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

Cahier de recherche/Working Paper **09-51**

Legal Liability when Individuals Have Moral Concerns

Bruno Deffains
Claude Fluet

Novembre /November 2009

Deffains: University Paris 2 and CNRS
bdeffains@u-paris10.fr
Fluet: Université du Québec à Montréal and CIRPÉE
fluet.claude-denys@uqam.ca

We thank Andrew Daughety, Lewis Kornhauser, William Neilson, Stéphane Pallage, Jennifer Reinganum and Veikko Thiele for helpful comments and suggestions. Funding from GIP "Droit et Justice" (France), FQRSC (Quebec) and SSHRC (Canada) is gratefully acknowledged.

Abstract:
We incorporate normative motivations into the economic model of accidents and tort rules. The social norm is that one should avoid harming others and should compensate if nevertheless harm is caused. To some extent, this is internalized through intrinsic moral concerns; moreover, those thought not to adhere to the norm are met with social disapproval. Moral and reputational concerns are not strong enough, however, for injurers to willingly compensate their victims. Absent legal liability, normative concerns induce precautions to prevent harm but precautions are then socially inefficient. By contrast, perfectly enforced legal liability crowds out informal incentives completely (e.g., individuals causing harm suffer no stigma) but precautions are then socially efficient. Under imperfectly enforced legal liability, formal legal sanctions and normative concerns are complements and interact to induce more precautions than under no-liability.

Keywords: Intrinsic motivations, social norms, esteem, strict liability, negligence, crowding out

JEL Classification: D8, K4, Z13

1 Introduction

Legal liability induces precautions to prevent accidental harm to third parties. In the economic model of tort rules, incentives to exercise care are purely 'external' and reduce to the 'implicit prices' set by legal sanctions.[1] Casual observation suggests that other motivations are usually also at work. Most people exercise some care out of moral considerations. As noted by Shavell (2004): "There are two fundamental reasons why individuals will often want to obey moral notions... One is that individuals have internal incentives to do so, namely, they feel virtuous if they adhere to them, and experience guilt if they do not. Second, individuals have external incentives to obey moral notions in that they will be praised by others for that behavior and admonished, scold, or otherwise punished for immoral behavior". In this paper, we incorporate moral concerns into the economic model of accidents and legal liability.

Other-regarding motivations, normative incentives and the like come in many forms. The recent economic literature, not to mention psychology or sociology, offers an abundant menu of notions: altruism and warm glow, status, fairness, inequity aversion, esteem and self-esteem, reciprocity, aversion to norm-breaking, to name only a few.[2] We consider the role of a social or moral norm that is particularly relevant in a tort context. Kaplow and Shavell (2002) remark that there is a strong social norm to avoid harming others and to compensate for the harm that one does cause. We take the existence of such a norm as given. Individuals feel guilt when they do not abide by it, but some individuals are intrinsically more morally concerned than others. Moreover, individuals earn social esteem if they are thought to have high moral concerns and suffer social disapproval if not.

The natural question is then how normative incentives interact with formal legal sanctions to influence behavior (see Posner, 2000, Kaplow and

[1] The basic model is due to Brown (1973) and has been developed, in particular, in Landes and Posner (1987) and Shavell (1987).

[2] See for instance Andreoni (1989, 1990), Glazer and Konrad (1996), Fehr and Schmidt (1999), Brennan and Pettit (2005), Bénabou and Tirole (2003, 2006), Frey (1997), Rabin (1993), Lopez-Perez (2008).

Shavell, 2007, and for a recent survey McAdams and Rasmusen, 2007). The issue is often formulated in terms of whether law and normative incentives are substitutes or complements.[3] On a more general level, the recent literature on informal incentives has much emphasized the possibility that extrinsic incentives crowd out intrinsic motivations (see the survey by Frey and Jegen, 2001). We contribute to this debate by discussing a framework with the following properties. First, there is a well defined notion of what constitutes socially efficient behavior. In our analysis, this is simply the efficient level of care of the standard economic model of accidents. Secondly, the explicit monetary incentives that we consider stem from the legal rules known to be efficient in the standard model where individuals have no normative concerns. Finally, the normative concerns that we consider have a natural role to play in the context of accidental harm.

We share with a recent strand of the literature, notably Bénabou and Tirole (2006), the idea that an individual's actions may signal something about his 'moral type'.[4] In our analysis, moral type refers to the extent to which one adheres to the social norm. Various actions or their consequences conceivably constitute signals about one's moral type: whether or not the individual engages in an activity that may cause harm; if he does, the extent to which he takes precautions to prevent harm; if precautions are not directly observable, the mere occurrence of harm may suggest low care, thereby indirectly signalling moral type; when harm is caused, the injurer could also go so far as to spontaneously compensate the victim. We assume, however, that preferences are such that moral and reputational concerns are not strong enough for injurers to willingly compensate their victims.[5] This

[3]For instance Bohnet et al. (2001), Lazzarini et al. (2004) and Zasu (2007).

[4]Signaling models in a similar vein are found in Bernheim (1994), Corneo (1997) and, closer to to Bénabou and Tirole's framework, Daughety and Reinganum (2009). The basic idea has of course often been expressed informally. See also Sen (1975) for interesting insights.

[5]This is not to deny that spontaneous compensation is often oberved. To quote the 'other' Adam Smith: "The person...who...has involuntary hurt another...naturally runs up to the sufferer to express his concern for what has happened. If he has any sensibility, he necessarily desires to compensate the damage." (Smith, 1790 [1976], p. 104).

means that our individuals are only 'imperfectly' morally concerned; in addition, their desire for social esteem is not strong enough for them to attempt to be perceived as 'perfectly' moral.

Three environments are considered: no-liability, perfectly enforced legal liability (either strict liability or the negligence rule) and imperfectly enforced legal liability. Absent legal liability, normative concerns induce some care to prevent harm but the precaution levels are then socially inefficient. By contrast, perfectly enforced legal liability crowds out informal incentives completely — e.g., individuals who caused harm suffer no social disapproval — but precautions are then socially efficient. Under imperfectly enforced legal liability (e.g., victims do not always sue), however, formal legal sanctions and normative concerns are complements and interact to induce more precautions than under no-liability. Although there is motivational crowding-out, there is no net crowding-out with respect to overall incentives. We complete the analysis with a welfare comparison of the different legal regimes — no-liability, strict liability, and the negligence rule. In particular, we discuss the extent to which the legal rules are consistent with the underlying social norm.

Section 2 presents the basic setup. The sections 3 and 4 discuss respectively the no-liability and liability regimes. Section 5 presents the welfare comparison. Section 6 concludes.

2 The model

Our starting point is the unilateral accident model of the law and economics literature. Some individuals, hereafter injurers, have access to an activity that provides a private benefit to themselves but imposes a risk of harm to others. Precautions reduce this risk but are costly. Given risk neutrality, the socially optimal precautions minimize the sum of precaution costs to the injurer and of the expected harm to victims. Absent legal liability, however, injurers would disregard the negative externality they generate.[6]

[6]See Shavell (2007) for a recent survey. In Shavell's terminology, injurers and victims are "strangers to one another", which rules out contractual agreements to prevent or

Setup. We depart from the basic model by introducing normative motivations. Society holds that harming others should be avoided; if one nevertheless causes harm, one should compensate the victim. To some extent, this social norm is internalized through intrinsic 'moral concerns': individuals suffer moral disutility (e.g., guilt) if they do not comply with the social norm. Moreover, those thought to have weak moral concerns are met with disapproval. If one's actions suggest a disregard for the social norm, the fear of social disapproval may then provide incentives. We refer to the latter source of incentives as 'reputational concerns'. Both moral and reputational concerns constitute normative motivations in the sense that they derive from one's allegiance to the social norm or one's attempt to signal allegiance to the norm.

In our version of the unilateral accident model, an injurer is concerned with his own material payoff, with compliance with the social norm and with social esteem. His utility function is

$$U(y, x, \delta, \theta_s; \theta) = y - \theta\delta x + \beta\theta_s. \tag{1}$$

The first term, y, is the injurer's pecuniary payoff; the second term, $\theta\delta x$, is moral disutility; the third, $\beta\theta_s$, is the utility from social esteem.

Moral disutility arises from not complying with the social norm, i.e., one has caused harm and has not compensated the harm: x is the amount of harm, $\delta \in \{0, 1\}$ is the action of compensating ($\delta = 0$) versus not compensating ($\delta = 1$) and θ is a parameter that captures the degree to which the individual adheres to the social norm. When the individual is responsible for uncompensated harm, the moral disutility is greater the greater the severity of the harm and the more one is morally concerned. The larger θ, the more is the individual willing to sacrifice private gain in order to comply with the social norm. Perfect adherence to the social norm would correspond to $\theta = 1$. The 'perfect' individual would in effect treat harm caused to others as his own and would always be willing to compensate.[7]

mitigate harm.

[7] Compliance with the social norm requires full compensation and injurers are assumed to have sufficient wealth to comply.

An individual's θ is private information and will be referred to as his type. Social esteem depends on one's perceived type or social image θ_s. This will depend on what information is publicly available about the individual and may well differ from the individual's true moral type. Someone thought to care much about harming others earns social approval which provides utility, i.e., the parameter β is positive.

The risk generating activity produces a private benefit of amount b and imposes on others a loss of amount L with probability p. The probability of accident depends on the injurer's precautions. To economize on notation, we use the probability of accident itself to describe precautionary behavior. A smaller p means more precautions. The cost of precautions is $c(p)$ with $c' < 0$ and $c'' > 0$. At the boundaries, $c(1) = c'(1) = 0$ and $c'(0) = -\infty$, i.e., the marginal precaution cost is nil at the no-precaution level $p = 1$ but totally eliminating the risk of harm is prohibitively costly. The net income of a potential injurer who does not engage in the activity is normalized to zero.

