C.I.R.P.É.E

Centre Interuniversitaire sur le Risque,
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **15-20**

Social Norms and Legal Design

Bruno Deffains
Claude Fluet

Septembre/September 2015

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

We thank the participants of the CESifo Law and Economics 2011 Workshop, the 4th Law and Economics Theory Conference at Berkeley, the PET and ALEA conferences in 2015, and of seminars at the universities of Aix-Marseille, Bonn, Haifa, Lausanne, Laval, Lorraine, Mannheim, Nanterre, Tel-Aviv, Toulouse, Vanderbilt and Yale. Financial support from SSHRC Canada is gratefully acknowledged (SSHRC 435-2023-1671).

Abstract:
We compare fault-based and strict liability offences in law enforcement when behavior is influenced by informal prosocial norms of conduct. Fault tends to be more effective than strict liability in harnessing social or self-image concerns. When enforcement relies on fines and assessing fault is not too costly, the optimal legal regime is fault-based with a standard consistent with the underlying social norm if convictions would seldom occur under optimal enforcement; otherwise liability should be strict. When sanctions are nonmonetary or when stigmatization imposes a deadweight loss, the legal standard may be harsher or more lenient than the social norm.

Keywords: Social preferences, regulatory offences, law enforcement, strict liability, fault, legal standard, compliance, deterrence

JEL Classification: D8, K4, Z13

1 Introduction

Illegal behavior ranges from crimes of great antiquity such as murder carrying strong moral opprobrium down to lesser 'quasi-crimes', e.g., misleading advertising, tax noncompliance or fishing out of season. An important issue in legal design is the categorization of offences. Should they be criminalized or qualified as mere violations? Legal systems have been dealing with this question since the mid 19th century owing to the multiplication of modern regulatory offences, e.g., in factory legislation or food and drug laws. There has been a resurgence of the issue in the wake of the recent criminal law reforms in many countries. It is also debated in new fields of law such as competition law, financial regulations and environmental protection legislation.

In their survey of the economic theory of public law enforcement, Polinsky and Shavell (2007) discuss the various policy choices facing the state, one of which concerns the sanctioning rule: "The rule could be *strict* in the sense that a party is sanctioned whenever he has been found to have caused harm (or expected harm). Alternatively, the rule could be *fault-based*, meaning that a party who has been found to have caused harm is sanctioned only if he failed to obey some standard of behavior or regulatory requirement." Whether there should be strict liability crimes is nevertheless contentious. To give but two examples, the Model Penal Code of the American Law Institute in the 1960s rejected the principle of strict liability in criminal law. By contrast, in the 1990s Australia reformed its criminal code squarely on the basis of the fault-based versus strict liability dichotomy.[1]

We analyze this legal design issue from the perspective of harnessing

[1] For a glimpse of the debates in other countries, see Law Reform Commission of Canada (1974), Faure and Heine (2005), Horder (2005), Simester (2005), Spencer and Pedain (2005), Wils (2007), and Law Commission (2010). For earlier influential discussions, see Kadish (1963) and Fitzgerald (1965). For recent assessments of the evolution in the US, see Singer (1989) and Brown (2012).

normative motivations. The standard model of legal enforcement is extended to incorporate social preferences and pre-existing socially efficient norms of conduct. At one extreme, the social norm has little salience (e.g., few people feel concerned) so that policy prescriptions are the same as in the standard model without social preferences. At the other extreme, the social norm has high salience. Individuals who are thought not to care meet strong disapproval, a source of disutility. We inquire how the salience of the social norm, together with social or self-image concerns with respect to deviations from the norm, affects the relative performance of fault-based versus strict liability offences from a deterrence and enforcement cost point of view.

As in the standard model, society faces a trade-off between enforcement costs and the deterrence of socially undesirable behavior. The law may be under-enforced because enforcement is costly. Nevertheless, some individuals behave efficiently from a social point of view. Some do so out of intrinsic moral or prosocial predispositions. Others have no such predispositions but would like people to believe that they do or perhaps would want to perceive themselves as having such concerns; that is, they care about social approval or self image. To the extent that informal motivations suffice, legal enforcement is of course superfluous. We consider situations where an individual's actions are only vaguely observable by one's reference group or would only be self-servingly recalled by the individual himself. However, convictions for offences provide hard information from which inferences can be drawn about the intrinsic predispositions of the individuals involved. Under either fault-based or strict liability offences, social and self image concerns therefore provide incentives to mimic the virtuous.

A basic result of our analysis is that fault-based offences tend to be more effective in harnessing image concerns. Legal sanctions are then more informative. A strict liability offence conveys that the offender committed a harmful action but says nothing about the circumstances in which the action was committed. A fault-based offence unambiguously reveals reprehensible

behavior, thereby providing more precise information about the individual's character. When the social norm has high salience with potentially strong stigmatization of violators, socially useful incentives are therefore provided by the signaling role of fault, allowing greater deterrence or lower enforcement costs. When the social norm has relatively low salience, however, it may be that strict liability does better in harnessing image concerns. The optimal legal regime and enforcement policy are interdependent and depend in a complex way on the underlying situation. When the norm has high salience and assessing fault is not too costly, the best regime is fault-based. If enforcement relies on fines, the optimal legal standard of fault then replicates the underlying social norm and convictions are rare events under the optimal enforcement policy. Otherwise, when the best regime is strict liability, convictions (and therefore offences) are frequent events. With nonmonetary sanctions such as imprisonment or when stigmatization effects entail a social deadweight loss, other considerations come into play. Under an optimal fault-based regime, the legal standard of fault may then be more lenient than the underlying social norm.

The dichotomy between fault-based and strict liability offences partly captures the distinction between "criminalized" offences and purely "regulatory" offences. In our analysis, the legal design problem is approached from a standard utilitarian perspective. The stigmatization effects of legal sanctions are considered for their incentive properties. Fault-based offences tend to do better for acts that are clearly bad from a moral or social point of view. When there is only a weak pre-existing norm, strict liability does as well and is less costly. Our analysis therefore provides an economic interpretation of the usefulness of the distinction between *malum in se* and *malum prohibitum*[2] for optimal legal design and enforcement (see the discussion in

[2] *Malum in se* means wrong or reprehensible in itself independently of regulations or laws. *Malum prohibitum* refers to conduct that is wrong only because it is prohibited by law. The difference is often decribed in terms of *iussum quia iustum* and *iustum quia iussum*, namely something that is commanded (*iussum*) because it is just (*iustum*) and

Dau-Schmidt, 1990).

Section 2 reviews some of the relevant literature. Section 3 presents the basic setup. Section 4 compares the incentives under different legal regimes and enforcement policies. Section 5 derives the implications for efficient legal design when enforcement relies on fines. Section 6 extends the analysis to nonmonetary sanctions and discusses the possibility that stigmatization entails a deadweight loss. Section 7 concludes. Proofs are in the Appendix.

2 Literature review

Our analysis belongs to a recent microeconomic literature on social preferences emphasizing that one's actions may reveal unobservable predispositions and that some predispositions are socially valued, hence social pressure may influence behavior through the individuals' image concerns (e.g., Bernheim 1994, Bénabou and Tirole 2006, 2011, Daughety and Reinganum, 2010). Numerous experimental or field studies show that image concerns are important motivators of prosocial behavior (Masclet *et al.* 2003, Dana *et al.* 2006, Ellingsen and Johannesson 2008, Andreoni and Bernheim 2008, Ariely *et al.* 2010, Funk 2010, Lacetera and Macis 2010, among others).

Another strand of literature deals with the interaction between formal legal sanctions and informal nonlegal sanctions. Part of this literature analyzes the substitutability between legal and nonlegal sanctions, stressing that stigma or loss of standing in a community may deter undesirable behavior just as or more effectively than formal legal sanctions (Macauley 1963; Ellickson 1991; Bernstein 1992). Another part discusses the potential complementarity between formal and informal sanctions, noting that legal penalties may influence the existence and impact of informal sanctions (Kahan, 1998, Posner 2000; Cooter 2000a, 2000b; Teichman 2005; Iacobucci

something that is just (*iustum*) because it is commanded (*iussum*).

2014). This also relates to the role of stigma and shaming penalties in relation to criminal activity; see Rasmusen (1996), Harel and Clement (2007), and Zasu (2007) among others.

The "norms and law" literature also discusses the relationship between morality and law. Posner (1997), Shavell (2002), and McAdams and Rasmusen (2007) provide a general discussion of legal sanctions versus informal motivations as regulators of conduct. Shavell compares the two in terms of the social costs of enforcement and the effectiveness in controlling behavior. He argues that, if the expected private gain from undesirable action and the expected harm due to the conduct are large, it is optimal to have law supplement morality and, if morality does not function well, law alone is optimal. Mialon (2014) analyzes the effectiveness of moral norms in an evolutionary context, showing that legal rules may be necessary when norms are easily swayed by social interaction in the long run. Legal design also bears a relation to the concept of "expressive law". According to this view even "mild law", i.e., law backed by small nondeterrent sanctions or weakly enforced, can have desirable effects on behavior; see Cooter (1998b), Tyran and Feld (2006), and Galbiati and Vertova (2008, 2014).

In a paper related to the present one, although in a civil litigation context, Deffains and Fluet (2013) model how tort rules and social pressure interact to provide incentives to take care. In their analysis, the extent to which liability rules are privately enforced and the characteristics of the tort rules themselves are taken as given, e.g., if found liable the injurer must pay compensatory damages to the victim. In the present paper we consider legal design together with optimal public law enforcement. The design problem is whether offences should be strict or fault-based and the determination of the legal standard of fault in the latter case. Enforcement concerns the detection of violations and the setting of sanctions, whether monetary or nonmonetary.

3 Set-Up

We start with a simple version of the economic model of public law enforcement.[3] The model analyzes the use of legal rules for preventing socially harmful behavior and of public agents to detect and punish offenders. In the standard model, individuals violate the law when their private net benefit from doing so is positive given the risk of legal sanctions. We use this framework to define strict liability and fault-based offences. Next we extend the framework to incorporate social preferences.

The standard model. Risk-neutral individuals obtain a private gain g from committing an act that causes an external harm of amount h. The private gain, equivalently the opportunity cost of not committing the act, varies between individuals and depends on the circumstances.[4] The probability distribution is $F(g)$ with density $f(g)$ on the support $[0, \overline{g}]$, where $\overline{g} > h$. Social welfare is the sum of the gains individuals obtain from committing the act less the harm they cause to others. Denoting behavior by $e \in \{0, 1\}$, where $e = 1$ means commission of the act, and denoting with $e(g)$ the behavior in the circumstance g, social welfare is

$$\int_0^{\overline{g}} e(g)\,(g-h)\,f(g)\,dg.$$

Socially optimal behavior is

$$e^*(g) = \begin{cases} 1 \text{ if } g \geq h, \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

The harmful act is assumed to be sometimes socially warranted, allowing a meaningful distinction between strict liability and fault-based legal regimes.

[3] Well known surveys are Polinsky and Shavell (2000, 2007). We differ by explicitly introducing a costs of ascertaining fault.

[4] Acts can be interpreted from different perspectives, namely acts of "omission" (not complying with some regulation, e.g., fire detectors) versus "positive" acts (driving through red lights).

The harmful act is qualified as a strict liability offence if it is illegal irrespective of circumstances. For the time being the sanction for violating the law is taken to be a fine s, a socially costless transfer of money. The probability of detecting harmful acts is p and the per capita enforcement cost is $c(p)$ with derivatives $c' > 0$, $c'' \geq 0$. An individual does not comply with the law if his private gain exceeds the expected fine, $g \geq ps$. For a given enforcement policy, welfare is therefore

$$\int_{ps}^{\overline{g}} (g - h)\, f(g)\, dg - c(p).$$

An optimal policy maximizes this expression with respect to the value of the fine and the probability of detection. Becker's (1968) maximum sanction principle applies: to economize on detection costs, the fine should be set at the highest feasible level, say the individuals' wealth or some given upper bound on allowable fines which we denote by s_M. Given the maximal fine, welfare is maximized with respect to the probability of detection. Assuming an interior solution, the first-order condition is

$$(h - ps_M)\frac{dF(ps_M)}{dp} = c'(p). \tag{2}$$

The left-hand side is the marginal social benefit from deterrence, the right-hand side is the marginal enforcement cost. The first-order condition requires $h > ps_M$, implying that optimal enforcement entails underdeterrence compared with first-best behavior. Some individuals, those for whom $ps_M \leq g < h$, will commit the harmful act even though it is not socially warranted. Optimal enforcement trades-off some inefficiency in behavior against savings in enforcement expenses.

