2019-02

Social Norms and Legal Design

Bruno Deffains
Claude Fluet

Février / February 2019

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

www.crrep.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. When society is only concerned with the trade-off between deterrence and enforcement costs, legal standards defining offences should align on 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 on 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.

KEYWORDS: Stigmatization, reputational sanctions, social norms, law enforcement, legal standard, compliance, deterrence.
JEL: D8, K4, Z13

Bruno Deffains : Université Panthéon Assas, CRED and Institut Universitaire de France - Bruno.Deffains@u-paris2.fr
Claude Fluet : Université Laval, CRREP and CRED - Claude.Fluet@fsa.ulaval.ca

We thank Robert Cooter, Jennifer Reinganum, and Kathryn Spier for very helpful comments on an earlier version and 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., 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."

Our contribution is to analyze this issue 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 the probability of detection or investigation, reputational sanctions are maximized when legal standards defining offences align on 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 on 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. 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 on 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.[1] The so-called Laws

[1] Links with the literature are also discussed in the last two sections.

and Norms literature has looked at closely related issues, more or less formally and from a variety of angles. McAdams (1997), Posner (2000), 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 overcriminalization, 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. Fluet and Mungan (2018) analyze the optimal burden of proof when care levels are observed with error and agents have reputational motivations. 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).[2]

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 distri-

[2] The idea is not new, considering for instance the organization of law merchant in the Middle Ages (see Milgrom, North and Weingast, 1990).

bution 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. 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). 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). Similar results in the U.S. show that reputational losses may be large, depending on the type of misconduct and whether those affected can penalize the firm (e.g., Karpoff and Lott, 1993; Alexander, 1996; Beatty et al., 1998; Karpoff et al., 2005, 2008). 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 circumstances. g denotes both the objective circumstances 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 g_L such that the action is illegal when $g < g_L$; 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

[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.

administrative guidelines. Under the regime g_L, an agent's expected utility in circumstance g is $(g - pf)e$ if $g < g_L$ and ge if $g \geq g_L$, where e is 0 or 1. The agent therefore engages in the action when $g \geq \min(g_L, pf)$.

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

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

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 $g_L \geq pf_m$; 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 and $g^* = p^* f_m$ 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 $g_L \geq g^*$, 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., $g_L = \infty$).

Social norms and reputational concerns. Social norms are non legal standards of conduct potentially supported by informal sanctions.[5] A norm is a threshold g_S. The action is viewed as inappropriate in circumstances

[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."

$g < g_S$, otherwise not. A natural interpretation is that, to some extent, one ought to take into account the potential harm to others. g_S reflects the socially acceptable trade-off in this respect, i.e., causing harm in order to avoid negligible private costs is blameworthy. It may also well be that the norm is very demanding, some threshold $g_S > h$. Societal standards of behavior or professional codes of conduct may be disproportionate compared to the apparent potential harm from noncompliance. Norms of conduct that seem excessive in the particular case may in fact be useful to screen types, e.g., norms of civility as remarked in Posner (2000).

There is a continuum of agents. They differ by their disposition to comply with the social norm. A proportion λ, referred to as type $t = 1$, 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 $t = 0$, 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.[6] 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

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

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, $\delta_0 = 0$ irrespective of circumstances; for the good citizen, δ_1 is positive and large if

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

$g < g_S$ 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, q_0 and q_1 are the marginal products of the bad and good types in future jobs, with $q_1 > q_0$. Given the belief μ, the wage paid is equal to the expected productivity

$$v(\mu) = (1-\mu)q_0 + \mu q_1 = q_0 + \mu(q_1 - q_0).$$

Alternatively, q_0 and q_1 are the purely hedonic utility of social esteem *per se*. Although the linear specification has been widely used[7], 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 set 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 circumstances faced by an agent nor his behavior. This assumption cuts off the direct social or market pressure on bad types to mimic the good types.[8] 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 G for "guilty", i.e., detected and convicted of an offence, or D for "detected and

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

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

dismissed", meaning that the agent was detected for engaging in the action but not found guilty of an offence. 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 G, D or N. The significance of these labels will depend on what they reveal about types at equilibrium, given the legal regime and the enforcement policy.

Continuous care levels. At equilibrium, agents will adopt threshold strategies. For some threshold g_t, a type-t agent causes harm only in circumstances $g \geq g_t$. Rather than considering actions conditional on realized circumstances, it is often more meaningful to view the agents as choosing ex ante a precaution level g_t. The larger the threshold, the wider the set of circumstances where potential harm is prevented. The cost of precaution level g_t is

$$K(g_t) = \int_0^{g_t} gz(g)\, dg.$$

The detection probability p is now interpreted as a probability of audit by the enforcement authorities. When an agent is audited and it is found that his conduct creates potential harm, a more thorough investigation takes place, which is public information. The investigation reveals the circumstance in which harm would occur. If $g_t < g_L$, where g_L is the legal or regulatory standard of precaution, the agent is "guilty" and the investigation yields the event G; otherwise, the investigation does not lead to prosecution, yielding the event D. The social norm g_S is a non legal standard of precaution which may differ from the legal requirement. For instance, prevailing legal requirements concerning privacy policies and safeguards for social media platforms may well fall short of users' expectation about the protection of personal data.[9]

[9] A striking example is the 2018 backlash suffered by Facebook because of the Cambridge Analytica data breach.

The reformulation in terms of care levels yields the same results as our benchmark model. We do not formally pursue this approach because it is slightly more cumbersome, but will refer to it in some interpretations.

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 beliefs conditional on the information G, D or N. We first determine 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_G, μ_D and μ_N. Abstracting from variables not affected by the agent's action, the expected utility of a type-0 agent (the bad citizen) who takes action e in the circumstance g is

$$\overline{u}_0 = \begin{cases} (g - pf)e + (1 - pe)v(\mu_N) + pev(\mu_G) \text{ if } g < g_L, \\ ge + (1 - pe)v(\mu_N) + pev(\mu_D) \text{ if } g \geq g_L. \end{cases} \qquad e = 0, 1. \qquad (4)$$

The expression includes the expected net material benefit from engaging in the action and the expected reputational utility. If the agent does not engage in the action or if he does but is not detected, the belief about his type will be μ_N, the posterior probability that he is a good citizen given "no news". If he engages in the action and is investigated, the belief is μ_G if the act is unlawful and μ_D if it is lawful.

Refraining from the action yields the utility $v(\mu_N)$. Engaging in the action when it is illegal (i.e., $g < g_L$) yields the expected utility

$$g - pf + (1 - p)v(\mu_N) + pv(\mu_G).$$

An offence is therefore committed if

$$g \geq p[f + v(\mu_N) - v(\mu_G)]. \qquad (5)$$

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 “guilty” labels.