In the standard model, θ and β are zero. Absent legal liability, an injurer's utility is then $U = b - c(p)$ and is maximized by taking no precautions. Social welfare, on the other hand, takes into account the loss suffered by victims and is therefore $W = b - c(p) - pL$. The socially efficient precaution level p^* minimizes $c(p) + pL$, the sum of precaution and expected accident costs. We will also refer to p^* as the efficient precaution level; its welfare significance when injurers have moral and reputational concerns is discussed further in section 5. In the standard model, it is socially efficient for potential injurers to engage in the risk generating activity if $W = b - c(p^*) - p^*L > 0$. To simplify, the gross benefit b is taken to be sufficiently large that exercising the activity is always socially warranted, even when injurers take no precautions. Thus, we assume $b > L$.

For future use, let

$$h(k) = \arg\min_p c(p) + kp, \text{ where } k \geq 0.$$

It is easily seen that $h(k) > 0$ and is strictly decreasing, with $h(0) = 1$. In terms of this function, the socially efficient precaution level is $p^* = h(L)$.

No spontaneous compensation. All injurers care equally about social esteem, i.e., they have the same parameter $\beta > 0$. The parameter θ is distributed according to the density $f(\theta)$ with support $[0, \theta_m]$ and mean value denoted by $\overline{\theta}$. We impose the restriction:

ASSUMPTION 1: $\theta_m < L/(\beta + L)$.

The condition defines an upper bound on the extent to which preferences depart from that of the standard model. Injurers put some weight on complying with the social norm, but they put greater weight on their own material payoff, i.e., $\theta < 1$ for all types. Moreover, their moral concerns and their desire for social esteem cannot simultaneously be too large. This rules out some forms of signaling behavior.

To see this, suppose that causing harm is public information and that there is no legal liability. Nevertheless, there is nothing to prevent injurers from willingly compensating their victims. They will not do so on purely moral grounds because the 'bad feelings' from not complying with the social norm is less painful to them than the money cost of compensating. But the action of compensating one's victim could provide a sufficient reputational benefit to make it worthwhile.

Let y be the injurer's *ex post* payoff and suppose harm has been caused. If a type-θ injurer does not compensate (action 1), $\delta = 1$ in (1) and the individual's utility is $U_1 = y - \theta L + \beta\theta_1$ where θ_1 is the social image (or updated expected type) of injurers who do not compensate. If he compensates (action 0), $\delta = 0$, i.e., the individual eliminates the moral disutility from not complying with the social norm. Because he thereby also transfers the victim's loss to himself, his utility is $U_0 = y - L + \beta\theta_0$ where θ_0 is the social image of injurers who compensate. The injurer will choose not to compensate his victim if $U_0 < U_1$, that is if

$$\beta\,(\theta_0 - \theta_1) < (1 - \theta)L. \tag{2}$$

The posterior beliefs about one's type must belong to the interval $[0, \theta_m]$, so that $\theta_0 - \theta_1 \leq \theta_m$. Assumption 1, which can be rewritten as $\beta\theta_m < (1-\theta_m)L$, implies that (2) always holds. Thus, we have:

Lemma 1 *At equilibrium, injurers do not voluntarily compensate their victims.*

Although our injurers differ from standard *homo economicus*, they behave the same following the occurrence of harm. By contrast, if injurers were to voluntarily compensate their victims, they would impose upon themselves the same penalties as under the strict liability legal regime. Anticipating this, they would therefore exert the same care to prevent harm as under strict liability.[8]

Assumption 1 also ensures that all types of potential injurers engage in the risk generating activity. Suppose that whether or not one is engaged in the activity is public information. If a type-θ injurer does not engage in the activity, his net income is zero and he inflicts no harm. His utility is $U_0 = \beta\theta_0$ where θ_0 is now the perceived type of those who do not engage. For simplicity, suppose that those who engage take no precautions so that they always cause harm (i.e., $p = 1$ and precaution costs are zero). Because he does not compensate, the type-θ injurer has utility $U_1 = b - \theta L + \beta\theta_1$ where θ_1 is the perceived type of those who engage. Given $b > L$,

$$U_1 = b - \theta L + \beta\theta_1 > (1 - \theta)L + \beta\theta_1 > \beta\theta_0 = U_0,$$

where the second inequality is the same as (2). Hence, all types engage. The argument extends to the case where those who engage prefer to take some precautions (see section 3).

Under a tort regime, an injurer may be legally obligated to compensate the victim. We take it that forced compensation has the same effect on moral disutility as if it had been voluntary. When compensation is made, albeit unwillingly, an injurer feels he has complied with the social norm.

[8] In the absence of further restrictions on the distribution of types, assumption 1 is necessary to rule out voluntary compensation. For instance, suppose there are only two types, $\theta = 0$ and $\theta = \theta_m$ and that assumption 1 does not hold. If $\beta\theta_m < L$, there exists a separating equilibrium (with $\theta_0 = \theta_m$ and $\theta_1 = 0$) where the high type voluntarily compensates and the low type does not. If $\beta\theta_m > L$ and the proportion of high types is sufficiently large, there exists a pooling equilibrium where both types compensate.

Posterior information. Throughout the paper, with a qualification in the case of negligence based liability, an injurer's precautions are private information. Precautions can affect one's reputation only through the occurrence (or non occurrence) of harm. This is in line with the view that tort law is an *ex post* harm-based mechanism for deterring undesirable behavior, by contrast with an *ex ante* act-based approach as with safety regulations.[9] When the tort regime is negligence, we will assume that some evidence about the injurer's precautions becomes available following the occurrence of harm and that courts are able to assess whether the injurer complied with the legal due care standard.

In our basic scenario, *ex post* public information about an injurer will take the form of a binary signal with outcome B or G. The notation is B for 'bad news' (i.e., unfavorable information) and G for 'good news'. The interpretation of these events will depend on the context. For instance, B may be "injurer has caused harm" or "injurer has caused harm and has been found negligent (hence is held legally liable under a negligence rule)".

The probability of these events depends on the injurer's precautions; we denote with $\varphi(p)$ the probability of B. The injurer's perceived type will be conditional on whether B or G occurred. Denoting society's posterior beliefs about one's type by θ_B and θ_G, an injurer's expected perceived type, as a function of his precaution level, is

$$\overline{\theta}_s(p) = \varphi(p)\theta_B + (1 - \varphi(p))\theta_G.$$

A general formulation for the expected utility of a type-θ injurer is then

$$\overline{U} = \overline{y}(p) - \theta\overline{x}(p) + \beta\left[\varphi(p)\theta_B + (1 - \varphi(p))\theta_G\right], \tag{3}$$

where $\overline{y}(p)$ is the injurer's expected material payoff and $\overline{x}(p)$ is the expected uncompensated harm for which he will be 'morally responsible'. The expectations are written as a function of the injurer's precautions. Expectations

[9] On the distinction between harm-based and act-based deterrence, see Shavell (1993). In a related context, Daughety and Reinganum (2009) analyze the effect of privacy versus publicity about one's actions.

depend on the legal regime because it determines whether it is the victim or the injurer who ultimately bears the accidental loss.

Society's posterior beliefs are part of an equilibrium. The requirements are as follows.

DEFINITION 1: *Let the injurers' expected utility be as in (3). A perfect Bayesian equilibrium consists of strategies* $p^e(\theta)$ *and of beliefs* θ_B^e *and* θ_G^e *such that*

(i) $p^e(\theta)$ *maximizes* $\overline{U} = \overline{y}(p) - \theta\overline{x}(p) + \beta[\varphi(p)\theta_B^e + (1-\varphi(p))\theta_G^e]$, $\theta \in [0, \theta_m]$;

(ii) if $\overline{\varphi}_B \equiv \int_0^{\theta_m} \varphi(p(\theta)) f(\theta)\, d\theta > 0$, *then*

$$\theta_B^e = \frac{\int_0^{\theta_m} \theta\varphi(p(\theta)) f(\theta)\, d\theta}{\overline{\varphi}_B}. \tag{4}$$

(iii) if $1 - \overline{\varphi}_B > 0$,

$$\theta_G^e = \frac{\int_0^{\theta_m} \theta[1 - \varphi(p(\theta))] f(\theta)\, d\theta}{1 - \overline{\varphi}_B}; \tag{5}$$

Beliefs satisfy Bayesian up-dating when the conditioning events have positive probability over the population of injurers. When a conditioning event has zero probability (e.g., when bad news never occur), the posterior belief is to some extent arbitrary, but must be consistent with (i.e., "support") the equilibrium strategies.

Of particular interest is $\Delta \equiv \theta_G^e - \theta_B^e$, the gap in social image between good and bad news, which we will refer to as the *reputational penalty*. When both B and G have positive probability, (5) and (4) can be combined to yield

$$\Delta = -\frac{\int_0^{\theta_m} \theta[\varphi(p^e(\theta)) - \overline{\varphi}_B] f(\theta)\, d\theta}{\overline{\varphi}_B(1 - \overline{\varphi}_B)}. \tag{6}$$

The integral in the numerator is the covariance between θ and $\varphi(p^e(\theta))$, a negative quantity when $\varphi(p^e(\theta))$ is decreasing in θ. Reputational concerns will provide incentives through the reputational penalty attached to bad news.

While injurers are not willing to compensate their victims *ex post*, they will want to take precautions *ex ante* to prevent the occurrence of harm. We

first consider the case where harm is not subject to legal liability. Incentives to take precautions then rely purely on moral and reputational concerns. Next, we introduce legal liability and examine how this combines with informal incentives.

3 No Liability

Causing harm often does not trigger legal liability. For instance, the harm is not subject to judicial sanction because it is trivial or part of the usual risks of life. Even when legal liability applies in principle, victims do not necessarily file suit. Judicial procedures may be too expensive compared to the stakes or there may not be enough evidence. Moreover, injurers are not always detected, e.g., damages to one's car in a parking lot.

No publicity benchmark. We start with the case where an injurer causing harm is never detected. Perhaps the victim knows the injurer, but this is not "public" information. Alternatively, the occurrence of harm is commonly observable but the identity of the injurer is unknown. In either case, there is no public information about the injurer's involvement in causing harm.