With fault-based offences individuals who cause harm are sanctioned only if they failed to obey some standard of behavior. The legal standard is in terms of the circumstances under which the harmful act is committed. An individual's private benefit must be above some threshold $\widehat{g}$ in order for him to avoid liability; otherwise, he is considered to be at fault. Committing the

harmful act is illegal when the circumstances are $g < \widehat{g}$, in which case the individual is subject to a fine if he is detected; when the circumstances are $g \geq \widehat{g}$, the harmful act does not constitute an offence. Individuals therefore commit the act when $g \geq \min(ps, \widehat{g})$.

Enforcement costs include the cost of detecting harmful acts and the additional cost k of assessing circumstances. For a given probability of detection, the enforcement cost is now

$$c(p) + kp\left[1 - F(\min(ps, \widehat{g}))\right]$$

where the second term is the per capita cost of assessing the circumstances of the harmful acts committed by undeterred individuals. The optimal policy consists in choosing the fine, the probability of detection and the legal standard so as to maximize

$$\int_{\min(ps,\widehat{g})}^{\overline{g}} (g-h)\, f(g)\, dg - c(p) - kp\left[1 - F(\min(ps, \widehat{g}))\right].$$

The maximum sanction principle still applies and it is easily seen that an optimal policy requires $\widehat{g} \geq ps_M$, otherwise enforcement costs could be reduced with no detrimental effect on deterrence. An interior solution yields the first-order condition

$$(h + pk - ps_M)\frac{dF(ps_M)}{dp} = c'(p) + k\left[1 - F(ps_M)\right]. \tag{3}$$

Optimal enforcement may now entail either underdeterrence or overdeterrence compared with first-best behavior. Overdeterrence would reduce the frequency of harmful acts and therefore the cost of ascertaining circumstances. As with strict liability offences, there is a trade-off between enforcement costs and some distortion of behavior.

If assessing circumstances involves no additional cost (that is, $k = 0$), enforcement costs are the same under fault-based and strict liability offences. Condition (3) reduces to (2) and welfare is therefore the same under either

legal regime. Strict liability and fault-based offences are then equally efficient. When $k > 0$ the optimal legal regime is strict liability. Thus, when both the legal regime and the enforcement policy are optimally chosen, welfare is maximized by strict liability offences unless assessing fault is costless, in which case the legal regime does not matter. Moreover, in the optimal policy individuals are underdeterred to some extent.

Another observation is that, under a fault regime, any standard $\widehat{g} \geq ps_M$ yields the level of deterrence ps_M. In other words, the legal standard is irrelevant. A final observation is that a standard above the upper bound of possible gains (i.e., $\widehat{g} \geq \overline{g}$) is equivalent to strict liability because committing the act is then illegal irrespective of possible circumstances.

Social preferences. In the standard model behavior depends on private costs and benefits as conventionally defined. We now consider normative motivations. There are two types of individuals. A proportion λ, referred to as type $t = 1$, is intrinsically motivated to behave in a socially responsible manner. Such individuals are "good citizens" with moral predispositions. The other group, referred to as type $t = 0$, has no such predispositions. However, prosocial predispositions are socially valued and those who are thought to be good citizens earn social esteem or status.

The utility of a type-t individual is

$$u_t = w - \gamma_t \max(e - e^*, 0) + \beta\mu, \quad t = 0, 1. \tag{4}$$

The first term, w, is net "material" payoff as in the standard model. In the middle term, the parameter γ_t is the disutility ("guilt") suffered when one causes external harm while deviating from the socially responsible behavior e^*. Misbehavior occurs when $e = 1$ and $e^* = 0$. For the good citizens, $\gamma_1 > 0$ and is sufficiently large to intrinsically motivate the individual[5]; for the bad citizen, $\gamma_0 = 0$ and the middle term vanishes. As defined here, the

[5] It suffices that $\gamma_1 \geq h$, i.e., the good citizen "internalizes" the harm he causes.

social (or moral) norm of conduct is what everyone should be doing given the circumstances, which amounts to a simple version of Kant's categorical imperative (see Brekke et al. 2003).

The third term in (4) is the utility from one's social image. β is a positive parameter and μ is society's belief about the individual's type. The belief will depend on information concerning the individual, i.e., μ equals the conditional expectation $E(t \mid I)$ where I denotes publicly available information. Given our definition of types, the conditional expectation is simply the posterior probability that the individual is a good citizen. All individuals care equally about social approval; β may be interpreted as the utility of being perceived as a good citizen, given that the utility of being perceived as a bad citizen is normalized to zero. The parameter captures both the importance individuals attach to social approval and the importance ("social pressure") society ascribes to being a good citizen in the case at hand. Both β and the proportion λ of good citizens reflect the salience of the social norm with respect to the situation (and therefore possible acts) considered. For instance, when the harmful act is widely viewed as particularly reprehensible, β will be large and presumably so will be λ.

It is useful to reflect on the assumptions made so far. Consider the possibility of a multiplicity of moral types $\gamma_t \geq 0$, as in Deffains and Fluet (2013). Different individuals then trade-off differently the moral disutility of acting bad against material payoffs. The basic logic of our analysis would nevertheless remain the same. Similarly the importance of image concerns could differ between individuals, i.e., they have different β's. As long as the β's are positive, the basic logic would still be unaffected. Obviously, if the bad citizens in our two-type set-up were characterized by $\beta = 0$, we would be back to the standard enforcement model with respect to regulating their behavior (the law being superfluous for the good citizen). To push things further, it could be that the bad citizen has $\beta < 0$, i.e., he enjoys being seen as bad; for instance, his reference group consists only of bad citizens.

Image concerns would then have antisocial effects and much of the results of our analysis would need to be reversed. Rather than seeking to harness image concerns, optimal legal design should seek to mute signalling effects; strict liability offences would then tend to perform better. Our assumptions disregard such possibilities. In effect, they describe a cohesive society with commonly shared notions of what is good, although some individuals lack inner moral strength.[6]

Welfare is defined in the usual way as the sum of utility over all individuals,

$$W = \int_0^{\overline{g}} \left[(1-\lambda)u_0(g) + \lambda u_1(g)\right] f(g)\, dg \tag{5}$$

where $u_t(g)$ is the utility (or expected utility) of a type-t individual in the circumstance g.

Before proceeding, we show that first-best behavior in the present set-up is the same as in the standard model without social preferences. Assume that individuals can both cause harm or suffer harm caused by others. Consider an omniscient regulator who can directly impose the action profile $e(g)$, $g \in [0, \overline{g}]$. The average net material payoff is then

$$w = w_0 + \int_0^{\overline{g}} e(g)\,(g-h)\, f(g)\, dg. \tag{6}$$

where w_0 is initial per capita wealth. Let the action profile $\widehat{e}(g)$ be welfare maximizing and suppose that the regulator has the option of either publicizing or preventing any information about the individuals' types. If an optimum entails that no information is disclosed, then $\widehat{e}(g)$ maximizes W subject to the resource constraint (6), given that beliefs satisfy $\mu = \lambda$ where λ is the prior belief about types. This implies $\widehat{e}(g) = e^*(g)$ as defined in (1).

[6] In the terminology of social network analysis with one's reference group consisting of one's "neighbors", we are assuming an integrated social network in the sense that the distribution of types among neighbors reflects the population distribution (see Bramoullé et al. 2012).

Combining (4) and (5), the first-best welfare then equals

$$W^* = w_0 + \int_h^{\overline{g}} (g - h)\, f(g)\, dg + \beta\lambda. \tag{7}$$

The action profile $e^*(g)$ would also be optimal when full or imperfect information about types is disclosed. First, because the omniscient regulator is able to independently control the flow of information, there would be no reason for him to distort behavior from the wealth maximizing action profile. Secondly, welfare would also be as in (7) because reputational benefits and losses simply cancel out.[7]

Offences and labeling. Society at large is assumed not to be able to directly observe the circumstances faced by an individual nor his behavior. The assumption prevents social pressure from bearing directly on individuals independently of the legal system. Otherwise one's type could be inferred directly from one's behavior. When β is large enough there would then be situations where the legal system plays no useful role. Direct informal reputational sanctions would suffice to induce socially appropriate behavior.

Public enforcers can detect harmful acts and can ascertain the circumstances; that is, they are able to enforce the law when offences are fault-based. Legal proceedings against offenders constitute public information from which society at large draws inferences about the individuals' type. For simplicity, we assume that the only information publicly available about an individual is either G for "guilty", in which case the individual is known to have been found guilty of an offence, or N for "no news". The latter means that either the individual did not commit an offence or that he did but was not convicted. The publicly available information affecting one's social image is therefore the binary signal $I \in \{N, G\}$. Society's belief about an individual will then be either $\overline{t}_N = E(t \mid N)$ or $\overline{t}_G = E(t \mid G)$.

[7] The result follows from the law of iterated expectations and the linearity of reputational utility in the beliefs about one's type, that is, $E(E(t \mid I)) = E(t) = \lambda$.

The significance of the signal depends on what the events "no news" and "guilty" reveal about one's type in the social equilibrium. This will depend on the legal regime, namely whether offences are strict or fault-based, the legal standard in the latter case, and on the enforcement policy. Generally speaking, the event "guilty" will be detrimental to one's reputation. Other things equal, individuals wish to avoid being labeled as offenders.

We have stressed social signaling which requires that convictions constitute public information. In practice convictions often receive little publicity. However, as in Bénabou and Tirole (2006), our framework can also be reinterpreted in terms of self-signaling. A simple formulation is Bodner and Prelec's (2003) dual-self model. In the latter, an individual's total utility is the sum of the "outcome utility" from choosing a particular course of action, which depends on one's fuzzily known true inner predispositions, and of a "diagnostic utility" whereby an individual draws inferences about his true self from information about his past behavior. For example, one may occasionally exceed the speed limit in a school zone or evade tax, but would feel shame from being labeled an offender even if convictions are not publicized.[8]

4 Equilibrium under a Given Regime

This section describes the equilibria under given legal regimes and enforcement policies. A perfect Bayesian equilibrium is characterized by the individuals' action profiles and the beliefs about individuals' type conditional on the "guilty" and "no news" events. The legal regime is defined by the standard of fault when committing the harmful act. The regime is fault-based if the standard is less than the upper bound of possible gains, otherwise liability is strict. The enforcement policy is defined by the sanction for unlawful conduct and the probability of detecting violations.

[8] See McAdams and Rasmusen (2007) on the distinction between guilt, social disesteem, and shame.

We proceed in three steps. First we derive the action profiles taking as given the posterior beliefs conditional on the "guilty" and "no news" events. Next we derive these beliefs as a function of action profiles. Finally we solve for the equilibrium wherein action profiles and beliefs are consistent with one another.

Incentives. Denote the sanctioning rule by $\delta(g, \widehat{g})$ where $\delta(g, \widehat{g}) = 1$ if $g < \widehat{g}$ and is otherwise zero. The expected utility of a type-t individual in the circumstance g is

$$u_t = w + e\left[g - p\delta(g, \widehat{g})s\right] - \gamma_t \max(e - e^*(g), 0) \\ + \beta\left[pe\delta(g, \widehat{g})\bar{t}_G + (1 - pe\delta(g, \widehat{g}))\,\bar{t}_N\right], \quad e \in \{0, 1\},\ t \in \{0, 1\}.$$

The first two terms comprise material payoff as conventionally defined. The first term, w, is the part of the individual's wealth that he takes as given. This consist of initial wealth minus the average harm caused by others plus the per capita tax to finance the enforcement policy (expenditures minus fines collected). The second term is the expected net material payoff from committing or not committing the harmful act. The third term is the moral disutility from committing the harmful act when it is socially unwarranted. The fourth term is the expected reputational utility. If the individual does not commit the act (i.e., $e = 0$) or if he would not be legally at fault when he does (i.e., $\delta(g, \widehat{g}) = 0$), the belief about his type will be $\bar{t}_N$ for sure, the posterior probability that he is a good citizen given "no news". If he unlawfully commits the act, he is detected with probability p and the belief about his type is then $\bar{t}_G$, the posterior probability conditional on "guilty". If he is not detected, the belief is again $\bar{t}_N$. These beliefs are determined at equilibrium but are taken as given by the individual.

