq \hat{g}_L. \tag{12}$$

where $\mu_D(\hat{g}_L)$ is as defined in (9). The inequality (12) means that the mere threat of prosecution does not yield greater deterrence than the risk of conviction. In the Appendix (see Lemma A1), we show that the condition holds if the legal standard satisfies $\hat{g}_L \geq g_c(p)$. The latter is the bad citizens' equilibrium threshold when the enforcement policy reduces to a mechanism for detecting and publicizing harmful acts. Such a policy can be interpreted as a strict liability regime with symbolic fines, e.g., public reprimands or shaming penalties. It is easily verified that $g_c(p)$ is increasing in p.

Proposition 1 *Let $\hat{g}_L \geq \min[pf, g_c(p)]$. There is a unique equilibrium. If $pf \geq \hat{g}_S$, then $g_G = g_B = pf$. Otherwise, $pf < g_B \leq \hat{g}_S = g_G$:*
(i) When $\hat{g}_L \leq \hat{g}_S$, then g_B is increasing in p and f and invariant in $\hat{g}_L$ so

long as $g_B < \hat{g}_L$; otherwise, $g_B = \hat{g}_L$.
(ii) When $\hat{g}_L > \hat{g}_S$, then $g_B < \hat{g}_S$ and is increasing in p and f and decreasing in $\hat{g}_L$.

The foregoing defines the function $g_B(p, f, \hat{g}_L)$ characterizing the bad citizens' equilibrium threshold. Legal design matters for deterrence only when the expected fine is less than the social norm, i.e., $pf < \hat{g}_S$. Convictions then provide information about one's type and reputational sanctions supplement formal sanctions. When $pf \geq \hat{g}_S$, the outcome is a pooling equilibrium. It is the same as in the standard model without informal motivations and reputational concerns.

<<COMP: Place Fig. 3 about here>>

Figure 3a provides an illustration for the case where the legal standard is less demanding than the social norm. The upward sloping line, henceforth the threshold function, is the relation between the bad citizens' best response threshold and the stigma from a conviction. The relevant portion is for $g_B \leq \hat{g}_L$. In the figure there are two legal standards, $\hat{g}_L^1$ and $\hat{g}_L^2$, yielding the equilibria E_1 and E_2 respectively. Under the less demanding standard $\hat{g}_L^1$, everyone complies with the law and the standard is binding. Strengthening the standard, e.g., increasing the standard up to $\hat{g}_L^2$, then increases deterrence. Under the more demanding standard $\hat{g}_L^2$, the equilibrium is an interior solution where some of the bad citizens do not comply with the law. A small increase in the standard then has no effect on deterrence.

In Figure 3b, the legal standard is more demanding than the social norm. Stigma curves are drawn for the standards $\hat{g}_L^1$ and $\hat{g}_L^2$. A more demanding standard such as $\hat{g}_L^2$ shifts the stigma curve downwards. More good citizens are convicted, i.e., those in circumstances in $(\hat{g}_S, \hat{g}_L^2)$ compared to those in circumstances in $(\hat{g}_S, \hat{g}_L^1)$ under the less demanding standard. As a result, convictions impose a smaller stigma, all else equal. The effect of strengthening the legal standard is now to reduce deterrence, i.e., equilibrium E_2 is to the left of equilibrium E_1.

Whether the legal standard is above or below the social norm, the stigma curves shift upwards when the probability of detection is increased. The intuition is that a larger probability of detection increases the significance of "no news", which is now more suggestive that one did not engage in the harmful action. A higher probability of detection also shifts the threshold function to the right. Thus, more detection provides greater deterrence, except when the legal standard is binding as with the standard $\hat{g}_L^1$ in Figure 3a. A larger fine only affect the threshold function, which shifts to the right. This also increases deterrence except when the lagal standard is binding.[11]

4 Optimal Policies

Reputational sanctions affect the trade-off between deterrence and enforcement costs. They bear on the optimal legal regime because offences can be designed to harness reputational effects. We first focus on this property, disregarding the possibility that the information from judicial verdicts has social value independently of its usefulness in providing incentives. Next, we incorporate the informational benefits to third parties in our welfare calculus. There is then a trade-off between deterrence, enforcement costs, and the provision of valuable information to third parties.

<<COMP: Place Fig. 4 about here>>

Deterrence maximizing legal regimes. We start with a preliminary result, taking the enforcement policy as given and comparing different legal designs in terms of deterrence.

Proposition 2 *Given any enforcement policy, as defined by the probability of detection and the fine:*
(i) if $pf < \hat{g}_S$, deterrence is maximized with the legal standard $\hat{g}_L = \hat{g}_S$;
(ii) if $pf \geq \hat{g}_S$, the legal standard is irrelevant for deterrence so long as $\hat{g}_L \geq pf$.

When $pf < \hat{g}_S$, different legal standards are not equivalent because they yield different reputational effects, which in turn affects incentives. Deterrence is then maximized when the legal standard aligns on the underlying norm. To illustrate, Figure 4 compares the bad citizens' equilibrium threshold under a strict liability regime (i.e., $\hat{g}_L = \infty$) and under a fault regime with legal standard equal to the social norm. When $pf \geq \hat{g}_S$, by contrast, reputational effects vanish because both the good and bad citizens' thresholds equal pf, irrespective of the legal regime.

Trade-off between deterrence and enforcement costs. Welfare is

$$W = (1-\lambda)\int_{g_B}^{\infty} (g-h)\, z(g)\, dg + \lambda \int_{g_G}^{\infty} (g-h)\, z(g)\, dg - c(p) \qquad (13)$$

where g_B and g_G are the equilibrium thresholds of the bad and good citizens respectively as characterized in the preceding section. This formulation abstracts from any additional social value attached to the information from verdicts.

Intuitively, given Proposition 2, a legal regime and enforcement policy with $\hat{g}_L > \hat{g}_S > pf$ cannot be optimal. Lowering the legal standard down to the social norm yields larger reputational sanctions. The same deterrence could therefore be reached with a smaller probability of detection, thereby reducing enforcement costs. Conversely, a policy with $\hat{g}_L \geq pf > \hat{g}_S$ can be optimal only because the social norm is too weak. The following proposition combines these two observations. Recall that g^* is the optimal threshold in the standard model resulting from the maximization of (1).

Proposition 3 *Suppose society is only concerned with deterrence and enforcement costs. Then the fine is set at the maximum feasible level f_m.*

There exists a threshold $\hat{g} < h$ such that:
(i) If $\hat{g}_S \geq \hat{g}$, then the standard $\hat{g}_L = \hat{g}_S$ is optimal and the probability of detection satisfies $p(f_m + v(\lambda) - v(0)) \leq \hat{g}_S$.
(ii) If $\hat{g}_S < \hat{g}$, then the probability of detection satisfies $pf_m = g^*$ and any standard $\hat{g}_L \geq g^*$ is optimal.

To illustrate, consider the optimal enforcement policy under a strict liability regime, exogenously imposed. Suppose this yields $g_B \leq \hat{g}_S$. Because bad citizens are also motivated by reputational concerns, it must then be that $pf_m < g_B$. Owing to the greater reputational effects, the same level of deterrence can therefore be obtained at lower cost under the legal regime defined by $\hat{g}_L = \hat{g}_S$ (see Figure 4). Conversely, suppose the optimal enforcement policy yields $g_B > \hat{g}_S$. Then $g_B = g_G = g^*$ because the trade-off between deterrence and enforcement costs is now the same as in the standard model. This can arise only when $\hat{g}_S < g^*$, and therefore $\hat{g}_S < h$. The latter inequality, however, is not by itself sufficient for a strict liability regime to be optimal. The benefit from smaller enforcement costs under the regime with legal standard replicating the social norm may be worth the loss from the smaller deterrence (now bounded above by $\hat{g}_S$ for both types). This can be optimal only if $\hat{g}_S$ is not too small.

In part (i) of Proposition 3, the case where $p(f_m + v(\lambda) - v(0)) \leq \hat{g}_S$ holds as a strict inequality corresponds to an interior solution where some of the bad citizens do not comply with the law. When the equality holds, the optimum is a corner solution with $g_B = \hat{g}_S$. Everyone then complies with the law and, therefore, with the social norm.

The proposition states that $\hat{g}_L = \hat{g}_S$ is optimal, but other standards may do as well. Consider Figure 3a and suppose the optimal policy yields the equilibrium E_2. It is then indifferent whether the legal standard is g_L^2 as shown in the figure or $\hat{g}_S$ because both standards yield the same equilibrium, given the enforcement policy. More generally, let g_B be the bad citizens' equilibrium threshold in the optimal policy. Then any standard

$\hat{g}_L \in [g_B, \hat{g}_S]$ yields the same equilibrium.