A type θ injurer then chooses his precaution level p to maximize

$$\overline{U} = b - c(p) - \theta pL + \beta\overline{\theta},$$

where $\overline{\theta}$ is the prior mean of types in the population. Because no information about the injurer is made public, his social image is given by the prior mean. Using the function defined in the previous section, the type-θ injurer chooses $p = h(\theta L)$. This is greater than the efficient p^* but less than unity: some precautions are taken because of moral concerns. Let $\overline{p}_M$ denote the average probability of harm in the no-publicity environment.[10]

Publicity. Consider now the case where the occurrence of harm and the causal relation to the injurer can become public information. Specifically,

[10] That is, $\overline{p}_M = \int_0^{\theta_m} h(\theta L) f(\theta)\, d\theta$.

suppose that, following an accident, the injurer's involvement is publicly known with probability $q > 0$.

Denote by B the event "an occurrence of harm is ascribed to the injurer". In other words, an accident has occurred and the injurer's involvement is common knowledge. In terms of the notation of the previous section, event B has probability $\varphi(p) = pq$. The event G is "no occurrence of harm is ascribed to the injurer", meaning that there is no information concerning the injurer's involvement in an accident. This event has probability

$$1 - p + p(1 - q) = 1 - pq.$$

Either harm has not occurred or it has occurred but has not been observed by society at large or has been observed but not related to the particular injurer.

The expected utility of a type-θ injurer is now

$$\overline{U} = b - c(p) - \theta pL + \beta[pq\theta_B + (1 - pq)\theta_G]. \tag{7}$$

This can be rewritten as

$$\overline{U} = b + \beta\theta_G - c(p) - p(\theta L + \beta q\Delta), \text{ where } \Delta \equiv \theta_G - \theta_B. \tag{8}$$

In equilibrium, as shown below, society's beliefs will satisfy $\theta_B < \overline{\theta} < \theta_G$. An injurer is stigmatized by event B while event G provides social esteem. Given the reputational penalty, the best-response function of a type-θ individual is

$$p_Z(\theta, \Delta) = h(\theta L + \beta q\Delta), \tag{9}$$

where the subscript stands for 'zero liability'. Compared to the no publicity case, reputational concerns now provide incentives.

Next we look for the equilibrium reputational penalty. It is obtained by substituting the injurers' best response functions (9) in lieu of $p^e(\theta)$ in the right-hand side of (6) and solving the resulting equation for Δ. Recalling that $\varphi(p) = qp$, define

$$\psi_Z(\Delta) \equiv -\frac{\int_0^{\theta_m} \theta[p_Z(\theta, \Delta) - \overline{p}_Z(\Delta)]f(\theta)\, d\theta}{\overline{p}_Z(\Delta)\,(1 - q\overline{p}_Z(\Delta))}, \tag{10}$$

where $\overline{p}_Z(\Delta)$ is the average best response over all types. At equilibrium, $\Delta = \psi_Z(\Delta)$.

Lemma 2 *Any solution to* $\Delta = \psi_Z(\Delta)$ *belongs to* $(0, \theta_m)$ *and there is at least one satisfying* $\psi'_Z(\Delta) < 1$.

The condition $\psi'_Z(\Delta) < 1$ characterizes a stable equilibrium. To see this, write $\Delta_s = \psi_Z(\Delta)$, where Δ_s is the reputational penalty consistent with the injurers' behavior and Δ is the penalty anticipated by injurers, perhaps erroneously. The anticipated penalty determines the injurers' precautions; in turn, the injurers' behavior determines society's posterior beliefs, which yields Δ_s. At equilibrium $\Delta_s = \Delta$. Suppose now that the penalty anticipated by injurers receives a small positive shock. Injurers will increase their precautions, which will change society's beliefs contingent on the events good or bad news. When $\psi'_Z(\Delta) \geq 1$, a new equilibrium will be reached where precautions remain higher than before the shock even after injurers have learned society's true beliefs.

`-- Figure 1 about here --`

We discard unstable equilibria, should they exist (as with point Q in figure 1). There remains the possibility that there is more than one stable equilibrium (say, E_1 and E_2 in the figure). While this raises interesting issues, it is not our main concern. In what follows, we loosely refer to "the" equilibrium; should there be multiple equilibria, we focus on the one with the largest reputational penalty.[11] The next proposition summarizes our results for the no-liability regime.

Proposition 1 *Under no-liability, individuals known to have caused harm are stigmatized and both moral and reputational concerns mitigate carelessness. All injurers exert less care than the socially efficient level. They exert greater care the more likely the publicity about involvement in causing harm.*

From (9), reputational incentives depend on the probability of publicity times the reputational penalty. The reputational penalty is itself a function

[11] It is well kown that social interaction models may exhibit multiple equilibria. See for instance Rasmusen (1996), Glaeser et al. (1996) or Bénabou and Tirole (2006).

of the probability of publicity and can be written as $\Delta(q)$. The proposition therefore states that $q\Delta(q)$ is increasing in q.[12] Nevertheless, moral and reputational concerns are never strong enough to induce efficient care, even when causing harm is always detected.

A greater likelihood of publicity need not increase the reputational penalty itself. One can show that $\Delta(q)$ is locally decreasing when ψ'_Z is negative and increasing when the sign is positive (as at E_1 and E_2 in figure 1). The effect of greater publicity on the reputational penalty depends on whether precautions are strategic substitutes or complements. In particular, precautions are strategic substitutes when accidents occur often and injurers are detected with sufficiently high probability. A greater probability of detection then means that bad news become even more banal. This reduces the reputational penalty, which is not to say that incentives to exert care are reduced.[13]

4 Liability

The tort rules are strict liability and negligence. Under a strict liability regime, injurers are liable for full compensatory damages irrespective of the precautions they have taken. Victims only need to prove causation. Under a negligence regime, they also need to prove the injurer's carelessness, i.e., that precautions did not meet the legal due care standard.[14]

We assume that a lawsuit imposes a small cost on plaintiffs, so that they sue only if it is worthwhile, but we otherwise disregard litigation costs. Non negligible litigation costs would impact on the socially efficient precaution

[12] This comparative static result obtains only at stable equilibria.

[13] Klement and Harel (2007) point out some limitations to the usefulness of stigma as a tool in crime prevention, e.g., in the context of shaming penalties. In their analysis, stigmatization effects may decrease as more crime are detected, so that deterrence may be reduced. See also Rasmusen (1996). In our analysis, the effect of greater publicity on the reputational penalty can go either way because it depends on the inferences drawn at equilibrium about an injurer's type; however, it cannot decrease to the point that better detection would reduce deterrence.

[14] Fault or negligence is the usual basis of liability.

level, an issue we want to abstract from (see Shavell, 2007). They would also give incentives to settle before trial. Even without such costs, however, injurers could favor confidential settlements — possibly allowing victims to extract 'hush money' — if stigmatization effects can thereby be avoided.[15] There are also many possibilities regarding what information is publicly available. For example, involvement in causing harm may be public information irrespective of lawsuits; alternatively, it may become known only if victims file suit or if causation is proved at trial. When causing harm is public information and no suit is filed, it may or may not be known whether victim and injurer reached an agreement; perhaps the victim did not pursue the matter because he did not possess sufficient evidence to prove negligence, which does not necessarily mean that the injurer actually complied with due care. For simplicity, we consider a litigation subgame where *ex post* public information effectively reduces to a binary outcome, as in the previous section. Extensions are discussed as we go on.

The setup is as follows. The occurrence of harm and the identity of the injurer are initially known only to the victim. The victim as plaintiff has the burden of proof if the case goes to trial. Following the occurrence of harm, a victim knows for sure whether he has sufficient evidence to succeed in court or whether he does not; this is common knowledge between the parties. Under the strict liability rule, either the victim has evidence demonstrating causation or he has no evidence at all. Under the negligence rule, the evidence comes in a 'bundle': if harm occurred and the injurer did not comply with due care, either the victim has evidence demonstrating this or he has no evidence whatsoever; in all other cases, the victim has nothing to show. Finally, secret settlements are not feasible: if payment has been extracted from an injurer, information is always leaked and it becomes publicly known that an agreement was reached. The filing of a lawsuit is public information. The outcome at trial, i.e., whether or not the defendant is held liable, is of course also public information.

[15] Confidential settlements by producers, to avoid sequential suits when there are many potential plaintiffs or to exploit consumer ignorance about the safety of a product, have been extensively analyzed by Daughety and Reinganum (1999, 2002, 2005).

In this simple framework, victims with a non viable case do nothing. Victims with a viable case file suit and pursue the case up to trial; equivalently, they settle for the amount of damages they would have obtained in court. In either case, the reputational effect on the injurer is the same. An out-of-court settlement imposes the same reputational penalty because settlements are common knowledge and the injurer would not have offered payment if the victim had no evidence; since the reputational penalty is the same, the injurer will not pay more than the damages he would have paid if the case had gone to trial (and the victim would not accept less). Conversely, if the victim does not have a viable case, filing suit is not worth the small filing cost: the case would not succeed in court and no settlement will be forthcoming, which in either case demonstrates that the plaintiff had nothing to show when he filed suit.

Perfectly enforced legal rules. Perfect enforcement corresponds to the elementary version of the economic model of torts. Under *strict liability*, the occurrence of harm and causation (hence the identity of the injurer) can always be proved and victims always sue. Injurers therefore know that, should they cause harm, they will have to pay compensatory damages. One's involvement in causing harm is then always public information.

The bad news event B is "injurer was sued, hence must have caused harm, and was (or would have been) found liable". This has probability p. The complementary event G is "injurer was not sued, hence did not cause harm". Injurers are forced to compensate their victim. Hence, although unwillingly, they comply with the underlying social norm. As a result, they suffer no moral disutility but bear the victims' losses. The expected utility of a type-θ injurer is therefore

$$\overline{U} = b - c(p) - pL + \beta[p\theta_B + (1-p)\theta_G].$$

Expected utility does not depend on the injurers' type and best responses are therefore type independent. It follows (see definition 1) that the events G and B do not provide information about type, i.e., at equilibrium $\theta^e_B = \theta^e_G = \overline{\theta}$. There is no reputational penalty from causing harm and all injurers choose the precaution level p^* minimizing $c(p) + pL$.