If the harmful act is not committed, utility is $w + \beta\bar{t}_N$ for either type. If it is committed and is lawful, that is $g \geq \widehat{g}$, the utility of the nonprosocial is $w + g + \beta\bar{t}_N$. Hence it will then be committed. In circumstances where

the act is unlawful, expected utility is

$$w + (g - ps) + \beta(p\bar{t}_G + (1-p)\bar{t}_N)$$

and the act is then committed if $g \geq p(s + \beta\Delta)$, where $\Delta \equiv \bar{t}_N - \bar{t}_G$ will be referred to as the legal stigma, i.e., the reputational loss from being convicted. The term ps is the standard material incentive to comply with the law, the term $p\beta\Delta$ is the reputational motive. Altogether a nonprosocial commits the harmful act when

$$g \geq \min[\widehat{g}, p(s + \beta\Delta)] \equiv g_0, \tag{8}$$

where g_0 is short-hand for the private gain threshold of nonprosocial individuals. In turn the threshold determines the proportion $F(g_0)$ of nonprosocial who do not commit the harmful act.

Good citizens are also motivated by legal sanctions and reputational concerns. In addition, their behavior reflects an intrinsic concern for complying with the *social* (as opposed to the *legal*) norm conduct. Given γ_1 sufficiently large, a good citizen never commits the harmful act in circumstances $g < h$. When $g \geq h$, the harmful act entails no guilt and the good citizen then behaves the same as the nonprosocial. The harmful act is therefore committed if

$$g \geq \max(h, g_0) \equiv g_1 \tag{9}$$

where g_1 is the gain threshold for good citizens. The proportion of good citizens who do not commit the harmful act is $F(g_1)$. The following summarizes the preceding discussion.

Lemma 1 *$g_0 \leq \widehat{g}$ and $g_1 = \max(h, g_0)$.*

We shall say that type-t individuals are underdeterred (resp. overdeterred) when the equilibrium threshold g_t is less (resp. greater) than the first best h. Whether individuals comply with the law is different. As

noted, the legal standard may differ from the first best (and social norm). A consequence of Lemma 1 is therefore that, if there is some overdeterrence (which requires $\widehat{g} > h$), then all individuals are equally overdeterred. Otherwise they either all efficiently behave or the good citizens do while bad citizens are underdeterred. Moreover, bad citizens never overcomply with the law but good citizens might (when $\widehat{g} < h$).

Beliefs. The conditions (8) and (9) define the best response functions of individuals of either type given the behavior of others. How others behave affects the payoffs from one's actions through its effect on the social significance of the "guilty-no news" events, as captured by the beliefs $\bar{t}_G$ and $\bar{t}_N$. The posterior beliefs are obtained from Bayes' rule given the frequency of convictions among good and bad citizens. From Lemma 1, g_1 is a function of g_0. Hence posterior beliefs and therefore the legal stigma can be written as a function of g_0.

Lemma 2 *If $\widehat{g} \leq h$, $\Delta \geq \lambda$ and is decreasing in g_0 down to $\Delta = \lambda$ when $g_0 = \widehat{g}$. If $\widehat{g} > h$ and $g_0 < g_1 = h$, $\Delta > 0$ and is decreasing in g_0. If $\widehat{g} > h$ and $g_0 = g_1 \geq h$, $\Delta = 0$.*

Unless both types behave the same, bad citizens are more likely to commit the harmful act. Therefore they are more likely to be convicted, implying that the event "guilty" is bad news concerning the individual's type compared to "no news". When the legal standard satisfies $\widehat{g} \leq h$, good citizens are never found guilty. A conviction then reveals perfectly that the individual is nonprosocial, so that $\bar{t}_G = 0$ and $\Delta = \bar{t}_N$. The more the nonprosocial behave like good citizens, the smaller $\bar{t}_N$. When everyone behaves the same, the event "no news" is uninformative because it occurs with certainty, so the posterior probability then equals the prior λ that an individual is a good citizen.[9] When $\widehat{g} > h$, as would be the case with a strict liability offence, both good and bad citizens will at times be convicted, hence $\bar{t}_G > 0$. As

[9] $\beta\Delta = \beta\lambda$ is the disutility from being perceived as a bad rather than an average citizen.

long as violating the law is more likely for bad citizens, $\bar{t}_N > \bar{t}_G$ and the legal stigma is positive.[10] When both types behave the same, the events "guilty" and "no news" are uninformative and posterior beliefs equal the prior in either case. The legal stigma then vanishes.

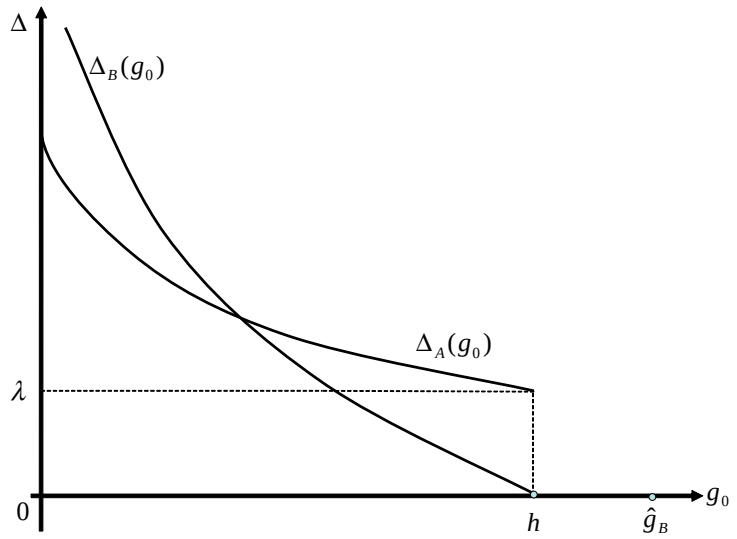

Fig. 1: Legal stigma curves

Figure 1 provides examples of the legal stigma as a function of the bad citizens' threshold under two different legal regimes. The probability of detection is the same in both regimes. In case A, the legal standard corresponds to the social norm, $\widehat{g}_A = h$, so that $g_0 \leq h$. The legal stigma is then bounded below by λ. In case B, the legal standard is above the social norm, $\widehat{g}_B > h$, hence $g_0 \leq \widehat{g}_B$. The legal stigma then vanishes when all the non-prosocial conform to the social norm (i.e., $g_0 \geq h$). When the non prosocial

The event "guilty" is then an out-of-equilibrium event with zero probablity, implying that $\bar{t}_G$ cannot be computed using Bayes' rule. The legal stigma at equilibrium is obtained from $\lim_{g_0 \uparrow \widehat{g}} \Delta = \lim_{g_0 \uparrow \widehat{g}} \bar{t}_N = \lambda$. This can also be rationalized in terms of Cho and Kreps' (1987) D1 criterion.

[10] Strict liability disregards circumstances. Good citizens will then sometimes efficiently choose not to comply with the law given their knowledge of circumstances. See Shavell (2012).

are only slightly underdeterred, the legal stigma under regime A is therefore larger than under regime B. As depicted, the curves intersect. This need not occur but it is a possibility when the nonprosocial are sufficiently underdeterred. We discuss this further in Section 6.

Equilibrium. An equilibrium consists of private gain thresholds and of a legal stigma that are mutually consistent.

Proposition 1 *Let the legal regime and enforcement policy satisfy* $\widehat{g} \geq ps$. *There is a unique equilibrium with* $g_0 \leq g_1$.
(i) If $ps \geq h$, *then* $g_0 = g_1 = ps$.
(ii) If $ps < \widehat{g} \leq h$, *then* $g_1 = h$, $g_0 \in (ps, \widehat{g}]$ *and is increasing in* p *and* s *so long as* $p(s + \beta\lambda) < \widehat{g}$, *otherwise* $g_0 = \widehat{g}$; *in either case,* g_0 *is increasing in* $\widehat{g}$.
(iii) If $ps < h < \widehat{g}$, *then* $g_1 = h$, $g_0 \in (ps, h)$ *and is increasing in* p *and* s, *it may be increasing or decreasing in* $\widehat{g}$.

Legal design matters for incentives only when $ps < h$. The Figures 2 to 4 illustrate different equilibria for this case. Good citizens then conform to the social norm. In the figures, $\Delta(g_0)$ is the legal stigma as a function of the bad citizens' threshold under a given legal regime and enforcement policy; $g_0(\Delta)$ is their threshold as a function of the legal stigma under the same regime and enforcement policy. The perfect Bayesian equilibrium is the intersection of the two curves (at point E).

Figure 2 compares the equilibria under two different legal standards satisfying $\widehat{g} < h$. With the standard $\widehat{g}_A$ the equilibrium is at E_A, a corner equilibrium where everyone complies with the law. With the standard $\widehat{g}_B > \widehat{g}_A$, we have an interior equilibrium at E_B. In either case, deterrence increases with a strengthening of the legal standard because this shifts the stigma curve to the right (so long as $\widehat{g} < h$). The intuition is that strengthening the standard increases the significance of the "no news" event.

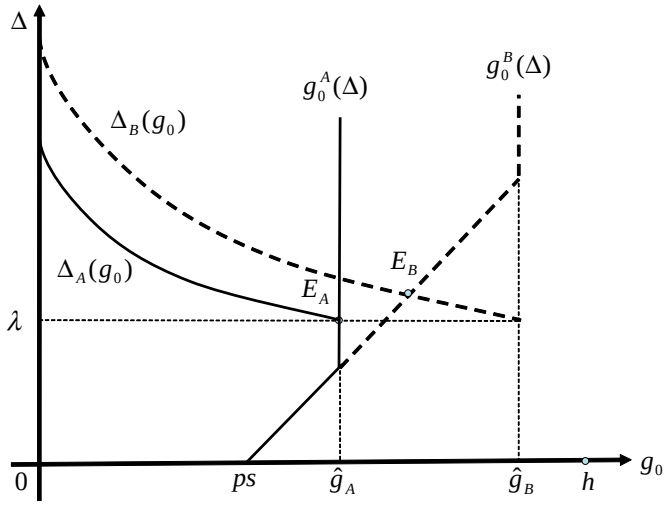

Fig. 2: Equilibria with $\widehat{g} < h$

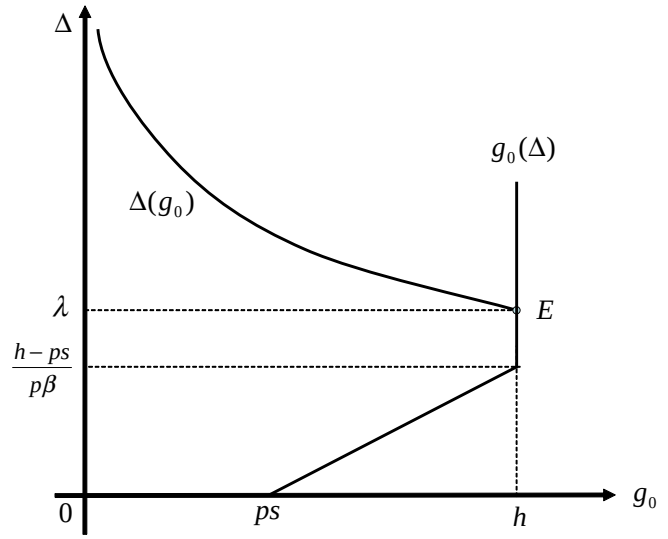

Fig. 3: A corner equilibrium with $\widehat{g} = h$

In Figure 3, the standard is the first-best $\widehat{g} = h$. In the case represented all individuals comply with the law and therefore are optimally deterred. This arises when $p(s+\beta\lambda) \geq h$. In the figure, the latter condition holds as a strict inequality, hence the detection probability could be reduced while still preserving first-best deterrence. Thus, under a fault-based regime, first-best

deterrence is feasible even though $ps < h$. Indeed, when $p\beta\lambda \geq h$, first-best deterrence obtains with purely symbolic convictions with no material legal sanctions. The figure illustrates the role of "mild law" as defined in Tyran and Feld (2006), i.e., law backed by nondeterrent sanctions.