To summarize, aligning the legal standard with the social norm may be optimal even though the norm is inefficient from a utilitarian point of view. If the norm is too demanding compared to the first best (i.e., $\hat{g}_S > h$), the good citizens will inefficiently refrain from the action. However, the bad citizens need not be overdeterred because the probability of detection can adjust to dampen incentives. A legal standard equal to the social norm may also be optimal when the latter is deficient, provided it is not too much so. Otherwise, part (ii) of Proposition 3 applies and the optimal policy is the same as in the standard model.

The value of information to third parties. In the foregoing, the information conveyed by verdicts is useful in generating incentives through reputational concerns. However, society may also benefit from telling good types from bad ones because this is important in other social or economic interactions. We now take this into account in designing the optimal policy.

Information about the individuals' type may be productive in shaping the future interactions with the individuals. Let $q_t(a)$ denote the total surplus of a future interaction with a type-t individual, given a vector a of decision variables describing measures governing the relationship. These actions may include the amount of relationship-specific investment in a formal or informal context, restrictions on conduct and monitoring, features of performance schemes, the sorting of individuals into different jobs or matches, and the like. We assume $q_G(a) > q_B(a)$ for all a. Relations with good citizens are always more valuable than with bad citizens. Good citizens are more valuable because they are more trustworthy owing to their disposition to internalize norms and to behave less opportunistically.

The optimized total surplus from a relationship with a type-t individual is $\varphi(t) \equiv \max_a q_t(a)$. When the type is unobservable and the individual is believed to be a good citizen with probability μ, the optimized expected surplus is

$$\varphi(\mu) \equiv \max_a (1-\mu) q_B(a) + \mu q_G(a). \tag{14}$$

It follows trivially that $\varphi(\mu)$ is increasing in μ and, in particular, that it is a convex function.[12]

To illustrate, suppose different jobs are represented by the scalar $a \in [0,1]$. Suppose $q_t(a)$ is strictly concave with $q_B(a)$ maximized at $a = 0$ and $q_G(a)$ maximized at $a = 1$, i.e., job $a = 0$ is the best match for the bad citizen and job $a = 1$ is the best match for the good citizen. From the comparative statics of problem (14), the optimal a will be increasing in μ. More trustworthy individuals have a comparative advantage in jobs higher on this particular job scale. Letting $a(\mu)$ denote the solution,

$$\varphi''(\mu) = -\frac{[q'_G(a(\mu)) - q'_B(a(\mu))]^2}{(1-\mu) q''_B(a(\mu)) + \mu q''_G(a(\mu))} > 0.$$

Posterior beliefs about one's type depend on publicly available information, say a signal X so that beliefs can be written as $\mu(X)$. From an ex ante point of view, $\mu(X)$ is a random variable with expected value equal to λ, the prior probability that an individual is a good citizen. If a signal Y is more informative about types than the signal X, then $\mu(Y)$ is a mean preserving spread of $\mu(X)$; see Ganuza and Penalva (2010). Because φ is convex, $E[\varphi(\mu(Y))] \geq E[\varphi(\mu(X))]$ and strictly so if φ is strictly convex. The expected value of future social or economic interactions is then larger under the more informative signal Y.

We use this framework to incorporate the productive value of the information from verdicts in our welfare calculus. Reputational utility is $v(\mu) = \rho\varphi(\mu)$ where $\rho \in (0,1]$ is the individuals' share of the surplus from a future interaction, hence $1-\rho$ is the counterparts' share. Publicly available information consists of the "convicted", the "detected and dismissed", and the "no news" events, i.e., the signal is $X \in \{C, D, N\}$. Averaged over all individuals, the value of future contractual interactions is

$$\overline{\varphi} \equiv P_G\varphi(\mu_C) + P_D\varphi(\mu_D) + P_N\varphi(\mu_N) \tag{15}$$

where P_C, P_D and P_N are the probabilities of the events.[13] Welfare is now redefined as

$$\overline{W} \equiv W + \overline{\varphi} \tag{16}$$

where W is as in equation (13).

It is worth emphasizing that the information from verdicts yields social benefits, in addition to being useful for deterrence, only if it leads to different actions conditional on the information. Suppose, to the contrary, that the solution to problem (14) is always the same action, say $\hat{a}$. Then

$$\varphi(\mu) = (1-\mu)q_0(\hat{a}) + \mu q_1(\hat{a})$$

is linear in the posterior belief about the individual's type. The information provided to third parties has no social value because $\overline{\varphi} = \varphi(\lambda)$ irrespective of the properties of the signal. Reputational effects are then purely redistributive and maximizing $\overline{W}$ is equivalent to maximizing W. We assume that φ is strictly convex, implying that different beliefs always entail different actions to govern a relationship.

The variables on the right-hand side of (15) depend only on p, $\hat{g}_L$ and g_B. To compare different public signals, it is therefore sufficient to study the properties of the function $\overline{\varphi}(g_B, p, \hat{g}_L)$. We keep g_B constant when considering changes in p or $\hat{g}_L$, i.e., we are considering the partial (or direct) effects of changes in these policy variables.

Lemma 2 *If $g_B \geq \hat{g}_S$, then $\overline{\varphi} = \varphi(\lambda)$. If $g_B < \hat{g}_S$, then $\overline{\varphi} > \varphi(\lambda)$ and is increasing in p, decreasing in g_B, increasing in $\hat{g}_L$ for $\hat{g}_L < \hat{g}_S$, and decreasing in $\hat{g}_L$ for $\hat{g}_L > \hat{g}_S$.*

Everything else equal, aligning the legal standard with the social norm improves the information provided to third parties. However, the lemma points to possible conflicts between deterrence and the provision of information. Consider an increase in the fine, a costless measure. When $g_B < h$, a larger fine increases W if it increases deterrence, as in the interior equilibrium E_2 of Figure 3a. However, by itself, more deterrence reduces $\overline{\varphi}$. Similar observations apply to changes in the legal standard and the probability of detection.

Proposition 4 *Suppose society trades off deterrence and enforcement costs against the informational value of verdicts. When $\hat{g}_S \geq \hat{g}$, then in an optimal policy:*
(i) either $\hat{g}_L = \hat{g}_S$ with a fine that need not be maximal;
(ii) or $\hat{g}_L > \hat{g}_S$ with the fine set equal to zero.

We consider only the situation where the social norm is not too deficient, i.e., $\hat{g}_S \geq \hat{g}$ where the latter is the threshold defined in Proposition 3. The optimal policy in a society unconcerned with the informational value of verdicts is then to set the legal standard equal to the social norm. When the value of information to third parties is also a concern, the optimal policy will seek to increase the signal value of the "no-news" event. This is accomplished by a larger probability of detection. However, greater publicity increases deterrence, which has a negative effect on the informational value of "no-news". It is then desirable to reduce the fine in order to dampen deterrence, which can only go so far as a fine equal to zero. When this constraint is binding, part (ii) of Proposition 4 applies. Dampening the deterrence effects of more detection is then obtained by increasing the legal standard above the social norm. Convictions are then noisier signals.[14]

To see this, write the function defined in (13) as $W(g_B, p)$, where we use the fact that $g_G = \max(\hat{g}_S, g_B)$. The optimal policy maximizes

$$\overline{W}(g_B, p, \hat{g}_L) \equiv W(g_B, p) + \overline{\varphi}(g_B, p, \hat{g}_L) \text{ where } g_B = g_B(p, f, \hat{g}_L).$$

In the solution, the probability of detection and the legal standard satisfy the first-order conditions:

$$\frac{\partial \overline{W}}{\partial p} = \left[W_{g_B} + \overline{\varphi}_{g_B}\right] \frac{\partial g_B}{\partial p} - c'(p) + \overline{\varphi}_p = 0, \tag{17}$$

$$\frac{\partial \overline{W}}{\partial \hat{g}_L} = \left[W_{g_B} + \overline{\varphi}_{g_B}\right] \frac{\partial g_B}{\partial \hat{g}_L} + \overline{\varphi}_{\hat{g}_L} = 0. \tag{18}$$

The fine depends on the sign of

$$\frac{\partial \overline{W}}{\partial f} = \left[W_{g_B} + \overline{\varphi}_{g_B}\right] \frac{\partial g_B}{\partial f}. \tag{19}$$

In the proof, we show that $\hat{g}_L$ cannot be binding in the solution.[15] It follows that condition (18) can only be satisfied with $\hat{g}_L \geq \hat{g}_S$. When the solution is $\hat{g}_L = \hat{g}_S$, condition (18) holds with $\partial g_B / \partial \hat{g}_L$ and $\overline{\varphi}_{\hat{g}_L}$ both equal to zero. Depending on the sign of (19), the fine may then be maximal, zero or some value in between. When the solution is $\hat{g}_L > \hat{g}_S$, condition (18) holds with $\partial g_B / \partial \hat{g}_L$ and $\overline{\varphi}_{\hat{g}_L}$ both negative, implying that $W_{g_B} + \overline{\varphi}_{g_B}$ is negative, so condition (19) then implies that the fine is zero.

