C.I.R.P.É.E

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

Cahier de recherche/Working Paper **13-21**

The Role of Social Image Concerns in the Design of Legal Regimes

Bruno Deffains
Claude Fluet

Août/August 2013

Deffains : Université Paris 2 and European Business School at Wiesbaden
bruno.deffains@u-paris2.fr
Fluet : Université du Québec à Montréal, Université Paris 2 and CIRPÉE
fluet.claude-denys@uqam.ca

We thank the participants of the CESifo Law and Economics Workshop in Munich and of seminars at the universities of Bonn, Lausanne, Lorraine, Mannheim, Nanterre, Vanderbilt and Yale. We are particularly grateful to John Asker, Cécile Bourreau-Dubois, Richard Brooks, Andrew Daughety, Dominique Demougin, Robert Ellickson, Andreas Engert, Henri Hansmann, Yolande Hiriart, Giovanni Immordino, Christine Jolls, Éric Langlais, Susan Rose-Ackerman, Jennifer Reinganum, Urs Scheizer, Gerhard Wagner, Abraham Wickelgre, and Ansgar Wohlschlegel.

Abstract:
We consider situations where legal liability yields insufficient incentives for socially efficient behavior, e.g., individuals who cause harm are not always sued or are unable to pay fully for harm done. Some individuals nevertheless behave efficiently because of intrinsic prosocial concerns. Others have no such concerns but would like people to believe that they do. We show that fault-based liability is generally more effective than strict liability in harnessing social image concerns. This extends to the case where courts can make mistakes. The rules of proof then affect the inferences drawn from court decisions and therefore the stigma attached to an adverse judgment. If fault is a rare event, plaintiffs or prosecutors should bear the burden of proving the defendant's fault; otherwise there are cases where defendants should prove compliance with the legal standard of behavior. Under either assignment of the burden of proof, incentives to comply are maximized by a standard of proof stronger than a mere preponderance of evidence.

Keywords: Normative motivations, prosocial behavior, fault, negligence, strict liability, tort law, public enforcement of law, burden of proof, standard of proof

JEL Classification: D8, K4, Z13

1 Introduction

In his *Theory of Moral Sentiments*, Adam Smith remarks that an individual found to have caused harm faces not only the possibility of a legal sanction — e.g., the damages he must pay — but also social disapproval or stigma. We inquire how a concern for social approval interacts with the incentives created by law and how this affects the relative performance of strict versus fault-based liability regimes. The distinction between negligence and strict liability is pervasive in private litigation, e.g., tort or contract law, but also arises in the public enforcement of laws and regulations.

To fix ideas, consider a situation where tort law yields insufficient incentives to control the risk of harming third parties. Injurers are partially judgment-proof or are not always sued, e.g., it is not always feasible to prove harm or identify the injurer. Some individuals nevertheless exert socially efficient care. They do so out of intrinsic moral or prosocial concerns. Other individuals have no such concerns but would like people to believe that they do; that is, they care about social approval. For instance, in a recent experimental study on liability rules (Angelova *et al.* 2012), half the subjects invested in safety measures even in the "No Liability" treatment and even though they could not be identified by the other subjects as having caused harm. Regarding the reputational motive, many experimental or field studies have also shown that social image concerns are important motivators of prosocial behavior (Dana *et al.* 2006, Ellingsen and Johannesson 2008, Andreoni and Bernheim 2008, Ariely *et al.* 2010, Funk 2010, Lacetera and Macis 2010, among others).

In our framework, an individual's actions are not directly observable by society at large. However, adverse court judgments provide public information from which inferences can be drawn about the individuals' actions and therefore about their intrinsic predispositions. Under either strict liability or the negligence rule, social image concerns are shown to provide the non prosocial individuals with some incentives to mimic the virtuous. The issue is how this influences the optimal design of liability regimes, when the objective is to induce socially efficient behavior.

A basic result is that fault-based liability is more effective than strict liability in harnessing reputational concerns. The reason is that trial outcomes are then more informative. Under strict liability, an adverse ruling merely ascertains that the defendant caused harm, not that he took inadequate precautions. Under the negligence rule, a liability ruling also ascertains that the defendant exerted inadequate care, thereby providing more precise information about his character. Socially useful incentives are therefore derived from the signaling role of "fault". Although the preceding argument is made with respect to the finding of negligence in a tort context, the same reasoning applies more generally.

To further explore the signalling role of fault, we extend the analysis to imperfect evidence about a defendant's actions. A complete characterization of a fault-based regime must now consider how courts deal with the risk of judicial error. The legal tools for this purpose are the assignment of the burden of proof — whether it is for the plaintiff to prove the defendant's fault or for the defendant to prove compliance with the legal standard of behavior — and the standard of proof that needs to be satisfied by the party with the burden of proof. The burden of proof assignment and the standard of proof affect the inferences drawn from trial outcomes; that is, they bear on the "meaning" or "significance" of a finding of fault. We show that when injurers have social image concerns, and by contrast with the results in Demougin and Fluet (2006, 2008), compliance with the legal due care standard is maximized by a standard of proof stronger than the common law preponderance of evidence. Roughly speaking, the assignment of the burden of proof depends on whether inadequate behavior is a frequent or infrequent event. When it seldom occurs, incentives to comply are maximized by assigning to the plaintiff (or prosecutor) the burden of proving the defendant's non-compliance.

A recent microeconomic literature has emphasized that one's actions may signal something about unobservable predispositions and that some predispositions are socially valued; see in particular Bernheim (1994), Bénabou and Tirole (2006, 2011), and Daughety and Reinganum (2010). Deffains and Fluet (2013) incorporate this approach in the unilateral accident model. The focus of that paper is the extent to which formal legal sanctions crowd-out or crowd-in informal motivations under different liability rules.[1] In the present paper, we also compare liability regimes but the emphasis is on the information content of judicial decisions and on prescriptions about the law of evidence under evidentiary uncertainty.

Our analysis is related to the legal literature on the interaction between laws and norms and on the role of signalling motives and imitative behavior; see in particular Kahan (1998) and Posner (1998, 2000). For a general discussion of legal sanctions versus informal motivation as regulators of conduct, see McAdams and Rasmusen (2007) and Shavell (2002). Our description of fault-based regimes also bears a relation to the concept of "expressive law". According to this view even "mild law", i.e., law backed by small sanctions or poorly enforced, can have desirable effects on behavior; see for instance Cooter (1998) and the discussion in Tyran and Feld (2006). Finally, our results are also related to the Law and Economics literature on the effect of stigma and shaming penalties in relation to criminal activity; see Rasmusen (1996), Harel and Clement (2007), and Zasu (2007) among others.

Section 2 presents the basic setup. Section 3 compares the incentives under strict liability and fault-based regimes with no judicial error. The next two sections introduce imperfect evidence about the defendant's behavior and derive the implications concerning the design of compliance maximizing legal regimes. Section 6 concludes. Proofs are in the Appendix.

[1] How material penalties and rewards affect informal motivations has been explored in a vast experimental and empirical literature. See Frey and Jegen (2001) for a survey.

2 The model

We start with a simple version of the economic model of liability for accidents (i.e., tort law). Private victims may sue individuals who have caused harm in order to obtain compensation. Later we show that, with minor modifications, the same model can be reinterpreted in terms of the public enforcement of laws and regulations.[2] Governmental agents then detect and prosecute violations which are sanctioned by fines.

Legal liability. Risk-neutral individuals are engaged in an activity which may impose an accidental loss of amount L on third parties. The risk of causing harm depends on the level of care which is $e = 0$ for low (or no) care and $e = 1$ for high care respectively. The probability of accident is $p(0) = p_l$ and $p(1) = p_h$ where $p_l > p_h$. The opportunity cost of low care is normalized to zero, that of high care is c distributed according to the differentiable cumulative function $G(c)$ with support $[0, \bar{c}]$. The interpretation is that the opportunity cost of care depends on the circumstances one may be facing.

Social welfare is maximized if, depending on the circumstances, individuals choose the level of care minimizing the sum of precaution costs and of expected harm. When the cost of care is c, the socially efficient action minimizes $ce + p(e)L$. It is therefore

$$e^*(c) = \begin{cases} 1 \text{ if } c \leq c^* \equiv (p_l - p_h)L, \\ 0 \text{ otherwise,} \end{cases} \tag{1}$$

where c^* is the critical cost level below which high care should be exerted.

We consider situations where legal liability does not always ensure socially efficient behavior. First, individuals causing harm are sued only with some probability q, e.g., they cannot always be identified or the victim has no proof. Secondly, injurers may not be able to pay fully for the harm they

[2]See Polinsky and Shavell (2007) for a survey of the economic model of the public enforcement of law.

caused. If found liable, the legal damages they will actually pay is $\min(w, L)$ where w denotes the defendant's defendant's liability limit, e.g., his assets.[3] When injurers are partially judgment-proof, strict liability is well known to provide insufficient incentives. By contrast, the negligence rule may then induce efficient behavior. The following assumption rules out this possibility.

ASSUMPTION 1: $p_l q \min(w, L) < \min(c^*, \bar{c})$.

Any combination of q and w satisfying the assumption is sufficient for our purpose. In Figure 1, the intervals denoted "strict liability", "negligence", and "either rule" indicate the range of c values for which the liability rules induce socially efficient precautions.

Consider first the strict liability rule. Injurers must then in principle pay for the harm they cause irrespective of circumstances. However, because he is not always sued or because of the limited liability constraint, an individual with cost of care c exerts high care only if $c \leq q(p_l - p_h)\min(w, L)$. Assumption 1 therefore implies that there will be circumstances where individuals take inadequate precautions.