If the action does not constitute an offence (i.e., $g \geq g_L$), expected utility is $g + (1-p)v(\mu_N) + pv(\mu_D)$. The agent therefore engages in the action if

$$g \geq p[v(\mu_N) - v(\mu_D)]. \tag{6}$$

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

The right-hand side of (5) and (6) defines possible thresholds for the bad citizen to engage in the action. Which one is effective depends on the legal standard. At equilibrium, $\mu_G \leq \mu_D \leq \mu_N$ as will be seen. The right-hand side of (5) is therefore larger than that of (6). The possibilities are illustrated in Figure 1.

Under the standards g_L^2 and g_L^3, condition (6) is always satisfied when the action is legal. With the standard g_L^3, the bad citizens' threshold is the right-hand side of condition (5). Bad citizens then sometimes do not comply with the law. With the standard g_L^2, condition (5) is never satisfied by illegal acts. The bad citizens' threshold then equals the legal standard g_L^2, i.e., bad citizens engage in the action only when it is legal. Finally, with the standard g_L^1, condition (6) is not satisfied in some circumstances where the action is legal. The bad citizens' threshold is then the right-hand side of condition (6). The threat of prosecution then provides deterrence, irrespective of conviction.

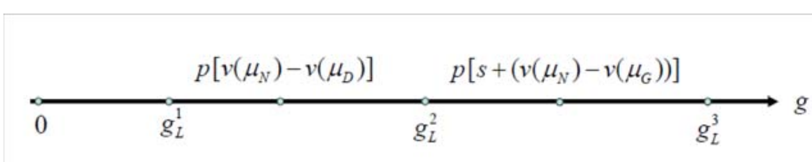


Figure 1. Thresholds and legal standards

Denote by g_0 and g_1 the bad and the good citizens' thresholds for engaging in the action. When $g < g_S$, a good citizen never engages in the action. When $g \geq g_S$, the action does not violate the social norm, so the good citizen behaves the same as the bad citizen. Therefore, $g_1 = \max(g_S, g_0)$.

Beliefs. How others behave affects the payoffs from one's actions through the effect on beliefs conditional on the events G, 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' threshold g_1 is a function of the bad citizens' threshold g_0, posterior beliefs can be written as a function of g_0. Recall that λ is the proportion of good citizens.

Lemma 1 *If $g_0 \geq g_L$, the event G never occurs and $\mu_D \leq \lambda \leq \mu_N$ where*

$$\mu_N = \frac{\lambda(1 - p + pZ(\max(g_S, g_0)))}{\lambda(1 - p + pZ(\max(g_S, g_0)) + (1 - \lambda)(1 - p + pZ(g_0))} \tag{7}$$

$$\mu_D = \frac{\lambda(1 - Z(\max(g_S, g_0))}{\lambda(1 - Z(\max(g_S, g_0)) + (1 - \lambda)(1 - Z(g_0))} \tag{8}$$

If $g_0 < g_L$, then $\mu_G \leq \mu_D \leq \lambda \leq \mu_N$ where μ_N is as defined in (7) and

$$\mu_G = \frac{\lambda \max\left[0, Z(g_L) - Z(\max(g_S, g_0))\right]}{\lambda \max\left[0, Z(g_L) - Z(\max(g_S, g_0))\right] + (1 - \lambda)\left[Z(g_L) - Z(g_0)\right]} \tag{9}$$

$$\mu_D = \frac{\lambda(1 - Z(\max(g_S, g_L))}{\lambda(1 - Z(\max(g_S, g_L)) + (1 - \lambda)(1 - Z(g_L))} \tag{10}$$

The numerator of (7) 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 (6) 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 (5) or it is equal to g_L. Thus,

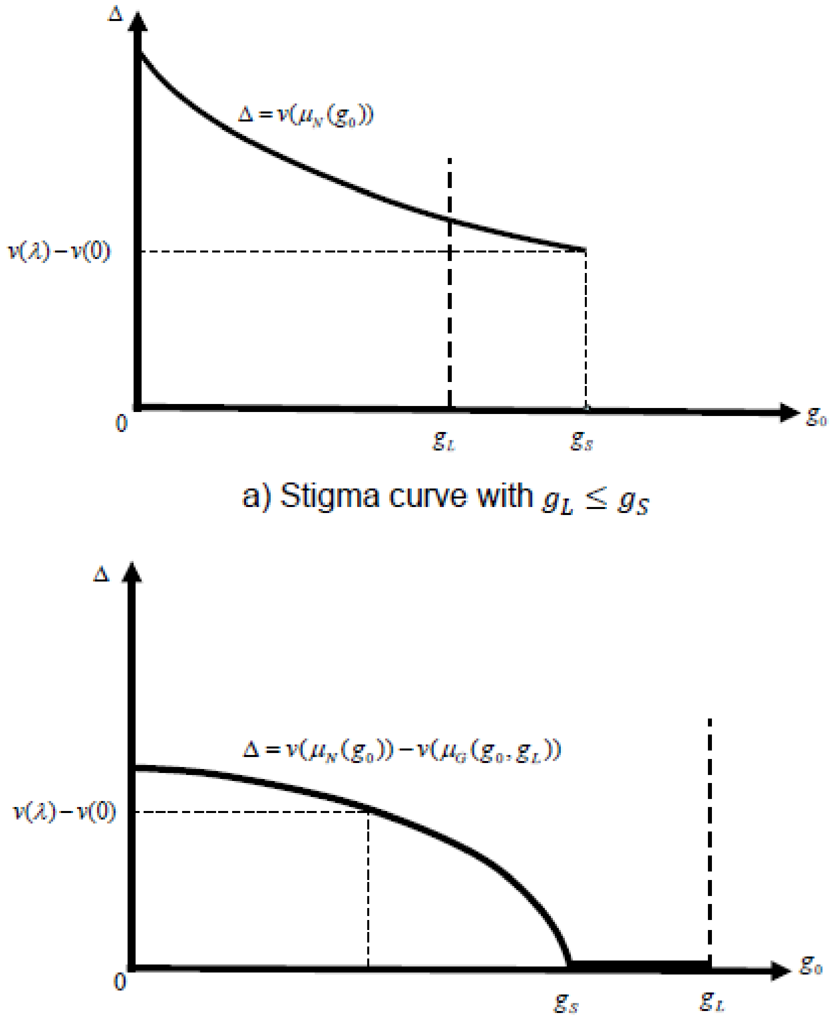
$$g_0 = \min[g_L, p(f + \Delta)]$$

where $\Delta \equiv v(\mu_N) - v(\mu_G)$ 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: $g_L \leq g_S$

Good citizens then always comply with the law. A conviction therefore reveals that one is for sure a bad citizen, hence $\mu_G = 0$. 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 $\mu_N(g_0)$. As more of the bad types abstain from the action, "no news" becomes less indicative that one is a good citizen, so $\mu_N(g_0)$ is decreasing. Figure 2a depicts the resulting stigma curve $\Delta = v(\mu_N(g_0)) - v(0)$. The relevant portion is for values of the bad citizens' threshold satisfying $g_0 \leq g_L$. If the legal standard is increased, a greater portion of the curve becomes relevant. When $g_L = g_S$, the legal standard is equal to the social norm. In the particular case where bad citizens always comply with the law, i.e., $g_0 = g_L$, 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]