Consider now the *negligence* rule. As in the standard model, the legal due care is taken to be the socially efficient level of precaution. An injurer is found negligent if the evidence shows that his precautionary behavior entailed $p > p^*$. Under perfect enforcement, the victim always has access to such evidence. Hence, injurers who do not comply with due care and cause harm always compensate their victims and this is publicly known. Should the case go to trial, we assume that the general public does not learn the actual level of care of an injurer found negligent; it only knows that precautions did not satisfy due care. For instance, the general public only pays attention to the trial outcome and has no time for the details; equivalently, the evidence submitted in court is "crude" and allows only to prove that $p > p^*$. Thus, the bad news event B is "injurer has been sued and was (or would have been) found negligent". Event G is the complementary event "injurer has not been sued", which means that either the injurer was not involved in causing harm or that he caused harm but complied with due care.

A type-θ injurer now has the expected utility

$$\overline{U} = \begin{cases} b - c(p) - \theta pL + \beta\theta_G \text{ if } p \leq p^*, \\ b - c(p) - pL + \beta[p\theta_B + (1-p)\theta_G] \text{ if } p > p^*. \end{cases} \tag{11}$$

The upper branch with $p \leq p^*$ is the expected utility when the precaution level satisfies due care. With probability one, the injurer's social image will then be θ_G. With probability p, the injurer will nevertheless cause noncompensated harm, which yields moral disutility. The lower branch with $p > p^*$ is for an injurer who does not comply with due care. With probability $1-p$, harm will not occur and the social image will again be θ_G. With probability p, the injurer will be sued and will pay damages L. He then suffers no moral disutility from having caused harm, but his social image is θ_B.

Consider first the injurer's best precaution level consistent with meeting the due care standard. Because $c(p) + pL$ is strictly convex and is minimized at p^*,

$$\frac{d\overline{U}}{dp} = -c'(p) - \theta L > -c'(p) - L \geq 0, \text{ for all } p \leq p^*.$$

Thus, precautions will never exceed due care. For precaution levels that do

not satisfy due care, and supposing that $\theta_B \leq \theta_G$,

$$\frac{d\overline{U}}{dp} = -c'(p) - L - \beta(\theta_G - \theta_B) < 0, \text{ for all } p > p^*.$$

Combining both results, the utility maximizing precaution level is therefore p^*, so that the injurer will never be found negligent. Note the discontinuity in the injurer's payoff when he decides to barely comply with due care rather than not comply (see figure 2).

-- Figure 2 about here --

Let $\varphi(p) = 0$ if $p \leq p^*$ and $\varphi(p) = p$ if $p > p^*$. The expected utility in (11) can then be rewritten as

$$\overline{U} = b - c(p) - \varphi(p)L - \theta(p - \varphi(p))L + \beta[\varphi(p)\theta_B + (1 - \varphi(p))\theta_G], \quad (12)$$

so that definition 1 can be applied directly. Because all injurers exercise due care, $\theta_G^e = \overline{\theta}$. The bad news event B never occurs, so that θ_B is an out-of-equilibrium belief. From the above argument, any $\theta_B \leq \theta_G^e = \overline{\theta}$ supports the equilibrium. The next proposition summarizes our results for perfectly enforced legal liability.

Proposition 2 *Suppose liability rules are perfectly enforced. Then, (i) under strict liability all injurers exert efficient care and injurers sued or found liable suffer no stigma; (ii) under the negligence rule with due care set at the efficient level, all injurers comply with due care and not having been sued or found negligent confers no prestige. Under either rule, moral and reputational concerns play no role in providing incentives to exercise care.*

When formal legal sanctions are introduced and enforcement is perfect, moral concerns either disappear (under strict liability) or are superfluous (under the negligence rule); reputational concerns play no role. Although there is 'motivational crowding-out', there is no net crowding out effect because all injurers take more precautions than under no liability.[16]

[16] Our results also contrast with Cooter and Porat (2001) where the issue is whether courts should deduct "nonlegal sanctions" from legal damages to avoid overdeterrence.

Our assumptions about the litigation subgame can be modified in many respects without affecting the result. For instance, it would not matter if secret settlements were allowed. Neither does it matter if involvement in causing harm is observed independently of lawsuits.

Imperfectly enforced strict liability. Victims now do not necessarily file suit because they cannot always prove causation. Let k denote the probability that, following the occurrence of harm, a victim has access to sufficient evidence. Injurers then know that, should they cause harm, with probability k they will pay damages and be known to have caused harm.

For a type-θ injurer, expected utility is

$$\overline{U} = b - c(p) - pkL - \theta p(1-k)L + \beta[p\theta_B + (1-pk)\theta_G],$$

The bad news event B, now with probability pk, has the same interpretation as under perfect enforcement. The event G is "injurer was not sued, hence did not cause harm or caused harm but there was no evidence to prove it". The best response function is

$$p_S(\theta, \Delta) = h[(k + \theta(1-k))L + \beta k\Delta], \tag{13}$$

where the subscript stands for strict liability. An injurer's precautions depend on his type, provided the probability of enforcement k is less than unity. The analysis is similar to that of no-liability. Being sued (and settling or being held liable if the case goes to trial) now imposes a reputational penalty. The equilibrium reputational penalty solves an equation such as (10) but with $p_S(\cdot)$ substituted for $p_Z(\cdot)$.

A natural comparison is with no-liability for the same probability of publicity about involvement in causing harm. From (9), a type-θ injurer's incentives to exert care under no-liability are given by $\theta L + \beta q \Delta_Z$, where Δ_Z is the equilibrium reputational penalty under no-liability. From (13), the injurer's incentives under strict liability are $(k + \theta(1-k))L + \beta q \Delta_S$, where Δ_S is the equilibrium reputational penalty under strict liability. When $k = q$, moral incentives are lower under strict liability because injurers will sometimes be forced to pay damages, but this is more than compensated by the expected

legal damages. However, reputational incentives are likely to be weaker, i.e., Δ_S may be smaller than Δ_Z.[17] Legal liability has a greater effect on the incentives of injurers with low moral concerns (a small θ) than on those with strong moral concerns. If precautions become more alike between types, the bad news-good news signal will be less informative about moral type. Overall, incentives to exert care are nevertheless greater under the strict liability regime, as shown in the next proposition.

Proposition 3 *When strict liability is imperfectly enforced, injurers found liable are stigmatized and both moral and reputational concerns mitigate carelessness. All injurers exert greater care than they would under no-liability with the same (or a smaller) probability of observing involvement in causing harm.*

If under no-liability injurers are "detected" with probability q, one would expect a move to a strict liability regime with negligible litigation costs to yield a probability of enforcement at least as large, i.e., $k \geq q$. This seems reasonable to the extent that the factors conducing to common knowledge about involvement in causing harm under no-liability would also entail sufficient evidence to prove causation. The probability of enforcement k may be strictly larger than q because victims have monetary incentives to "go public" under a strict liability regime, which they do not under no-liability. Note that it does not matter if involvement in causing harm can become public information independently of lawsuits.

Two remarks are in order. First, by contrast with the result for no-liability in proposition 1, the level of care under strict liability need not be monotonically increasing in k. In other words, a small increase in the probability of enforcement may reduce precautions, a form of marginal net crowding out. The reason is that reputational incentives may be sufficiently reduced.[18]

[17] In particular, the reputational penalty under strict liability tends to zero as q approaches unity, but remains strictly positive under no-liability.

[18] This effect can arise only "locally". Average precautions are of course "globally" increasing in k. When $k = 0$, $\overline{p}_S = \overline{p}_M$, the no-publicity benchmark; when $k = 1$, $\overline{p}_S = p^*$.

Secondly, with high but less than perfect enforcement, some injurers may well exert more care than the socially efficient level. Write $\Delta_S(k)$ for the equilibrium reputational penalty in (13). Under perfect enforcement, $\Delta_S(1) = 0$ and injurers exerts socially efficient care. When $k < 1$, $\Delta_S(k) > 0$. As k is reduced from the perfect enforcement level, albeit not by too much, the occurrence of harm becomes informative. The fear of stigmatization may then more than compensate the decrease in formal incentives. The effect of a change in k in the neighborhood of perfect enforcement is

$$\left.\frac{\partial p_S(\theta, \Delta_S(k))}{\partial k}\right|_{k=1} = h'(L)\left((1-\theta)L + \beta\Delta_S'(1)\right).$$

where $\Delta_S'(1) < 0$. Precautions are greater with slightly less than perfect enforcement when the above expression is positive.[19]

Imperfectly enforced negligence rule. As before, due care is assumed to be set at the socially efficient precaution level. A victim now sues a negligent injurer only with probability k. The bad news event B has the same interpretation as with perfect enforcement. Event G is "injurer did not cause harm, or caused harm but complied with due care, or did not comply but the victim could not prove it".

The expected utility of a type-θ injurer can be written as in (12) but with $\varphi(p) = 0$ if $p \leq p^*$, $\varphi(p) = kp$ if $p > p^*$. More explicitly,

$$\overline{U} = \begin{cases} b - c(p) - \theta pL + \beta\theta_G \text{ if } p \leq p^*, \\ b - c(p) - p(k + \theta(1-k))L + \beta[pk\theta_B + (1-pk)\theta_G] \text{ if } p > p^*. \end{cases} \tag{14}$$

We first seek to characterize the pattern of compliance with due care. Write the expected utility in (14) as $\overline{U}(p, \theta)$ and define

$$\overline{U}_C(\theta) \equiv \max_{p \leq p^*} \overline{U}(p, \theta) = b - c(p^*) - \theta p^* L + \beta\theta_G. \tag{15}$$

This is the maximum utility level reached by a type-θ injurer who complies with the legal due care standard. Similarly, let

$$\overline{U}_{NC}(\theta) \equiv \sup_{p > p^*} \overline{U}(p, \theta). \tag{16}$$

[19] One can show that this can arise only if $h(\cdot)$ is sufficiently convex.

