Legal Liability when Individuals Have Moral Concerns

Bruno Deffains
Université Paris 2 and European Business School at Wiesbaden

Claude Fluet
Université du Québec à Montréal, Université Paris 2 and CIRPEE
fluet.claude-denys@uqam.ca

Abstract. We incorporate normative motivations into the unilateral precaution model of tort. Individuals have moral concerns about causing harm and would like others to believe that they do. In the absence of legal liability, causing harm suggests low concerns and is therefore damaging to one's social image, which feeds back into incentives to take precautions. These nevertheless remain suboptimal when informal motivations are not strong enough for injurers to willingly compensate victims ex post. By contrast, perfectly enforced legal liability crowds out informal motivations completely (e.g., tortfeasors suffer no disesteem) but precautions are then efficient. Under imperfect enforcement, informal motivations and legal sanctions complement one another. With strict liability, individuals held liable suffer disesteem, there is some motivational crowding-out but no net crowding-out with respect to overall incentives. Under the negligence rule, there is motivational crowding-in when image concerns induce bunching on the legal due care standard. (*JEL* D8, K4, Z13)

For helpful comments and suggestions, we thank Andrew Daughety, Nicolas Jacquemet, Lewis Kornhauser, David Myatt, William Neilson, Stéphane Pallage, Jennifer Reinganum, Eli Spiegelman, Veikko Thiele, Jean Tirole, as well as a the co-editor, Bentley McLeod, and three referees. Funding from GIP "Droit et Justice" (France), FQRSC (Quebec) and SSHRC (Canada) is gratefully acknowledged.

1 Introduction

The risk of lawsuits induces precautions to prevent accidental harm to third parties. In the economic model of legal liability, the incentives to exercise care are the monetary penalties set by tort rules.[1] Casual observation suggests that other motivations are often also at work. Most people exercise some care out of intrinsic concerns about hurting others or because they fear social disapproval if they are thought not to mind. In this paper, we augment the standard, unilateral model of tort to include such concerns.

We consider the role of a moral prescription that seems particularly relevant in a tort context. Kaplow and Shavell (2002) remark that there is a strong norm to avoid harming others and to compensate for the harm that one does cause. The prescription is a variant of the so-called 'golden rule', i.e., "do unto others as you would have them do unto you". We take this prescription as given. Individuals are assumed to feel guilt or bad conscience when they do not abide by it. This provides an internalized motivation to prevent damaging events. However, individuals are assumed to be heterogenous in this respect, i.e., some are intrinsically more motivated than others.

Individuals also have a preference for social approval. They care about the esteem earned if they are believed to have good predispositions and the disesteem or shame if not. Intrinsic predispositions are unobservable but one's actions may provide some information. In addition to intrinsic motivation, compliance with the prescription not to cause harm may therefore to some extent also be enforced by approval or reproach from others.[2]

The issue is how such 'normative motivations' interact with formal legal sanctions to influence behavior. It is often formulated in terms of whether law and informal incentives are substitutes or complements.[3] The behavioral economics literature has much emphasized the possibility that pecuniary incentives may undermine informal motivations; e.g., the much quoted study by

Gneezy and Rustichini (2000) on the crowding-out effect of fines and the survey by Frey and Jegen (2001). If crowding-out effects are sufficiently strong, legal liability in the tort context could well be counterproductive and reduce precautions to prevent accidental harm. On the other hand, it could be that informal motivations and legal sanctions combine to generate too much incentives. For instance, Cooter and Porat (2001) ask whether courts should deduct 'nonlegal sanctions' from legal damages to avoid overdeterrence.

We first consider two benchmarks: no liability versus perfectly enforced legal liability (both strict liability and the negligence rule). In the absence of legal liability, injurers take precautions, if at all, solely out of moral or image concerns. In fact, when such concerns are sufficiently strong, injurers could go so far as to willingly compensate their victims *ex post*, thereby imposing upon themselves the same penalties as under a strict liability legal regime. Anticipating this, they would therefore choose *ex ante* to exert efficient care. To allow a role for legal liability, we introduce an upper bound on the extent to which preferences differ from the standard model. The upper bound rules out spontaneous compensation at equilibrium. Under no liability, individuals then take suboptimal precautions but nevertheless exert some care to avoid causing harm. By contrast, perfectly enforced legal liability crowds out informal incentives completely, e.g., tortfeasors incur no social disapproval. However, precautions are then socially efficient, as in the standard model. There is no overdeterrence: when legal liability is introduced, informal incentives either disappear or lose their bite.