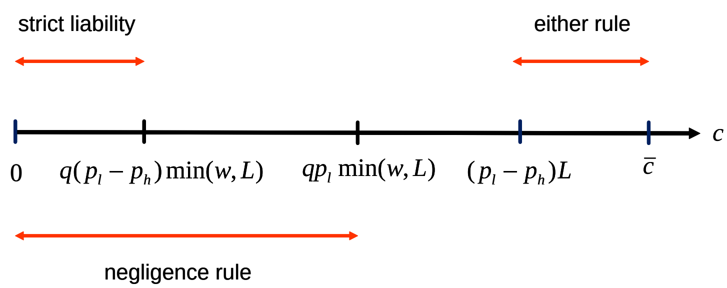

Figure 1. Inefficiencies

Under the negligence rule, courts are assumed to be able to verify the circumstance c and to set the legal due care standard at the socially effi-

[3] A similar set-up is employed in Shavell (1984) to analyze the joint use of liability and regulation. Note that the cost of care does not affect the individuals' liability limit. Either the cost is non pecuniary or the liability limit satisfies $w \leq w_0 - \bar{c}$, where w_0 is initial wealth and $\bar{c}$ is the upper bound on precaution costs.

cient level. Accordingly, an individual faces a risk of liability only when he exerts low care and $c \leq c^*$. Because of the dilution of incentives due to the chance of not being sued or to his partial judgment-proofness, an individual complies with due care only if $c \leq p_l q \min(w, L)$. Again there will be circumstances where precautions are inadequate, although inefficient care will be less frequent than under strict liability (unless p_h is zero).

When $\bar{c} \leq c^*$ or equivalently $G(c^*) = 1$, high care is socially efficient in all circumstances. If in addition $p_h = 0$, there is then no difference between strict liability and a fault-based regime. We rule out this possibility.

ASSUMPTION 2: $G(c^*) < 1$ or $p_h > 0$.

Social image concerns. So far we have described the standard framework where behavior depends only on private costs and benefits as conventionally defined. We now consider informal motivations. We assume that there are two types of potential injurers. Some potential injurers are "good citizens" with prosocial predispositions. They seek to behave in a socially or morally responsible manner by comparing their opportunity cost of care with the expected harm they impose on others.[4] Such individuals, referred to as type $\theta = 1$, choose the socially efficient level of care irrespective of legal legal sanctions. They exert high care when $c \leq c^*$ and low care otherwise. There is a known proportion λ of such individuals.

Secondly, individuals who are thought to be intrinsically prosocial earn social esteem, a source of utility. For those individuals who are not prosocial, referred to as type $\theta = 0$, behavior is determined by the utility function $u = w_n + \beta\overline{\theta}_I$ where w_n is net final wealth, β is a positive parameter and $\overline{\theta}_I \equiv E(\theta \mid I)$ is the belief of society at large about the individual's type conditional on the information I. Given our definition of types, $\overline{\theta}_I$ is simply the posterior probability that the individual is intrinsically prosocial and $\beta\overline{\theta}_I$ is

[4] This is a simple version of Kant's Categorical Imperative, as in Brekke et al. (2003). See Section 3 for a formulation where prosocial individuals suffer guilt when they deviate from the socially appropriate behavior.

the utility derived by the individual from society's beliefs about his intrinsic predispositions. In our analysis, much will depend on what information is available in society at large.

An individual's type is private information. For society at large, so are the circumstances faced by the individual, his chosen level of care and whether he caused harm, except insofar as these can be inferred from legal proceedings against the individual or from court judgments. Specifically, we assume that the only information "publicly" available about an individual — that is, in society at large — is either B for "bad news", which refers to the case where the individual is known to have been liable for harm done under the prevailing liability rule, or G for "no news", i.e. , the individual is not known to have been liable. As will become clear, no news is "good news" in the sense that one's social image is then more favorable than following bad news.

Adverse reputational effects imply that injurers would favor confidential settlements (and would be willing to pay hush money, see Daughety and Reinganum 1999). For simplicity we consider a simple litigation subgame where confidentiality is not feasible. Following the occurrence of harm, with probability q victims have access to all the evidence required under the prevailing liability rule; this is common knowledge between the parties. Initiating procedures involves a small cost, but litigation costs are otherwise negligible. Under a strict liability rule, the victim knows that he will succeed in court, hence a suit is filed. Under the negligence rule, a suit is filed only if the evidence shows that the injurer would be found negligent. We assume the following: first, lawsuits are public information (as would be the outcome at trial); second, if the case does not go to trial and payment has been extracted by the plaintiff, it becomes publicly known that such an agreement has been reached. Under these assumptions, out-of-court settlements have the same reputational effects as trials and the parties are indifferent between settling or going to trial.

The meaning of good and bad news depends on the liability regime. The prevailing liability rule is common knowledge and society makes the correct inferences from judicial procedures and trial outcomes. Under strict liability, bad news B about an individual means that he caused harm, was subsequently sued (which arises only with probability q) and was or would have been held liable in court. Under the negligence rule, bad news means that the individual caused harm, was sued and was or would have been held liable, hence was at fault.

Public enforcement of law. It will sometimes be useful to reinterpret our findings in a public enforcement of law framework. Plaintiffs are replaced by public agents in charge of enforcing the law, e.g., the police, inspectors or prosecutors. We give two examples.

Consider first a regulation against environmental spills. A spill creates harm of amount L and occurs with probability p_h or p_l depending on precautions. The fine for a spill is s. To allow for imperfect enforcement, Assumption 1 is rewritten as $p_l q \min(w, s) < \min(c^*, \overline{c})$ where q is the probability of detecting and prosecuting the individual responsible. Suppose $\overline{c} \leq c^*$, i.e., the high precaution level should always be chosen. Under a strict liability regulation, all spills are in principle sanctioned. Under a fault-based regime, the regulation is that high precautions should be taken. Spills are sanctioned only when the regulation has not been complied with; otherwise they are considered as an unavoidable risk.

Consider next a situation where $p_h = 0$ and $p_l = 1$, i.e., $e = 0$ can be interpreted as committing an action that causes harm (or expected harm, e.g. speeding on highways) versus not committing the action. Suppose $\overline{c} > c^*$ so that the action is socially warranted in some circumstances. A strict liability regime defines the action as unlawful irrespective of circumstances; q is the probability of detecting violations of the law. Under a fault-based regime, the action is unlawful only in the circumstances $c \leq c^*$.[5]

[5]Strict liability disregards circumstances. Prosocial individuals then sometimes effi-

3 Strict versus Fault-Based Liability

By assumption prosocial individuals always take adequate precautions, so we only need to examine the behavior of the non prosocial. Let $\alpha(e,c)$ denote the probability that an individual is found liable given that he caused harm and is sued. The probability is determined by the liability rule and may depend on the defendant's level of care and the circumstances. Under strict liability, $\alpha(e,c) = 1$ irrespective of care and of circumstances. Under the negligence rule, $\alpha(e,c) = 1$ if $e = 0$ and $c \leq c^*$ and is otherwise zero.

Incentives. Given the cost of care c, the expected utility of a non prosocial as a function of his care level is

$$\begin{aligned}\overline{u} &= p(e)q\alpha(e,c)[w_0 - \min(w,L) + \beta\overline{\theta}_B)] \\ &+ (1 - p(e)q\alpha(e,c))[w_0 + \beta\overline{\theta}_G] - ce, \quad e \in \{0,1\}, \end{aligned} \tag{2}$$

where $\overline{\theta}_B$ and $\overline{\theta}_G$ are society's beliefs about the individual's type conditional on bad and good news respectively. These beliefs are determined at equilibrium but are taken as given by the individual. High care is exerted if and only if

$$c \leq q\left[p(0)\alpha(0,c) - p(1)\alpha(1,c)\right]\left[\min(w,L) + \beta\Delta\right], \tag{3}$$

where $\Delta \equiv \overline{\theta}_G - \overline{\theta}_B$ will be referred to as the reputational penalty associated with bad news.

Under the strict liability rule, the condition (3) reduces to

$$c \leq q(p_l - p_h)[\min(w,L) + \beta\Delta] \equiv c_S(\Delta). \tag{4}$$

The right-hand side is the cost threshold below which a non prosocial exerts high care under strict liability (hence the subscript S). The critical cost level is written as a function of the reputational penalty yet to be determined.

ciently choose not to comply with the law given their knowledge of circumstances. See Shavell (2012).

Under the negligence rule, $\alpha(1, c) = 0$ for all values of c, $\alpha(0, c) = 1$ if $c \leq c^*$ and is zero otherwise. Therefore the condition (3) cannot be satisfied when $c > c^*$, i.e., the individual then necessarily exerts low care. He exerts high care if and only if

$$c \leq \min\{qp_l[\min(w, L) + \beta\Delta], c^*\} \equiv c_N(\Delta), \tag{5}$$

where the right-hand side is the cost threshold below which a non prosocial exerts high care under the negligence rule (hence the subscript N). Note that the threshold cannot be above the efficient c^*.

The proportion of the non prosocial exerting high care is $G(c_r)$ where r denotes the liability regime. It will be useful to focus instead on the ratio, denoted y, of the population of non prosocial exerting high care over those who *should* be exerting high care. This ratio will be referred to as the compliance rate and is defined by $y \equiv G(c_r)/G(c^*)$. As a function of the reputational penalty, the compliance rate satisfies

$$y = \psi_r(\Delta) \equiv \frac{G(c_r(\Delta))}{G(c^*)}, \quad r = S, N. \tag{6}$$

When the reputational penalty is nil, the compliance rate is the same as in the standard model without social preferences. Assumption 1 then ensures $\psi_S(0) < \psi_N(0) < 1$. Under either liability rule, the compliance rate is increasing in the reputational penalty. Under strict liability, for a sufficiently large penalty, overcompliance (i.e., $y > 1$) is consistent with (4) and (6), although this will never arise at equilibrium as shown below. Under the negligence rule, the compliance rate is also increasing in the reputational penalty, but only up to the socially efficient level $y = 1$.