This is the most an injurer can obtain when he does not comply. Finally, let

$$\widehat{p}(\theta) \equiv \begin{cases} h[(k+\theta(1-k))L + \beta k\Delta] \text{ if this is larger than } p^*, \\ p^* \text{ otherwise.} \end{cases} \tag{17}$$

It is then easily seen that

$$\overline{U}_{NC}(\theta) = b - c(\widehat{p}(\theta)) - \widehat{p}(\theta)[(k+\theta(1-k))L + \beta k\Delta] + \beta\theta_G. \tag{18}$$

When the condition in the top row of (17) holds, problem (16) has the 'interior' solution

$$\widehat{p}(\theta) = h[(k+\theta(1-k))L + \beta k\Delta] > p^*.$$

When the condition does not hold, the injurer wants to get as close as possible to due care. The injurer chooses not to comply when $\overline{U}_{NC}(\theta) > \overline{U}_C(\theta)$. With $\Delta \geq 0$, this obviously can arise only when $\widehat{p}(\theta) > p^*$.

Lemma 3 *Let $\Delta \geq 0$. If a type θ' injurer complies with due care, so does a type $\theta'' \geq \theta'$.*

Both $\overline{U}_C(\theta)$ and $\overline{U}_{NC}(\theta)$ are decreasing in θ. Either all injurers comply, none does, or high types comply while low types do not, i.e., the curves cross at most once as shown in figure 3 (see the proof). The threshold for the latter case is denoted by $\widehat{\theta}$.

`-- Figure 3 about here --`

Next we derive two critical values for the probability of enforcement. In the standard model without moral or reputational concerns, it is well known that the negligence rule may yield efficient care even when enforcement is less than perfect.[20] Let k_2 be the solution to

$$b - c(p^*) = \max_p b - c(p) - kpL. \tag{19}$$

[20]This contrasts with strict liability. The argument is usually made with respect to the injurers' judgment-proofness. The damages effectively paid are then constrained by the injurer's wealth, but inducing first-best precautions remains feasible provided the injurer is not too poor (see Shavell, 1987). The reason is the discontinuity in the expected payoff when the injurer decides to barely comply (recall figure 2).

When the probability of enforcement is k_2, the injurer in the standard model is just indifferent between complying and not complying with due care; p solving the right-hand side is then greater than p^*. Clearly, $k_2 < 1$ and injurers strictly prefer exercising due care whenever $k > k_2$. This will also be the case in our setup.

Consider now an injurer who cares for his reputation but has no moral concerns (i.e., $\theta = 0$). Let k_1 be the solution to

$$b - c(p^*) + \beta\overline{\theta} = \max_p b - c(p) - kpL + \beta(1 - pk)\overline{\theta}. \tag{20}$$

The maximand on the right-hand side is the expected utility of the 'completely immoral type' who anticipates $\theta_G = \overline{\theta}$ if he is not found negligent and $\theta_B = 0$ if he is. When the probability of enforcement is k_1, he is just indifferent between complying and not. Note that $k_1 < k_2$.

Proposition 4 *Suppose the negligence rule with due care set at the efficient level is enforced with probability $k > 0$. Then all injurers exert more care than under no-liability with the same (or a smaller) probability of publicity. There exists $k_0 < k_1 < k_2 < 1$ such that:*
(i) When $k \geq k_2$, the equilibrium is the same as under the perfectly enforced negligence rule.
(ii) When $k_1 \leq k < k_2$, all injurers also comply with due care and there is no prestige from not having been found negligent. Moral concerns play no role, but reputational concerns provide incentives to comply.
(iii) When $k_0 \leq k < k_1$, there is a threshold $\widehat{\theta}(k)$ such that injurers with $\theta \geq \widehat{\theta}(k)$ comply with due care, injurers with $\theta < \widehat{\theta}(k)$ do not. Moral and reputational concerns influence the decision to comply and, for non-compliers, they mitigate carelessness.
(iv) When $k < k_0$, all injurers take less than due care and the outcome is the same as under strict liability with the same probability of enforcement.

Figure 4 provides an illustration. The heavily drawn curve denoted $\overline{p}_N(k)$ is the average probability of harm under the negligence regime as a function

of the probability of enforcement. The other curves represent the average probability of accident under strict liability and no-liability.

-- Figure 4 about here --

When the probability of enforcement is large, namely greater than k_1, all injurers exert due care, which amounts to a 'pooling equilibrium'. Because even the completely immoral type exercises due care, moral incentives play no role in providing incentives. Moreover, $\theta_G^e = \overline{\theta}$, i.e., injurers avoiding liability earn no prestige. The belief θ_B is for events that do not occur at equilibrium. When $k \geq k_2$, any $\theta_B \leq \overline{\theta}$ supports the equilibrium, hence reputational concerns then play no role either. However, when $k < k_2$, the equilibrium is supported only with $\theta_B \leq \overline{\theta}_B(k) < \overline{\theta}$, where $\overline{\theta}_B(k)$ is an upper bound that is increasing in the probability of enforcement. In particular, $\overline{\theta}_B(k_1) = 0$. At the threshold k_1, the unique out-of-equilibrium belief is $\theta_B^e = 0$. Injurers with $\theta = 0$ are then induced to comply with due care only because of the threat of being seen as completely immoral if they are found negligent. Thus, reputational concerns provide useful incentives.

When enforcement is smaller than k_1, both moral and reputational incentives play a role. When $k \geq k_0$, the more morally concerned types — those with θ above some threshold — comply with due care. Intuitively, the more morally concerned the individual, the 'less costly' it is to comply. Less morally concerned injurers do not exert due care. Nevertheless, their carelessness is mitigated by both moral and reputational concerns, as under an imperfectly enforced strict liability regime. Finally, when enforcement is less than k_0, no one complies so that the negligence rule has the same effect overall as strict liability.

Note that the pattern of equilibria would remain the same qualitatively if involvement in causing harm were public information independently of lawsuits. The event G would then be partioned into two events, say G_1 and G_2, where G_1 means "did not cause harm" and G_2 means "caused harm but was not sued". When the tort rule is sufficiently well enforced for all injurers to comply, the equilibrium beliefs are $\theta_{G_1}^e = \theta_{G_2}^e = \overline{\theta}$. When not all injurers

comply, it is straightforward to see that $\theta^e_{G_1} > \overline{\theta}$, but whether G_2 would be good or bad news depends on the probability of enforcement.

5 Welfare

The different legal regimes (no-liability, strict liability and the negligence rule) were compared in terms of how close the injurers' precautions were to p^*, the efficient precaution level of the standard model where injurers only care about their pecuniary payoff. When injurers have moral and reputational concerns, however, it is not clear that p^* is still an appropriate target. We now provide a welfare comparison in an explicit utilitarian framework.

We also discuss a related but more intricate issue. In our analysis, the moral or social norm was exogenously given: one should care about harm caused to others. We then examined how formal legal incentives interact with moral (and the derived reputational) concerns to deter careless behavior. We did not consider the degree to which the 'legal norm' was consistent with or differed from the 'social norm', nor the possibility that the 'legal norm' could influence the individuals' moral preferences.

Comparison. As shown in he previous sections, formal legal sanctions may partially or totally crowd out informal incentives, although this does not reduce overall incentives. The costs of enforcing formal sanctions would therefore naturally bear on the comparison of legal regimes. Nevertheless, we will continue to abstract from such costs.

For simplicity, suppose that injurers can also be victims; that is, they can themselves suffer harm caused by another agent. For instance, the risk generating activity under consideration is an everyday activity which everyone engages in.[21] Let $\overline{U}_j(\theta)$ be the expected utility of a type-θ injurer at equilibrium as defined in the previous sections, where $j = Z, S, N$ denotes the legal regime. Because individuals are potentially both injurer and victim,

[21] Alternatively, one could have two classes of agents, potential victims and potential injurers, and sum utility over both victims and injurers. The conclusions would be unaffected.

expected utility is now $\overline{U}_j(\theta) - \overline{D}_j$ where $\overline{D}_j$ is the expected loss that the individual faces due to the actions of others, net of the legal damages that may eventually be awarded. Total welfare is

$$W_j = \int_0^{\theta_m} \left(\overline{U}_j(\theta) - \overline{D}_j\right) f(\theta)\, d\theta, \quad j = Z, S, N.$$

Under no-liability, a victim's harm is never compensated. Hence, $\overline{D}_Z = \overline{p}_Z L$ where $\overline{p}_Z$ is the average probability of accident at equilibrium. Under strict liability enforced with probability k, an individual suffers harm caused by others with probability $\overline{p}_S$, but is then compensated by legal damages with probability k. Hence, $\overline{D}_S = (1-k)\overline{p}_S L$. Finally, under the negligence rule enforced with probability k, an individual faces

$$\overline{D}_N = \int_0^{\theta_m} [p_N^e(\theta) - \varphi(p_N^e(\theta))] L f(\theta)\, d\theta,$$

where $\varphi(p) = 0$ if $p \leq p^*$, $\varphi(p) = kp$ if $p > p^*$.

Summing the utilities over the whole population yields

$$W_Z = b - \overline{p}_Z L - \int_0^{\theta_m} [c(p_Z^e(\theta)) + \theta p_Z^e(\theta) L]\, f(\theta)\, d\theta + \beta\overline{\theta},$$

$$W_S = b - \overline{p}_S L - \int_0^{\theta_m} [c(p_S^e(\theta)) + \theta(1-k) p_S^e(\theta) L]\, f(\theta)\, d\theta + \beta\overline{\theta},$$

$$W_N = b - \overline{p}_N L - \int_0^{\theta_m} [c(p_N^e(\theta)) + \theta(p_N^e(\theta) - \varphi(p_N^e(\theta)) L]\, f(\theta)\, d\theta + \beta\overline{\theta}.$$

On the right-hand side, the first term is the gross benefit from the risk generating activity; the second term is the average accidental loss; the integral sums the precaution costs and the moral disutility from causing non compensated harm; the last term is the average utility from social esteem. Note that reputational benefits and penalties cancel out (see also Bénabou and Tirole, 2006).