[10] Because everyone complies with the law, the event "guilty" is off-equilibrium, so μ_G cannot be computed from Bayes' rule. We take the limit as g_0 approaches g_S from the left. This can also be rationalized in terms of Cho and Kreps' (1987) D1 criterion.



a) Stigma curve with $g_L \leq g_S$

b) Stigma curve with $g_L > g_S$

Figure 2. Reputational sanctions

Case 2: $g_L > g_S$

Both good and bad citizens are now at times convicted, therefore $\mu_G > 0$. From Lemma 1, μ_G depends on g_0 and g_L, which we write as $\mu_G(g_0, g_L)$. This is increasing in g_0 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_0. So long as violating the law is more likely for bad citizens, the stigma $\Delta = v(\mu_N(g_0)) - v(\mu_G(g_0, g_L))$ is positive and decreasing in g_0. When $g_0 \geq g_S$, everyone behaves the same. The events "guilty" 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_0 below g_S) when the legal standard is increased because $\mu_G(g_0, g_L)$ is increasing in 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_0, Δ) to the system of equations

$$g_0 = \min[g_L, p(f + \Delta)], \tag{11}$$

$$\Delta = v(\mu_N(g_0)) - v(\mu_G(g_0, g_L)), \tag{12}$$

provided the solution satisfies

$$g_L \geq p[v(\mu_N(g_0)) - v(\mu_D(g_L))]. \tag{13}$$

where $\mu_D(g_L)$ is as defined in (10). Condition (13) means that the mere threat of prosecution does not yield greater deterrence than the risk of conviction.[11]

Lemma 2 *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 (7) and (10) by setting $g_0 = g_L = g$. Then $g_c(p)$ is unique and satisfies $g_c(p) < g_S$. At equilibrium, $g_0 = g_c(p)$ if $g_L < g_c(p)$ while g_0 and Δ solve (11) and (12) if $g_L \geq g_c(p)$.*

The bad citizens' threshold is bounded below by some critical value, denoted $g_c(p)$. When $g_L < g_c(p)$, the legal standard and the fine play no

[11] This is the case with the standards g_L^2 and g_L^3 in Figure 1.

role. The enforcement regime then amounts to a mechanism for detecting and publicizing harmful acts.[12] Such a policy is equivalent to a strict liability regime with symbolic fines, e.g., public reprimands or shaming penalties. To see this, observe that, when $g_L < g_c(p)$, no one is ever found guilty. However, some agents (those for whom $g > g_c(p)$) will engage in the action and will be detected, in which case they are labeled as "detected and dismissed". This event yields the same posterior beliefs as the event "guilty" under a strict liability regime inducing the equilibrium threshold $g_0 = g_c(p)$, which occurs when strict liability is "punished" with a fine equal to zero.[13] Note that $g_c(p)$ need not be small. If reputation is very valuable and the probability of detection is large, substantial deterrence may be achieved even with a zero fine. We return to these points in Section 4.

Proposition 1 *Let $g_L \geq \min[pf, g_c(p)]$. There is a unique equilibrium. If $pf \geq g_S$, then $g_1 = g_0 = pf$. Otherwise, $g_1 = g_S$ and $pf < g_0 \leq g_S$:*
(i) When $g_L \leq g_S$, then g_0 is increasing in p and f and invariant in g_L so long as $g_0 < g_L$; otherwise, $g_0 = g_L$.
(ii) When $g_L > g_S$, then $g_0 < g_S$ and is increasing in p and f and decreasing in g_L.

The foregoing defines the function $g_0(p, f, g_L)$ characterizing the bad citizens' equilibrium threshold. Legal design matters for deterrence only when the expected fine is less than the social norm, i.e., $pf < g_S$. Convictions then provide information about one's type and reputational sanctions supplement formal sanctions. When $pf \geq g_S$, the outcome is a pooling equilibrium. It is the same as in the standard model without informal motivation and reputational concerns.

[12] It is easily verified that $g_c(p)$ is increasing in p.

[13] Write μ_G in (9) as $\mu_G(g_0, g_L)$. Under strict liability, $g_L = \infty$ so that $Z(g_L) = 1$. The equilibrium when the fine is nil is g_0 solving $g_0 = p[\mu_N(g_0) - \mu_G(g_0, \infty)]$. Because (9) and (8) yield the same expression when $g_L = \infty$, it follows that $g_0 = g_c(p)$.