Information and beliefs. We now turn to the determination of the reputational penalty. Using Bayes' rule, the posterior beliefs — and therefore the reputational penalty — can be expressed as a function of the compliance rate. The function differs between liability regimes.

Lemma 1 *Let* $p^* \equiv G(c^*)p_h + (1 - G(c^*))p_l$*. Given the compliance rate* y*, the reputational penalty satisfies* $\Delta = \varphi_r(y) \equiv \overline{\theta}^r_G(y) - \overline{\theta}^r_B(y)$*,* $r = S, N$*. Under strict liability,*

$$\overline{\theta}^S_B(y) = \frac{\lambda p^*}{p^* + (1-\lambda)(1-y)G(c^*)(p_l - p_h)},$$
$$\overline{\theta}^S_G(y) = \frac{\lambda(1 - qp^*)}{1 - q[p^* + (1-\lambda)(1-y)G(c^*)(p_l - p_h)]}.$$

Under the negligence rule, $\overline{\theta}^N_B(y) = 0$ *for all* y *and*

$$\overline{\theta}^N_G(y) = \frac{\lambda}{1 - (1-\lambda)(1-y)G(c^*)p_l q}.$$

Both $\varphi_S(y)$ *and* $\varphi_N(y)$ *are strictly decreasing functions, with* $\varphi_S(1) = 0$ *and* $\varphi_N(1) = \lambda$*.*

Note that p^* is the average probability of accident (over all potential circumstances) under the socially appropriate level of care; Assumption 2 implies $p^* > 0$. Under the negligence rule, bad news reveals perfectly that the injurer is non prosocial, hence $\overline{\theta}^N_B(y) = 0$. Under strict liability, this only becomes more likely so. When all the non prosocial exert efficient care, bad and good news provide no information at all under strict liability. Everyone then behaves the same and therefore faces the same probability of an adverse judgment. As a result, posterior beliefs do not differ from the prior, i.e., $\overline{\theta}^S_B(1) = \overline{\theta}^S_G(1) = \lambda$. By contrast, when all individuals comply with due care under the negligence rule, good news is uninformative because it occurs with certainty, but bad news would remain perfectly revealing.[6] A reputational penalty that is strictly decreasing in the compliance rate implies strategic substitutability: when more individuals exert efficient care, the incentive to avoid bad news becomes smaller.

[6] Bad news is then an out-of-equilibrium event with zero probablity, hence $\overline{\theta}^N_B(1)$ cannot be computed using Bayes' rule. The reputational penalty is then obtained from $\varphi_N(1) \equiv \lim_{y\to 1}\varphi_N(y) = \lambda$. The belief $\overline{\theta}^N_B(1) = 0$ can also be rationalized in terms of Cho and Kreps' (1987) D1 criterion.

The good news-bad news events constitute a binary signal about the individuals' type. An important consideration for what follows is whether the binary signal under the negligence regime is more informative than the signal under strict liability. Obviously, learning whether an individual has or has not caused harm is less informative than also learning whether he was at fault following the occurrence of harm. However, the good and bad news events considered here do not contain such detailed information. From Lemma 1 (and given Assumption 2), bad news represents more *unfavorable* information under the negligence rule than under strict liability, i.e., $\overline{\theta}_B^N(y) < \overline{\theta}_B^S(y)$ for all y. By contrast, good news (i.e., "no news" in the present context) need not constitute more *favorable* information. Loosely speaking, the intuition is that under negligence "no news" may provide little information if there is little scope for finding fault (i.e., $G(c^*)$ is small), while under strict liability "no news" may provide substantial information when p_l is large compared to p_h.

Lemma 2 *The good news-bad news signal under the negligence rule is more informative than under strict liability, implying* $\overline{\theta}_G^N(y) \geq \overline{\theta}_G^S(y)$ *for all* y*, if and only if*

$$qp_l \leq \frac{p_h}{G(c^*)p_h + (1 - G(c^*))p_l}. \tag{7}$$

Condition (7) is satisfied when $G(c^*) = 1$, i.e., when high care is socially warranted in all potential circumstances. By continuity, it is also satisfied when $G(c^*)$ is not too small and $qp_l < 1$. Condition (7) ensures that

$$\overline{\theta}_B^N(y) < \overline{\theta}_B^S(y) < \overline{\theta}_G^S(y) \leq \overline{\theta}_G^N(y), \quad \text{for all } y. \tag{8}$$

When (7) is not satisfied, the weak inequality on the right-hand side of (8) does not hold for some values of the compliance rate. The binary signals under the negligence rule and strict liability are then non comparable in terms of the usual "more informative than" ranking. Observe that (8) is sufficient, albeit not necessary, for the reputational penalties to satisfy $\varphi_N(y) > \varphi_S(y)$

for all y. Later in this section we provide an example where the preceding inequality does not hold at equilibrium.

Equilibrium. An equilibrium is a compliance rate and a reputational penalty that simultaneously solve $y = \psi_r(\Delta)$ and $\Delta = \varphi_r(y)$. Denote an equilibrium by $E_r = (y_r, \Delta_r)$. Figure 2 provides an example for both the strict liability and negligence rules. Figure 3 provides yet another example with a larger β, i.e., non prosocial individuals care more about social approval. The figure illustrates the case where β is large enough for everyone to comply with due care under the negligence rule.

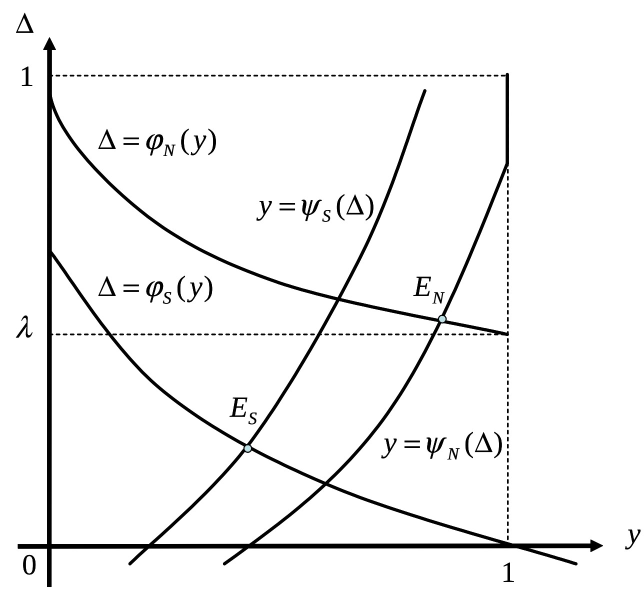

Figure 2. Equilibria

Proposition 1 *Under either regime there is a unique equilibrium* (y_r, Δ_r) *with* $\Delta_r > 0$*, where* $r = S$ *refers to strict liability and* $r = N$ *to the fault-based (or negligence) regime. In all cases,* $y_S < 1$.
(i) When the condition (7) holds, $y_S < y_N$.
(ii) When the condition does not hold, there are cases where $y_S > y_N$ *and* $\Delta_S > \Delta_N$.

(iii) $y_S < y_N$ and $\Delta_S < \Delta_N$ when λ or β are large. When β is sufficiently large, $y_N = 1$ and $\Delta_N = \lambda$.

The negligence rule does better than strict liability when it provides a more informative signal or when there is a large proportion of prosocial individuals or esteem concerns are important. In particular, the negligence rule induces the first-best rate of compliance when esteem concerns are strong enough. By contrast, strict liability always induces undercompliance.

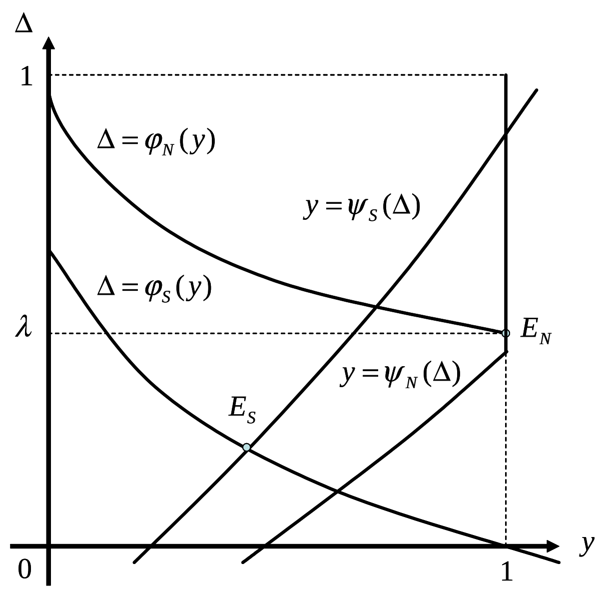

Figure 3. First best under the negligence rule

Corollary 1 *Under either regime, the compliance rate is increasing in w, q or β; the reputational penalty is decreasing in w and β. Under the fault-based regime, the reputational penalty and compliance rate are increasing in λ; under strict liability, they are increasing (decreasing) in λ when λ is small (large).*

Relaxing the sources of inefficiency, i.e., increasing the probability of suit or reducing the extent of judgment-proofness, increases compliance despite the possibility of some "motivational crowding-out" because of a smaller reputational penalty. Similarly, a greater concern for social image improves compliance. The second part of the corollary considers how the externality due to esteem concerns varies with the number of prosocial individuals. In a fault-based regime, a greater proportion of virtuous individuals shifts the φ_N curve upwards and therefore induces a greater number of the non prosocial to exert efficient care. Under strict liability, the effect is ambiguous: if the proportion of virtuous individuals is sufficiently small, more of them shifts the φ_S curve upwards, thereby increasing the frequency of socially efficient behavior; the opposite obtains when the proportion of virtuous individuals is sufficiently large.