It is instructive to compare the legal regimes for the case where $q \leq k = 1$. Under no-liability, individuals exert less than efficient care. Under either strict liability or negligence, all individuals exert the precaution level p^*. Substituting in the above expressions then yields

$$W_S = b - p^* L - c(p^*) + \beta\overline{\theta} \equiv W^*, \tag{21}$$

$$W_N = b - p^*L - c(p^*) - \overline{\theta}p^*L + \beta\overline{\theta} = W^* - \overline{\theta}p^*L. \tag{22}$$

Under strict liability, individuals bear accidental harm as injurers but not as victims. Under the negligence rule, it is the opposite since no one is found negligent. Under this rule, however, individuals also suffer moral disutility from inflicting non compensated harm. It is clear that welfare cannot be greater than W^* as defined in (21): the sum of precaution and accident costs cannot be made smaller and individuals suffer no moral disutility from causing harm.

Corollary 1 *When $q \leq k = 1$, $W_Z < W_N < W_S$.*

Strict liability yields greater welfare because it forces injurers to compensate their victims, thereby eliminating the moral disutility from causing harm. It is as if strict liability forced injurers to purchase a clear conscience, something they would not do spontaneously. Under negligence, by contrast, both victims and injurers suffer from the occurrence of harm, hence welfare is smaller. Finally, welfare is greater under negligence than under no-liability because average wealth is larger and because the moral cost of imposing non compensated harm is smaller; both results follow from the fact that $p_Z^e(\theta) > p^*$ for all types.

When enforcement is imperfect, the welfare comparison is not as straightforward, although some results emerge easily. For instance, suppose all individuals are underdeterred under strict liability. Applying proposition 2, $p_Z(\theta) > p_S(\theta) > p^*$ for all θ and it is readily seen that $W_Z < W_S$. Similarly, $W_Z < W_N$. It is not clear, however, how negligence compares with strict liability with the same probability of enforcement. Suppose $k \geq k_1$ as defined in proposition 4. Under negligence, all individuals then exercise efficient care. Wealth is therefore greater under negligence than under strict liability. On the other hand, the average moral disutility could be smaller under strict liability.

Legal versus social norms. In the above analysis, legal liability reduced to a pure system of external penalties contingent on some evidence.

These penalties determined the 'price' of carelessness but they had no symbolic significance and expressed no values. We now inquire about the values underlying legal liability.

The social norm we postulated was that harming others should be avoided and that one should compensate for the harm that one does cause. Perfect conformity with this social norm would arise with $\theta = 1$. In a less than ideal world, 'perfect' individuals do not exist. However, strict legal liability can in principle (when enforcement is perfect) induce individuals to behave in perfect conformity with the social norm. Thus, one could say that strict liability 'expresses' perfectly the underlying social norm.

This is not so with the negligence rule we considered. Under this rule, the connotation is that there is no 'legal wrongdoing' when harm occurs but the injurer's actions conformed to the legal due care standard. In our analysis, moral values were unaffected by the legal norm; the individuals' conception of 'moral wrongdoing' remained based solely on the preexisting social norm that one should not cause uncompensated harm. As a result, morally concerned individuals suffer disutility from causing harm even if they are not legally 'culpable'. Note that this effect would not arise if individuals were 'perfect'. If all were characterized by $\theta = 1$, injurers would always spontaneously compensate their victims even when there is no legal obligation to do so and all would exercise due care, i.e., average welfare would be W^* irrespective of the quality of enforcement.

According to the expressive theory of law, legal rules have normative power in the sense that they affect behavior not only by shaping the material payoffs, but also by directly influencing people's motives. If law expresses values or has social meaning, it could change the individuals' perception of the social norm.[22] One possibility is that the legal norm of due care modifies the interpretation of wrongdoing, in the sense that individuals who comply with the legal rule of conduct experience no moral disutility from causing

[22] On the expressive theory of law, see Kahan (1997) and Cooter (1998). Tyran and Feld (2006) and Galbiati and Vertova (2007) discuss experiments on the direct behavioral effects of legal obligations, independently of sanctions. See also McAdams and Nadler (2005).

harm.

Specifically, suppose again that due care under the negligence rule is p^*. Because of the legal norm defined by due care, the individual's utility in (1) is now

$$U = y - \theta\delta(p)x + \beta\theta_s,$$

where $\delta(p) = 0$ if $p \leq p^*$ (or if the individual compensates the harm he caused) and $\delta(p) = 1$ otherwise. Expected utility under the negligence rule is as before except that θ is replaced with $\theta\delta(p)$. Summing over all types then yields

$$W_N = b - \overline{p}_N L - \int_0^{\theta_m} \left[c(p_N^e(\theta)) + \theta\delta(p_N^e(\theta))(p_N^e(\theta) - \varphi(p_N^e(\theta))L\right] f(\theta)\, d\theta + \beta\overline{\theta},$$

where $p_N^e(\theta)$ now denotes the equilibrium level of care of type θ under the new moral preferences. The equilibria under no-liability or strict liability would of course remain the same.

Corollary 2 *Suppose individuals have no moral concerns about causing uncompensated harm when they comply with the legal due care standard under the negligence rule. If due care is set at the efficient precaution level and $k \geq k_1$ as defined in proposition 4, then all individuals comply and $W_N = W^* \geq W_S$, with strict inequality when $k < 1$.*

Everything else equal, compared to the situation where moral preferences are unaffected by the legal norm, complying with due care is now more desirable when $\theta > 0$ and it remains the same when $\theta = 0$. Hence, the threshold k_1 for overall compliance remains unchanged. Welfare under the negligence rule is now greater because moral concerns are in line with the legal norm of conduct. Because the negligence rule has the potential to implement efficient care even when enforcement is imperfect, it can now dominate strict liability.

6 Concluding remarks

The social norm considered in this paper relates to the time-honored ethical principle of the Silver Rule, often stated as "Do not do to others as you would

not have them do to you". Kant remarked that 'negative' duties (what he termed *perfect* duties), such as the duties not to cheat, steal or harm others, were in principle enforceable by law (see White, 1999a). In a well functioning civil community (as opposed to an ethical one), individuals would follow their perfect duties purely out of self-interest, i.e., to avoid sanctions. This is the situation considered in the standard economic model of socially efficient tort rules. By contrast, legal enforcement would not be needed if the perfect duties were followed out of ethical considerations.

It has often been argued that 'Kantian behavior' can improve outcomes in some situations, e.g., in public good problems. See Laffont (1975) for an earlier statement and Bilodeau and Gravel (2004) for a critical assessment. White (2009b) makes the point (see also Wolfesperger, 1999) that the usual interpretations of the categorical imperative in economics does not distinguish between Kant's *perfect* and *imperfect* duties. The latter require a general 'positive' attitude, but prescribe no precise course of action, hence may not suffice for an efficient outcome in complex situations. In the simple accident model, however, overall allegiance to the negative moral duty "not to harm others" is sufficient for socially efficient behavior.

We showed that moral concerns and legal sanctions are complements when both of them are imperfect. Thus, if the substantive laws are efficient but imperfectly enforced, allegiance to the moral duty improves efficiency, even if allegiance is only imperfect. Conversely, if allegiance to the moral duty is imperfect, appropriate legal rules also improve efficiency, even if their enforcement is imperfect. In particular, there is no net crowding-out of incentives due to the introduction of formal sanctions. This contrasts with some analytical results in the literature, as well as with many empirical and experimental observations.

Our no crowding-out result has a simple structure. First, we did not consider an arbitrary set of moral preferences and monetary rewards. As emphasized, in its perfect form, each set of incentives was on its own conducive to efficiency. Moreover, as discussed in the preceding section, there is a deep affinity between the two forms of incentives. Secondly, and crucially, our result was derived under a parameter restriction that, in particular, re-

duced the scope of reputational concerns. Net crowding-out can arise only through the effect on reputational incentives at equilibrium.

Absent our parameter restriction, there will be situations where some individuals willingly compensate their victims under no-liability, i.e., these individuals end up causing no harm (except to themselves) and will take precautions as under perfectly enforced strict liability. Either this will be true for all individuals or only for the more morally concerned types. In the latter case, less morally concerned individuals know that, if they are seen as having caused harm, the bad news signal will be "caused harm and did not compensate the victim". Net crowding-out could occur for these individuals following the introduction of imperfectly enforced legal sanctions. Indeed, one can provide examples where under no-liability some of these individuals will exercise more than the efficient level of care. Perfectly enforced strict liability would then induce them to exercise no more than the efficient level of care, a net crowding-out result that in this case turns out to be socially beneficial.

One limitation of our analysis is that social interactions arise only through reputational effects, as in Bénabou and Tirole (2006). It may well be that 'intrinsic motivation' itself depends on one's perception of the extent to which others adhere to the social norm, a form of reciprocity or conditional cooperation (see for instance Frey and Torgler, 2007). Similarly, the desire to be perceived as virtuous presumably depends on the importance of virtue to others. An interesting extension of the present analysis would therefore be to incorporate this second channel of social interactions into the accident model.