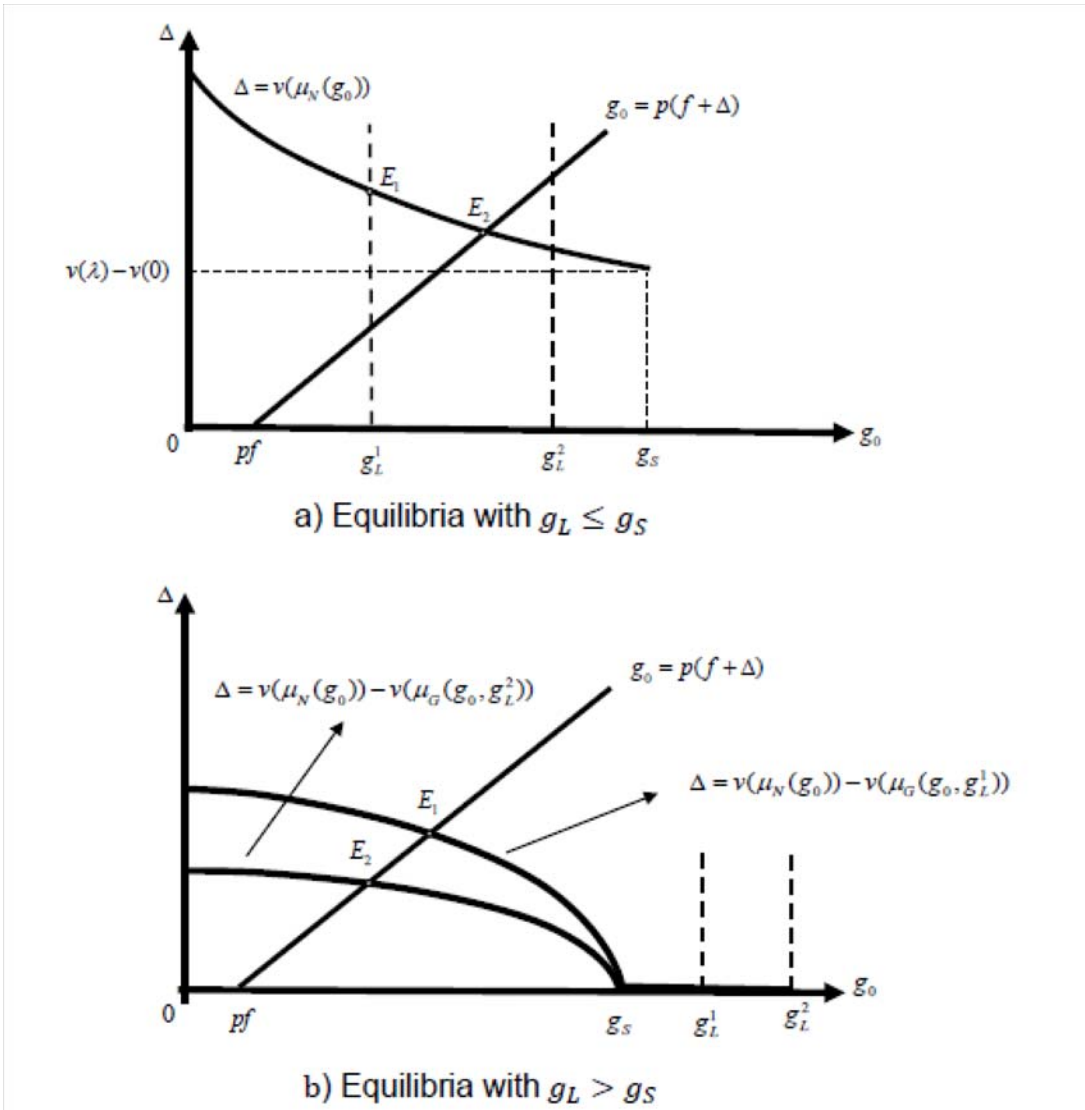


Figure 3. Equilibria

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' threshold and the stigma from a conviction (the relevant portion is for $g_0 \leq g_L$). In the figure there are two legal standards, g_L^1 and g_L^2, yielding the equilibria E_1 and E_2 respectively. Under the standard g_L^1, everyone complies with the law and the standard is binding. Strengthening the standard (i.e., increasing g_L) then increases deterrence. Under the more demanding standard 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 g_L^1 and g_L^2. A more demanding standard shifts the stigma curve downwards. More good citizens are convicted, so convictions impose a smaller stigma. The effect is to reduce deterrence.

Whether the legal standard is above or below the social norm, the stigma curves shift upwards when the probability of detection is increased. A higher probability of detection also shifts the threshold function to the right. Thus, more detection provides greater deterrence.[14]

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.

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

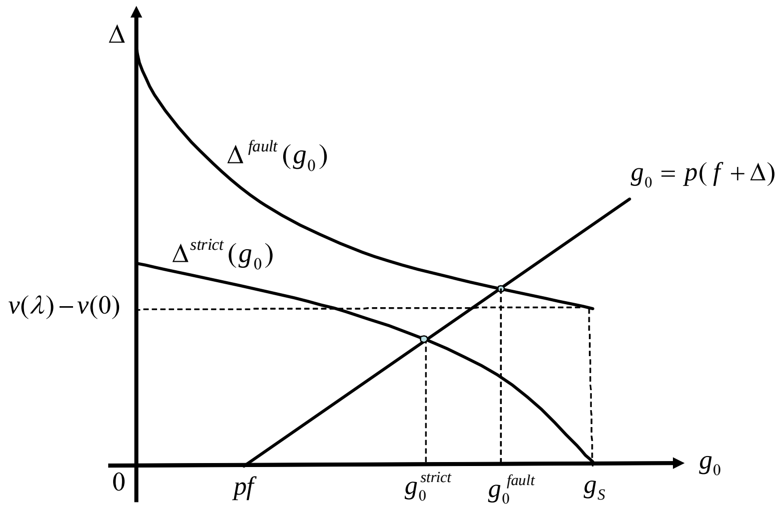


Figure 4. Strict versus fault-based offence with stanfard $g_L = g_S$

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 with $pf < g_S$, deterrence is maximized with the legal standard $g_L = g_S$. When $pf \geq g_S$, the legal standard is irrelevant for deterrence so long as $g_L \geq pf$.*

When $pf < g_S$, 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., $g_L = \infty$) and under a fault regime with legal standard equal to the social norm. When $pf \geq g_S$, 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

$$W = (1-\lambda)\int_{g_0}^{\infty}(g-h)\,z(g)\,dg + \lambda\int_{g_1}^{\infty}(g-h)\,z(g)\,dg - c(p) \qquad (14)$$

where g_0 and g_1 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 $g_L > g_S > pf$ cannot be optimal. Lowering the legal standard down to the social norm yields larger reputational sanctions. The same level of deterrence could therefore be reached with a smaller probability of detection, thereby reducing enforcement costs. Conversely, a policy with $g_L \geq pf > g_S$ can be optimal only because the social norm is too weak. Indeed, it is then irrelevant in the welfare maximizing policy. The following proposition combines these two observations. Recall that $g^* < h$ 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. There exists a threshold $\hat{g} < h$ such that:*
(i) If $g_S \geq \hat{g}$, then the standard $g_L = g_S$ is optimal and the probability of detection satisfies $p(f_m + v(\lambda) - v(0)) \leq g_S$.
(ii) If $g_S < \hat{g}$, then the probability of detection satisfies $pf_m = g^*$ and any standard $g_L \geq g^*$ is optimal.

To illustrate, consider the optimal enforcement policy under a strict liability regime, exogenously imposed. Suppose this yields $g_0 \leq g_S$. Because bad citizens are also motivated by reputational concerns, it must then be that $pf_m < g_0$. Owing to the greater reputational effects, the same level of deterrence can therefore be obtained at lower cost under the legal regime defined by $g_L = g_S$ (see Figure 4). Conversely, suppose the optimal enforcement policy yields $g_0 > g_S$. Then $g_0 = g_1 = g^*$ because the trade-off between deterrence and enforcement costs is now the same as in the standard model. This can arise only when $g_S < g^*$, and therefore $g_S < h$. 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 g_S for both types). This can be optimal only if g_S is not too small.

In part (i) of Proposition 3, the case where $p(f_m + v(\lambda) - v(0)) \leq g_S$ 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 $g_0 = g_S$. Everyone then complies with the law and, therefore, with the social norm.

The proposition states that $g_L = g_S$ 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 g_S because both standards yield the same equilibrium, given the enforcement policy. More generally, let g_0 be the bad citizens' equilibrium threshold in the optimal policy. Then any standard $g_L \in [g_0, g_S]$ yields the same equilibrium.