When the optimal policy is characterized by a positive fine, the optimal legal standard replicates the social norm. As before, letting g_B be the bad citizens' equilibrium threshold, any standard $\hat{g}_L \in [g_B, \hat{g}_S]$ would yield the same deterrence but only $\hat{g}_L = \hat{g}_S$ is optimal because this provides more information to third parties. When the fine is positive but less than maximal, then $W_{g_B} = -\overline{\varphi}_{g_B}$ and $\overline{\varphi}_p = c'(p)$, where the latter follows from (17). The optimal fine then trades off the marginal benefit from more deterrence against the marginal informational loss. In turn, the marginal informational benefit from greater detection equals the marginal detection cost.

Optimal deterrence. When deterrence is the only concern, some underdeterrence (i.e., $g_B < h$) is optimal in order to economize on enforcement costs, as in the standard model without reputational sanctions. This is not necessarily so when verdicts convey valuable information to third parties.

Corollary 1 *When the law only aims at deterrence, bad citizens are underdeterred. When $\hat{g}_S > h$ and the law also aims at providing valuable information to third parties, bad citizens may be over-deterred. The optimal fine is then equal to zero.*

Overdeterrence of the bad citizens means that W_{g_B} is negative, implying that (19) is negative, so that the fine is zero. With W_{g_B} negative, condition (17) implies

$$\overline{\varphi}_p + \overline{\varphi}_{g_B} \frac{\partial g_B}{\partial p} > 0, \tag{20}$$

which can hold only if $g_B < \hat{g}_S$. Thus, over-deterrence can only arise when the social norm is more demanding than the first-best level. Improving future allocative decisions, through the information conveyed to third parties, is then marginally more valuable than improving current allocative decisions.

5 Discussion

Comparison with the literature. It is useful to compare our framework with one often used in modelling the stigma from convictions.[16] In the latter, agents differ in the benefit from committing an illegal act or in the cost of complying with regulations. Using our notation, an agent's unobservable type is g. For instance, large values reflect impulsiveness or the idiosyncratic gains from criminal activity; in the commercial context, they reflect some form of organizational failure. g is negatively correlated with the agent's productivity in future interactions with third parties. In most of this literature, illegal acts are strict liability offences. For a given expected fine, some individuals violate the law and others comply. Offenders therefore signal that they have a large g, thereby triggering reputational sanctions.

In the present framework, g is not specific to the individual. It describes the private material benefit from engaging in some action in the various circumstances agents happen to face. What distinguishes agents is their willingness to sacrifice current material interest in order to comply with some background social norm. For simplicity, we assumed a two-type population with types $t \in \{B, G\}$, but our analysis can be recast in terms of a continuum of types. For example, types are $t \in [0, 1]$ with $t\hat{g}_S$ as the "willingness to pay" of a type-t agent in order to comply with the norm. The net benefit from engaging in the action is then

$$b_t = \begin{cases} g - t\hat{g}_S \text{ if } g < \hat{g}_S \\ g \text{ if } g \geq \hat{g}_S \end{cases}$$

An agent who engages in the action, when the observable cost g of not doing so is small, reveals that he is a low type and is viewed unfavorably. This reformulation yields essentially the same results as our two-type set-up. In particular, when $pf \geq \hat{g}_S$, all types behave the same and no inferences can be drawn from violations of the law. When $pf < \hat{g}_S$, convictions are more likely for low types.

Informal enforcement and overdeterrence. In our analysis, there is no direct informal enforcement of underlying norms because behavior is not directly observable. Information is available only through public enforcement. Our results continue to hold if we allow informal enforcement, so long as society's directly available information is sufficiently imperfect.

Suppose the contrary, i.e., behavior and circumstances are perfectly observable independently of public enforcement. Redefine the event C as "engaged in the action in circumstances $g < \hat{g}_S$"; the event D as "engaged in the action in circumstances $g \geq \hat{g}_S$"; and the event N as "did not engage in the action". These events are all that matters for reputational sanctions or for conveying valuable information to third parties. Borrowing from Lemma 1, posterior beliefs satisfy $\mu_C = 0$, $\mu_D = \lambda$ and

$$\mu_N(g_B) = \frac{\lambda Z(\max(\hat{g}_S, g_B)))}{\lambda Z(\max(\hat{g}_S, g_B)) + (1-\lambda)Z(g_B)} \tag{21}$$

Legal design no longer plays any informational role. Without loss of generality, we therefore focus on strict liability offences.

Although one's conduct is directly observable, let us assume that legal authorities must nevertheless rely on formal ("verifiable") auditing of behavior. They audit with some probability p at a cost. As before, public signals are uninformative when the expected fine $pf \geq \hat{g}_S$, because all agents then behave the same. When $pf < \hat{g}_S$, the bad citizens' equilibrium threshold solves

$$g_B = \min[\hat{g}_S, pf + v(\mu_N(g_B)) - v(0)]. \quad (22)$$

Stigma effects now occur with certainty, but the fine is only imposed with some probability. Although legal design plays no role, the enforcement policy affects the information provided by public signals through the effect of the expected fine on the equilibrium g_B.

The informal enforcement of norms suggests the possibility of overdeterrence compared with the first-best utilitarian level. There are two possible cases.

Case 1: $\hat{g}_S \leq h$ or $v(\lambda) - v(0) \leq h$

This corresponds to situations where the social norm is not too demanding or reputational sanctions are not too strong. Overdeterrence cannot arise if $\hat{g}_S \leq h$. When $\hat{g}_S > h$, however, the solution to (22) may yield overdeterrence for some arbitrary $pf > 0$. The possibility that reputational sanctions cause overdeterrence has been much discussed, in particular with respect to corporate liability.[17] It has been suggested that fines should be reduced to avoid this possibility. However, in the present context, it turns out that the optimal policy imposes the maximal fine and involves no overdeterrence. First, only the expected fine matters, so any level of expected fine should be obtained with the smallest feasible audit probability.[18] Secondly, if society were only concerned with deterrence and enforcement costs, it would choose pf sufficiently small so that overdeterrence does not arise. If society is also concerned with providing information to third parties, it would choose pf even smaller (perhaps with a zero probability of audit) so as to reduce deterrence.

Case 2: $\hat{g}_S > h$ and $v(\lambda) - v(0) > h$

In this case, the social norm is very demanding and reputational sanctions are large. Even when the expected fine is nil, the solution to (22) yields overdeterrence. Too much publicity, together with an excessive background norm, inefficiently distorts behavior. The optimal policy is then simply to do nothing, assuming that publicity cannot be prevented.

Judicial error. We assumed that, when an individual is audited, conduct and circumstances are assessed without error. As is well known, judicial error reduces the incentive effects of legal sanctions (Kaplow and Shavell 1994). It will also reduce the information from verdicts. We tentatively discuss how the risk of error affects the optimal policies.

Suppose that, following an audit, the authorities obtain imperfect signals about the agent's conduct and about circumstances. To start, consider the optimal policy when there are no social norms and reputational concerns. First, the signal about circumstances should be disregarded, which amounts to a strict liability regime. The reason is that conditioning sanctions on circumstances is anyway inessential in the standard model. Secondly, conduct should be assessed on a maximum likelihood basis, i.e., the agent is deemed to have engaged in the action if this "hypothesis" has greater likelihood given the evidence at hand.[19] This decision rule maximizes incentives to comply with the law, thereby allowing any given level of deterrence to be achieved at minimal audit costs.[20]

The foregoing policy is no longer optimal when agents have reputational concerns. In a related model, Fluet and Mungan (2018) show that the optimal decision rule trades-off the direct deterrence effects, taking the level of sanctions as given, and the effect on the level of reputational sanctions which indirectly add to deterrence. The optimal evidence threshold for finding liability depends on the frequency of illegal behavior, e.g., stronger evidence is required when illegal behavior is uncommon. We conjecture that a similar result would obtain in the present context.