Appendix

Proof of lemma 2. From (9), the best responses $p_Z(\theta, \Delta)$ are non zero and strictly decreasing in θ. Hence, $0 < \overline{p}_Z(\Delta) < 1$, i.e., both B and G have positive probabilities, and $0 < \psi_Z(\Delta) < \theta_m$ for all $\Delta \in [0, \theta_m]$ where $\psi_Z(\Delta)$ is a continuous function. Defining $\xi(\Delta) = \psi_Z(\Delta) - \Delta$, it follows that any

solution to $\xi(\Delta) = 0$ belongs to $(0, \theta_m)$. Such a solution exists since $\xi(0) > 0$, $\xi(\theta_m) < \theta_m$ and $\xi(\Delta)$ is continuous. Moreover, there must be one (e.g., the one with the smallest Δ) at which the $\psi_Z(\Delta)$ curve cuts the forty-five degree line from above, which is equivalent to $\psi'_Z(\Delta) < 1$ at the solution. QED

Proof of proposition 1. The first claim follows directly from $p_Z(\theta, \Delta) = h(\theta L + q\beta\Delta)$, given that $\Delta > 0$ by lemma 2. We now turn to the two other claims.

a) We first show that $\overline{p}_Z > p^*$. At equilibrium, $p_Z(\theta, \Delta) = h(\theta L + \beta q\Delta)$ where $\Delta < \theta_m$ by lemma 2. Because h is decreasing, for all types $\theta \in [0, \theta_m]$, $h(\theta L + \beta q\Delta) > h(\theta_m L + \beta\theta_m) > h(L) = p^*$, where the last inequality follows from assumption 1.

b) To show that $\overline{p}_Z$ is decreasing in q, substitute for the best responses $h(\theta L + \beta q\Delta)$ in (10) and write ψ_Z explicitly as $\psi_Z(\Delta, q)$. Let $x \equiv q\Delta$ and define

$$\psi(x, q) \equiv -\frac{\int_0^{\theta_m} \theta[h(\theta L + \beta x) - \overline{h}(x)] f(\theta)\, d\theta}{\overline{h}(x)\left(1 - q\overline{h}(x)\right)}$$

where $\overline{h}(x)$ is the average best response over all types. Observe that $\psi(x, q) \equiv \psi_Z\left(\frac{x}{q}, q\right)$ or equivalently $\psi(q\Delta, q) \equiv \psi_Z(\Delta, q)$ Condition (10) can then be rewritten as

$$x = q\psi(x, q), \tag{23}$$

where x solving (23) is an equilibrium expected reputational penalty and depends on q. Differentiating totally with respect to q yields

$$\frac{dx}{dq} = \frac{\psi(x, q) + q\psi_q(x, q)}{1 - q\psi_x(x, q)}.$$

The numerator is positive since $\psi(x, q) > 0$ and the function is increasing in q. The denominator is positive at a stable equilibrium since

$$q\psi_x(x, q) = \frac{d\psi(q\Delta, q)}{d\Delta} \equiv \frac{\partial\psi_Z(\Delta, q)}{\partial\Delta} < 1.$$

Hence, $x(q) \equiv q\Delta(q)$ is increasing in q, implying that all injurers exert more effort. QED

Proof of proposition 3. We prove only the last claim. Incentives are strictly greater under strict liability if

$$[k + \theta(1-k)]\, L + k\beta\Delta_S > \theta L + q\beta\Delta_Z$$

or equivalently

$$k(1-\theta)L + \beta(k\Delta_S - q\Delta_Z) > 0.$$

With $k \geq q$,

$$\begin{aligned} k(1-\theta)L + \beta(k\Delta_S - q\Delta_Z) &\geq k\left[(1-\theta)L + \beta(\Delta_S - \Delta_Z)\right] \\ &> k\left[(1-\theta_m)L - \beta\theta_m\right] \\ &> 0. \end{aligned}$$

The last inequality follows from assumption 1. The second-to-last from $\theta \leq \theta_m$ and the fact that Δ_S, $\Delta_Z \in (0, \theta_m)$. QED.

Proof of lemma 3. Let

$$\begin{aligned} \zeta(\theta) &\equiv \overline{U}_{NC}(\theta) - \overline{U}_C(\theta) \\ &= c(p^*) + \theta p^* L - c(\widehat{p}(\theta)) - \widehat{p}(\theta)\left[(k + \theta(1-k))L + k\beta\Delta\right]. \end{aligned}$$

We show that $\zeta(\theta) = 0$ has at most one solution, say $\widehat{\theta}$, and that $\zeta(\theta) > 0$ if $\theta < \widehat{\theta}$ and $\zeta(\theta) \leq 0$ otherwise. Applying the envelope theorem,

$$\zeta'(\theta) = p^* L - \widehat{p}(\theta)(1-k))L$$

and therefore

$$\zeta''(\theta) = -\widehat{p}'(\theta)(1-k))L,$$

where $\widehat{p}'(\theta) < 0$ when $\widehat{p}(\theta) > p^*$ and is zero when $\widehat{p}(\theta) = p^*$. Hence, $\zeta''(\theta) \geq 0$, implying that $\zeta(\theta)$ is a convex function.

Let us extend the range of possible values for θ to $[0, 1]$ and look for solutions in this interval. Because $\zeta(\theta)$ is convex, it is either monotonic or first decreasing and then increasing. In the first case, there is at most one solution in the interval; in the second case, there can be at most two. Now, $\widehat{p}(1) = p^*$ so that $\Delta \geq 0$ implies $\zeta(1) \leq 0$, where the equality holds only

if $\Delta = 0$. Therefore, if there are two solutions, one must be $\theta = 1$ (when $\Delta = 0$), implying that there is at most one solution in $[0, \theta_m]$. In the latter interval, a solution does not exist if $\zeta(\theta)$ is everywhere increasing because in that case $\zeta(\theta_m) < 0$. Otherwise, if a solution exists, it must be that $\zeta(\theta)$ is decreasing at the solution. QED

Proof of proposition 4. Part (i) follows from the argument in the text and lemma 3. To show (ii), let $k \in [k_1, k_2]$ and define

$$v(\theta_B, k) \equiv \max_p b - c(p) - kpL + \beta[pk\theta_B + (1 - pk)\overline{\theta}], \quad \theta_B \leq \overline{\theta}.$$

This is the expected utility of the complete egoist, if found liability with probability k, when bad news yields θ_B and good news yields $\overline{\theta}$. The function is increasing in θ_B and decreasing in k. Let $\overline{\theta}_B$ solve

$$v(\overline{\theta}_B, k) = b - c(p^*) + \beta\overline{\theta}, \tag{24}$$

where the right-hand side is the utility of complying when good news yields $\overline{\theta}$. The complete egoist complies if $\theta_B \leq \overline{\theta}_B$. From (19), $\overline{\theta}_B = \overline{\theta}$ when $k = k_2$. From (20), $\overline{\theta}_B = 0$ when $k = k_1$. Expressed as a function of k, $\overline{\theta}_B(k)$ is increasing and continuous. Thus, when $k \in [k_1, k_2)$, there is an upper bound $\overline{\theta}_B(k) < \overline{\theta}$ such that beliefs satisfying $\theta_B \leq \overline{\theta}_B(k)$ induce compliance when $\theta = 0$; by lemma 3, such beliefs induce overall compliance.

When $k < k_1$, equation (24) has no solution. In equilibrium some injurers will therefore not comply and both B and G will have positive probability, implying $\Delta \in (0, \theta_m)$. If some injurers comply, lemma 3 implies the existence of a type threshold $\widehat{\theta}(k) < \theta_m$ as stated in (iii). For k sufficiently close to k_1, an equilibrium with such a threshold necessarily exists. For k sufficiently small, however, it does not. For instance, define

$$u(k) \equiv \max_p b - c(p) - p(k + \theta_m(1 - k))L + \beta(1 - pk)\theta_m.$$

This is the expected utility of the high type $\theta = \theta_m$, if found liability with probability k, when bad news yields $\theta_B = 0$ and good news yields $\theta = \theta_m$, i.e., the anticipated reputational penalty is $\Delta = \theta_m$. The function is strictly decreasing in k. Let k_c be the solution to

$$u(k_c) = b - c(p^*) - \theta_m p^* L + \beta\theta_m. \tag{25}$$

The right-hand side is the expected utility of the same injurer if he complies. It is easily verified that (25) has a solution satisfying $0 < k_c < k_1$. Because at equilibrium we must in fact have $\Delta < \theta_m$, even the high type $\theta = \theta_m$ would not comply when $k \leq k_c$. Thus, there exists some k_0, with $k_c < k_0 < k_1$, as stated in (iii) and (iv). QED

References

[1] Bénabou, R. and J. Tirole (2003), "Intrinsic and Extrinsic Motivation", *Review of Economic Studies*, 70, 489-520.

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

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

[4] Bilodeau, M. and N. Gravel (2004), "Voluntary Provision of a Public Good and Individual Morality", *Journal of Public Economics*, 88, 645-666.

[5] Bohnet I., B. Frey and S. Huck (2001) "More order with less law: on contract enforcement, trust and crowding", *American Political Science Review*, 95, 131-144.

[6] Brennan, G. and P. Pettit (2005), *The Economy of Esteem: An Essay on Civil and Political Society*, Oxford University Press.

[7] Brown, J. P. (1973), "Toward an economic theory of liability", *Journal of Legal Studies* 2, 323-349.

[8] Cooter, R. (2000), "Expressive Law and Economics", *Journal of Legal Studies*, 27, 585-608.

[9] Cooter, R. and A. Porat (2001), "Should Courts Deduct Nonlegal Sanctions from Damages?", *The Journal of Legal Studies*, 30, 401–422.

[10] Corneo, G. (1997), "The Theory of the Open Shop Union reconsidered", *Labour Economics*, 4, 71-84.

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

[12] Daughety, A. and J. Reinganum (2002), "Informational Externalities in Settlement Bargaining: Confidentiality and Correlated Culpability", *RAND Journal of Economics*, 33, 587-604.

[13] Daughety, A. and J. Reinganum (2009), "Secrecy and Safety", *American Economic Review*, 95, 1074-1091.

[14] Daughety, A. and J. Reinganum (2009), "Public Goods, Social Pressure, and the Choice Between Privacy and Publicity", mimeo.

[15] Fehr, E. and Schmidt (1999), "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics*, 115, 817-868.

[16] Frey B. (1997), *Not Just for the Money: An Economic Theory of Personal Motivation*, Cheltenham, Edward Elgar.

[17] Frey, B. and R. Jegen (2001), "Motivation Crowding Out Theory", *Journal of Economic Surveys* 15(5), 589-611.

[18] Frey, B. and B. Torgler (2007), "Tax Morale and Conditional Cooperation", *Journal of Comparative Economics*, 35, 136-159.

[19] Galbiatti, R. and P. Vertova (2008), "Obligations and Cooperative Behaviour in Public Good Games", *Games and Economic Behavior*, 64, 146-170.

[20] Glaeser, E.L., B. Sacerdote and J. Scheinkman (1996), "Crime and Social Interactions", *Quarterly Journal of Economics*, 111, 507-548.

[21] Glazer, A. and K. Konrad (1996), "A Signaling Explaination for Charity", *American Economic Review*, 86, 1019-1028.

[22] 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.