To summarize, aligning the legal standard on 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., $g_S > h$), 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 $q_1(a) > q_0(a)$ 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

$$\varphi(\mu) \equiv \max_a (1-\mu)q_0(a) + \mu q_1(a). \tag{15}$$

It follows trivially that $\varphi(\mu)$ is increasing in μ and, in particular, that it is a convex function.[15]

To illustrate, suppose different jobs are represented by the scalar $a \in [0,1]$. Suppose $q_t(a)$ is strictly concave with $q_0(a)$ maximized at $a = 0$ and $q_1(a)$ 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 (15), the optimal a will be increasing in μ. More trustworthy individuals have a comparative advantage in jobs higher

[15] This is well known from decision theory. See for instance Gollier (2001), chapter 24.

on this particular job scale. Letting $a(\mu)$ denote the solution,

$$\varphi''(\mu) = -\frac{[q_1'(a(\mu)) - q_0'(a(\mu))]^2}{(1-\mu)q_0''(a(\mu)) + \mu q_1''(a(\mu))} > 0.$$

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 "no news", the "guilty", and the "detected and dismissed" events, i.e., the signal is $X \in \{G, D, N\}$. The probabilities are[16]

$$P_G = \lambda p \max\left[0, Z(g_L) - Z(\max(g_S, g_0))\right] + (1-\lambda)p\left[Z(g_L) - Z(g_0)\right], \quad (16)$$

$$P_D = \lambda p(1 - Z(\max(g_S, g_L)) + (1-\lambda)p(1 - Z(g_L)). \quad (17)$$

$$P_N = \lambda(1 - p + pZ(\max(g_S, g_0)) + (1-\lambda)(1 - p + pZ(g_0)). \quad (18)$$

Averaged over all individuals, the value of future contractual interactions is

$$\overline{\varphi} \equiv P_G\varphi(\mu_G) + P_D\varphi(\mu_D) + P_N\varphi(\mu_N) \quad (19)$$

where μ_G, μ_D and μ_N are as defined in Lemma 1. Welfare is now redefined as

$$\overline{W} \equiv W + \overline{\varphi} \quad (20)$$

[16] See the denominators in the expressions of Lemma 1. We consider only the case where $g_0 \leq g_L$.

where W is as in equation (14).

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 (15) 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.[17] 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. Because maximizing $\overline{W}$ is then equivalent to maximizing W, the only issue from a utilitarian point of view is the trade-off between deterrence and enforcement costs. 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 (19) depend only on p, g_L and g_0. To compare different public signals, it is therefore sufficient to study the properties of the function $\overline{\varphi}(g_0, p, g_L)$. We keep g_0 constant when considering changes in p or g_L, i.e., we are considering the partial (or direct) effects of changes in these policy variables.

Lemma 3 *If $g_0 \geq g_S$, then $\overline{\varphi} = \varphi(\lambda)$. If $g_0 < g_S$, then $\overline{\varphi} > \varphi(\lambda)$ and is increasing in p, decreasing in g_0, increasing in g_L for $g_L < g_S$, and decreasing in g_L for $g_L > g_S$.*

Everything else equal, aligning the legal standard on the social norm improves the information provided to third parties. The lemma nevertheless points to possible conflicts between deterrence and the provision of information. Consider an increase in the fine, a costless measure. When $g_0 < h$, a larger fine increases W if it increases deterrence.[18] However, by itself, more

[17] Obviously, this is also the case if there is no action to take, i.e., $\varphi(\mu) \equiv (1-\mu)q_0 + \mu q_1$ where q_0 and q_1 are exogenously given.

[18] This is the case at an interior equilibrium such as E_2 of Figure 3a

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 $g_S \geq \hat{g}$, then in an optimal policy*
(i) either $g_L = g_S$ with a fine that need not be maximal;
(ii) or $g_L > g_S$ with the fine set equal to zero.

We consider only the situation where the social norm is not too deficient, i.e., $g_S \geq \hat{g}$ 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 compared to the "guilty" 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 verdicts. It is then optimal to reduce the fine in order to dampen deterrence. When information is very valuable and one's reputation is therefore also very valuable, the fine is zero. Dampening the deterrence effects of more detection is then obtained by increasing the legal standard above the social norm.[19]

To see this, write the function defined in (14) as $W(g_0, p)$, where we use the fact that $g_1 = \max(g_S, g_0)$. The optimal policy maximizes

$$\overline{W}(g_0, p, g_L) \equiv W(g_0, p) + \overline{\varphi}(g_0, p, g_L) \text{ where } g_0 = g_0(p, f, g_L).$$

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

$$\frac{\partial \overline{W}}{\partial p} = \left[W_{g_0} + \overline{\varphi}_{g_0}\right] \frac{\partial g_0}{\partial p} - c'(p) + \overline{\varphi}_p = 0, \tag{21}$$

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

$$\frac{\partial \overline{W}}{\partial g_L} = \left[W_{g_0} + \overline{\varphi}_{g_0}\right] \frac{\partial g_0}{\partial g_L} + \overline{\varphi}_{g_L} = 0. \tag{22}$$

The fine depends on the sign of

$$\frac{\partial \overline{W}}{\partial f} = \left[W_{g_0} + \overline{\varphi}_{g_0}\right] \frac{\partial g_0}{\partial f}. \tag{23}$$

In the proof, we show that g_L cannot be binding in the solution.[20] Applying Lemma 3, it follows that condition (22) can only be satisfied with $g_L \geq g_S$. When the solution is $g_L = g_S$, condition (22) holds with $\partial g_0/\partial g_L$ and $\overline{\varphi}_{g_L}$ both equal to zero. The sign of (23) may then be positive, negative or nil. Therefore, the fine may be maximal, zero or some value in between. When the solution is $g_L > g_S$, condition (22) holds with $\partial g_0/\partial g_L$ and $\overline{\varphi}_{g_L}$ both negative, implying that $W_{g_0} + \overline{\varphi}_{g_0}$ is negative. (23) 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_0 be the bad citizens' equilibrium threshold, any standard $g_L \in [g_0, g_S]$ would yield the same equilibrium. However, only $g_L = g_S$ is optimal because this provides more information to third parties. When the fine is positive but less than maximal, then $W_{g_0} = -\overline{\varphi}_{g_0}$ and $\overline{\varphi}_p = c'(p)$, where the latter follows from (21). The optimal fine trades off the marginal benefit from more deterrence against the marginal informational loss. The marginal informational benefit from greater detection equals the marginal detection cost.

An optimal policy sacrifices some deterrence so as to generate more informative signals. With the legal standard equal to the social norm, reducing the fine in order to dampen the deterrence effect of greater detection can only go so far as a fine equal to zero. When this constraint is binding, part (ii) of Proposition 4 applies. The marginal benefit from more deterrence is

[20] This excludes the possibility that $g_0(p, f, g_L) \equiv g_L$ for small changes in p or f. Therefore, excluding this case, Proposition 1 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 $g_L > g_S$.

now smaller than the associated informational loss. Dampening deterrence is achieved by a legal standard above the social norm, i.e., convictions are noisier signals. Such a policy should not be interpreted as necessarily implying low deterrence.[21] When information about individuals' type is very valuable, there are also strong reputational sanctions.