6 Summary and Concluding Remarks

Violating the law need not elicit social disapproval or reputational sanctions. Offences have reputational effects insofar as they signal non-adherence to underlying norms. We take these norms as given and consider the informational role of offences, focusing on the "belief shaping role of the law" (Shapira 2016, p. 1247) as opposed to its value or preference shaping role (Cooter 1998). Enforcing the law may reveal whether given background norms were violated. How much so depends on the design of legal obligations. We analyzed the implications in the so-called specific deterrence context where detecting misconduct is costly.

We consider norms that exhibit some congruence with utilitarian welfare, although they need not be efficient in this respect, i.e., they may be more or less demanding than what would maximize utilitarian welfare. For instance, social media users may expect very high standards concerning private data protection, perhaps above what would be justified on a cost-benefit basis. Conversely, individuals may be too lenient towards underreporting of income to avoid paying taxes. We study two possible channels of interaction between legal obligations and background norms.

First, legal obligations should be designed so as to efficiently harness reputational motivations, because this economizes on public enforcement costs. Violations of norms, rather than violations of the law *per se*, are correlated with undesirable characteristics. Offences are less noisy signals of norm violations when legal standards align with prevailing norms. The proviso is that background norms must not be too deficient, otherwise too little deterrence would ensue. The second channel of interaction concerns the informational value of verdicts to third parties, which may yield social benefits in addition to the deterrence effect of the information. In a utilitarian framework, the information from verdicts has social value only if it is conducive to productive actions, by contrast with purely redistributive reputational effects.

As a general rule, harnessing reputational motivations for deterrence purposes and providing valuable information to third parties yield the same prescription regarding the design of offences. However, deterrence relies on the threat of sanctions, whether legal or reputational. Deterrence is therefore consistent with little information being revealed at equilibrium, e.g., there is no screening of types in a pooling equilibrium where everyone complies with the law. Thus, there is a trade-off between deterrence in the particular case and providing valuable information for future allocative decisions. As noted by Iacobucci (2014, p. 189), focusing on deterrence is too narrow: "It may be socially preferable in some circumstance to adjust legal penalties to allow actors to reveal their type than to adjust legal sanctions to promote optimal deterrence."

Conveying information to third parties translates into greater detection effort, so that violating the law earns greater publicity. Because misconduct is more often revealed, avoiding prosecution is also a more significant signal. The significance of "no prosecution" is then an essential element of the optimal policy. We show that it generally relies on non maximal fines, in contrast to the Beckerian principle. The optimal fine mitigates the deterrent effect of publicity and trades off deterrence against the provision of information. When providing information is very valuable, the optimal fine is nil. Rasmusen (1996) remarks that a legal sanction equal to zero is often a reasonable approximation of how the law operates, e.g., probation or community work. Similarly, formal legal sanctions for corporate misconduct are often dwarfed by market reputational sanctions. As a practical matter, regulatory authorities often rely on mere public reprimands. We find that symbolic legal sanctions are optimal only if providing information to third parties is an essential concern.

Our results can be extended and qualified in many ways. For instance, the point has often been made that the stigma attached to criminal records lowers the opportunity costs of future crimes. This is the argument behind ban-the-box legislations and it extends to the future disincentive effects of reputational losses in general. We made the simplifying assumption of a two-phase game: first, agents decide whether to undertake the harmful action; next, a matching value with third parties is obtained and agents earn their reputational payoff. Suppose this this set-up is replicated in a two-period setting. At the beginning of each period, agents decide whether to undertake the action; at the end of each period, a matching value ensues. A cursory analysis suggests that muting reputational effects (e.g., keeping convictions secret) cannot be optimal. However, the dynamics introduces new trade-offs with respect to the timing of the information provided to third parties. For instance, when the legal sanction in the first period is not maximal, it should be greater in the second period for agents twice convicted (as in Funk, 2004) because there is no point in mitigating deterrence for agents whose type is known. Another extension is to relax our informational assumptions. In our main analysis, we discard any direct enforcement of norms because information about behavior relies solely on public enforcement. The more realistic case is where society has some information and public enforcement can provide more reliable information.

Appendix

Proof of Lemma 1. Take g_B as given and satisfying $g_B < \hat{g}_L$. The fraction of bad citizens violating the law is then $Z(\hat{g}_L) - Z(g_B)$. If $\max(\hat{g}_S, g_B) \geq \hat{g}_L$, the good citizens never violate the law; otherwise, a fraction $Z(\hat{g}_L) - Z(\max(\hat{g}_S, g_B))$ does. Both categories of individuals are detected with probability p. Applying Bayes' rule then yields (8). For an individual labelled N, either the act was not committed or it was but was not detected. The fraction of bad citizens in this situation is

$$Z(g_B) + (1-p)(1 - Z(g_B) = 1 - p + pZ(g_B).$$

Similarly, for the good citizen the fraction is

$$1 - p + pZ(\max(\hat{g}_S, g_B)).$$

Applying Bayes' rule then yields (6). Finally, for bad citizens the probability of event D is $p[1 - Z(\hat{g}_L)]$. For good citizens it is $p[1 - Z(\max(\hat{g}_S, \hat{g}_L)$. Applying Bayes' rule then yields (9). It is straightforward to verify that $\mu_C \leq \mu_D \leq \lambda \leq \mu_N$. A similar argument applies to the case $g_B \geq \hat{g}_L$. ■

Lemma A1 *Let $g_c(p)$ solve $g = p[v(\mu_N(g)) - v(\mu_D(g))]$ where $\mu_N(g)$ and $\mu_D(g)$ are obtained from (6) and (9) by setting $g_B = \hat{g}_L = g$. Then $g_c(p)$ is unique and satisfies $g_c(p) < \hat{g}_S$. At equilibrium, $g_B = g_c(p)$ if $\hat{g}_L < g_c(p)$ while g_B and Δ solve (10) and (11) if $\hat{g}_L \geq g_c(p)$.*

Proof. Let g_B solve (10) and (11). Suppose first that $g_B < \hat{g}_L$. Then

$$\begin{aligned} g_B &= p[f + v(\mu_N(g_B)) - v(\mu_C(g_B, \hat{g}_L))] \\ &\geq p[v(\mu_N(g_B)) - v(\mu_C(g_B, \hat{g}_L))] \\ &\geq p[v(\mu_N(g_B)) - \mu_D(\hat{g}_L)] \end{aligned}$$

where the last inequality follows from Lemma 1. It follows that (12) holds.

Next suppose that $g_B = \hat{g}_L$. If $\hat{g}_L > \hat{g}_S$, then $g_B = g_G$ and $\mu_N = \mu_D = \lambda$ by Lemma 1 again. Hence (12) is trivially satisfied. So let $\hat{g}_L \leq \hat{g}_S$. Define

$$\varphi(\hat{g}) = \hat{g} - p[v(\mu_N(\hat{g})) - v(\mu_D(\hat{g}))]$$

where $\mu_N(\hat{g})$ is obtained by setting $g_B = \hat{g}$ in (6) and $\mu_D(\hat{g})$ is obtained by setting $\hat{g}_L = \hat{g}$ in (9). From the expressions in Lemma 1, $\varphi(\hat{g})$ is increasing. Moreover, $\mu_N(\hat{g}) > \lambda > \mu_D(\hat{g})$ if $\hat{g}_L < \hat{g}_S$ and $\mu_N(\hat{g}) = \mu_D(\hat{g}) = \lambda$ if $\hat{g}_L = \hat{g}_S$, hence $\varphi(0) < 0$ and $\varphi(\hat{g}_S) > 0$. It follows that there exists $g_c < \hat{g}_S$ as stated. We write $g_c(p)$ to emphasize that it is a function of p. ■

Proof of Proposition 1. Either $g_B = g_G > \hat{g}_S$ or $g_B \leq g_G = \hat{g}_S$. By Lemma 1, the first case implies $\Delta(g_B, \hat{g}_L) \equiv v(\mu_N(g_B)) - v(\mu_C(g_B, \hat{g}_L)) =$