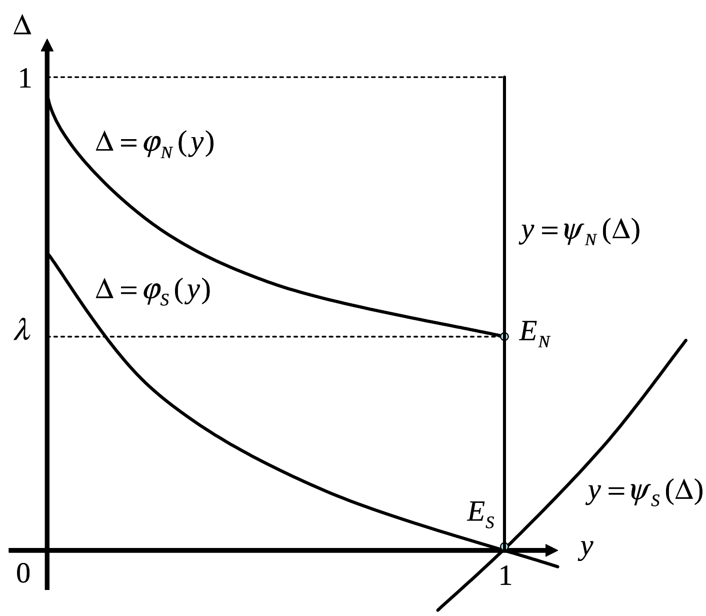

Figure 4. Relaxing Assumption 1

We briefly describe the equilibria when Assumption 1 is relaxed. Suppose that individuals causing harm are always sued and have sufficient assets to fully compensate for the harm done. In the standard model without social image concerns, the first best then obtains under either strict liability or the negligence rule; thus $\psi_N(0) = \psi_N(0) = 1$. The equilibria with social image concerns are represented in Figure 4. Under strict liability, the equilibrium is $(y_S, \Delta_S) = (1, 0)$; an adverse court judgment imposes no reputational penalty and the non prosocial are motivated solely by formal legal incentives. Under the negligence rule, the equilibrium is $(y_N, \Delta_N) = (1, \lambda)$. Although having been found negligent would impose a reputational penalty, the incentives provided by reputational concerns are superfluous with respect to inducing compliance with due care.

A case where strict liability does better. Let $p_h = 0$, $p_l = 1$ and $G(c^*) < 1$ as in the public enforcement of law example discussed in Section 2. Specifically, individuals must decide between committing or not committing a harmful action such as speeding. The condition (7) then does not hold. Lemma 1 nevertheless implies that $\varphi_N(y) > \varphi_S(y)$ in a neighborhood of full compliance. However, whether the inequality holds everywhere now depends on the parameters. Let $q > 1/2$ and $G(c^*) > (1-q)/q$. It can be shown that $\varphi_N(y) < \varphi_S(y)$ in a neighborhood of $y = 0$ if

$$\lambda < \frac{1}{2}\left(1 - \frac{1-q}{G(c^*)q}\right).$$

The curves then cross at

$$y_0 = \frac{1}{2(1-\lambda)}\left(1 - 2\lambda - \frac{1-q}{G(c^*)q}\right).$$

When $p_h = 0$, the compliance rate function is the same for both liability regimes in the relevant range, so we draw a single function ψ in Figure 6. The situation represented is one where formal legal incentives have little bite because $\min(w, s)$ is small, i.e., the fine is small ('mild law') or individuals have little wealth anyway. As shown in the figure, the equilibrium compliance rate and reputational penalty are larger under strict liability than under the fault-based regime. Nevertheless, with stronger concerns for social image yielding a compliance function such as $\widehat{\psi}$ in the figure, the fault-based regime provides greater incentives than strict liability. With even stronger social image concerns, full compliance would be achieved under the fault regime.

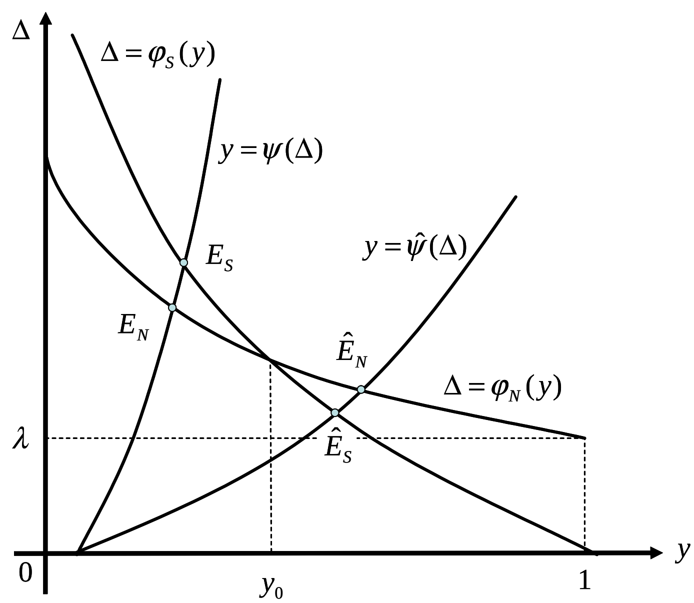

Figure 5. A case with $p_h = 0$, $p_l = 1$.

In the situation represented in Figure 6, not having been prosecuted is more favorable information under strict liability than under a fault-based regime. This is true at any compliance rate but the discrepancy gets larger at small compliance rates, then yielding a larger reputational penalty under strict liability. When liability regimes can only implement mediocre compliance rates, strict liability does better because of the larger reputational penalty. At higher compliance rates, the reputational penalty becomes small under strict liability but is bounded below by the prior λ under the fault-based regime. When higher compliance rates are attainable, the fault-based regime then does better. The intuition is that, under strict liability, escaping liability is a relatively rare and therefore meaningful event when y is small, given that $p_h = 0$, $p_l = 1$ and q is large. Fault-based liability makes this event more banal and therefore less meaningful.

Welfare. In comparing liability regimes, we took for granted that the socially appropriate behavior in the circumstance c was the level $e^*(c)$ minimizing the sum of prevention costs and expected harm, as in the standard model without image concerns. Accordingly, the liability regimes were ranked on the basis of the equilibrium compliance rates. We now derive this approach from fundamentals.

Let us write the individuals' utility as

$$u_\theta = w_n - \theta\gamma \max(\widehat{e} - e, 0) + \beta\overline{\theta}_I, \quad \theta = 0, 1. \tag{9}$$

For the non prosocial, $\theta = 0$ and the utility function is the same as before, with w_n as net wealth. For prosocial individuals, $\theta = 1$ and the middle term is the disutility (e.g., guilt) suffered when the level of care is less than the socially optimal $\widehat{e}$ (the optimal level will depend on the circumstances c); γ is large so that a prosocial always chooses the social standard of behavior. Welfare is the sum of utility over all individuals:

$$W = \int_0^{\overline{c}} [\lambda\overline{u}_1(c) + (1-\lambda)\overline{u}_0(c)]\ dG, \tag{10}$$

where $\overline{u}_\theta(c)$ is type θ's expected utility in the circumstance c.

Individuals can both cause harm or suffer harm caused by others. Consider an omniscient regulator who can directly impose the action $e(c)$, $c \in [0, \overline{c}]$, on all individuals. The average net wealth is then

$$\overline{w}_n = w_0 - \int_0^{\overline{c}} [ce(c) + p(e(c))L]\ dG, \tag{11}$$

where w_0 denotes the individuals' initial wealth. Let $\widehat{e}(c)$ be welfare maximizing and assume that the regulator can choose whether or not to publicize information about the individuals' types. Suppose first that an optimum entails that no information is disclosed. Then $\widehat{e}(c)$ maximizes W subject to the resource constraint (11) and to beliefs satisfying $\overline{\theta}_I = \lambda$ for all individuals. Clearly, this implies $\widehat{e}(c) = e^*(c)$ as defined in (1). Welfare then equals

$$W^* = w_0 - \int_0^{\overline{c}} [ce^*(c) + p(e^*(c))L]\ dG + \beta\lambda.$$

Now the same result would obtain with full or partial disclosure of information about types because the reputational benefits and losses simply cancel out. Therefore $e^*(c)$ is indeed the socially optimal level of care in the circumstance c.

When legal liability yields a second best with undercompliance, welfare is easily seen to equal

$$W = W^* - \int_{c_r}^{c^*} [(p_l - p_h)L - c] \, dG$$

where c_r solves $G(c_r) = y_r$ and y_r is the equilibrium compliance rate under the legal regime r. Abstracting from administrative costs, the best liability regime is therefore the one with the highest compliance rate.

4 Judicial Error

Fault-based liability is generally more efficient than strict liability at transforming the externality due to social image concerns into incentives to exert socially appropriate care. We now inquire whether this remains so when the greater informational requirements of such regimes can only be partially satisfied, so that trial outcomes are noisy signals. We consider the case where circumstances can be ascertained without error, hence there is no uncertainty about the legal standard of behavior, but courts can make mistakes in assessing whether the defendant complied with due care. They can erroneously rule against the defendant (a "false positive" or type I error) or erroneously rule in his favor (a "false negative" or type II error).