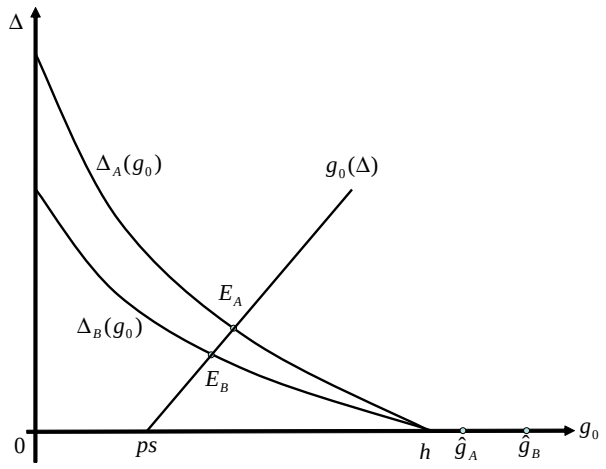

Fig. 4: Equilibria with $\widehat{g} > h$

In Figure 4, $\widehat{g} > h$ and both types will then sometimes not comply with the law. First-best deterrence then cannot be achieved with $ps < h$. Strengthening the legal standard further may shift the legal stigma curve upwards or downwards. As before, a stronger standard increases the significance of "no news". However, it also reduces the significance of the "guilty" outcome because more good citizens are convicted, so the net effect on deterrence is ambiguous. In the situation represented in Figure 4, weakening the standard from $\widehat{g}_B$ to $\widehat{g}_A$ shifts the stigma curve upwards, so deterrence increases.

Increasing the probability of detecting illegal acts increases the significance of "no news", with no effect on the significance of the "guilty" event.[11] In the figures, the legal stigma curves shifts (or rather rotate) upwards. Because offenders are now more likely to be convicted, a larger probability of

[11] The conditional expectation $\bar{t}_G$ does not depend on p while $\bar{t}_N$ is increasing in p. See the proof of Lemma 2.

detection also shifts the deterrence curve to the right. Thus, greater detection unambiguously increases deterrence, except at corner solutions where all bad citizens are efficiently deterred as in Figure 2. A larger fine increases deterrence because it shifts the deterrence curves to the right.[12]

5 Optimal Legal Regime and Enforcement

Welfare equals the first best W^* as defined in (7) minus the loss from socially inefficient behavior and minus the per capita enforcement expenditure:

$$\begin{aligned} W &= W^* - \{(1-\lambda)\int_{g_0}^{h}(h-g)\,f(g)\,dg + \lambda\int_{g_1}^{h}(h-g)\,f(g)\,dg\} \\ &\quad -C(p,g_0,g_1,\widehat{g}) \end{aligned} \tag{10}$$

where g_0 and g_1 are equilibrium thresholds as derived in Section 4. The expression inside the brackets is the loss from inefficient behavior; the formulation allows for the possibility of overdeterrence[13]. The third term is the enforcement cost function. When the legal standard $\widehat{g} = \overline{g}$, the regime is strict liability. Enforcement expenditures then reduce to the cost of detecting harmful acts:

$$C(p,g_0,g_1,\overline{g}) = c(p). \tag{11}$$

When $\widehat{g} < \overline{g}$, offences are fault-based. Enforcement expenditures then include the cost of assessing the circumstances of the harmful acts that are detected:

$$C(p,g_0,g_1,\widehat{g}) = c(p) + pk\left[1-(1-\lambda)F(g_0)-\lambda F(g_1)\right], \quad \widehat{g} < \overline{g}. \tag{12}$$

Deterrence maximizing legal design. We first take the enforcement policy as given and compare different legal designs in terms of deterrence.

[12] Greater detection has an ambiguous effect on the equilibrium legal stigma. A negative effect may be interpreted as greater legal enforcement partially crowding out informal motivations; a positive effect reflects complementarity between legal enforcement and informal sanctions. By contrast, a larger fine always reduces the legal stigma.

[13] When $g_t > h$, $\int_{g_t}^{h}(h-g)\,f(g)\,dg = \int_{h}^{g_t}(g-h)\,f(g)\,dg > 0$.

Proposition 2 *When $ps < h$, deterrence of the nonprosocial is maximized by either strict liability or the fault-based regime with the standard $\widehat{g} = h$.*

The result contrasts with the irrelevance of the legal standard in the standard model without social preferences. Strict liability and fault-based offences are no longer equivalent because they yield different legal stigmas, which in turn affects incentives. Moreover, if deterrence is maximized with a fault regime, the efficient legal standard equals the social norm. When $ps \geq h$, the standard is irrelevant provided that $\widehat{g} \geq ps$. Individuals then behave as in the standard model and are either efficiently deterred or equally overdeterred.

The intuition for Proposition 2 is that deterrence of the nonprosocial increases with the legal standard when $\widehat{g} < h$. If it can be increased further still with $\widehat{g} > h$, deterrence reaches its maximum at the upper bound $\widehat{g} = \overline{g}$. For an enforcement policy satisfying $ps < h$, deterrence is therefore maximized either with the legal standard $\widehat{g} = h$ or with the standard $\widehat{g} = \overline{g}$, where the latter amounts to strict liability.

Legal design and reputational incentives. Figure 5 illustrates why one regime may perform better. Regime A is fault-based with the standard $\widehat{g} = h$, regime B is strict liability. The fine and the probability of detection are the same, hence enforcement need not be optimal.

We compare two situations, L and H, which differ in the intensity of reputational concerns with $\beta_L < \beta_H$. In situation L the deterrence curve is not very sensitive to beliefs about one's type and deterrence under either legal regime is relatively low. As shown, it is greater under strict liability. Situation H yields the opposite. The social norm has high salience and individuals are very sensitive to reputational penalties. Deterrence is then relatively high and is greater with the fault regime.

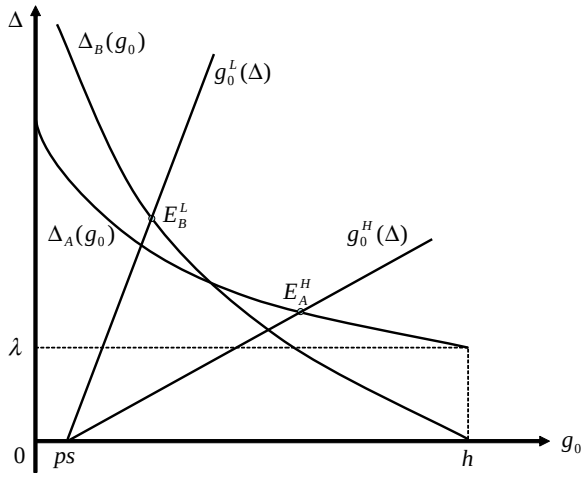

Figure 5. Stigma effects of legal design when $\beta_L < \beta_H$

The foregoing presumes that the stigma curves intersect. As remarked in Section 3, this need not occur. Convictions are more revealing about intrinsic predispositions in a fault-based than in a strict liability regime. However, while the event "guilty" constitutes *more unfavorable* news about an individual's type under the fault regime, the event "no news" is not *more favorable* under fault than under strict liability. Because reputational incentives depend on the difference in beliefs between the "guilty" and "no news" events, the stigma curves may intersect.

Lemma 3 *The stigma curves under strict liability and the fault regime with the legal standard $\widehat{g} = h$ intersect at most once and do so if and only if*

$$p > \frac{1}{1 + (1 - 2\lambda)F(h)}. \tag{13}$$

The condition (13) in the lemma cannot be satisfied if $p \leq 1/2$ or if $\lambda \geq 1/2$. Put differently, the situation depicted in Figure 5 cannot arise when good citizens constitute a majority or when a majority of harmful acts are undetected. Given $ps < h$, a fault regime then always induces greater deterrence than strict liability.

Optimal policies. The optimal legal regime and enforcement policy must be jointly chosen. Which policy is best depends on the underlying situation, e.g., the proportion of good citizens, the salience of the social norm, the likelihood of the circumstances under which harmful acts would be socially warranted, and the cost of detecting offenders and assessing circumstances.

Proposition 3 *Under an optimal legal regime and enforcement policy, the fine is maximal and the probability of detection satisfies* $ps_M < h$.
(i) If liability is fault-based, the legal standard is $\widehat{g} = h$, *the probability of detection satisfies* $p(s_M + \beta\lambda) \leq h$ *and the nonprosocial are underdeterred or efficiently deterred; convictions constitute a rare event,*

$$p(1-\lambda)(F(h) - F(g_0)) < \frac{1}{2}. \tag{14}$$

(ii) If liability is strict, the nonprosocial are underdeterred.
When ascertaining circumstances is not too costly,
(iii) liability is fault-based if $\lambda \geq 1/2$ *or if* s_M *or* β *are sufficiently large; if liability is strict, convictions (and therefore offences) constitute a frequent event,*

$$p\left[1 - \lambda F(h) - (1-\lambda)F(g_0)\right] \geq \frac{1}{2}. \tag{15}$$

The left-hand side of (14) is the frequency of convictions under a fault regime. The left-hand side of (15) is the frequency of convictions under strict liability.[14]

The maximum sanction principle still holds for the usual reason: a larger fine allows the same level of deterrence to be achieved with a smaller probability of detection, thus saving on enforcement costs. By contrast with the standard model, however, a fault regime may now be optimal even though assessing fault is costly and first-best deterrence may be optimal.

[14] Everyone commits the harmful act in circumstances $g \geq h$; for g in $[g_0, h)$ the non prosocial also commit the wrongful act.

Overdeterrence is never optimal. It would require $ps_M > h$, in which case reputational incentives vanish and strict liability does as well as the fault regime and strictly better if $k > 0$. But then this is dominated by strict liability with $ps_M = h$ which in turn is dominated by strict liability with some degree of underdeterrence and possibly also, if k is not too large, by the fault regime with either first-best deterrence or some degree of underdeterrence.

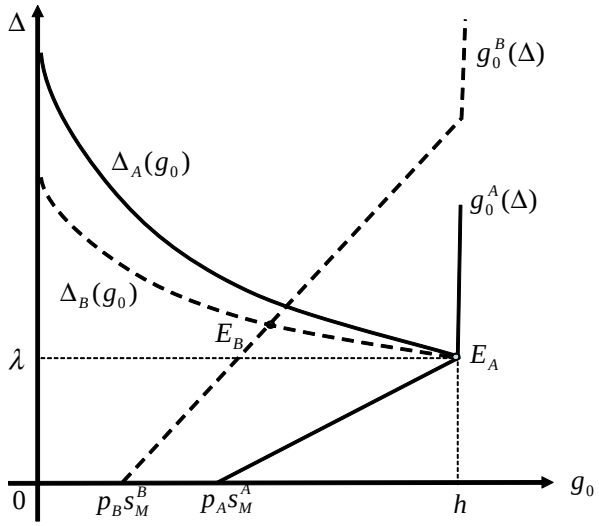

Fig. 6: Possible optima under a fault regime

In Figure 6, the optimal regime is fault-based when the maximal fine is s_M^A. As shown, the outcome is the corner equilibrium E_A. The probability of detection is then p_A such that $p_A(s_M^A + \beta\lambda) = h$ and enforcement satisfies the corner condition

$$p_A k f(h) \left(\frac{dg_0}{dp}\right)^-_{p=p_A} \geq c'(p_A) + k\left[1 - F(h)\right]. \tag{16}$$

The right-hand side is the increase in detection and fault assessment costs from a marginal increase in the probability of detection. The left-hand side is the resulting savings in fault assessment costs due to a marginal increase in deterrence up to $g_0 = h$. The derivative in (16) is a left derivative; the

right derivative is zero because increasing detection slightly beyond p_A has no effect on deterrence, as is obvious from Figure 6. Note that a corner solution as in (16) cannot arise if ascertaining fault is costless, i.e., $k = 0$.