Next we examine imperfectly enforced legal liability, e.g., victims do not always bring suit because they have insufficient evidence to prevail in court. We show that normative motivations and formal legal sanctions then complement one another, e.g., an individual held liable faces both legal damages and disesteem. Thus, informal and formal incentives interact to induce more precautions than under no liability. Under the strict liability regime, there

is some crowding-out of informal incentives but no net crowding-out with respect to overall incentives. Under the negligence rule, there may be motivational crowding-in. Because of the signal sent by a negligence ruling, image concerns tend to induce bunching on the legal due care standard. Thus, when enforcement is imperfect, the negligence rule may do much better than strict liability because of the individuals' concern for social approval. 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 moral norm. We also discuss causation requirements under the negligence rule and how the analysis can be extended to the bilateral precaution framework.

We share with a strand of literature the idea that an individual's actions may signal something about his unobservable predispositions and that some predispositions are more socially esteemed than others. In Bernheim (1994), individuals seek to signal that they do not have extreme predispositions, hence the possibility of an endogenous conformity norm. In Bénabou and Tirole (2006) and Daughety and Reinganum (2010), they seek to signal that they have good pro-social characteristics by contributing to a public good. This leads them to contribute more than if their actions were unobservable. In our analysis, predispositions refer to one's 'moral type', by which we mean an individual's sensitivity to the prescription not to cause uncompensated harm. By contrast with the above literature, however, an individual's actions (i.e., precautions to avoid accidental harm) are not directly observable by the public at large. They affect public perception about one's predispositions only through what can be inferred from the occurrence of damaging events or from court rulings (e.g., liable for damages versus not liable).

Section 2 presents the basic setup. In section 3, we analyze no liability and perfectly enforced liability; in section 4, imperfectly enforced liability. Section 5 discusses the results from a utilitarian point of view. Section 6

concludes.

2 The model

Our starting point is the unilateral accident framework of the law and economics literature. Some individuals, hereafter injurers, engage in an activity that may harm third parties. In the standard model, injurers care only about their private costs and benefits as conventionally defined. We briefly review the role of tort rules in this context.[4] Next we introduce normative motivations.

The standard model. The risk generating activity provides the private benefit b; with probability p it imposes on others the loss L. Both b and L are pecuniary or are monetary equivalents; agents are risk neutral. The probability of accident depends on the injurer's precautions. A smaller p means more precautions; the cost is $c(p)$ with $c' < 0$ and $c'' > 0$. At the boundaries, $c(1) = c'(1) = 0$ and $c'(0) = -\infty$. Totally eliminating the risk of harm is prohibitively costly but the marginal precaution cost is nil at the no-precaution level.

Absent legal liability, the payoff from the activity is $y = b - c(p)$; not engaging in the activity yields a payoff normalized to zero. All potential injurers therefore engage and exert no care to avoid harm. Given the loss suffered by victims, social welfare is $W \equiv b - c(p) - pL$. The socially efficient precaution level p^* minimizes $c(p) + pL$, the sum of precaution and expected accident costs. Engaging in the risk generating activity is socially efficient if $b - c(p^*) - p^*L > 0$.

For future use, let $P(t)$ be the level of care minimizing $c(p) + tp$, where $t \geq 0$ is some given parameter. The function $P(t)$ is strictly decreasing, with $P(0) = 1$. The socially efficient precaution level is $p^* = P(L)$.

The tort rules that we will consider are *strict liability* and *negligence*. Under strict liability, injurers are liable for full compensatory damages irrespective of the precautions they have taken. The victim then only needs to prove causation. Litigation costs are assumed to be negligible. An injurer's expected payoff is therefore $\overline{y} \equiv b - c(p) - pL$, which induces efficient precautions. Under the negligence regime, victims need to prove the injurer's carelessness, i.e., that precautions did not meet the legal due care standard. Due care is assumed to be set at the socially efficient level, meaning that negligence is found when the injurer's precautionary behavior entailed $p > p^*$. The expected payoff from the activity is then

$$\overline{y} = \begin{cases} b - c(p) \text{ if } p \leq p^*, \\ b - c(p) - pL \text{ if } p > p^*. \end{cases} \tag{1}$$

This is maximized by complying with due care.

Under either liability rule, potential injurers exert the first-best level of care. As is well known, however, the negligence rule is deficient in controlling activity levels, by contrast with strict liability (see, e.g., Shavell 1987). Under the latter, injurers engage in the risk generating activity only if the gross private benefit is larger than the full social cost $c(p^*) + p^*L$. Under the negligence rule, injurers incur only the cost of care and therefore inefficiently engage in the activity when $c(p^*) < b < c(p^*)+p^*L$. For simplicity, we assume that the negligence rule is appropriate for the activity under consideration, i.e., b is large and the issue is whether the activity is exercised with sufficient care.[5]