0. Thus, it can arise only if $pf > \hat{g}_S$ and the equilibrium is then simply $g_B = g_G = pf$. A policy with $pf \leq \hat{g}_S$ therefore yields the second case. The relevant domain for g_B is then the interval $[pf, \min(\hat{g}_L, \hat{g}_S)]$. If $pf = \min(\hat{g}_L, \hat{g}_S)$, the equilibrium is trivially $g_B = pf$, so let $ps < \min(\hat{g}_L, \hat{g}_S)$. The equilibrium g_B is then a solution to

$$g_B = \min\left[\hat{g}_L, p(f + \Delta(g_B, \hat{g}_L))\right] \tag{A1}$$

Equivalently, g_B solves

$$\psi(g_B) \equiv \min\left[\hat{g}_L, p(f + \Delta(g_B, \hat{g}_L))\right] - g_B = 0,\ g_B \in [pf, \min(\hat{g}_L, \hat{g}_S)], \tag{A2}$$

where $\psi(g_B)$ is a continuous function. By Lemma 1, $\Delta(pf, \hat{g}_L) > 0$ and therefore $\psi(pf) > 0$. For the case $\hat{g}_L \leq \hat{g}_S$, obviously $\psi(\hat{g}_L) \leq 0$. Because $\Delta(g_B, \hat{g}_L)$ is strictly decreasing in g_B in the relevant domain, so is $\psi(g_B)$ and the equilibrium is therefore unique and satisfies $g_B > pf$. For the case $\hat{g}_L > \hat{g}_S$, $\Delta(\hat{g}_S, \hat{g}_L) = 0$ so that $\psi(\hat{g}_L) = pf - \hat{g}_S < 0$. Again $\psi(g_B)$ is strictly decreasing, ensuring uniqueness with $g_B > 0$.

(i) For $pf < \hat{g}_L \leq \hat{g}_S$, the above argument shows that $g_G = \hat{g}_S$ and $g_B \in (pf, \hat{g}_L]$. If $p(f + \Delta(\hat{g}_L, \hat{g}_L)) \geq \hat{g}_L$, the equilibrium satisfies $g_B = \hat{g}_L$. Otherwise $g_B < \hat{g}_L$ and solves $g_B = p(f + \Delta(g_B, \hat{g}_L; p))$ where we now take into account that Δ depends on p. Differentiating totally with respect to p yields

$$\frac{dg_B}{dp} = \frac{s + \Delta + p\Delta_p}{1 - p\Delta_{g_B}} > 0. \tag{A3}$$

From the expressions in Lemma 1, Δ_{g_B} is negative while Δ_p is positive because $\partial\mu_N/\partial p > 0$.

(ii) For $pf < \hat{g}_S < \hat{g}_L$, the argument is similar except that the solution now satisfies $g_B \in (pf, \hat{g}_S)$. We now have

$$\frac{\partial g_B}{\partial \hat{g}_L} = \frac{p\beta\Delta_{\hat{g}_L}}{1 - p\beta\Delta_{g_B}} < 0, \tag{A4}$$

where $\Delta_{\hat{g}_L}$ is negative because $\partial\mu_C/\partial\hat{g}_L > 0$. ■

Proof of Proposition 2. From Proposition 1, $g_G = \hat{g}_S$ whenever $pf < \hat{g}_S$. For any $\hat{g}_L \le \hat{g}_S$, a sufficiently small value of pf yields an equilibrium $g_B < \hat{g}_L$ which is therefore constant in $\hat{g}_L$; by contrast, a sufficiently large value yields the equilibrium $g_B = \hat{g}_L$, hence g_B is then increasing in $\hat{g}_L$; in either case, deterrence is maximized by $\hat{g}_L = \hat{g}_S$. For $\hat{g}_L > \hat{g}_S$, g_B is monotonically decreasing in $\hat{g}_L$. Under any enforcement policy, the deterrence maximizing standard is therefore $\hat{g}_L = \hat{g}_S$. ■

Proof of Proposition 3. Let

$$W(\hat{g}_S) \equiv \max_{p,f,\hat{g}_L} (1-\lambda) \int_{g_B}^{\infty} (g-h)\, z(g)\, dg + \lambda \int_{\max(g_B,\hat{g}_S)}^{\infty} (g-h)\, z(g)\, dg - c(p) \tag{A5}$$

where $g_B(p, f, \hat{g}_L)$ is the function defined in Proposition 1. Clearly, $f = f_m$. Let $g^* = p^* f_m$ be the threshold resulting from the maximization of (1) in the standard model and denote by W^* the maximized value.

Fact 1: if the solution of (A5) satisfies $g_B > \hat{g}_S$, then $g_B = g_G = g^*$. Thus, $g_B > \hat{g}_S$ implies $\hat{g}_S < g^*$. Equivalently, $\hat{g}_S \ge g^*$ implies $g_B \le \hat{g}_S$.

Fact 2: if $g_B \le \hat{g}_S$, then it is easily seen from Proposition 2 that the optimal policy sets $\hat{g}_L = \hat{g}_S$, so that $g_B \le g_G = \hat{g}_S$. A probability of detection satisfying $p[f_m + v(\lambda) - v(0)] > \hat{g}_S$ would then induce $g_B = \hat{g}_S$ but could be reduced without affecting deterrence. Therefore $p[f_m + v(\lambda) - v(0)] \le \hat{g}_S$ as claimed.

The regimes described in (i) and (ii) are therefore the only two possibilities. $\hat{g}_S \ge g^*$ is sufficient for case (i), so we need only examine the outcome for $\hat{g}_S \in [0, g^*]$. Define

$$V(\hat{g}_S) \equiv \max_{p} (1-\lambda) \int_{g_B(p,f_m,\hat{g}_S)}^{\infty} (g-h)\, z(g)\, dg + \lambda \int_{\hat{g}_S}^{\infty} (g-h)\, z(g)\, dg - c(p). \tag{A6}$$

Then

$$W(\hat{g}_S) = \begin{cases} V(\hat{g}_S) \text{ if } V(\hat{g}_S) \ge W^*, \\ W^* \text{ otherwise.} \end{cases}$$

From the above discussion, $V(g^*) > W^*$ and $V(0) < W^*$. Therefore, by the intermediate value theorem, there exists $\hat{g} \in (0, g^*)$ such that $W(\hat{g}) = V(\hat{g})$. We show that $\hat{g}$ is unique because $V(\hat{g}_S)$ is strictly increasing. Using the envelope theorem,

$$\frac{dV(\hat{g}_S)}{d\hat{g}_S} = (1-\lambda)(h-g_B)z(g_B)\frac{\partial g_B}{\partial \hat{g}_S} + \lambda(h-\hat{g}_S)z(\hat{g}_S) > 0.$$

The sign follows from $g_B(p, f_m, \hat{g}_S) \leq \hat{g}_S \leq g^* < h$ and from

$$\frac{\partial g_B}{\partial \hat{g}_S} = \left.\frac{\partial g_B(p, f_m, \hat{g}_L)}{\partial \hat{g}_L}\right|_{\hat{g}_L = \hat{g}_S} \geq 0,$$

where the strict inequality holds only when the legal standard is binding. ■

Proof of Lemma 2. Rewrite the publicly observable signal as $X \in \{\mu_C, \mu_D, \mu_N\}$. By Lemma 1, when $g_B \geq \hat{g}_S$, then $\mu_C = \mu_D = \mu_N = \lambda$, therefore $\overline{\varphi} = \varphi(\lambda)$. Henceforth, let $g_B < \hat{g}_S$, in which case $\mu_C < \mu_D < \mu_N$. We compare X with the signal $X' \in \{\mu'_C, \mu'_D, \mu'_N\}$ resulting from a change in $\hat{g}_L$, p or g_B. Let H and H' be the cdf's of X and X' respectively, which by construction have the same mean, and define $S \equiv H' - H$. We show that S changes sign only once, a sufficient condition for the distributions to be ranked in terms of an increase or decrease in risk (Rothschild and Stiglitz 1970).

Consider the change from $\hat{g}_L$ to $\hat{g}'_L > \hat{g}_L$. By Lemma 1 and using (**??**) to (**??**), when $\hat{g}'_L \leq \hat{g}_S$, then $P'_C > P_C$, $P'_D < P_D$, $P'_N = P_N$, and $\mu'_C = \mu_C$, $\mu'_D > \mu_D$, $\mu'_N = \mu_N$. Therefore, $P'_C = P_C + \alpha$ and $P'_D = P_D - \alpha$ for some positive $\alpha < P_D$. Then

$$S(x) \equiv H'(x) - H(x) = \begin{cases} 0 & \text{if } x < \mu_C, \\ \alpha & \text{if } \mu_C \leq x < \mu_D, \\ -(P_D - \alpha) & \text{if } \mu_D \leq x < \mu'_D, \\ 0 & \text{if } \mu'_D \leq x. \end{cases}$$

Because $S(x)$ is first positive and then negative, X' has greater risk than X and is therefore more informative.