Optimal deterrence. When deterrence is the only concern, some underdeterrence (i.e., $g_0 < h$) 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 under-deterred. When $g_S > h$ 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_0} is negative, implying that (23) is negative, so that the fine is zero. With W_{g_0} negative, condition (21) implies

$$\overline{\varphi}_p + \overline{\varphi}_{g_0} \frac{\partial g_0}{\partial p} > 0. \tag{24}$$

The informational benefit of marginally increasing publicity is positive, even taking into account the countervailing effect due to greater deterrence. Condition (24) can hold only if $g_0 < g_S$. Thus, over-deterrence can only arise when the social norm is more demanding than the first-best level. When (24)

[21] Let g_L and p denote the optimal policy with a fine equal to zero when $g_L > g_S$. Consider a small change defined by an increase in p and a compensating increase in g_L so as to keep g_0 constant, i.e., dp and dg_L satisfy

$$dg_0 = \frac{\partial g_0}{\partial p} dp + \frac{\partial g_0}{\partial g_L} dg_L = 0.$$

Substituting from the first-order conditions (21) and (22), the change in the value of information to third parties is $d\overline{\varphi} = c'(p)dp > 0$. Loosely speaking, improving the information to third parties can be obtained without sacrificing too much deterrence.

holds, improving future allocative decisions, through the information conveyed to third parties, is 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.[22] 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.[23] 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 $t \in \{0, 1\}$, but our analysis can be recast in terms of a continuum of types. For example, types are $t \in [0, 1]$ with tg_S as the "willingness to pay" of a type-t agent in order to comply with the norm. The net benefit

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

[23] Exceptions are Deffains and Fluet (2013) and Fluet and Mungan (2018), but this distinction is irrelevant for the present argument.

from engaging in the action is then

$$b_t = \begin{cases} g - tg_S \text{ if } g < g_S \\ g \text{ if } g \geq g_S \end{cases}$$

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 $pf \geq g_S$, all types behave the same and no inferences can be drawn from violations of the law. When $pf < g_S$, 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 G as "engaged in the action in circumstances $g < g_S$"; the event D as "engaged in the action in circumstances $g \geq g_S$"; 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 $\mu_G = 0$, $\mu_D = \lambda$ and

$$\mu_N(g_0) = \frac{\lambda Z(\max(g_S, g_0)))}{\lambda Z(\max(g_S, g_0)) + (1-\lambda)Z(g_0)} \tag{25}$$

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 $pf \geq g_S$, because all agents then behave the same. When $pf < g_S$, the bad citizens' equilibrium threshold solves

$$g_0 = \min[g_S, pf + v(\mu_N(g_0)) - v(0)]. \tag{26}$$

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_0.

We discuss the optimal policy when it is characterized by $pf < g_S$ and the achieved level of deterrence solves (26). The informal enforcement of norms suggests the possibility of overdeterrence compared with the first-best utilitarian level. There are two possible cases.

Case 1: $g_S \leq h$ 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 $g_S \leq h$. When $g_S > h$, however, the solution to (26) may yield overdeterrence for some $pf > 0$. The possibility that reputational sanctions cause overdeterrence has been much discussed, in particular with respect to corporate liability.[24] It has been suggested that fines should be reduced to avoid this possibility. However, in the present context, 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.[25] 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 yield an even lower equilibrium g_0.

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

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

Case 2: $g_S > h$ 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 (26) 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.[26] This decision rule maximizes incentives to comply with the law, thereby allowing any given level of deterrence to be achieved at minimal audit costs.[27]

The foregoing policy is no longer optimal when agents have reputational concerns. In a related model, Fluet and Mungan (2018) show that the optimal rule trades-off the direct deterrence effects, taking the level of sanctions

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

[27] 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.

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.[28] We conjecture that a similar result will obtain in the present context, with a rule for imposing sanctions that depends on imperfect information about both conduct and circumstances. Taking into account the productive value of information conveyed to third parties will introduce additional considerations.

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 analyze 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.

[28] There is a similarity with policies that trade-off deterrence and judicial error, as in Demougin and Fluet (2005).

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. To maximize the reputational effects of offences, legal obligations should therefore align on prevailing norms whenever possible. Offences are then less noisy signals of norm violations. 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. As noted by Rasmusen (1996), providing information to third parties may yield social benefits that are distinct from 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. Productive actions bear on the allocative efficiency of future social or economic interactions, e.g., trading or matching decisions.

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. For instance, there is no screening of types in a pooling equilibrium where everyone complies with the law, and therefore with the social norm if the law aligns on the norm. 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 information conveyed to third parties. We show that optimal enforcement 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 shown in Karpoff et al. (2008) and Armour et al. (2017). As a practical matter, regulatory authorities may rely on mere public reprimands. To some, this suggests that reputational sanctions are sufficient for appropriate deterrence in the case at hand. However, this interpretation is unsatisfactory in a pure public enforcement context. We find that symbolic legal sanctions are optimal only if providing information to third parties is an essential concern. If the law only sought deterrence, fines should be positive (and maximal) irrespective of reputational sanctions, because this economizes on enforcement costs.

Our set-up delivers sharp results, perhaps too much so. We briefly discuss extensions that may qualify the results. First, 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. The argument extends to the future disincentive effects of reputational losses in general. Rasmusen (1996, p. 539) aptly summarizes the policy dilemma: "The trade-off is between the beneficial effect of secrecy on recidivism and the harmful effects on deterrence of first crimes and on allocative efficiency." Allocative efficiency refers to the value of information conveyed to third parties. 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. Let us replicate this set-up 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 together with the agent's reputational payoff. A cursory analysis suggests that muting reputational effects (i.e., keeping convictions secret) cannot be optimal. Suppose the legal standard aligns on the social norm. Then a conviction in the first period destroys one's reputation. Hence, a convicted agent has no incentive to comply with the law in the second period except for the threat of a legal sanction. An agent not convicted in the first period is in a different pool at the beginning of the second period. Because he is in a "good" pool, this agent still has reputational incentives to comply with the law. Greater deterrence in the first period increases the pool of such agents and therefore reduces enforcement costs (or increases deterrence) in the second period. However, this is at the cost of less information provided to third parties at the end of the first period. Thus, the dynamics introduces new trade-offs with respect to deterrence costs or the deterrent effect of law and 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. In the preceding section, we discuss the opposite assumption of freely available perfect information about behavior. The more realistic case is where society has some information and public enforcement can provide more reliable information. This corresponds to the informational role of law discussed in Shapira (2016). More reliable information improves deterrence and the allocative efficiency of reputational sanctions. However, public enforcement may itself depend on imperfect evidence. In the preceding section, we remarked that the optimal decision rules for convictions (i.e., the appropriate evidence thresholds) need to trade-off the direct deterrent effect of sanctions, the effect on reputational sanctions, and the value of information provided to third parties. These effects are intricate and it is not clear, for instance, that a rule which increases reputational sanctions also improves the allocative value of information. We leave this to future research.

Appendix

Proof of Lemma 1. Take g_0 as given and satisfying $g_0 < g_L$. The fraction of bad citizens violating the law is then $Z(g_L) - Z(g_0)$. If $g_1 = \max(g_S, g_0) \geq g_L$, the good citizens never violate the law; otherwise, a fraction $Z(g_L) - Z(\max(g_S, g_0))$ does. Both categories of individuals are detected with probability p. Applying Bayes' rule then yields (9). 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

$$Z(g_0) + (1-p)(1 - Z(g_0) = 1 - p + pZ(g_0).$$

Similarly, for the good citizen the fraction is

$$1 - p + pZ(\max(g_S, g_0)).$$

Applying Bayes' rule then yields (7). Finally, for bad citizens the probability of event D is $p[1 - Z(g_L)]$. For good citizens it is $p[1 - Z(\max(g_S, g_L)$. Applying Bayes' rule then yields (10). It is straightforward to verify that $\mu_G \leq \mu_D \leq \lambda \leq \mu_N$. A similar argument applies to the case $g_0 \geq g_L$. ■

Proof of Lemma 2. Let g_0 solve (11) and (12). Suppose first that $g_0 < g_L$.