The equilibrium at E_B illustrates an interior optimum where the maximum fine is smaller and is capped at s_M^B. The optimal probability of detection is then p_B such that $p_B(s_M^B + \beta\lambda) < h$, hence the nonprosocial are underdeterred. The marginal social benefit from greater detection is smaller than in case A because of the smaller fine. The first-order condition is now

$$(h + p_B k - g_0) f(g_0) \left(\frac{dg_0}{dp} \right)_{p=p_B} = c'(p_B) + k \left[1 - (1 - \lambda) F(g_0) - \lambda F(h) \right] \tag{17}$$

Whether the optimum is interior or at a corner, fault-based offences may do better than strict liability even though assessing fault is costly because of the larger legal stigma attached to convictions. The condition (14) in Proposition 3 is then necessary. If it did not hold deterrence could be increased by switching to strict liability under the same enforcement policy, as this would then yield a larger stigma (see the proof). Because this would also save on the fault assessment costs, strict liability would be unambiguously better.

Part (iii) of the proposition provides sufficient conditions for a fault-regime to be optimal when assessing fault is not too costly. The condition that good citizens are sufficiently numerous follows directly from Lemma 3. Even when good citizens are not a majority, fault-based offences do better when either the maximum permissible fine or image concerns are sufficiently large. In either case, appropriate deterrence (including first-best deterrence) can be achieved with a relatively small probability of detection, which ensures that the fault regime is deterrence maximizing. Condition (15) in Proposition 3 requires $p > 1/2$ and is necessary for strict liability to be optimal when assessing fault is not too costly. If the condition did not hold, switching to fault-based offences would increase deterrence under the

same enforcement policy. The condition need not hold when ascertaining circumstances imposes significant costs. The optimality of strict liability then follows solely from the fact that assessing fault is too costly.

6 Costly Sanctions and Stigmatization

We consider two extensions of the foregoing analysis. First, we inquire how the optimal policies differ when enforcement relies on nonmonetary sanctions rather than fines. Next we relax the assumption that reputational consequences only serve to motivate and examine the possibility that they also entail a social cost.

Nonmonetary sanctions. We take imprisonment as an example but the results would carry over to other forms of nonmonetary sanctions such as community service or suspension of a licence. Let s now denote the length of prison sentence, with s_M as the maximum allowable. The disutility is assumed to be proportional to the sentence. While a fine is a pure transfer involving no social costs, imprisonment is a net loss in the utilitarian calculus. In addition society may bear a resource cost which we represent by qs where q is the administrative cost per unit of sentence. Nonmonetary sanctions are in practice rarely imposed for strict liability offences. Even with fault-based offences, they make sense only because they extend the range of sanctions, given that effective fines are capped due to the individuals' limited wealth.

Consider first the standard framework without social preferences. Compared with the formulation in Section 3, enforcement costs are now augmented by the addition of

$$\int_{\min(ps,\widehat{g})}^{\widehat{g}} p(s+qs)f(g)\,dg,$$

the per capita costs of the sanctions imposed on convicted offenders. Whether the regime is strict liability or fault-based, the optimal policy is again to set

the sanction at the maximum permissible level.[15] Under strict liability and by contrast with the case of fines, the optimal probability of detection may now entail overdeterrence relative to first-best behavior. The possibility arises because increasing deterrence may reduce prison costs: while offenders are more likely to be sanctioned, there will also be a smaller number of them. Under a fault regime, the optimal legal standard satisfies $\widehat{g} = ps_M$ and everyone then complies with the law (see Shavell 1987). Undeterred individuals should not be found to be at fault even though they behave inefficiently from a social point of view, otherwise unnecessary sanction costs are incurred. When sanctions are socially costly, the advantage of fault-based liability is therefore to rely on the threat of sanctions while avoiding the cost of actually imposing them.

When assessing fault imposes no additional cost, a fault-based regime is clearly preferable. Indeed, fault is preferable if $k < (1+q)s_M$, i.e., the cost of assessing fault is smaller than the social cost of the nonmonetary sanction. With $\widehat{g} = ps_M$, deterrence under a fault regime is the same as under strict liability with the enforcement policy ps_M. Enforcement costs are

$$c(p) + kp[1 - F(ps_M)] < c(p) + (1+q)s_M p[1 - F(ps_M)]$$

where the left-hand side refers to the fault regime and the right-hand side to strict liability. In an optimal fault regime there may also be overdeterrence because increasing deterrence reduces fault assessment costs.

We now turn to the model with social preferences. The properties of the equilibria derived in Section 4 remain the same. The only change is with

[15] For any given level of deterrence ps, detection costs are reduced if s is raised and p is reduced proportionnaly, with no effect on sanction costs. One could also consider a combination of fines and imprisonment. See Polinsky and Shavell (2007).

respect to the social welfare function which is now

$$\begin{aligned} W &= W^* - (1-\lambda)\{\int_{g_0}^{h}(h-g)\,f(g)\,dg + p\int_{\min(g_0,\widehat{g})}^{\widehat{g}}(s+qs)f(g)\,dg\} \\ &\quad -\lambda\{\int_{g_1}^{h}(h-g)\,f(g)\,dg + p\int_{\min(g_1,\widehat{g})}^{\widehat{g}}(s+qs)f(g)\,dg\} \\ &\quad -C(p,g_0,g_1,\widehat{g}). \end{aligned} \tag{18}$$

The terms inside each set of brackets are the social loss from inefficient behavior (again allowing for the possibility of overdeterrence) and the social cost of sanctions imposed on detected offenders. We focus on the characteristics of an optimal fault-based regime.

Proposition 4 *With nonmonetary sanctions, when the optimal legal regime is fault-based the legal standard may be above or below the social norm. When* $\widehat{g} > h$*, the sanction is maximal,* $ps_M = \widehat{g}$ *and all comply with the law. When* $\widehat{g} \leq h$*, either the sanction is maximal,* $p(s_M + \beta\lambda) = \widehat{g}$ *and all comply with the law; or* $s \leq s_M$*,* $p(s+\beta\lambda) < \widehat{g}$ *and some of the nonprosocial do not comply.*

Compared with the case of fines, the difference is that the legal standard may now differ from the underlying social norm. Compared with nonmonetary sanctions in the standard model, the difference is that the sanctions may actually be imposed. Another difference is that the sanction may be less than maximum.

When $ps_M = \widehat{g} > h$, everyone complies with the law and is equally overdeterred. Sanctions are then at the maximum allowable level because they are never actually imposed. The possibility of a standard $\widehat{g} < h$ arises because, while a higher standard would increase deterrence, the number of convictions would also increase. Specifically, $\partial[F(\widehat{g}) - F(g_0)]/\partial\widehat{g} > 0$; see Figure 2. This would increase welfare when sanctions consist of fines, but with costly sanctions more convictions imply that sanction costs increase. At corner solutions where everyone complies with the law and $\widehat{g} \leq h$, sanctions

are again maximal because they serve only as a threat. When not everyone complies, however, they may be less than maximal.

To see the latter, recall that the equilibrium threshold of the nonprosocial solves

$$g_0 = p(s + \beta\Delta(g_0, p)), \tag{19}$$

where the legal stigma is written as a function of g_0 and of the probability of detection. For a given level of deterrence, the trade-off between the probability of detection and the sanction is

$$\eta \equiv -\frac{s}{p}\frac{dp}{ds}\bigg|_{g_0=\text{ct}} = \frac{s}{s + \beta\Delta_p} \tag{20}$$

where Δ_p denotes the partial derivative. When there are no reputational concerns, $\beta = 0$ and therefore $\eta = 1$. This yields the argument in the standard model, i.e., it is always desirable to increase the sanction and reduce the probability of detection proportionally. When $\beta > 0$ but the optimal policy entails overdeterrence, the same argument applies because all individuals are then equally overdeterred, hence the legal stigma vanishes and $\Delta_p = 0$. However, when $\widehat{g} \leq h$ and the nonprosocial undercomply with the law, the legal stigma is positive and $\Delta_p > 0$, implying $\eta < 1$. Increasing the sanction and reducing the probability of detection so as to keep deterrence constant then reduces detection costs but also increases sanction costs. The net effect may be to increase costs.

Socially costly stigmatization. Historically one of the main arguments against strict liability offences was the risk of stigmatizing respectable entrepreneurs.[16] Similar arguments have been made in the discussions accompanying the recent criminal law reforms. One way to capture social aversion to stigmatization risks is to express reputational utility as a con-

[16] See Paulus (1977) on the debates about "welfare offences" to counter food adulteration in mid 19th century Britain.

cave function of the beliefs about one's type.[17] To facilitate comparison with our previous formulation, we write the reputational term as $\beta v(\mu)$ where v is increasing and strictly concave with $v(0) = 0$ and $v(1) = 1$. The overall utility function of a type-t individual is now

$$u_t = w - \gamma_t \max(e - e^*, 0) + \beta v(\mu), \quad t = 0, 1. \tag{21}$$

An omniscient utilitarian regulator would impose the same wealth maximizing action profile $e^*(g)$, but he would not disclose information about the individuals' type because reputational gains and losses no longer cancel out. Hence the first-best welfare is

$$W^* = w_0 + \int_h^{\overline{g}} (g - h)\, f(g)\, dg + \beta v(\lambda). \tag{22}$$

All of the results of Section 4 continue to hold provided the legal stigma is rewritten as $\Delta = v(\bar{t}_N) - v(\bar{t}_G)$. Assuming that enforcement relies on fines, welfare is now

$$\begin{aligned} W = W^* - \{(1-\lambda) \int_{g_0}^{h} (h - g)\, f(g)\, dg + \lambda \int_{g_1}^{h} (h - g)\, f(g)\, dg\} \\ -\beta[v(\lambda) - \lambda \overline{v}_1 - (1-\lambda)\overline{v}_0] - C(p, g_0, g_1, \widehat{g}) \end{aligned} \tag{23}$$

where $\beta \overline{v}_t$ is the average reputational utility of a type-t individual,

$$\overline{v}_t = p \max[F(\widehat{g}) - F(g_t), 0] v(\bar{t}_G) + \{1 - p \max[F(\widehat{g}) - F(g_t), 0]\} v(\bar{t}_N), \quad t = 0, 1. \tag{24}$$

In equation (23) the term inside the curly brackets is the loss from inefficient behavior. The third term is the deadweight loss from stigmatization. In the first best, reputational utility is $\beta v(\lambda)$ for all individuals. Under a given legal regime and enforcement policy, the average reputational utility is equal to $\beta \overline{v}$ where

$$\overline{v} \equiv (1 - \lambda)\overline{v}_0 + \lambda \overline{v}_1.$$

[17] A similar approach has been used in models of self-signalling, see Köszegi (2006) and Dal Bó and Terviö (2013).

Because v is concave, $\overline{v} \leq v(\lambda)$ with strict inequality unless everyone behaves the same. Social aversion to stigmatization therefore introduces a trade-off between the usefulness of legal stigma for motivating appropriate behavior and the deadweight loss from stigmatization.[18]

Proposition 5 *Under stigmatization aversion and with sanctions consisting of fines, the optimal legal regimes and enforcement policies are as in Proposition 3 except for the legal standard of fault which satisfies* $\widehat{g} \leq h$.

In a strict liability regime, both good and bad citizens will at times choose not to comply with the law. When the nonprosocial are underdeterred, $\bar{t}_N > \bar{t}_G$ at equilibrium and therefore $\overline{v} < v(\lambda)$, meaning that there is a social loss from legal stigmatization. This loss could be reduced by increasing the probability of detection because the equilibrium $\overline{v}$ is increasing in p. Indeed the loss would vanish if the nonprosocial were made to behave like good citizens with an expected fine $ps_M \geq h$. However, it is easily shown that the loss from stigmatization is of the second order when enforcement is marginally reduced from the level ensuring first-best deterrence. Hence, under strict liability, bad citizens are optimally underdeterred as in Proposition 3.

Under a fault regime, the optimal enforcement policy is similar to part (i) of Proposition 3 except that the legal standard of fault may now be below the social norm, i.e., the law is more accommodating than the underlying social norm. In a corner solution, the deadweight loss from stigmatization vanishes because everyone complies with the law. The legal standard may be less than the first-best because strengthening the standard reduces the average reputational utility: a stronger standard increases deterrence, but some of the nonprosocial then no longer comply with the law. A standard

[18] A parallel can be made with Polinsky and Shavell (1979) who discuss the optimal fine and enforcement policy when individuals are risk averse with respect to income. Optimal enforcement may then be non stochastic.