When $\hat{g}_L > \hat{g}_S$, the change from $\hat{g}_L$ to $\hat{g}'_L > \hat{g}_L$ implies $P'_C > P_C$, $P'_D < P_D$, $P'_N = P_N$, and $\mu'_C > \mu_C$, $\mu'_D = \mu_D$, $\mu'_N = \mu_N$. Again, $P'_C = P_C + \alpha$ and $P'_D = P_D - \alpha$ for some positive $\alpha < P_D$. In this case,

$$S(x) \equiv H'(x) - H(x) = \begin{cases} 0 & \text{if } x < \mu_C, \\ -P_C & \text{if } \mu_C \leq x < \mu'_C, \\ \alpha & \text{if } \mu'_C \leq x < \mu_N \\ 0 & \text{if } \mu_N \leq x \end{cases}$$

Now X' has les risk than X and is therefore less informative. A similar argument applies to changes in p or g_B. ■

Proof of Proposition 4. To complete the argument in the text, we need to consider the possibility that an optimal regime with zero fine involves a binding standard $\hat{g}_L < \hat{g}_S$, i.e., $g_B = \hat{g}_L < pv(\mu_N(g_B))]$. Such a possibility is compatible with condition (18). As in part (ii) of the proposition, $W_{g_B} + \overline{\varphi}_{g_B} < 0$. However, we now have $\partial g_B / \partial \hat{g}_L > 0$ and $\overline{\varphi}_{\hat{g}_L} > 0$. We show that this cannot be optimal.

Suppose such a standard, denoted by $\hat{g}_L^1 < \hat{g}_S$, and let $g_B^{**} = \hat{g}_L^1$ be the associated equilibrium threshold. We assume $\hat{g}_L^1 > g_c(p)$ as defined in Lemma 2, otherwise the standard would have no effect. Note that $g_c(p)$ satisfies:

$$g_c(p) = p[v(\mu_N(g_c(p))) - v(\mu_C(g_c(p), \infty))]. \tag{A7}$$

To see the equivalence with the equation in Lemma A1, set $\hat{g}_L = \infty$ in (8) and $g_B = g_c(p)$ in (8) and (7), so that $\mu_C(g_B, \infty) = \mu_D(g_B)$. We show: (i) that the equilibrium $g_B^{**} = \hat{g}_L^1$ can also be implemented by a zero fine policy with the same p and with some standard $\hat{g}_L^2 > \hat{g}_S$; (ii) that the latter policy provides a more informative signal.

According to claim (i), there exists $\hat{g}_L^2 > \hat{g}_S$ solving

$$g_B^{**} = p[v(\mu_N(g_B^{**})) - v(\mu_C(g_B^{**}, \hat{g}_L^2))]. \tag{A8}$$

Define

$$\psi(g_B^{**}, \hat{g}_L^2) \equiv g_B^{**} - p[v(\mu_N(g_B^{**})) - v(\mu_C(g_B^{**}, \hat{g}_L^2))].$$

By assumption,

$$\psi(g_B^{**}, \hat{g}_S) = g_B^{**} - pv(\mu_N(g_B^{**})) < 0,$$

Because ψ is increasing in g_B^{**} and given (A7),

$$\psi(g_B^{**}, \infty) = g_B^{**} - p[v(\mu_N(g_B^{**})) - v(\mu_C(g_B^{**}, \infty))] > 0 \text{ for } g_B^{**} > g_c(p).$$

Hence, there exists $\hat{g}_L^2 > \hat{g}_S$ solving (A8). Because ψ is also increasing in its second argument, $\hat{g}_L^2$ is unique.

We now prove claim (ii). Denote the support of the policy with the binding standard $\hat{g}_L^1$ by $\{\mu_D, \mu_N\}$, where $\mu_D < \lambda < \mu_N$, and let the probabilities be P_D and P_N. For the policy with the standard $\hat{g}_L^2$, the support is $\{\mu'_C, \mu'_D, \mu'_N\}$ where $\mu'_C < \mu'_D = \lambda$ and $\mu'_N = \mu_N$. The probabilities satisfy $P'_N = P_N$ and $P_C + P'_D = P_D$, i.e., the probability mass initially at μ_D has been redistributed over μ'_C and μ'_D. Because $\mu'_D > \mu_D$, this constitutes a mean preserving spread if $\mu'_C < \mu_D$. From (9) in Lemma 1, given $\hat{g}_L^1 = g_B^{**}$,

$$\mu_D = \frac{\lambda(1 - Z(\hat{g}_S))}{\lambda(1 - Z(\hat{g}_S)) + (1 - \lambda)(1 - Z(g_B^{**}))}$$

From (8),

$$\mu'_C = \frac{\lambda \left[Z(\hat{g}_L^2) - Z(\hat{g}_S)\right]}{\lambda \left[Z(\hat{g}_L^2) - Z(\hat{g}_S)\right] + (1 - \lambda) \left[Z(\hat{g}_L^2) - Z(g_B^{**})\right]}$$

Therefore, $\mu'_C < \mu_D$. ■

Notes

[1]The idea is not new, considering for instance the organization of the so-called Law Merchant in the Middle Ages (see Milgrom, North and Weingast, 1990).

[2]Reputational losses are measured as the drop in firm value in excess of the cost of legally imposed penalties (together with compensation awards

and remedial measures) and have been shown to depend on the type of misconduct and whether those affected can penalize the firm (Karpoff and Lott, 1993; Alexander, 1996; Beatty et al., 1998; Karpoff et al., 2005, 2008). In a U.K. study, Armour et al. (2017) find that reputational losses for financial misconduct are nearly nine times the size of legal sanctions, when misconduct harms so-called related parties (e.g., customers, suppliers or investors).

[3]We borrow from Polinsky and Shavell (2007) in this respect.

[4]Obviously, acts can be interpreted from different perspectives, e.g., acts of omission such as not making a full stop at an intersection versus positive acts such as discharging pollutants in a river.

[5]Our definition of norms is consistent with McAdams' (1997, p. 340), as "informal social regularities that individuals feel obligated to follow because of an internalized sense of duty, because of fear of external non-legal sanctions, or both."

[6]As remarked by Posner (2000), norms of conduct that seem excessive compared to the apparent harm from noncompliance may in fact be useful to screen types.

[7]A tendency to violate a particular norm may suggest a disposition to violate norms generally (Posner and Rasmusen, 1999).

[8]See Rasmusen (1966) Bénabou and Tirole (2006, 2011), Daughety and Reinganum (2010), Deffains and Fluet (2013), Iacobucci (2014), and Mungan (2016a), among others.

[9]In Section 5, we discuss the consequences of relaxing this assumption.

[10]Because everyone complies with the law, the event "convicted" is off-equilibrium, so μ_C cannot be computed from Bayes' rule. We take the limit as g_B approaches $\hat{g}_S$ from the left. This can also be rationalized in terms of Cho and Kreps' (1987) D1 criterion.

[11]Note that informal sanctions may well decrease when expected formal sanctions increase (through a higher fine or more detection), but never to the point of reducing deterrence. This is a standard result; see Mungan (2016a).

[12]This is well known from decision theory. See for instance Gollier (2001),

chapter 24.

[13]The probabilities are the denominators in the relevant expressions of Lemma 1. We consider only the case where $g_B \leq \hat{g}_L$.

[14]The limiting case of the latter policy is the strict liability regime with symbolic fines alluded to in Section 3.

[15]Proposition 1 then implies that $\partial g_0/\partial p$ and $\partial g_0/\partial f$ are both positive, while $\partial g_0/\partial g_L$ is nil when $g_L \leq g_S$ and negative when $g_L > g_S$.

[16]See Rasmusen (1996), Harel and Klement (2007), Bénabou and Tirole (2011), Iacobucci (2014), and Mungan (2016a, 2016b).

[17]See Fischel and Sykes (1996) and Khanna (1996).

[18]This contrasts with Cooter and Porat (2001) who discuss private enforcement in a tort context where the "audit probability" is not a policy variable.

[19]Likelihood is defined as in classical statistics, disregarding prior probabilities.

[20]See Demougin and Fluet (2006). There is no overdeterrence because the audit probability is a decision variable. However, the chilling of desirable acts would be an issue if, as in Kaplow (2011), there are "benign acts" which could be confused with the action considered here.