The risk of judicial error about a defendant's actions implies that the negligence rule will have elements of strict liability. Moreover, a complete description of the legal regime must now take into account how the judicial system trades-off type I and type II errors. Different trade-offs define different legal regimes. In practice, how a tribunal deals with the risk of error is determined by the prevailing rules of proof, by which me mean the burden of proof assignment and the standard of proof. The party with the burden of proof needs to persuade the court that he is entitled to a judgment in his favor, otherwise the default decision is that he looses the case. The standard of proof refers to the weight of evidence needed to discharge the burden.

Burden and standard of proof. Because harm occurs more often under low care, its mere occurrence provides some information about an individual's behavior.[7] Any additional information that might be used to assess behavior is summarized by the random variable x with cumulative distribution functions $F_h(x)$ and $F_l(x)$ that depend on the defendant's level of care. The distributions have continuously differentiable density functions, denoted $f_h(x)$ and $f_l(x)$, and the same support $[\underline{x}, \overline{x}]$. The "invariant support" condition means that no realization x perfectly reveals the defendant's care level.

Assumption 3: $f_l(x)/f_h(x)$ is strictly decreasing with $f_l(\overline{x})/f_h(\overline{x}) = 0$.

The distributions satisfy the monotone likelihood ratio property (MLRP) with the convention that a small x is more indicative of low care. The values of x should be interpreted as summarizing particular realizations of the potential evidence, i.e., $x' < x''$ means that the evidence underlying x' is more unfavorable to the defendant than the evidence underlying x''. By itself, x merely reflects the ranking of potential realizations of the evidence. The strength of the evidence is related to the ratio f_l/f_h. The condition that f_l/f_h goes to zero means that values of x approaching the upper bound of the support are tantamount to perfectly informative evidence.[8]

The plaintiff (or public prosecutor) has the burden of proving the occur-

[7] Indeed, when $p_h = 0$ and $p_l > 0$, the mere occurrence of harm provides perfect information. In what follows, we assume $p_h > 0$.

[8] The condition is not essential but it simplifies the exposition by eliminating the possibility of corner solutions in what follows. The invariant support condition is not essential either. If the supports of f_h and f_l overlap only partly, the evidence will sometimes (but not always) reveal care perfectly.

rence of harm and the injurer's identity. As before, this can either be done without ambiguity (with probability q) or not at all. When this requirement is satisfied, a suit is feasible. Both the plaintiff and the defendant are then assumed to have access to the additional evidence x about the injurer's behavior, as well as to perfect evidence concerning the circumstances faced by the defendant. The complete evidence eventually submitted to the court therefore comprises the occurrence of harm and the injurer's identity, the circumstances and the realization x. When $c > c^*$ the individual who has caused harm would not be found negligent, so there is no point in suing. When $c \leq c^*$ there are two possibilities:

(i) If, as is usually the case, the plaintiff has the burden of proving the defendant's negligence, he succeeds only if he can submit x such that

$$\frac{p_l f_l(x)}{p_h f_h(x)} > k \tag{12}$$

where k is the standard of proof that must be satisfied to discharge the burden. The expression on the left-hand side is the likelihood ratio of low care versus high care on the part of the defendant, given the evidence "the defendant caused harm and the additional evidence is x". The condition (12) states that evidence must be adduced showing that inadequate care is k times more likely than due care.

(ii) If the defendant bears the burden of proving that he complied with due care, he avoids liability only if he can submit x such that

$$\frac{p_h f_h(x)}{p_l f_l(x)} > k. \tag{13}$$

The interpretation is similar except that the left-hand side is now the relative likelihood of high versus low care. To escape liability, given that he is known to have caused harm, the defendant must show that compliance with due care is k times more likely than non compliance.

In the above formulation, court rulings are based purely on the evidence pertaining to the particular case before the court. Rulings are independent of the views or "priors" the court may hold about the general prevalence of high care among the population of individuals known to have caused harm and with cost of care c. The contested issue is the particular defendant's action. Both in common law and in civil law, priors in the form of a "known" (e.g., at equilibrium) proportion of similar defendants exerting high care or low care in similar circumstances would not be considered as relevant or admissible evidence.[9]

We consider standards of proof satisfying $k \geq 1$. The case $k = 1$ is the common law *preponderance of evidence* standard. For the party with the burden of proof, it then suffices to show that the relevant evidence gives greater weight to his contention, however slightly; that is, the party with the burden of proof need only prove his claim on a "more likely than not" basis. A threshold $k > 1$ means a stronger standard. For instance, it is sometimes said that $k = 3$ roughly conveys the standard of *clear and convincing evidence* (see Schauer and Zeckhauser 1996).

For a given standard of proof, the assignment of the burden of proof yields different *evidentiary thresholds* for the court to rule in favor of the plaintiff or the defendant. In Figure 6, $\widehat{x}_P(k)$ is the evidentiary threshold when the plaintiff bears the burden of proof. The defendant is then found negligent if $x < \widehat{x}_P(k)$, which corresponds to condition (12). When the defendant bears the burden of proving compliance with due care, the requirement is defined by condition (13) and the evidentiary threshold is $\widehat{x}_D(k)$. The defendant then escapes liability only if $x > \widehat{x}_D(k)$. For the preponderance of evidence standard, the evidentiary threshold does not depend

[9] Or would amount to statistical discrimination. For a discussion, see Demougin and Fluet (2005, 2006) and the references therein. Decisions based on (12) amount to computing the posterior probabilities of low versus high care on the basis of 'neutral normative prior'. See Posner (1999, p. 47): "Ideally we want the trier of fact to work from prior odds of 1 to 1 that the plaintiff or prosecutor has a meritorious case. A substantial departure from this position, in either direction, marks the trier of fact as biased." See Kaplow (2011) for a more general discussion.

on the assignment of the burden of proof; it is denoted $\widehat{x}_E$ in the figure.

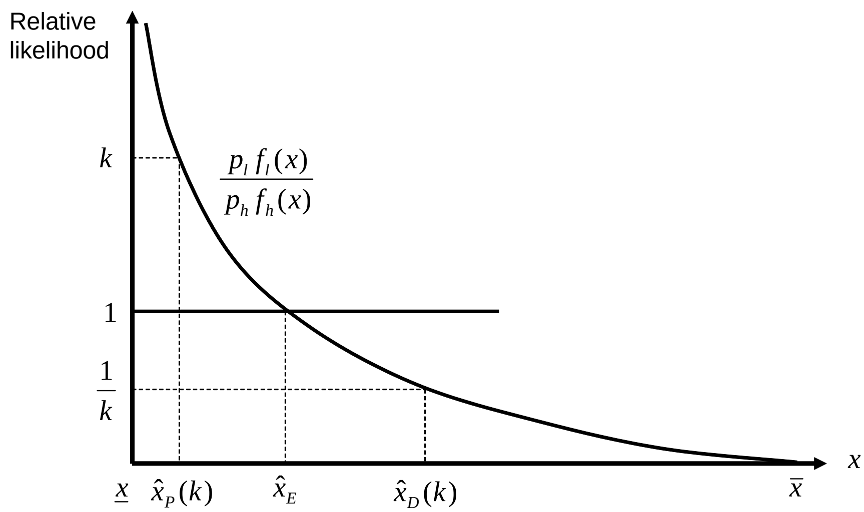

Fig. 6. Likelihood ratios

The liability risk differential. Let $\widehat{x}$ be the evidentiary threshold for some assignment of the burden and some standard of proof. Conditional on the occurrence of harm and a suit being filed, the probability of being found liable is $\alpha_j \equiv F_j(\widehat{x})$ depending on the care level $j = h, l$. When the defendant complied with due care, the probability of a type I error is α_h; when he exerted inadequate care, the probability of a type II error is $1 - a_l$. For any evidentiary threshold, the monotone likelihood ratio property implies that $\alpha_h < \alpha_l$ except when the threshold is at the bounds of the support, in which case the equality holds.

It is useful to express α_l as a function of the type I error $\alpha_h = F_h(\widehat{x})$, i.e., $\alpha_l(\alpha_h) \equiv F_l\left(F_h^{-1}(\alpha_h)\right)$.

Lemma 3 *$\alpha_l(\alpha_h)$ is strictly concave with $\alpha_l(0) = 0$, $\alpha_l(1) = 1$ and*

$$\alpha_l'(\alpha_h) = \frac{f_l(F_h^{-1}(\alpha_h))}{f_h(F_h^{-1}(\alpha_h))},\ \alpha_h \in [0, 1]. \tag{14}$$

Ex ante, given the possibility of court error, the liability risk differential between low and high care is $\delta = q(p_l\alpha_l - p_h\alpha_h)$. Written as a function of the type I error, the liability risk differential is $\delta(\alpha_h) = q(p_l\alpha_l(\alpha_h) - p_h\alpha_h)$ and is therefore concave in α_h. Figure 7 provides an illustration. Observe that the differential is zero when $\alpha_h = 0$ and that it is equivalent to the one under strict liability when $\alpha_h = 1$, except for the fact that injurers now escape liability when c is above c^*.