$\widehat{g} < h$ is also possible in an interior solution where some of the nonprosocial do not comply with the law.

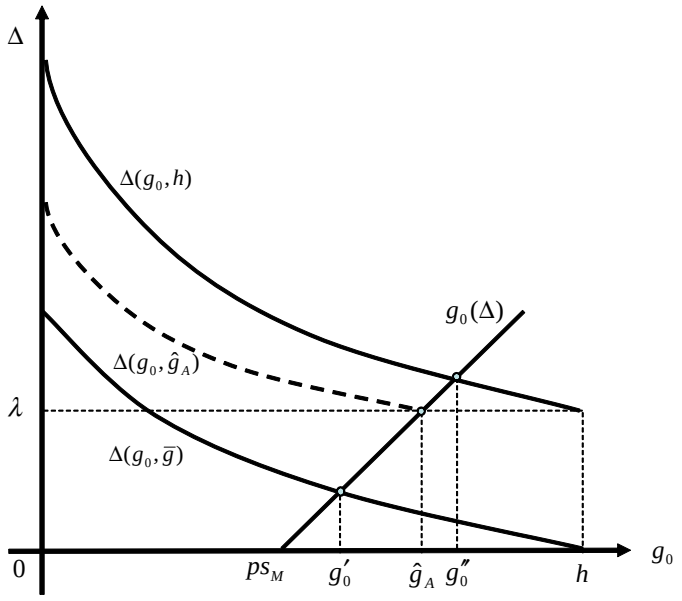

Fig. 7: Legal standards and stigmatization

Figure 7 illustrates the advantage of a fault regime together with the possibility of a standard less than the social norm. Suppose the probability of detection p is the best enforcement policy under a strict liability regime, yielding the deterrence curve $g_0(\Delta)$. We denote the stigma curves under the same probability of detection as $\Delta(g_0, \widehat{g})$. The curve for strict liability is $\Delta(g_0, \overline{g})$ and the equilibrium deterrence level is then g_0'. Switching to a fault regime with the standard $\widehat{g} = h$ would increase deterrence up to g_0''. However, this need not increase welfare because the loss from stigmatization may be much greater than in the initial situation under strict liability.[19] Nevertheless, as shown in Figure 7, a fault regime does unambiguously better than strict liability (assuming assessing fault is costless) if the legal standard

[19] It is easy to provide numerical examples where, at comparable levels of deterrence, the social loss from stigmatization is substantially larger under a fault regime than under strict liability.

is weakened to $\widehat{g}_A$, which corresponds to the stigma curve $\Delta(g_0, \widehat{g}_A)$. All individuals then comply with the law, so there is no stigmatization.

7 Concluding Remarks

Violating the law does not have the same social meaning under strict liability and fault-based offences. The latter is a stronger signal about one's character. Fault-based offences will therefore usually perform better in harnessing reputational concerns for the purpose of motivating socially appropriate behavior. Nevertheless, the result does not always follow because the social meaning and incentive effects depend on the frequency of convictions.

In many situations, socially unwarranted behavior will be a rare event because most individuals are socially minded. Convictions may also be rare because the enforcement policy achieves substantial deterrence. A fault-based regime that seeks to harness reputational incentives should aim at reducing apparent unlawfulness. Not finding fault may then be banal, therefore posterior beliefs conditional on "no news" do not differ too much from the prior. But then convictions yield substantial disesteem. By contrast, when convictions would be a frequent event under a fault regime, offences are banal. A strict liability regime would then perform better because it increases the significance of "no news".

The argument is reminiscent of Bénabou and Tirole's (2006, 2011) discussion of how acceptable behavior arises from the interplay of "honor" and "stigma". High stigma is attached to a behavior that "is just not done", only the worst type will do it. Alternatively, when "everyone does it", the same behavior carries little stigma. But then "not doing it" yields prestige. In the case of legal regimes, whether a conviction imposes significant stigma or whether "no conviction" confers significant honor depends on the underlying situation but also on the legal regime itself together with enforcement possibilities.

Our analysis emphasized the information conveyed by offences under different legal regimes given a pre-existing efficient social norm. One could also remark that different regimes have different "expressive content". In our analysis, the underlying social norm was that individuals should be socially minded and behave accordingly. Under a fault regime, the norm can be "expressed" by the duty or obligation with respect to which fault is defined. Indeed, we found that, when enforcement relies on fines, the optimal legal standard of fault is identical to the social norm. Strict liability is fuzzier in this respect. However, when enforcement relies on socially costly legal sanctions such as imprisonment or when stigmatization entails a social deadweight loss, the optimal legal norm may differ from the underlying social norm. When legal sanctions are costly, the standard of fault may be more lenient or harsher than the social norm. To mitigate the deadweight loss from stigmatization, the standard may be more lenient than the social norm.

Strict liability and fault-based offences differ in other ways with respect to expressive content. In particular, when individuals are imperfectly informed of the harm they may cause, a legal standard of behavior conveys information. Its prescriptive content helps socially minded individuals to coordinate on socially appropriate conduct. Imitative behavior due to social or self image concerns may then induce some bunching by the nonprosocial on the socially appropriate behavior.

Appendix A

Proof of Lemma 1. The claim follows directly from (8) and (9). ■

Proof of Lemma 2. Applying Bayes' rule,

$$\bar{t}_N = \frac{\lambda\left[1 - p\max(F(\widehat{g}) - F(g_1), 0)\right]}{1 - p\left[\lambda\max(F(\widehat{g}) - F(g_1), 0) + (1-\lambda)(F(\widehat{g}) - F(g_0))\right]}, \tag{25}$$

$$\bar{t}_G = \frac{\lambda \max(F(\widehat{g}) - F(g_1), 0)}{\lambda \max(F(\widehat{g}) - F(g_1), 0) + (1-\lambda)(F(\widehat{g}) - F(g_0))} \tag{26}$$

where (26) is undefined when $g_0 = g_1 = \widehat{g}$.

If $g_0 < \widehat{g} \leq h = g_1$, $\bar{t}_G = 0$ and therefore

$$\Delta = \bar{t}_N = \frac{\lambda}{1 - p(1-\lambda)(F(\widehat{g}) - F(g_0))}. \tag{27}$$

This is decreasing in g_0 with $\Delta = \lambda$ when $g_0 = \widehat{g}$. If $g_0 \leq g_1 = h < \widehat{g}$,

$$\begin{aligned} \Delta &= \frac{\lambda \left[1 - p(F(\widehat{g}) - F(h))\right]}{1 - p\left[\lambda(F(\widehat{g}) - F(h)) + (1-\lambda)(F(\widehat{g}) - F(g_0))\right]} \\ &\quad - \frac{\lambda(F(\widehat{g}) - F(h))}{\lambda(F(\widehat{g}) - F(h)) + (1-\lambda)(F(\widehat{g}) - F(g_0))}, \end{aligned} \tag{28}$$

which is positive and decreasing in g_0 with $\Delta = 0$ when $g_0 = h$. If $h < g_0 = g_1 < \widehat{g}$, (25) and (26) yield $\bar{t}_N = \bar{t}_G = \lambda$ so that $\Delta = 0$. For $h < g_0 = g_1 = \widehat{g}$, we take the limit of the preceding result, so that again $\Delta = 0$. ■

Proof of Proposition 1. Let $\widehat{g} \geq ps$. We first show uniqueness of the equilibrium. From Lemma 1 either $g_0 = g_1 > h$ or $g_0 \leq g_1 = h$. By Lemma 2, the first case implies $\Delta = 0$. Thus, it arises only if $ps > h$ and the equilibrium is then simply $g_0 = g_1 = ps$. A policy with $ps \leq h$ therefore yields the second case. The relevant domain for g_0 is then the interval $[ps, \min(\widehat{g}, h)]$. If $ps = \min(\widehat{g}, h)$, the equilibrium is trivially $g_0 = ps$, so let $ps < \min(\widehat{g}, h)$. From (8) the equilibrium g_0 is a solution to

$$g_0 = \min\left[\widehat{g}, p(s + \beta\Delta(g_0, \widehat{g}, p))\right] \tag{29}$$

where $\Delta(g_0, \widehat{g}, p)$ is defined by (27) or (28) for the cases $\widehat{g} \leq h$ or $\widehat{g} > h$ respectively. Equivalently, the equilibrium g_0 solves

$$\varphi(g_0) \equiv \min\left[\widehat{g}, p(s + \beta\Delta(g_0, \widehat{g}, p))\right] - g_0 = 0,\ g_0 \in [ps, \min(\widehat{g}, h)], \tag{30}$$

where $\varphi(g_0)$ is a continuous function. By Lemma 2, $\Delta(ps, \widehat{g}, p) > 0$ and therefore $\varphi(ps) > 0$. For the case $\widehat{g} \leq h$, $\Delta(\widehat{g}, \widehat{g}, p) = \lambda > 0$ and therefore $\varphi(\widehat{g}) \leq 0$. Because $\Delta(g_0, \widehat{g}, p)$ is strictly decreasing in g_0 in the relevant

domain, so is $\varphi(g_0)$ and the equilibrium is therefore unique. For the case $\widehat{g} > h$, $\Delta(h, \widehat{g}, p) = 0$ and $\varphi(h) < 0$. Again $\varphi(g_0)$ is strictly decreasing, ensuring uniqueness.

(i) The claim for the case $ps \geq h$ follows directly from the above argument.

(ii) For $ps < \widehat{g} \leq h$, the above argument shows that $g_1 = h$ and $g_0 \in (ps, \widehat{g}]$. If $p(s + \beta\lambda) \geq \widehat{g}$, $\varphi(\widehat{g}) = 0$ and the equilibrium satisfies $g_0 = \widehat{g}$. If $p(s + \beta\lambda) < \widehat{g}$, $\varphi(\widehat{g}) < 0$ and the equilibrium is $g_0 < \widehat{g}$ solving $g_0 = p(s + \beta\Delta(g_0, \widehat{g}, p))$. Differentiating totally with respect to $\widehat{g}$ and p yields

$$\frac{\partial g_0}{\partial \widehat{g}} = \frac{p\beta\Delta_{\widehat{g}}}{1 - p\beta\Delta_{g_0}}, \tag{31}$$

$$\frac{dg_0}{dp} = \frac{s + \beta\Delta + p\beta\Delta_p}{1 - p\beta\Delta_{g_0}}. \tag{32}$$

From (27), Δ_{g_0} is negative while Δ_p and $\Delta_{\widehat{g}}$ are positive. Hence (31) and (32) are both positive. To complete the argument, when $p(s + \beta\lambda) \geq \widehat{g}$, $g_0 = \widehat{g}$ and is then also increasing in $\widehat{g}$.

(iii) For $ps < h < \widehat{g}$, the argument is similar except that the solution now satisfies $g_0 \in (ps, h)$. Differentiating (29) totally with respect to $\widehat{g}$ and p again yields (31) and (32) but with Δ now defined by (28). The signs of Δ_{g_0} and Δ_p are as before but that of $\Delta_{\widehat{g}}$ is now ambiguous (see the proof of Proposition 2). Thus (32) is positive but the sign of (31) is ambiguous.

How g_0 varies with s is derived similarly and is left to the reader. ■

Proof of Proposition 2. Let $g_0(\widehat{g})$ denote the equilibrium as derived in Proposition 1 for some $ps < h$, so that $g_0(\widehat{g}) \leq g_1 = h$. The deterrence maximizing legal regime solves $\max_{\widehat{g}} g_0(\widehat{g})$. We consider separately the possibility that the solution satisfies $\widehat{g}^* \leq h$ or $\widehat{g}^* > h$.