Then

$$\begin{aligned} g_0 &= p[f + v(\mu_N(g_0)) - v(\mu_G(g_0, g_L))] \\ &\geq p[v(\mu_N(g_0)) - v(\mu_G(g_0, g_L))] \\ &\geq p[v(\mu_N(g_0)) - \mu_D(g_L)] \end{aligned}$$

where the last inequality follows from Lemma 1. It follows that (13) holds.

Next suppose that $g_0 = g_L$. If $g_L > g_S$, then $g_0 = g_1$ and $\mu_N = \mu_D = \lambda$ by Lemma 2 again. Hence (13) is trivially satisfied. So let $g_L \leq g_S$. 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 $g_0 = \hat{g}$ in (7) and $\mu_D(\hat{g})$ is obtained by setting $g_L = \hat{g}$ in (10). From the expressions in Lemma 1, $\varphi(\hat{g})$ is increasing. Moreover, $\mu_N(\hat{g}) > \lambda > \mu_D(\hat{g})$ if $g_L < g_S$ and $\mu_N(\hat{g}) = \mu_D(\hat{g}) = \lambda$ if $g_L = g_S$, hence $\varphi(0) < 0$ and $\varphi(g_S) > 0$. It follows that there exists $g_c < g_S$ as stated. We write $g_c(p)$ to emphasize that it is a function of p. ■

Proof of Proposition 1. Either $g_0 = g_1 > g_S$ or $g_0 \leq g_1 = g_S$. By Lemma 1, the first case implies $\Delta(g_0, g_L) \equiv v(\mu_N(g_0)) - v(\mu_G(g_0, g_L)) = 0$. Thus, it can arise only if $pf > g_S$ and the equilibrium is then simply $g_0 = g_1 = pf$. A policy with $pf \leq g_S$ therefore yields the second case. The relevant domain for g_0 is then the interval $[pf, \min(g_L, g_S)]$. If $pf = \min(g_L, g_S)$, the equilibrium is trivially $g_0 = pf$, so let $ps < \min(g_L, g_S)$. The equilibrium g_0 is then a solution to

$$g_0 = \min\left[g_L, p(f + \Delta(g_0, g_L))\right] \tag{27}$$

Equivalently, g_0 solves

$$\psi(g_0) \equiv \min\left[g_L, p(f + \Delta(g_0, g_L))\right] - g_0 = 0,\ g_0 \in [pf, \min(g_L, g_S)], \tag{28}$$

where $\psi(g_0)$ is a continuous function. By Lemma 1, $\Delta(pf, g_L) > 0$ and therefore $\psi(pf) > 0$. For the case $g_L \leq g_S$, obviously $\psi(g_L) \leq 0$. Because

$\Delta(g_0, g_L)$ is strictly decreasing in g_0 in the relevant domain, so is $\psi(g_0)$ and the equilibrium is therefore unique and satisfies $g_0 > pf$. For the case $g_L > g_S$, $\Delta(g_S, g_L) = 0$ so that $\psi(g_L) = pf - g_S < 0$. Again $\psi(g_0)$ is strictly decreasing, ensuring uniqueness with $g_0 > 0$.

(i) For $pf < g_L \leq g_S$, the above argument shows that $g_1 = g_S$ and $g_0 \in (pf, g_L]$. If $p(f + \Delta(g_L, g_L)) \geq g_L$, the equilibrium satisfies $g_0 = g_L$. Otherwise $g_0 < g_L$ and solves $g_0 = p(f + \Delta(g_0, g_L; p))$ where we now take into account that Δ depends on p. Differentiating totally with respect to p yields

$$\frac{dg_0}{dp} = \frac{s + \Delta + p\Delta_p}{1 - p\Delta_{g_0}} > 0. \tag{29}$$

From the expressions in Lemma 1, Δ_{g_0} is negative while Δ_p is positive because $\partial\mu_N/\partial p > 0$.

(ii) For $pf < g_S < g_L$, the argument is similar except that the solution now satisfies $g_0 \in (pf, g_S)$. We now have

$$\frac{\partial g_0}{\partial g_L} = \frac{p\beta\Delta_{g_L}}{1 - p\beta\Delta_{g_0}} < 0, \tag{30}$$

where Δ_{g_L} is negative because $\partial\mu_G/\partial g_L > 0$. ■

Proof of Proposition 2. From Proposition 1, $g_1 = g_S$ whenever $pf < g_S$. For any $g_L \leq g_S$, a sufficiently small value of pf yields an equilibrium $g_0 < g_L$ which is therefore constant in g_L; by contrast, a sufficiently large value yields the equilibrium $g_0 = g_L$, hence g_0 is then increasing in g_L; in either case, deterrence is maximized by $g_L = g_S$. For $g_L > g_S$, g_0 is monotonically decreasing in g_L. Under any enforcement policy, the deterrence maximizing standard is therefore $g_L = g_S$. ■

Proof of Proposition 3. Let

$$W(g_S) \equiv \max_{p,f,g_L} (1-\lambda) \int_{g_0}^{\infty} (g-h)\, z(g)\, dg + \lambda \int_{\max(g_0, g_S)}^{\infty} (g-h)\, z(g)\, dg - c(p) \tag{31}$$

where $g_0(p, f, g_L)$ 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 (31) satisfies $g_0 > g_S$, then $g_0 = g_1 = g^*$. Thus, $g_0 > g_S$ implies $g_S < g^*$. Equivalently, $g_S \geq g^*$ implies $g_0 \leq g_S$.

Fact 2: if $g_0 \leq g_S$, then it is easily seen from Proposition 2 that the optimal policy has $g_L = g_S$, so that $g_0 \leq g_1 = g_S$. A probability of detection satisfying $p[f_m + v(\lambda) - v(0)] > g_S$ would then induce $g_0 = g_S$ but could be reduced without affecting deterrence. Therefore $p[f_m + v(\lambda) - v(0)] \leq g_S$ as claimed.