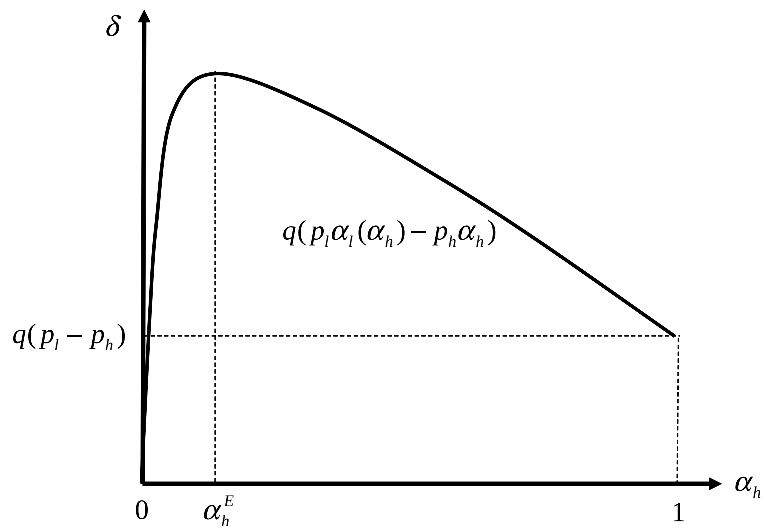

Fig. 7. Liability risk differential

Lemma 4 *Under Assumption 3, $\delta(\alpha_h)$ has a strict interior maximum at $\alpha_h = \alpha_h^E \equiv F_h(\widehat{x}_E)$ where $\widehat{x}_E$ solves*

$$\frac{p_l f_l(\widehat{x}_E)}{p_h f_h(\widehat{x}_E)} = 1.$$

The lemma states that, irrespective of the burden of proof assignment, the liability risk differential is maximized if courts decide the case on the basis of the preponderance of evidence standard of proof. Assigning the burden of proof to the plaintiff (resp. the defendant) and using a standard of proof stronger than preponderance would yield a type I error smaller (resp. larger) than α_h^E.

5 Efficient Rules of Proof in Fault-Based Regimes

We now consider how to design the rules of proof in order to maximize incentives to comply with due care. In the present set-up this is equivalent to maximizing welfare. From the preceding section, a liability regime can be summarized by the type I error α_h in rulings of negligence.

Bad news and good news are defined as before. Replicating the approach in Section 3, it is easily seen that a non prosocial individual with cost of care c exerts high care if and only if

$$c \leq \min\{q(p_l\alpha_l(\alpha_h) - p_h\alpha_h)[\min(w, L) + \beta\Delta], c^*\} \equiv c_N(\Delta, \alpha_h). \quad (15)$$

The interpretation is the same as for the condition (5) of Section 3. The critical cost below which an individual exerts high care is now written as a function of the rules of proof.[10] Accordingly, the compliance rate now satisfies

$$y = \psi_N(\Delta, \alpha_h) \equiv \frac{G(c_N(\Delta, \alpha_h))}{G(c^*)}. \quad (16)$$

As before, the function is strictly increasing in the reputational penalty (as long as $y < 1$).

Clearly, when the non prosocial have no social image concerns, i.e., $\beta = 0$, $\psi_N(\Delta, \alpha_h)$ does not depend on the reputational penalty and therefore compliance is maximized by maximizing the liability risk differential, thus by setting $\alpha_h = \alpha_h^E$.[11] When social image matters, however, there is an additional consideration because the legal regime will also affect the reputational penalty.

[10] When the evidence about the defendant's precautions is perfectly informative, $\alpha_h = 0$ and $\alpha_l = 1$. The condition (15) then reduces to (5).

[11] Compliance is then smaller than under perfectly informative evidenve because $q[p_l\alpha_l(\alpha_h^E) - p_h\alpha_h^E] < qp_l$, where the right-hand side is the liability risk differential without judicial error. As is well known, judicial error reduces incentives (see Polinsky and Shavell 1989).

Lemma 5 *Under the negligence regime with type I error α_h, the reputational penalty satisfies $\Delta = \varphi_N(y, \alpha_h)$ where*

$$\varphi_N(y,\alpha_h) \equiv \frac{\lambda[1 - G(c^*)qp_h\alpha_h]}{1 - G(c^*)q[(\lambda + (1-\lambda)y)p_h\alpha_h + (1-\lambda)(1-y)p_l\alpha_l(\alpha_h)]} - \frac{\lambda p_h \alpha_h}{(\lambda + (1-\lambda)y)p_h\alpha_h + (1-\lambda)(1-y)p_l\alpha_l(\alpha_h)}.$$

The function is decreasing in y with $\varphi_N(1, \alpha_h) = 0$.

For a given legal regime, an equilibrium is a solution to $y = \psi_N(\Delta, \alpha_h)$ and $\Delta = \varphi_N(y, \alpha_h)$. As before, the solution is unique. We denote the equilibrium by $(y_N(\alpha_h), \Delta_N(\alpha_h))$.

Choosing the best liability regime requires maximizing $y_N(\alpha_h)$ with respect to its argument. Let us define

$$\pi(\alpha_h, y) \equiv G(c^*)q[(\lambda + (1-\lambda)y)p_h\alpha_h + (1-\lambda)(1-y)p_l\alpha_l(\alpha_h)]. \tag{17}$$

The expression is the proportion of individuals found liable under the regime α_h when the rate of compliance is y. At equilibrium under this regime, the proportion of individuals found negligent is $\pi(\alpha_h, y_N(\alpha_h))$. We will say that finding negligence is a rare event if the proportion of individuals found to be at fault is less than one half; conversely, it is a frequent event if the proportion is greater than one half. We can now state the following.

Proposition 2 *Suppose the negligence regime maximizes compliance with due care. Then the plaintiff bears the burden of proving negligence (resp. the defendant bears the burden of proving compliance with due care) if finding negligence is a rare (resp. frequent) event. In either case the standard of proof is stronger than preponderance of evidence.*

The intuition is a simple one. Suppose α_h^* is compliance maximizing. Consider a marginal increase in the type I error; that is, it now becomes easier for the plaintiff to prove the defendant's negligence or it becomes more difficult for the defendant to prove that he complied with due care. Suppose this shifts the ψ_N curve to the right in the neighborhood of the equilibrium. Observe that this can arise only when α_h^* is below α_h^E, the evidentiary threshold under the preponderance of evidence standard. In other words, the compliance maximizing regime is then characterized by the plaintiff bearing the burden of proof and by a standard of proof stronger than preponderance of evidence. Now, if the φ_N curve also shifts upwards in a neighborhood of the equilibrium, then compliance will increase and therefore α_h^* cannot have been compliance maximizing. Thus, it must be that the φ_N curve shifts downwards, i.e., an efficient regime trades-off the effects on the liability risk differential and on the reputational penalty. More generally, at a compliance maximizing regime, a small change in the type I error must have effects of opposite signs on the ψ_N and φ_N curves. As shown in the Appendix, whether the φ_N curve shifts upwards or downwards depends on the frequency of negligence rulings.

Corollary 2 *When $qG(c^*)(\lambda p_h + (1-\lambda)p_l) \leq 1/2$, maximizing compliance requires the plaintiff to bear the burden of proving the defendant's negligence and to do so to a standard greater than preponderance of evidence.*

The corollary provides a straightforward sufficient condition. The expression in the corollary is an upper bound for the proportion of individuals found to be at fault under any regime. Hence the condition in the Corollary follows trivially from Proposition 2. As a particular case, the condition is satisfied when $p_l \leq 1/2$. Thus, when the occurrence of harm seldom arises even under low care, the plaintiff should be assigned the burden of proof.

Sufficient conditions for the defendant to bear the burden of proof are not as straightforward to characterize. We describe one possible case. Suppose $q = 1$ so that undercompliance is solely due to the inability to pay fully for the harm done. Suppose further that $G(c^*) = 1$, i.e., high care is the due care standard in all circumstance. Let $\widehat{y}_N$ be the equilibrium compliance rate when the evidence is perfectly informative as in Section 3. The corresponding proportion of individuals found negligent is

$$\widehat{\pi}_N = (1 - \lambda)(1 - \widehat{y}_N)p_l.$$

There will be cases where $\widehat{\pi}_N > 1/2$. Fault is then a frequent event because most injurers are not prosocial, few of them are induced to comply with due care even with perfect evidence (e.g., they have small wealth and do not care too much about social image), and harm occurs often when inadequate care is taken. A similar outcome can arise when the evidence is imperfectly informative.

Table 1 presents two examples. In the first p_l is less than one half, hence the burden of proof is on the plaintiff. The standard of proof differs significantly from the preponderance of evidence standard, the more so the greater the proportion of prosocial individuals in the population. In the second example, the probability of causing harm under either level of care is larger and p_l is above one half. When the proportion of prosocial individuals is small ($\lambda = .2$), the defendant now bears the burden of proof. Indeed, he must then satisfy a standard of proof close to clear and convincing evidence in order to discharge the burden. When the proportion of prosocial individuals is larger, fault again becomes a rare event and the burden of proof is assigned to the plaintiff.

Table 1

Example 1: $p_h = .0.1$, $p_l = 0.4$, $\alpha_h^E = .336$, $\alpha_l(\alpha_h^E) = .983$

λ	Burden	k^*	α_h^*	$\alpha_l(\alpha_h^*)$	y	Δ
.2	Plaintiff	1.47	.306	.974	.14	.24
.4	Plaintiff	1.67	.297	.970	.19	.43
.6	Plaintiff	2.22	.274	.959	.23	.55
.8	Plaintiff	3.31	.241	.937	.24	.60

Example 2: $p_h = .0.4$, $p_l = 0.8$, $\alpha_h^E = .283$, $\alpha_l(\alpha_h^E) = .964$

λ	Burden	k^*	α_h^*	$\alpha_l(\alpha_h^*)$	y	Δ
.2	Defendant	3.35	.375	.990	.34	.27
.4	Plaintiff	1.37	.257	.949	.36	.44
.6	Plaintiff	1.82	.233	.930	.39	.50
.8	Plaintiff	2.47	.207	.902	.37	.47

Note: $q = 1$, $w = .2$, $\beta = .8$, $G(c) = c$ for $c \in [0, 1]$, L is large enough for $c^* > 1$, $F_j(x) = 1 - \exp(-\nu_j x)$ with $\nu_h = 1$ and $\nu_l = 10$.