If $\widehat{g}^* \leq h$, the function $g_0(\widehat{y})$ satisfies part (ii) of Proposition 1 and is strictly increasing, therefore $\widehat{g}^* = h$. If $\widehat{g}^* > h$, the function $g_0(\widehat{g})$ satisfies part (iii) of Proposition 1, hence $g_0(\widehat{g}) < h$. Either $\widehat{g}^* = \overline{g}$ or $\widehat{g}^*$ is an interior

solution in $(h, \overline{g})$. In the latter case, recalling (31), the solution must satisfy the first-order condition

$$\left.\frac{\partial g_0(\widehat{g})}{\partial \widehat{g}}\right|_{\widehat{g}=\widehat{g}^*} = \frac{p\beta\Delta_{\widehat{g}}(g_0(\widehat{g}^*), \widehat{g}^*)}{1 - p\beta\Delta_{g_0}(g_0(\widehat{g}^*), \widehat{g}^*)} = 0, \tag{33}$$

where the right-hand-side is as in (31) but with Δ defined as in (28). The second-order necessary condition is that

$$\left.\frac{\partial^2 g_0(\widehat{g})}{\partial \widehat{g}^2}\right|_{\widehat{g}=\widehat{g}^*} = \frac{p\beta\Delta_{\widehat{g}\widehat{g}}(g_0(\widehat{g}^*), \widehat{g}^*)}{1 - p\beta\Delta_{g_0}(g_0(\widehat{g}^*), \widehat{g}^*)} \tag{34}$$

be non positive, where the expression is obtained given that (33) holds and therefore $\Delta_{\widehat{g}}(g_0(\widehat{g}^*), \widehat{g}^*) = 0$. Because the denominator in (34) is positive, the second-order condition requires $\Delta_{\widehat{g}\widehat{g}}(g_0(\widehat{g}^*), \widehat{g}^*) \leq 0$. From (28),

$$\Delta_{\widehat{g}\widehat{g}}(g_0(\widehat{g}^*), \widehat{g}^*) = \frac{2p\lambda(1-\lambda)(F(h) - F(g_0(\widehat{g}^*)))}{[\Phi(1 - p\Phi]^2} > 0 \tag{35}$$

where

$$\Phi = F(\widehat{g}) - \lambda F(h) - (1-\lambda)F(g_0(\widehat{g}^*)).$$

Thus, the necessary condition does not hold, implying that the corner solution $\widehat{g}^* = \overline{g}$ is the only possibility. ■

Proof of Lemma 3. Solve (27) and (28) in the proof of Lemma 2 for the value of g_0 consistent with the same Δ under either strict liability or the fault regime with $\widehat{g} = h$. This yields

$$F(g_0) = \frac{1}{2(1-\lambda)}\left((1-2\lambda)F(h) - \frac{1-p}{p}\right). \tag{36}$$

The equation has a solution $F(g_0) > 0$ only if the condition in Lemma 3 holds. ■

Proof of Proposition 3. We first show that $ps < h$. Suppose to the contrary that $ps \geq h$. Proposition 1 then implies $g_0 = g_1 = ps$. If assessing

fault is costless the optimal regime is then indifferent, otherwise it must be strict liability. In either case, using (10),

$$\frac{\partial W}{\partial p} = (1-\lambda)\,(h-g_0)\,f(g_0)\frac{\partial g_0}{\partial p} + \lambda\,(h-g_1)\,f(g_1)\frac{\partial g_1}{\partial p} - c'(p). \qquad (37)$$

For $ps > h$, $\partial g_0/\partial p > 0$ and $\partial g_1/\partial p > 0$ so that (37) is negative. At $ps = h$, the preceding derivatives are discontinuous. Taking the left derivative,

$$\left.\frac{\partial W}{\partial p}\right|^{-}_{ps=h} = -\,c'(p) < 0.$$

Thus, an optimal policy entails $ps < h$. By Proposition 1, this implies $g_0 \leq g_1 = h$.

Next we show that $s = s_M$. From (10), under any legal regime, a policy change that reduces p with no change in g_0 is beneficial because

$$\left.\frac{\partial W}{\partial p}\right|_{g_0=\text{ct}} = -\,C_p < 0.$$

If $\widehat{g} > h$, p and s solve

$$g_0 = p(s + \beta\Delta(g_0, \widehat{g}, p)). \qquad (38)$$

where $\Delta_p > 0$. If $\widehat{g} \leq h$, either $g_0 < \widehat{g}$ and p and s solve (38); or $g_0 = \widehat{g}$ and $p(s + \beta\lambda) \geq \widehat{g}$. In all of these cases, it is possible to reduce p and increase s with no change in g_0. Hence the optimal fine is maximal.

Finally, we now show that, for any enforcement policy with $ps_M < h$, the optimal legal regime is deterrence maximizing. From (10),

$$\frac{\partial W}{\partial g_0} = (1-\lambda)(h-g_0)f(g_0) - C_{g_0} > 0.$$

The sign follows because $C_{g_0} \leq 0$; it is zero under strict liability and is negative if the regime is fault-based and $k > 0$. Because welfare is strictly increasing in g_0 for all $g_0 \leq h$, Proposition 2 implies that the optimal regime is either strict liability or the fault regime with $\widehat{g} = h$.

(i) Suppose the optimal regime is fault-based. The possibility of first-best deterrence with $p(s_M + \beta\lambda) = h$ is discussed in the text. Otherwise,

$p(s_M + \beta\lambda) < h$ and $g_0 < h$. We now prove (14). If fault-based liability is optimal for the probability of detection p, we must have

$$\left.\frac{\partial g_0}{\partial \widehat{g}}\right|_{\widehat{g}=h}^{+} = \frac{p\beta\Delta_{\widehat{g}}(g_0, h, p)}{1 - p\beta\Delta_{g_0}(g_0, h, p)} \leq 0. \tag{39}$$

where the expression is a right-derivative and is the same as (33) in the proof of Proposition 2. Now, the inequality must be strict because, as shown in the proof of the same proposition,

$$\left.\frac{\partial^2 g_0}{\partial \widehat{g}^2}\right|_{\widehat{g}=h}^{+} > 0,$$

again a right-derivative. If (39) held as an equality, g_0 would be increasing in $\widehat{g}$ in a neighborhood of h, implying that $\widehat{g} = h$ is not deterrence maximizing. Therefore $\Delta_{\widehat{g}}(g_0, h, p) < 0$. From (28), the latter inequality reduces to condition (14).

(ii) $ps_M < h$ implies $g_0 \leq g_1 = h$. Under strict liability g_0 solves

$$\varphi(g_0) \equiv p(s_M + \beta\Delta(g_0, \overline{g}, p)) - g_0 = 0.$$

Recalling that $\Delta(h, \overline{g}, p) = 0$, $\varphi(h) \equiv ps_M - h < 0$. Because $\varphi(g_0)$ is strictly decreasing, the equilibrium satisfies $g_0 < h$.

(iii) If $k = 0$, for any p the best regime is the one that maximizes deterrence. By Lemma 3, this is always the fault regime if $\lambda \geq 1/2$. When s_M or β are sufficiently large, one can obtain $g_0 = h$ with $p \leq 1/2$ satisfying $p(s_M + \beta\lambda) = h$, e.g., when $s_M \geq 2h$ or $\beta\lambda \geq 2h$. An optimal policy then necessarily satisfies $p \leq 1/2$. Condition (13) in Lemma 3 requires $p > 1/2$. Therefore, for $p \leq 1/2$, the fault regime dominates strict liability in terms of deterrence. By continuity, the same argument applies if k is positive but not too large. To conclude, we prove (15). Using the same argument as in (i), if the optimal regime is strict liability with $\widehat{g} = \overline{g}$, it must be the case that $\Delta_{\widehat{g}}(g_0, \overline{g}, p) \geq 0$, otherwise deterrence would be maximized with the fault regime $\widehat{g} = h$. From (28), this is equivalent to condition (15). ■

Before proving the next propositions, we derive a result for corner equilibria.

Lemma 4 *Let* $p^*(s_M+\beta\lambda) = \widehat{g}^* \leq h$ *so that* $g_0(p, \widehat{g}, s_M) = \widehat{g}^*$ *when* $(p, \widehat{g}) = (p^*, \widehat{g}^*)$. *Then the right and left derivatives satisfy*

$$\left.\frac{\partial g_0}{\partial \widehat{g}}\right|^{+}_{(p^*,\widehat{g}^*)} = \frac{p^{*2}\beta(1-\lambda)f(\widehat{g}^*)}{1+p^{*2}\beta(1-\lambda)f(\widehat{g}^*)}, \quad \widehat{g}^* < h, \tag{40}$$

$$\left.\frac{\partial g_0}{\partial \widehat{g}}\right|^{-}_{(p^*,\widehat{g}^*)} = 1, \tag{41}$$

$$\left.\frac{\partial g_0}{\partial p}\right|^{+}_{(p^*,\widehat{g}^*)} = 0, \tag{42}$$

$$\left.\frac{\partial g_0}{\partial p}\right|^{-}_{(p^*,\widehat{g}^*)} = \frac{s_M+\beta\lambda}{1+p^{*2}\beta(1-\lambda)f(\widehat{g}^*)}. \tag{43}$$

Proof: By Proposition 1, when $\widehat{g} < h$ and $p(s+\beta\lambda) \geq \widehat{g}$, $g_0 = \widehat{g}$ which implies (41) and (42). When $p(s+\beta\lambda) < \widehat{g}$, $\partial g_0/\partial \widehat{g}$ and $\partial g_0/\partial p$ satisfy (31) and (32) respectively where Δ satisfies (27). From the latter it is easily seen that

$$-\Delta_{g_0} = \Delta_{\widehat{g}} = p(1-\lambda)f(\widehat{g}) \text{ and } \Delta_p = 0 \text{ for } g_0 = \widehat{g} \leq h. \tag{44}$$

Substituting in (31) and (32) then yields (40) and (43). ■

Proof of Proposition 4. We only discuss the case where the optimal policy does not overdeter.

At a corner solution, $p(s+\beta\lambda) = \widehat{g} \leq h$. Increasing s and reducing p while preserving the preceding equality has no effect on sanctions (which are not incurred) but reduces the enforcement cost

$$c(p) + pk[\lambda(1-F(h)) + (1-\lambda)(1-F(\widehat{g})].$$

Hence the sanction must be maximal. To see that $\widehat{g} < h$ is then a possibility, differentiate welfare in (18) with respect to $\widehat{g}$. At a corner solution $p(s_M + \beta\lambda) = \widehat{g}^* < h$,

$$\frac{\partial W}{\partial \widehat{g}} = (1-\lambda)f(\widehat{g}^*)\left[(h+kp-\widehat{g}^*)\frac{\partial g_0}{\partial \widehat{g}} - p(1+q)s_M\left(1-\frac{\partial g_0}{\partial \widehat{g}}\right)\right].$$

Substituting from Lemma 4 yields

$$\text{sign}\left.\frac{\partial W}{\partial \widehat{g}}\right|_{\widehat{g}^*}^{+} = \text{sign}\left\{p(1-\lambda)(h+kp-\widehat{g}^*)f(\widehat{g}^*)\beta-(1+q)s_M\right\}.$$

If k is not too large, this is negative for $\widehat{g}^*$ sufficiently close to h.

At an interior solution where $g_0 < \widehat{g}$, the argument for the possibility that $\widehat{g} < h$ is similar. To see that $s < s_M$ is then a possibility, set $k = 0$ for simplicity. At the policy $(p, \widehat{g}, s)$,

$$\begin{aligned}\frac{\partial W}{\partial p} &= (1-\lambda)\left[(h+p(1+q)s-g_0)f(g_0)\frac{\partial g_0}{\partial p}-(1+q)s\int_{g_0}^{\widehat{g}} f(g)\,dg\right]\\ &\quad -c'(p)\\ &= 0. \qquad (45)\end{aligned}$$

The derivative with respect to s is

$$\frac{\partial W}{\partial s} = (1-\lambda)\left[(h+p(1+q)s-g_0)f(g_0)\frac{\partial g_0}{\partial s}-p(1+q)\int_{g_0}^{\widehat{g}} f(g)\,dg\right]. \quad (46)$$

Substituting from (45) in (46) yields

$$\frac{\partial W}{\partial s} = \frac{\theta p c'(p)}{s}-(1-\theta)(1-\lambda)p(1+q)\int_{g_0}^{\widehat{g}} f(g)\,dg \qquad (47)$$

where

$$\theta \equiv \frac{s\partial g_0/\partial s}{p\partial g_0/\partial p} = \frac{s}{s+\beta\Delta+p\beta\Delta_p}.$$

The expression follows from (31) and (32); (27) implies $\Delta_p > 0$ when $g_0 < \widehat{g}$. Substituting back in (47) yields

$$\text{sign}\frac{\partial W}{\partial s} = \text{sign}\left[c'(p)-\beta(\Delta+p\Delta_p)(1-\lambda)(1+q)\int_{g_0}^{\widehat{g}} f(g)\,dg\right].$$

Thus, the sign may be negative. ■

Proof of Proposition 5. We only prove (i) that $\widehat{g} < h$ is possible when the optimal regime is fault-based, (ii) that there is underdeterrence when the optimal regime is strict liability.