References

[1] Acemoglu, D., and M.O. Jackson. 2017. "Social Norms and the Encorcement of Laws," 15 *Journal of the European Economic Association* 245-295.

[2] Agan, A., and S. Starr. 2017. "The Effect of Criminal Records on Access to Employment," 107 *American Economic Review: Papers and Proceedings* 560–564.

[3] Alexander, C. R. 1996. "On the Nature of the Reputational Penalty for Corporate Crime: Evidence," 42 *Journal of Law and Economics* 489-526.

[4] Andreoni, J., and B.D. Bernheim. 2009. “Social Image and the 50-50 Norm” 77 *Econometrica* 1607-1636.

[5] Ariely, D., A. Bracha and S. Meier. 2009. “Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially,” 99 *American Economic Review* 544-555.

[6] Armour, J., Mayer, C., and A. Polo. 2017. “Regulatory Sanctions and Reputational Damages in Financial Markets,” 52 *Journal of Financial and Quantitative Analysis* 1429-1448.

[7] Beatty, R.P., H. Bunsis, and J. R.M. Hand. 1998. “The Indirect Economic Penalties in SEC Investigations of Underwriters,” 50 *Journal of Financial Economics* 151-186.

[8] Becker, G. 1968. “Crime and Punishment: An Economic Approach,” 76 *Journal of Political Economy* 169-217.

[9] Bénabou, R., and J. Tirole. 2006. “Incentives and Prosocial Behavior,” 96 *American Economic Review* 1652-1678.

[10] Bénabou, R., and J. Tirole. 2011. “Laws and Norms,” NBER, Working paper 17579.

[11] Bernheim, B.D. 1994. “A Theory of Conformity,” 102 *Journal of Political Economy* 905-953.

[12] Charness, G., and M. Rabin. 2002. “Understanding Social Preferences with Simple Tests,” 117 *Quarterly Journal of Economics* 817-869.

[13] Charness, G., L. Rigotti, and A. Rustichini. 2016. “Social Surplus Determines Cooperation Rates in the One-Shot Prisoner’s Dilemma,” 100 *Games and Economic Behavior* 113-124.

[14] Cho, I. K., and D. Kreps. 1987. “Signaling games and stable equilibria,” 102 *Quarterly Journal of Economics* 179-221.

[15] Cooter, R. 1998. "Expressive Law and Economics," 27 *Journal of Legal Studies*, 585-608.

[16] Cooter, R., and A. Porat. 2001. "Should Courts Deduct Nonlegal Sanctions from Damages?" 30 *Journal of Legal Studies* 401-422.

[17] Dana, J., D.M. Cain, and R. Dawes. 2006. "What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games," 100 *Organizational Behavior and Human Decision Processes* 193-201.

[18] Daughety, A., and J. Reinganum. 2010. "Public Goods, Social Pressure, and the Choice Between Privacy and Publicity," 2 *American Economic Journal: Microeconomics* 191-222.

[19] Deffains, B., and C. Fluet. 2013. "Legal Liability when Individuals Have Moral Concerns," 29 *Journal of Law, Economics, and Organization* 930-955.

[20] Demougin, D., and C. Fluet. 2006. "Preponderance of Evidence," 50 *European Economic Review* 963-976.

[21] Ellickson, R. C. 1991. *Order Without Law: How Neighbors Settle Disputes*. Cambridge, Mass.: Harvard University Press.

[22] Engelmann, D., and M. Strobel. 2004. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," 94 *American Economic Review* 857-869.

[23] Fischel, D., and A. Sykes. 1996. "Corporate Crime," 25 *Journal of Legal Studies* 319-350.

[24] Fluet, C. and M. Mungan (2018), "The Signal-Tuning Function of Liability Regimes," George Mason University – Antonin Scalia Law School, Working paper.

[25] Funk, P. 2004. "On the Effective Use of Stigma as a Crime-Deterrent," 48 *European Economic Review* 2004.

[26] Funk, P. 2010. "Social Incentives and Voter Turnout: Theory and Evidence," 8 *Journal of the European Economic Association* 1077-1103.

[27] Ganuza, J.-J., and J. S. Penalva. 2010. "Signal-Ordering Based on Dispersion and the Supply of Private Information in Auctions," 78 *Econometrica* 1007-1030.

[28] Gollier, C. 2001. *The Economics of Risk and Time*. Cambridge, Mass.: MIT Press.

[29] Harel, A., and A. Klement. 2007. "The Economics of Stigma: Why More Detection of Crime May Result in Less Stigmatization," 36 *Journal of Legal Studies* 355-378.

[30] Iacobucci, E.M. 2014. "On the Interaction Between Legal and Reputational Sanctions," 43 *Journal of Legal Studies* 189-207.

[31] Kahan, D.M., and E.A. Posner. 1999. "Shaming White-Collar Criminals: A Proposal for Reform of the Federal Sentencing Guidelines," 42 *Journal of Law and Economics* 365-392.

[32] Kaplow, L. 2011. "On the Optimal Burden of Proof," 119 *Journal of Political Economy* 1104-1140.

[33] Kaplow, L., and S. Shavell. 1994. "Accuracy in the Determination of Liability," 37 *Journal of Law and Economics* 1-15.

[34] Karpoff, J. M. 2012. "Does Reputation Work to discipline Corporate Misconduct?" in M.L. Barnett and T.G. Pollock, eds., *Oxford Handbook of Corporate Reputation*. Oxford: Oxford University Press.

[35] Karpoff, J.M., and J.R. Lott. 1993. "The Reputational Penalty Firms Bear from Committing Criminal Fraud," 36 *Journal of Law and Economics* 757-802.

[36] Karpoff, J.M., J.R. Lott, and E.W. Wehrly. 2005. "The Reputational Penalties for Environmental Violations: Empirical Evidence," 48 *Journal of Law and Economics* 653-675.

[37] Karpoff, J. M., D. Scott Lee, and G.S. Martin. 2008. "The Cost to Firms of Cooking the Books," 43 *Journal of Financial and Quantitative Analysis* 581-611.

[38] Khanna, V. 1996. "Corporate Criminal Liability: What Purpose Does it Serve?" 109 *Harvard Law Review* 1477-1537.

[39] Masclet, D., C. Noussair, S. Tucker, and M.C. Villeval. 2003. "Monetary and Non-Monetary Punishment in the Voluntary Contribution Mechanism," 93 *American Economic Review* 366-380.

[40] McAdams, R.H. 1997. "The Origin, Development, and Regulation of Norms," 96 *Michigan Law Review* 338-433.

[41] McAdams, R.H., and E. B. Rasmusen. 2007. "Norms in Law and Economics," in A.M. Polinsky and S. Shavell, eds., *Handbook of Law and Economics*, Vol. 1. New York: North-Holland.

[42] Milgrom, P.R., C.N. Douglass, and B.R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs," 2 *Economics and Politics* 1-23.

[43] Mungan, M. 2016a. "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.

[44] Mungan, M. 2016b. "Stigma Dilution and Over-Criminalization," 18 *American Law and Economics Review* 88-121.

[45] Parsons, C., J. Sulaeman and S. Titman. 2018. "The Geography of Financial Misconduct," 73 *The Journal of Finance* 2087-2137.

[46] Polinsky, M.A., and S. Shavell. 2000. "The Economic Theory of Public Enforcement of Law," 38 *Journal of Economic Literature* 45-76.

[47] Polinsky, M.A., and S. Shavell. 2007. "The Theory of Public Enforcement of Law," in A.M. Polinsky and S. Shavell, eds., *Handbook of Law and Economics*, Vol. 1. New York: North-Holland.

[48] Posner, E.A. 2000. *Law and Social Norms*. Cambridge, Mass.: Harvard University Press.

[49] Posner, R. A., and E. Rasmusen. 1999. "Creating and Enforcing Norms, with Special Reference to Sanctions," 19 *International Review of Law and Economics* 369-382.

[50] Rasmusen, E. 1996. "Stigma and Self-Fulfilling Expectations of Criminality," 39 *Journal of Law and Economics* 519–43.

[51] Rothschild, M., and J.E. Stiglitz. 1970. "Increasing Risk: I. A Definition," 2 *Journal of Economic Theory* 225-243.

[52] Shapira, R. 2016. "Reputation through Litigation: How the Legal System Shapes Behavior by Producing Information," 91 *Washington Law Review* 1193-1252.