6 Concluding Remarks

An adverse court judgment does not have the same social meaning in strict liability and fault-based regimes. In either case, the meaning also differs depending on the proportion of virtuous individuals in the population and the extent to which formal legal sanctions underdeter. When assessing a defendant's actions is subject to error, the meaning of a finding of fault also depends on the risk of type I and type II judicial errors and therefore on the rules of proof.

In many situations, accidental harm or violations of the law will be rare events and so will be legal suits. When the evidence is imperfect, a fault-based regime that seeks to maximize incentives to comply with the legal standard of behavior should then make it relatively difficult to find fault. This is achieved by assigning to the plaintiff (or prosecutor) the burden of proving the defendant's fault and imposing a standard of proof stronger than preponderance of evidence. The intuition is that not finding fault is then banal, i.e., posterior beliefs do not differ much from the prior. By contrast, a finding of fault yields substantial disesteem. Making it harder still to find fault increases the reputational sanction and therefore the incentives to comply. On the other hand, when accidental harm and suits are frequent events, not having been found liable may provide significant prestige. The reputational gain — hence the incentives to comply with due care — can be increased by making it relatively difficult to escape liability. The best regime is then one that imposes on the defendant the burden of proving that he complied with due care and to do so to a standard stronger than preponderance of evidence. Obviously, this reasoning abstract from other trade-offs such as litigation costs or differences between plaintiff and defendant in their access to the evidence (see Bernardo et al. 2000, Hay and Spier 1997, Demougin and Fluet 2008, and Shin 1998).

Our results are reminiscent of Bénabou and Tirole's (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", i.e., 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 trial outcomes under the negligence rule, whether the finding of fault imposes significant stigma or whether not finding fault confers significant honor depends on the underlying situation, but to some extent can also be influenced by the liability regime for the purpose of increasing incentives to comply with the legal standard of behavior.

Our analysis focused on reputational concerns and the information conveyed by judicial outcomes under different liability regimes. A finding of fault is a more precise signal about one's character than being found liable under a strict liability regime. 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. This norm is perfectly "expressed" by a fault-based regime, i.e., by the duty or obligation with respect to which fault is defined. Strict liability is fuzzier in this respect. In the accident model, it merely prescribes that individuals compensate victims for harm done. Strict liability and fault-based regimes may also differ in other ways with respect to expressive content. When individuals are imperfectly informed of the harm they may cause, the legal standard of due care under a fault-based regime conveys information, as in D'Antoni and Galbiati (2007). The prescriptive content of fault may then help the prosocial themselves to coordinate on the socially appropriate behavior (see Cooter 1998). Imitative behavior due to reputational concerns then induces some bunching by the non prosocial.

Appendix

Proof of Lemma 1. Let us write the good news-bad news events as G_r and B_r, depending on the liability regime $r = S, N$. Under strict liability, the probability that a prosocial is found liable is

$$\Pr\left(B_S \mid \theta = 1\right) = G(c^*)p_h q + [1 - G(c^*)]p_l q = p^* q.$$

For a non prosocial, using (6), the probability is

$$\begin{aligned}\Pr(B_S \mid \theta=0) &= G(c_S(\Delta))p_h q + [1-G(c_S(\Delta))]p_l q \\ &= yG(c^*)p_h q + (1-yG(c^*))p_l q \\ &= p^* q + (1-y)G(c^*)(p_l - p_h)q\end{aligned}$$

Hence,

$$\begin{aligned}\Pr(B_S) &= \lambda \Pr(B_S \mid \theta=1) + (1-\lambda)\Pr(B_S \mid \theta=0) \\ &= p^* q + (1-\lambda)(1-y)G(c^*)(p_l - p_h)q.\end{aligned}$$

Applying Bayes' rule,

$$\begin{aligned}\overline{\theta}_B^S = \Pr(\theta=1 \mid B_S) &= \frac{\lambda \Pr(B_S \mid \theta=1)}{\Pr(B_S)} \\ &= \frac{\lambda p^*}{p^* + (1-\lambda)(1-y)G(c^*)(p_l - p_h)}\end{aligned}$$

and

$$\begin{aligned}\overline{\theta}_G^S = \Pr(\theta=1 \mid G_S) &= \frac{\lambda(1-\Pr(B_S \mid \theta=1))}{1-\Pr(B_S)} \\ &= \frac{\lambda(1-qp^*)}{1-q[p^* + (1-\lambda)(1-y)G(c^*)(p_l - p_h)]}.\end{aligned}$$

Under the negligence rule, prosocial individuals are never found negligent so that $\Pr(B_N \mid \theta=1)=0$. For the non prosocial the probability is

$$\begin{aligned}\Pr(B_N \mid \theta=0) &= [G(c^*) - G(c_N(\Delta))]p_l q \\ &= (1-y)G(c^*)p_l q.\end{aligned}$$

Hence

$$\begin{aligned}\Pr(B) &= \lambda \Pr(B \mid \theta=1) + (1-\lambda)\Pr(B \mid \theta=0) \\ &= (1-\lambda)(1-y)G(c^*)p_l q.\end{aligned}$$

Applying Bayes' rule, $\overline{\theta}_B^N = \Pr(\theta=1 \mid B) = 0$ while

$$\overline{\theta}_G^N = \Pr(\theta=1 \mid G_S) = \frac{\lambda}{1-(1-\lambda)(1-y)G(c^*)p_l q}.$$

The rest of the proof is left to the reader.

Proof of Lemma 2. Given the information provided by bad or good news and taking y as given, the likelihood ratio of $\theta = 1$ relative to $\theta = 0$ is

$$\boldsymbol{l}_r(A) = \frac{\Pr(A \mid \theta = 1)}{\Pr(A \mid \theta = 0)}, \quad A \in \{G_r, B_r\}, \; r = S, N.$$

The probabilities on the right-hand side are derived in the proof of Lemma 1. The binary signal under the liability regime r' is more informative than the signal under regime r if $\boldsymbol{l}_{r'}(\cdot)$, considered as a random variable, is a mean-preserving spread of $\boldsymbol{l}_r(\cdot)$. Likelihood ratios have the same mean (equal to unity), so the MPS condition is satisfied if $\boldsymbol{l}_{r'}(B_{r'}) \le \boldsymbol{l}_r(B_r)$ and $\boldsymbol{l}_{r'}(G_{r'}) \ge \boldsymbol{l}_r(G_r)$ with at least one strict inequality. Clearly, $\boldsymbol{l}_N(B_N) \le \boldsymbol{l}_S(B_S)$. For $y < 1$, $\boldsymbol{l}_N(G_N) \ge \boldsymbol{l}_S(G_S)$ is easily seen to be equivalent to condition (7). That $\boldsymbol{l}_N(G_N) \ge \boldsymbol{l}_S(G_S)$ implies $\overline{\theta}_G^N(y) \ge \overline{\theta}_G^S(y)$ is left to the reader.

Proof of Proposition 1. Uniqueness of the equilibrium follows from $y = \psi_r(\Delta)$ and $\Delta = \varphi_r(y)$ being increasing and decreasing functions respectively, $r = S, N$. Because $\psi_S(0) < 1$ by Assumption 1 and $\varphi_S(1) = 0$ by Lemma 1, the $\varphi_S(y)$ and $\psi_S(\Delta)$ curves can intersect only at some $y_S < 1$, implying $\Delta_S > 0$. For the negligence rule, $\varphi_N(y)$ and $\psi_N(\Delta)$ can intersect either at $y_N < 1$ with $\Delta_N > \lambda$ or at $y_N = 1$ with $\Delta_N = \lambda$.

(i) To see that $y_S < y_N$, suppose the contrary, i.e., $y_S = \psi_S(\Delta_S) \ge \psi_N(\Delta_N) = y_N$. Recalling (4) and (5), the inequality can hold only if $\Delta_S \ge \Delta_N$. However, this yields a contradiction because (7) implies $\varphi_S(y) < \varphi_N(y)$ and both functions are decreasing, implying $\Delta_S = \varphi_S(y_S) < \varphi_S(y_N) < \varphi_N(y_N) = \Delta_N$.

(ii) See the example in the text.

(iii) It is easily verified that, for all y, $\varphi_S(y) \to 0$ and $\varphi_N(y) \to \lambda$ as $\lambda \to 1$. By continuity, for λ large, we therefore have $\varphi_S(y) < \lambda \le \varphi_N(y)$ for all y, implying $y_S < y_N$ and $\Delta_S < \Delta_N$. For the negligence rule, $y_N = 1$ and $\Delta_N = \lambda$ if β satisfies $p_l q[\min(w, L) + \beta\lambda] \ge c^*$, while it is always the case that $y_S < 1$ as shown above. By continuity, for β large, we will therefore have $y_S < y_N$. An increase in β shifts the ψ_S curve to the right with no effect on φ_S. Therefore, for a sufficiently large β, $\varphi_S(y_S) < \lambda$ implying $\Delta_S < \Delta_N$.