(i) When $\widehat{g} \leq h$, $g_0 \leq g_1 = h$. Setting $k = 0$ for simplicity, welfare in (23) reduces to

$$W = W^* - (1-\lambda)\int_{g_0}^{h} (h-g)\, f(g)\, dg - \beta[v(\lambda) - \overline{v}] - c(p)$$

where

$$\overline{v} = [\lambda + (1-\lambda)(1 - p(F(\widehat{g}) - F(g_0))]\, v(\overline{t}_N) \tag{48}$$

and $\overline{t}_N$ satisfies (27).

Consider the set of corner policies $(p, \widehat{g})$ with $g_0 = \widehat{g}$ and

$$p(s_M + \beta\lambda) = \widehat{g}. \tag{49}$$

A necessary condition for a policy in this set to be optimal is

$$\frac{dW}{dp} = (1-\lambda)(h-\widehat{g})f(\widehat{g})(s_M + \beta\lambda) - c'(p) = 0. \tag{50}$$

Let $(p^*, \widehat{g}^*)$ satisfy (49) and (50) and note that $\widehat{g}^* < h$. For this policy to be optimal it must also not be beneficial to move to an interior solution by marginal independent changes in either p or $\widehat{g}$.

The gain from a marginal change in $\widehat{g}$ while keeping $p = p^*$ is

$$\frac{\partial W}{\partial \widehat{g}} = (1-\lambda)(h-\widehat{g}^*)f(\widehat{g}^*)\frac{\partial g_0}{\partial \widehat{g}} + \beta\frac{\partial \overline{v}}{\partial \widehat{g}} \tag{51}$$

where the notations refer to either the right or left derivatives. From (48) and (27)

$$\begin{aligned}
\frac{\partial \overline{v}}{\partial \widehat{g}} &= -p^*(1-\lambda)f(\widehat{g}^*)\left(1 - \frac{\partial g_0}{\partial \widehat{g}}\right) v(\lambda) + v'(\lambda)\frac{\partial \overline{t}_N}{\partial \widehat{g}} \\
&= -p^*(1-\lambda)f(\widehat{g}^*)\left(1 - \frac{\partial g_0}{\partial \widehat{g}}\right)\left(v(\lambda) - \lambda v'(\lambda)\right).
\end{aligned} \tag{52}$$

Substituting from (52) in (51) and recalling Lemma 4,

$$\left.\frac{\partial W}{\partial \widehat{g}}\right|^{-}_{(p^*, \widehat{g}^*)} = (1-\lambda)(h-\widehat{g}^*)f(\widehat{g}^*)\frac{\partial g_0}{\partial \widehat{g}} > 0,$$

i.e., reducing $\widehat{g}$ from $\widehat{g}^*$ is not beneficial.

$$\text{sign } \frac{\partial W}{\partial \widehat{g}}\bigg|^{+}_{(p^*,\widehat{g}^*)} = \text{sign } \left\{p^*(1-\lambda)(h-\widehat{g}^*)f(\widehat{g}^*) - \left(v(\lambda) - \lambda v'(\lambda)\right)\right\}.$$

Hence, increasing $\widehat{g}$ from $\widehat{g}^*$ is not beneficial if

$$p^*(1-\lambda)(h-\widehat{g}^*)f(\widehat{g}^*) \leq v(\lambda) - \lambda v'(\lambda), \tag{53}$$

where the right-hand side is positive by the concavity of v.

Similarly the gain from a marginal change in p while keeping $\widehat{g} = \widehat{g}^*$ is

$$\frac{\partial W}{\partial p} = (1-\lambda)(h-\widehat{g}^*)f(\widehat{g}^*)\frac{\partial g_0}{\partial p} + \beta\frac{\partial \overline{v}}{\partial p} - c'(p^*) \tag{54}$$

where

$$\frac{\partial \overline{v}}{\partial p} = p^*(1-\lambda)f(\widehat{g}^*)\left[v(\lambda) - \lambda v'(\lambda)\right]\frac{\partial g_0}{\partial p}. \tag{55}$$

Then

$$\frac{\partial W}{\partial p}\bigg|^{+}_{(p^*,\widehat{g}^*)} = -c'(p^*) < 0.$$

Substituting from (55) and (50) in (54) and again using Lemma 4,

$$\text{sign } \frac{\partial W}{\partial p}\bigg|^{-}_{(p^*,\widehat{g}^*)} = \text{sign } \left\{-p^*(1-\lambda)(h-\widehat{g}^*)f(\widehat{g}^*) + \left(v(\lambda) - \lambda v'(\lambda)\right)\right\}.$$

Thus (53) also ensures that it is not beneficial to marginally change the probability of detection.

(ii) Under strict liability with the policy $ps_M = h$, the effect of a marginal decrease in p is given by the left derivative

$$\frac{\partial W}{\partial p}\bigg|^{-}_{ps_M=h} = \beta\frac{\partial \overline{v}}{\partial p} - c'(p) \tag{56}$$

where

$$\begin{aligned}\overline{v} &= p\left[(1-\lambda)(1-F(g_0)+\lambda(1-F(h))\right]v(\bar{t}_G) \\ &\quad + \left\{1 - p\left[(1-\lambda)(1-F(g_0)+\lambda(1-F(h))\right]\right\}v(\bar{t}_N).\end{aligned}$$

$\bar{t}_N$ and $\bar{t}_G$ are defined in (25) and (26) with $\widehat{g} = \overline{g}$ and g_0 solves (29). It is then easily verified that

$$\left.\frac{\partial \overline{v}}{\partial p}\right|^{-}_{ps_M=h} = 0,$$

hence (56) is negative, implying that p should be reduced from the full deterrence level. ■

References

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

[2] 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.

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

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

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

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

[7] Bernstein, L. (1992), "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry." *Journal of Legal Studies* 21, 115-157.

[8] Bodner, R. and D. Prelec (2003). "Self-Signaling and Diagnostic Utility in Everyday Decision Making." In I. Broca and J. Carillo (eds), *The Psychology of Economics*, Vol. I, Oxford University Press.

[9] Bramoullé Y., Currarini S., Jackson M.O., Pin P., Rogers B.W. (2012), "Homophily and Long-Run Integration in Social Networks." *Journal of Economic Theory*, 147, 1754-1786.

[10] Brekke, K.A., Kverndokk, S. and K. Nyborg (2003), "An Economic Model of Moral Motivation." *Journal of Public Economics* 87, 1967-1983.

[11] Brown, D. K. (2012). "Criminal Law Reform and the Persistence of Strict Liability." *Duke Law Journal* 62, 285-338.

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

[13] Cooter, R. (1998a), "Models of Morality in Law and Economics: Self-Control and Self-Improvement for the Bad Man of Holmes", *Boston University Law Review*, 78, 903-930.

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

[15] Cooter, R. (2000a), "Do Good Laws Make Good Citizens? An Economic Analysis." *Virginia Law Review* 86, 1577–1601.

[16] Cooter, R. (2000b), "Three Effects of Social Norms on Law: Expression, Deterrence and Internalization." *Oregon Law Review,* 79, 1–22.

[17] Dal Bó, E. and M. Terviö (2013), "Self-Esteem, Moral Capital and Wrong-Doing." *Journal of the European Economic Association* 11, 599-633.

[18] 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.

[19] Dau-Schmidt, K.G. (1990), "An Economic Analysis of the Criminal Law as a Preference-Shaping Policy." *Duke Law Journal*, 1-38.

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

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

[22] Ellingsen, T. and M. Johannesson (2008), "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98, 990-1008.

[23] Ellickson, R. C. (1991), *Order without Law: How Neighbors Settle Disputes*, Harvard University Press.

[24] Faure, M. and G. Heine (2005), *Criminal Enforcement of Environmental Law in the European Union*, Kluwer International.

[25] Fitzgerald, P.J. (1965), "Real Crimes and Quasi-Crimes". *Natural Law Forum* 10, 21-53.

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

[27] Galbiati, R. and P. Vertova P. (2008), "Obligations and Cooperative Behavior in Public Good Games." *Games and Economics Behavior* 64, 146-70.

[28] Galbiati R. and P. Vertova P. (2014), "How Laws Affect Behavior: Obligations, Incentives and Cooperative Behavior." *International Review of Law and Economics* 38, 48-57.

[29] 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.

[30] Horder, J. (2005), "Whose Values Should Determine When Liability is Strict?". In A. Simester (ed.), *Apraising Strict Liability*, Oxford University Press.

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

[32] Kadish, S.H. (1963). "Some Observations on the Use of Criminal Sanctions in Enforcing Economic Regulations." *University of Chicago Law Review* 30, 423-442.

[33] Kahan, D.M. (1998), "Social Meaning and the Economic Analysis of Crime." *Journal of Legal Studies* 27, 709-622.

[34] Köszegi, B. (2006), "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4, 673-707.

[35] Lacetera, N. and M. Macis (2010), "Social Image Concerns and Prosocial Behavior: Field Evidence from a Nonlinear Incentive Scheme." *Journal of Law, Economics, and Organization* 76, 225-237.

[36] Law Commission (2010), *Criminal Liability in Regulatory Contexts*, Consultation Paper No 195, London.

[37] Law Reform Commission of Canada (1974), *Studies on Strict Liability*, Information Canada, Ottawa.

[38] 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.

[39] Macaulay, S. (1963), "Non-Contractual Relations in Business," *American Sociological Review*, 28, 55-70.

[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] Mialon, S. (2014), "Declining Moral Standards and The Role of Law." Mimeo.

[42] Paulus, I. (1977). "Strict Liability: Its Place in Public Welfare Offences." *Criminal Law Quarterly* 20, 445-467.

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

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

[45] 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.

[46] Posner, R. (1997), "Social Norms and the Law: An Economic Approach", *American Economic Review* 87, 365-369.

[47] Posner, E. (1998), "Symbols, Signals, and Social Norms in Politics and the Law." *Journal of Legal Studies* 27, 765-798.

[48] Posner, E. (2000), *Law and Social Norms*. Harvard University Press.

[49] Rasmusen, E. (1996), "Stigma and Self-Fulfilling Expectations of Criminality." *Journal of Law and Economics* 39, 519-544.

[50] Shavell, S. (1987). "The Optimal Use of Nonmonetary Sanctions as a Deterrent." *American Economic Review* 77, 584–592.

[51] Shavell, S. (2002), "Law versus Morality as Regulators of Conduct." *American Law and Economics Review* 4, 227-257.

[52] Shavell, S. (2012), "When is Complying with the Law Socially Desirable?" *Journal of Legal Studies* 41, 1-36.378-405.

[53] Simester A. (2005), "Is Strict Liability always wrong?" In A. Simester (ed.), *Apraising Strict Liability*, Oxford University Press.

[54] Singer, R. (1989), "The Resurgence of Mens Rea: The Rise and Fall of Strict Criminal Liability." *Boston College Law Review* 30, 337-408.

[55] Spencer, J.R. and A. Pedain (2005), "Approaches to Strict and Constructive Liability in Continental Criminal Law." In A. Simester (ed.), *Apraising Strict Liability*, Oxford University Press.

[56] Teichman (2005), "Sex, Shame, and the Law: An Economic Perspective on Megan's Laws." *Harvard Journal on Legislation* 42, 355–415.

[57] Tyran, J. and L. Feld (2006), "Achieving Compliance When Legal Sanctions are Non-Deterrent." *Scandinavian Journal of Economics* 108, 135-156.

[58] Wils, W. (2006). "Is Criminalization of EU Competition Law the Answer?" In Cseres, K.J., Schinkel, M.-P. and F. Vogelaar (eds.), *Criminalization of Competition Law Enforcement*, Edward Elgar Publishing.

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