[53] Shavell, S. 2002. "Law versus Morality as Regulator of Conduct," 4 *American Law and Economic Review* 227-257.

[54] Uggen, C., M. Vuolo, S. Lageson, E. Ruhland, and H. K. Whitham. 2014. "The Edge of Stigma: An Experimental Audit of the Effects of Low-Level Criminal Records on Employment," 52 *Criminology* 627–54.

[55] Zasu, Y. 2007. "Sanctions by Social Norms and the Law: Substitutes or Complements?" 36 *Journal of Legal Studies* 379-396.

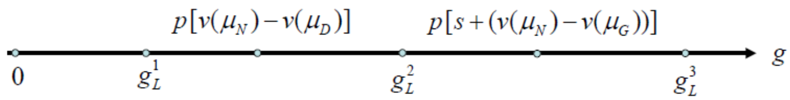


Figure 1. Thresholds and legal standards

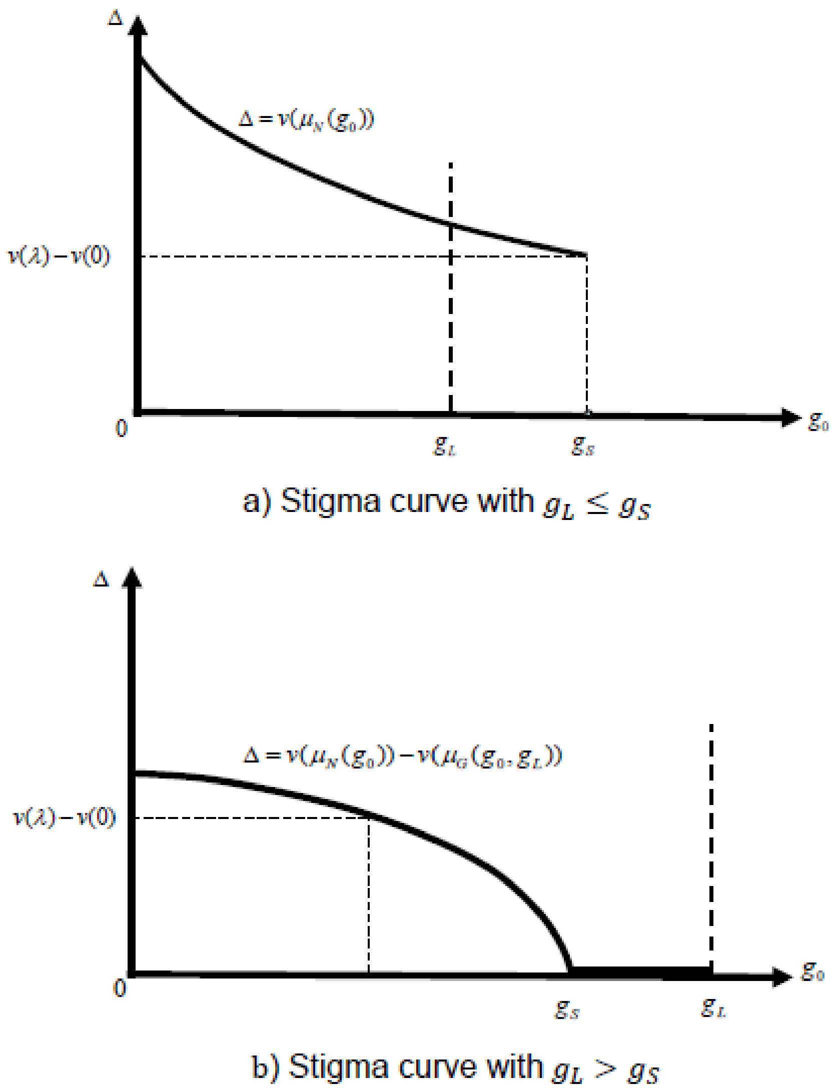
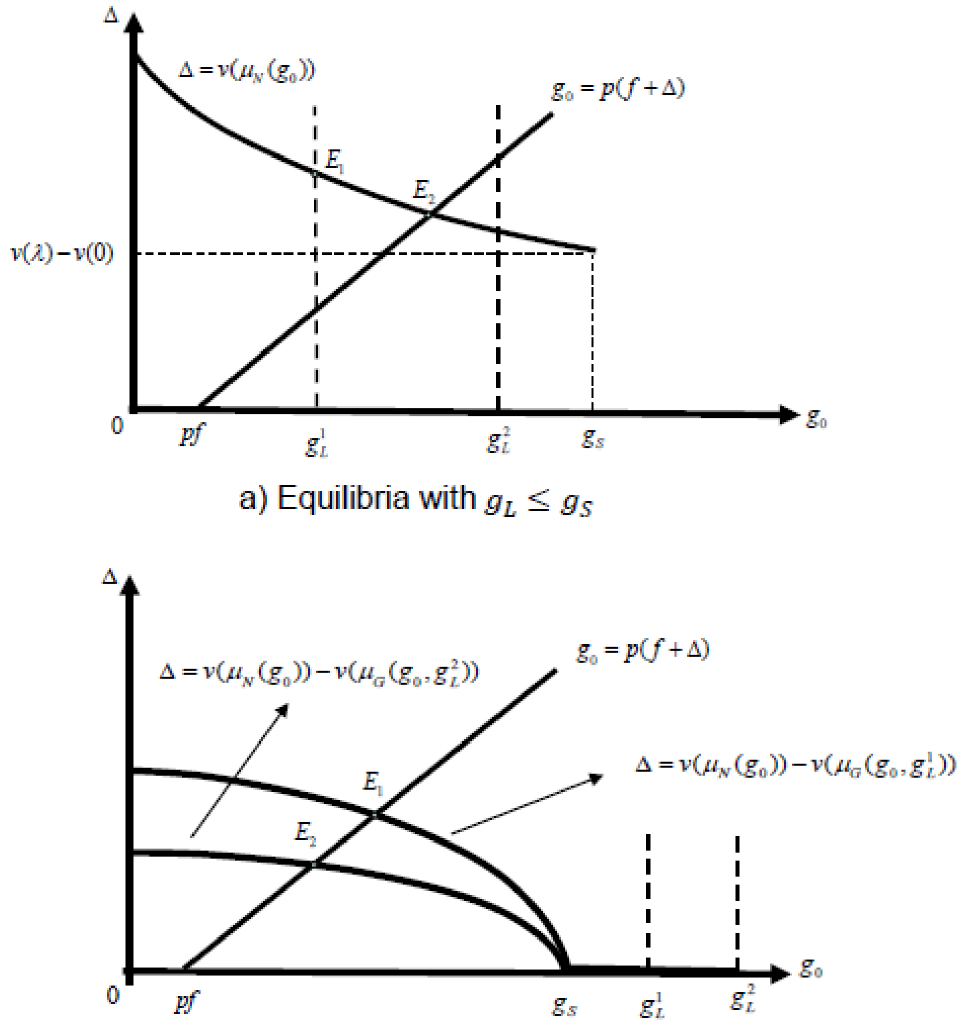


Figure 2. Reputational sanctions



a) Equilibria with $g_L \leq g_S$

b) Equilibria with $g_L > g_S$

Figure 3. Equilibria

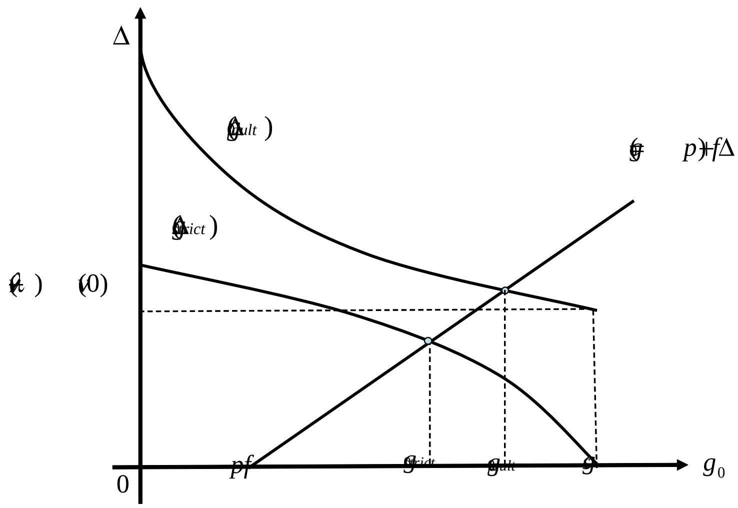


Figure 4. Strict versus fault-based offence with stanfard $g_L = g_S$