Proof of Corollary 1. An increase in w or β shifts the ψ_r curves to the right with no effect on the φ_r curves. An increase in q shifts the ψ_r curves to the right and the φ_r curves upwards, so that y_r increases but the effect on Δ_r is ambiguous. A change in λ affects only the φ_r curves. For the negligence rule, an increase in λ shifts the φ_N curve upwards, hence both y_N and Δ_N increases. For the strict liability rule, the penalty function can be expressed as

$$\varphi_S(y) = \frac{\lambda(1-\lambda)(1-y)G(c^*)q(p_l - p_h)}{\pi(y)(1-\pi(y))}$$

where

$$\pi(y) \equiv p^* + (1-\lambda)(1-y)G(c^*)q(p_l - p_h).$$

It follows that

$$sign\left(\frac{\partial\varphi_S(y)}{\partial\lambda}\right) = sign\left(\frac{\partial}{\partial\lambda}\left(\frac{\lambda(1-\lambda)}{\pi(y)(1-\pi(y))}\right)\right).$$

Now

$$\frac{\partial}{\partial\lambda}\left(\frac{\lambda(1-\lambda)}{\pi(y)(1-\pi(y))}\right) = \frac{\pi(y)(1-\pi(y))(1-2\lambda) + \lambda(1-\lambda)(1-2\pi(y))(d\pi(y)/d\lambda)}{\pi(y)^2(1-\pi(y))^2}.$$

When λ is close to zero, the second term in the numerator is negligible and the first term is positive. When λ is close to unity, the second term in the numerator is negligible but the first term is negative.

Proof of Lemma 3. Equation (14) follows directly from the definition of $\alpha_l(\alpha_h)$. Differentiating once more,

$$\alpha_l''(\alpha_h) = \frac{d}{d\alpha_h}\left(\frac{f_l(F_h^{-1}(\alpha_h))}{f_h(F_h^{-1}(\alpha_h))}\right) < 0,$$

where the sign follows from the MLRP in Assumption 3.

Proof of Lemma 4. Using Lemma 3,

$$\delta'(\alpha_h) = qp_h \left(\frac{p_l \alpha_l'(\alpha_h)}{p_h} - 1 \right) = qp_h \left(\frac{p_l f_l(x)}{p_h f_h(x)} - 1 \right) \text{ where } x = F_h^{-1}(\alpha_h).$$

By Assumption 3, and noting that $f_l(\underline{x})/f_h(\underline{x}) > 1$,

$$\delta'(0) = qp_h \left(\frac{p_l f_l(\underline{x})}{p_h f_h(\underline{x})} - 1 \right) > 0$$

and

$$\delta'(1) = qp_h \left(\frac{p_l f_l(\overline{x})}{p_h f_h(\overline{x})} - 1 \right) = -\, qp_h < 0.$$

The maximum of $\delta(\alpha_h)$ is therefore interior, satisfying the first-order condition stated in the lemma. The maximum is strict because of the strict concavity of $\delta(\alpha_h)$.

Proof of Lemma 5. The argument is the same as in Lemma 1, but noting that an individual exerting high care is now found liable with the probability $p_h q \alpha_h G(c^*)$; an individual exerting low care is found liable with the probability $p_l q \alpha_l G(c^*)$.

Proof of Proposition 2. A compliance maximizing regime solves

$$\max_{y,\Delta,\alpha_h} y \quad \text{s.t.} \quad y \leq \psi_N(\Delta, \alpha_h) \text{ and } \Delta \leq \varphi_N(y, \alpha_h).$$

The Lagrangian is

$$\mathcal{L} = y + \mu \left[\psi_N(\Delta, \alpha_h) - y \right] + \nu \left[\varphi_N(y, \alpha_h) - \Delta \right]$$

where μ and ν are non negative multipliers. The necessary conditions for an interior maximum $\alpha_h \in (0, 1)$ are

$$\frac{\partial \mathcal{L}}{\partial y} = 1 - \mu + \nu \frac{\partial \varphi_N(y, \alpha_h)}{\partial y} = 0, \tag{18}$$

$$\frac{\partial \mathcal{L}}{\partial \Delta} = \mu \frac{\partial \psi_N(\Delta, \alpha_h)}{\partial \Delta} - \nu = 0, \tag{19}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_h} = \mu \frac{\partial \psi_N(\Delta, \alpha_h)}{\partial \alpha_h} + \nu \frac{\partial \varphi_N(y, \alpha_h)}{\partial \alpha_h} = 0 \qquad (20)$$

Noting that $\partial \psi_N / \partial \Delta > 0$, (18) and (19) imply that both multipliers are strictly positive. It therefore follows from (20) that $\partial \psi_N / \partial \alpha_h$ and $\partial \varphi_N / \partial \alpha_h$ must be of opposite signs. From (15) and (16),

$$sign\left(\frac{\partial \psi_N(\Delta, \alpha_h)}{\partial \alpha_h}\right) = sign\left(\delta'(\alpha_h)\right).$$

The penalty function in Lemma 5 can be rewritten as

$$\varphi_N(y, \alpha_h) = \frac{G(c^*)\lambda(1-\lambda)(1-y)\delta(\alpha_h)}{\pi(\alpha_h, y)\left(1 - \pi(\alpha_h, y)\right)}$$

where $\pi(\alpha_h, y)$ is defined as in (17). It is easily verified that

$$sign\left(\frac{\partial \varphi_N(y, \alpha_h)}{\partial \alpha_h}\right) = sign\left[\delta'(\alpha_h) - \frac{\delta(\alpha_h)\left(1 - 2\pi(y, \alpha_h)\right)}{\pi(\alpha_h, y)\left(1 - \pi(\alpha_h, y)\right)} \frac{\partial \pi(y, \alpha_h)}{\partial \alpha_h}\right]$$

where $\partial \pi / \partial \alpha_h > 0$. In the solution α_h^*, $\partial \psi_N / \partial \alpha_h$ and $\partial \varphi_N / \partial \alpha_h$ are therefore of opposite signs only if

$$\delta'(\alpha_h^*)\left(1 - 2\pi(y_N(\alpha_h^*), \alpha_h^*)\right) \leq 0$$

or equivalently $\delta'(\alpha_h^*) \gtreqless 0$ if $\pi(y_N(\alpha_h^*), \alpha_h^*) \lesseqgtr \frac{1}{2}$. Recalling Lemma 3, that is, $\delta'(\alpha_h) \gtreqless 0$ if $\alpha_h \lesseqgtr \alpha_h^E$, then completes the proof.

Proof of Corollary 2. In the regime α_h, the proportion of individuals found negligent is

$$\begin{aligned}
&\pi(\alpha_h, y_N(\alpha_h)) \\
&\quad = G(c^*)q\left[(\lambda + (1-\lambda)y_N(\alpha_h))p_h\alpha_h + (1-\lambda)(1 - y_N(\alpha_h))p_l\alpha_l(\alpha_h)\right]
\end{aligned}$$

Because $\alpha_h \leq \alpha_l(\alpha_h) \leq 1$, in any regime

$$\begin{aligned}
\pi(\alpha_h, y_N(\alpha_h)) &\leq G(c^*)q\left[(\lambda p_h \alpha_h + (1-\lambda)p_l\alpha_l(\alpha_h)\right] \\
&\leq G(c^*)q\left[(\lambda p_h + (1-\lambda)p_l\right].
\end{aligned}$$

Hence the condition in the corollary ensures that finding negligence is a rare event in the compliance maximizing regime.

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

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

[5] Bernardo, A.E. , Talley, E. and Welch, I. (2000). "A theory of legal presumptions." *Journal of Law, Economics, and Organization* 16, 1-49.

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

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

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

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

[10] 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.

[11] D'Antoni, M. and R. Galbiati (2007), "A Signalling Theory of Non-monetary Sanctions." *International Review of Law and Economics* 27, 204-218.

[12] Daughety, A. and J. Reinganum (1999), "Hush Money."*RAND Journal of Economics* 30, 661-678.

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

[14] Deffains, B. and C. Fluet (2013), "Legal Liability when Individuals Have Moral Concerns." *Journal of Law, Economics, and Organization*, forthcoming.

[15] Demougin, D. and C. Fluet (2005), "Deterrence versus judicial error: a comparative view of standards of proof." *Journal of Institutional and Theoretical Economics* 161(2), 193-206.

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

[17] Demougin, D. and C. Fluet (2008), "Rules of proof, courts, and incentives." *RAND Journal of Economics* 39, 20-40.

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

[19] Frey, B. and R. Jegen (2001), "Motivation Crowding Out Theory." *Journal of Economic Surveys* 15, 589-611.

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

[21] 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.

[22] Hay, B.L. and K.E. Spier (1997), "Burdens of Proof in Civil Litigation: An Economic Perspective" *Journal of Legal Studies* 26, 413–431.

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

[24] Kaplow, L. (2011). On the Optimal Burden of Proof. *Journal of Political Economy* 119, 1004-1040.

[25] 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.

[26] 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, New York: North-Holland.

[27] Polinsky, M.A. and S. Shavell (1989), "Legal error, litigation, and the incentive to obey the law." *Journal of Law, Economics and Organization* 5, 99–108.

[28] 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.

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

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

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

[32] Schauer, F. and Zeckhauser, R. (1996), “On the Degree of Confidence for Adverse Decisions.” *Journal of Legal Studies* 25, 27–52.

[33] Shavell, S. (1984), “A Model of the Optimal Use of Liability and Safety Regulation.” *RAND Journal of Economics* 15, 271-280.

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

[35] Shavell, S. (2012), “When is Complying with the Law Socially Desirable?” *Journal of Legal Studies* 41, 1-36.

[36] Shin, H.S. (1998), “Adversarial and inquisitorial procedures in arbitration.” *RAND Journal of Economics*, 29, 378-405.

[37] Smith, A. (1759), *The Theory of Moral Sentiments*. Reedited (1997), Washington, D.C., Regnery Publishing.

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

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