[23] Kahan, D. (1997), "Social Influence, Social Meaning, and Deterrence", *Virginia Law Review*, 83, 349-395.

[24] Kaplow, L. and S. Shavell (2002), *Fairness versus Welfare*, Cambridge, Harvard University Press.

[25] Kaplow, L. and S. Shavell (2007), "Moral Rules, the Moral Sentiments, and Behavior: Toward a Theory of an Optimal Moral System", *Journal of Political Economy*, 115, 494-514.

[26] Laffont, J.-J. (1975), "Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics", *Economica*, 42, 430-437.

[27] Landes, W. and R. Posner (1987), *The Economic Structure of Tort Law*, Cambridge, Harvard University Press.

[28] Lazzarini S., G. Miller and T. Zenger (2004) "Order with some law: complementarity versus substitution of formal and informal arrangements", *Journal of Law, Economics and Organization*, 261- 281.

[29] Lopez-Perez R. (2008), "Aversion to Norm-Breaking: A Model", *Games and Economic Behavior*, 237-267.

[30] McAdams, R.H. and J. Nadler (2005), "Testing the Focal Point Theory of Legal Compliance: The Effect of Third-Party Expression in an Experimental Hawk/Dove Game," *Journal of Empirical Legal Studies*, 2, 87-113.

[31] McAdams, R.H. and E. B. Rasmusen (2005), "Norms in Law and Economics", in *Handbook of Law and Economics*, edited by A. Mitchell Polinsky and Steven Shavell, North Holland.

[32] Rabin, M. (1993), "Incorporating Fairness into Game Theory", *American Economic Review*, 83, 1281-1302.

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

[34] Sen, A. (1997), "Maximization and the Act of Choice", *Econometrica*, 65, 745-779.

[35] Shavell, S. (1984), "A Model of the Optimal Use of Liability and Safety Regulations", *RAND Journal of Economics*, 15, 271-280.

[36] Shavell, S. (1987), *Economic Analysis of Accident Law*, Harvard University Press, Harvard.

[37] Shavell, S. (1993), "The Optimal Structure of Law Enforcement", *Journal of Law and Economics*, 36, 255-287.

[38] Shavell, S. (2004), "Law versus Morality as Regulators of Conduct", *American Law and Economics Review*, 4, 227-257.

[39] Shavell, S. (2007), "Liability for Accidents", in A.M. Polinsky and S. Shavell, eds., *Handbook of Law and Economics*, Vol. 1, Amsterdam, North Holland.

[40] Smith, A. (1790), *The Theory of Moral Sentiments*, reprint of the 6th ed., Cambridge, Cambridge University Press, 1976.

[41] Tyran, J.-R. and L. P. Feld (2006), "Achieving Compliance when Legal Sanctions are Non-Deterrent", *Scandinavian Journal of Economics* 108(1), 135-156.

[42] White, M.D. (2009a), "Adam Smith and Immanuel Kant: On Markets, Duties, and Moral Sentiments", *Forum for Social Economics*, forthcoming.

[43] White, M.D. (2009b), "Kantian Ethics and the Prisoners' Dilemma", *Eastern Economic Journal*, forthcoming.

[44] Wolfelsperger, A. (1999), "Sur l'existence d'une solution 'kantienne' du problème des biens collectifs", *Revue Economique*, 50, 879-901.

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

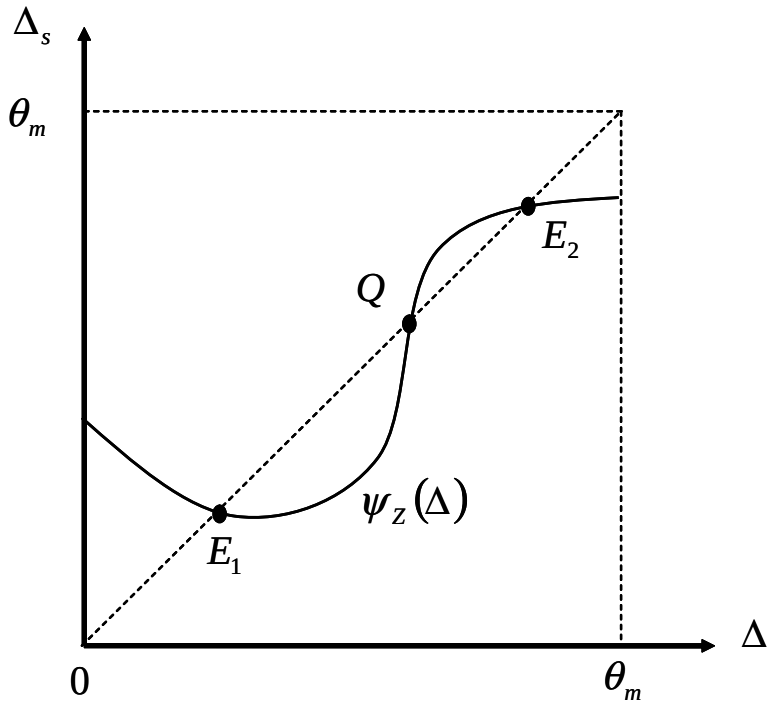

Figure 1. Reputational penalty

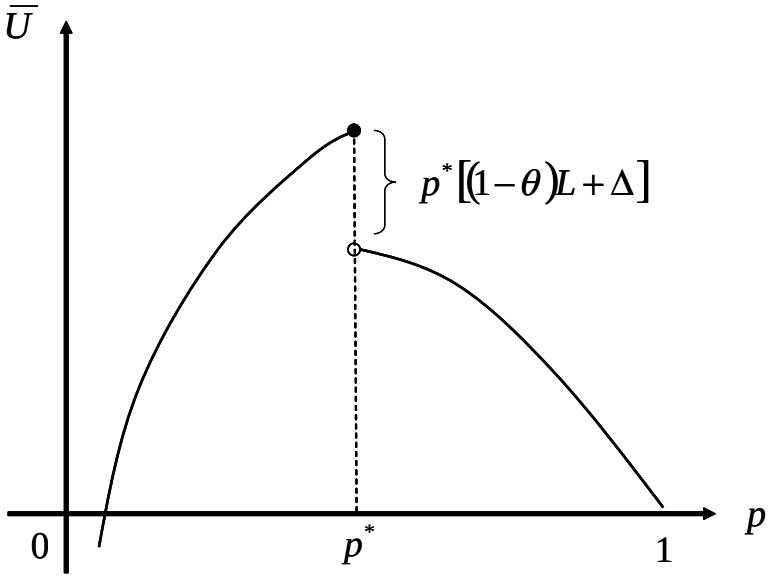

Figure 2. Negligence rule

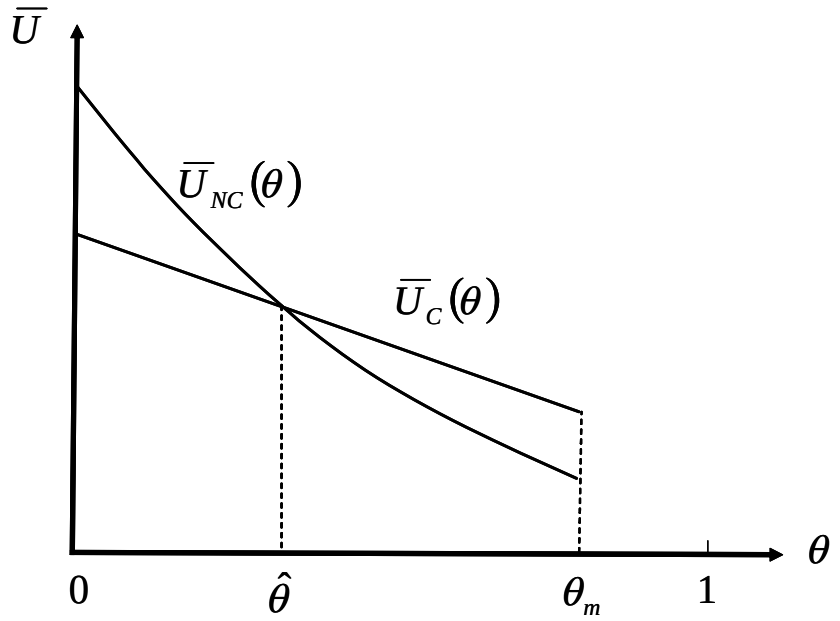

Figure 3. Compliance with due care

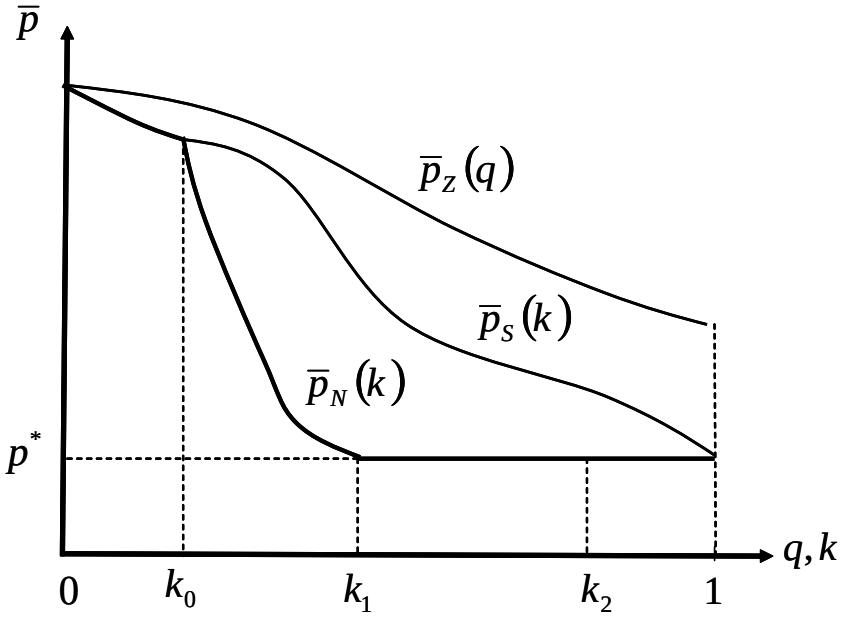

Figure 4. Liability regimes