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

$$V(g_S) \equiv \max_p (1 - \lambda) \int_{g_0(p, f_m, g_S)}^{\infty} (g - h)\, z(g)\, dg + \lambda \int_{g_S}^{\infty} (g - h)\, z(g)\, dg - c(p). \tag{32}$$

Then

$$W(g_S) = \begin{cases} V(g_S) \text{ if } V(g_S) \geq W^*, \\ W^* \text{ otherwise.} \end{cases}$$

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 $V(g_S)$ is strictly increasing. Using the envelope theorem,

$$\frac{dV(g_S)}{dg_S} = (1 - \lambda)(h - g_0) z(g_0) \frac{\partial g_0}{\partial g_S} + \lambda (h - g_S) z(g_S) > 0.$$

The sign follows from $g_0(p, f_m, g_S) \leq g_S \leq g^* < h$ and from

$$\frac{\partial g_0}{\partial g_S} = \left. \frac{\partial g_0(p, f_m, g_L)}{\partial g_L} \right|_{g_L = g_S} \geq 0,$$

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

Proof of Lemma 3. Rewrite the publicly observable signal as $X \in \{\mu_G, \mu_D, \mu_N\}$. By Lemma 1, when $g_0 \geq g_S$, then $\mu_G = \mu_D = \mu_N = \lambda$, therefore $\overline{\varphi} = \varphi(\lambda)$. Henceforth, let $g_0 < g_S$, in which case $\mu_G < \mu_D < \mu_N$. We compare X with the signal $X' \in \{\mu'_G, \mu'_D, \mu'_N\}$ resulting from a change in g_L, p or g_0. 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 g_L to $g'_L > g_L$. By Lemma 1 and using (16) to (18), when $g'_L \leq g_S$, then $P'_G > P_G$, $P'_D < P_D$, $P'_N = P_N$, and $\mu'_G = \mu_G$, $\mu'_D > \mu_D$, $\mu'_N = \mu_N$. Therefore, $P'_G = P_G + \alpha$ and $P'_D = P_D - \alpha$ for some positive $\alpha < P_D$. Then

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

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

When $g_L > g_S$, the change from g_L to $g'_L > g_L$ implies $P'_G > P_G$, $P'_D < P_D$, $P'_N = P_N$, and $\mu'_G > \mu_G$, $\mu'_D = \mu_D$, $\mu'_N = \mu_N$. Again, $P'_G = P_G + \alpha$ and $P'_D = P_D - \alpha$ for some positive $\alpha < P_D$. In this case,

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

Now X' has les risk than X and is therefore less informative. A similar argument applies to changes in p or g_0. ■

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 $g_L < g_S$, i.e., $g_0 = g_L < pv(\mu_N(g_0))]$. Such a possibility is compatible with condition (22). As in part (ii) of the proposition, $W_{g_0} + \overline{\varphi}_{g_0} < 0$. However, we now have $\partial g_0 / \partial g_L > 0$ and $\overline{\varphi}_{g_L} > 0$. We show that this cannot be optimal.

Suppose such a standard, denoted by $g_L^1 < g_S$, and let $g_0^{**} = g_L^1$ be the associated equilibrium threshold. We assume $g_L^1 > g_c(p)$ as defined in Lemma 2, otherwise the standard would have no effect. Note that $g_c(p)$ satisfies:

$$g_c(p) = p[v(\mu_N(g_c(p))) - v(\mu_G(g_c(p), \infty))]. \tag{33}$$

To see the equivalence with the definition of Lemma 2, set $g_L = \infty$ in (9) and $g_0 = g_c(p)$ in (9) and (8), so that $\mu_G(g_0, \infty) = \mu_D(g_0)$. We show: (i) that the equilibrium $g_0^{**} = g_L^1$ can also be implemented by a zero fine policy with the same p and with some standard $g_L^2 > g_S$; (ii) that the latter policy provides a more informative signal.

According to claim (i), there exists $g_L^2 > g_S$ solving

$$g_0^{**} = p[v(\mu_N(g_0^{**})) - v(\mu_G(g_0^{**}, g_L^2))]. \tag{34}$$

Define

$$\psi(g_0^{**}, g_L^2) \equiv g_0^{**} - p[v(\mu_N(g_0^{**})) - v(\mu_G(g_0^{**}, g_L^2))].$$

By assumption,

$$\psi(g_0^{**}, g_S) = g_0^{**} - pv(\mu_N(g_0^{**})) < 0,$$

Because ψ is increasing in g_0^{**} and given (33),

$$\psi(g_0^{**}, \infty) = g_0^{**} - p[v(\mu_N(g_0^{**})) - v(\mu_G(g_0^{**}, \infty))] > 0 \text{ for } g_0^{**} > g_c(p).$$

Hence, there exists $g_L^2 > g_S$ solving (34). Because ψ is also increasing in its second argument, g_L^2 is unique.

We now prove claim (ii). Denote the support of the policy with the binding standard 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 g_L^2, the support is

$\{\mu'_G, \mu'_D, \mu'_N\}$ where $\mu'_G < \mu'_D = \lambda$ and $\mu'_N = \mu_N$. The probabilities satisfy $P'_N = P_N$ and $P_G + P'_D = P_D$, i.e., the probability mass initially at μ_D has been redistributed over μ'_G and μ'_D. Because $\mu'_D > \mu_D$, this constitutes a mean preserving spread if $\mu'_G < \mu_D$. From (10) in Lemma 1, given $g^1_L = g^{**}_0$,

$$\mu_D = \frac{\lambda(1 - Z(g_S))}{\lambda(1 - Z(g_S)) + (1 - \lambda)(1 - Z(g^{**}_0))}$$

From (9),

$$\mu'_G = \frac{\lambda \left[Z(g^2_L) - Z(g_S)\right]}{\lambda \left[Z(g^2_L) - Z(g_S)\right] + (1 - \lambda) \left[Z(g^2_L) - Z(g^{**}_0)\right]}$$

Therefore, $\mu'_G < \mu_D$. ■

References

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

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

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

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

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

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

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

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

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

[10] Bénabou, R. and J. Tirole (2011), "Laws and Norms." NBER wp 17579.

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

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

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

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

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

[16] Cooter, R. and A. Porat (2001), "Should Courts Deduct Nonlegal Sanctions from Damages?" *Journal of Legal Studies* 30, 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." *Organizational Behavior and Human Decision Processes* 100, 193-201.

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

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

[20] Demougin, D. and C. Fluet (2005), "Deterrence versus Judicial Error: A Comparative View of Standards of Proof." *Journal of Institutional and Theoretical Economics* 16, 193-206.

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

[22] Ellickson, R. C. (1991), *Order Without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press.

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

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

[25] Fluet, C. and M. Mungan (2018), "The Signal-Tuning Function of Liability Regimes." Unpublished manuscript.

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

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

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

[29] Gollier, C. (2001), *The Economics of Risk and Time*, Cambridge: MIT Press.

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

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

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

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

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

[35] Karpoff, J. M. (2012), "Does Reputation Work to discipline Corporate Misconduct?" In: Barnett, M.L. and T.G. Pollock (eds.), *Oxford Handbook of Corporate Reputation*. Oxford, Oxford University Press, 361-383.

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

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

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

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

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

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

[42] McAdams, R.H. and E. B. Rasmusen (2007), "Norms in Law and Economics." In Polinsky, A. M. and S. Shavell (eds.), *Handbook of Law and Economics*, Vol. 1, North-Holland.

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

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

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

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

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

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

[49] Posner, E. A. (2000), *Law and Social Norms*. Cambridge: Harvard University Press.

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

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

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

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

[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." *Criminology* 52, 627–54.

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