We also need to consider situations where legal liability is only imperfectly enforced. Victims sometimes do not file suit because they cannot prove causation or negligence. Let q denote the probability that a victim has access to sufficient evidence. Under strict liability an injurer's expected payoff is now $\overline{y} \equiv b - c(p) - pqL$. The induced care level is then $p = P(qL)$ and is

suboptimal when $q < 1$. Under the negligence rule, the expected payoff is

$$\overline{y} = \begin{cases} b - c(p) \text{ if } p \leq p^*, \\ b - c(p) - pqL \text{ if } p > p^*. \end{cases} \tag{2}$$

Because of the discontinuity in the payoff function, efficient precautions may be induced provided the probability of enforcement is sufficiently large. Let q_N be the solution to

$$b - c(p^*) = \max_p b - c(p) - q_N pL. \tag{3}$$

When the probability of enforcement is q_N, the injurer is just indifferent between complying and not complying with due care. Clearly, $q_N < 1$ and injurers prefer complying whenever $q > q_N$. When $q < q_N$, they do not comply and behave as under strict liability, i.e., $p = P(qL)$.

Moral and image concerns. We now depart from the standard model by introducing normative motivations. The moral prescription is that harming others should be avoided; if one nevertheless causes harm, one should compensate the victim. To some extent, this prescription is internalized through intrinsic 'moral concerns'. Individuals suffer disutility (e.g., guilt or bad conscience) when they do not conform. Individuals also care about social esteem and would like others to believe that they have high moral concerns. Both moral and esteem concerns constitute normative motivations in the sense that they derive from one's allegiance to the moral prescription or one's attempt to signal allegiance.

The utility function is now

$$U(y, x, e) = y - \theta x + \beta e, \tag{4}$$

where y is net material payoff, x is the harm that one causes, and e is social esteem; θ and β are positive parameters. All injurers are assumed to care equally about social esteem, i.e., they have the same parameter β. However,

they differ in the extent of their moral concerns. An individual's θ is private information and denotes his 'moral type', by which we mean his sensitivity to the moral prescription. The distribution of types is common knowledge and is described by the density $f(\theta)$ with support $[\theta_l, \theta_h]$ and average value $\overline{\theta}$. Social esteem depends on society's perception of one's moral type, i.e., $e = \overline{\theta}_I \equiv E(\theta \mid I)$ where I denotes the information publicly available about the individual; the conditional expectation $E(\theta \mid I)$ is society's updated belief about the individual's type. An individual will be said to *earn esteem* (*suffer disesteem*) when his perceived type is above (below) the population average.[6]

No spontaneous compensation. In the standard model, θ and β are both zero and utility reduces to the net material payoff y. When the parameters are positive, however, there will be situations where all injurers exert efficient care even without legal liability. This occurs when those who cause harm willingly compensate their victim. In the standard model there is nothing to prevent injurers from doing so, but the issue is not raised because injurers would never willingly compensate. We will introduce parameter restrictions ruling out this type of behavior in the present set-up as well.

For instance, suppose that information about individuals is limited to their involvement (or non involvement) in causing harm of amount L and to whether compensation has been paid. *Ex post*, if a type-θ injurer does not compensate (action 0), his utility is $U_0 = b - c(p) - \theta L + \beta\overline{\theta}_0$ where $\overline{\theta}_0$ is the perceived moral type of injurers who do not compensate. If the injurer compensates (action 1), he eliminates the 'bad feelings' from not complying with the moral prescription because in the end no harm will have been inflicted, i.e., $x = 0$ in the notation of equation (4). Because the injurer thereby also transfers the victim's loss to himself, his utility is $U_1 = b - c(p) - L + \beta\overline{\theta}_1$ where $\overline{\theta}_1$ is the perceived type of injurers who compensate. When the lower bound of moral types satisfies $\theta_l \geq 1$, there is an essentially unique equilibrium in which all injurers voluntarily compensate their victims.

As a result, the action of compensating reveals nothing about one's moral type; hence, at equilibrium $\overline{\theta}_1 = \overline{\theta}$ where $\overline{\theta}$ is the prior mean.[7] *Ex ante*, injurers anticipate that they will compensate. Therefore they choose the precaution level that maximizes $b - c(p) - pL + \beta\overline{\theta}$, yielding the efficient p^*.

The same outcome arises even if the upper bound of moral types satisfies $\theta_h < 1$ provided the esteem concern parameter β is sufficiently large. Now injurers will not compensate on purely moral grounds because the 'bad feelings' from not complying with the moral prescription is less painful to them than the money cost of compensating. However, injurers care a lot about the disesteem they would suffer by signalling their disregard for the prescriptive rule.[8] Again, the action of compensating will be anticipated *ex ante* so that all injurers will exert the efficient precaution level. As in the previous example, formal legal liability would serve no purpose. The following parameter restriction rules out spontaneous compensation.

ASSUMPTION 1: $\theta_l = 0$, $\theta_h < L/(\beta + L)$.

The condition defines an upper bound on the extent to which preferences depart from the standard model. Injurers put some weight on complying with the moral prescription (some put zero or negligible weight), but they all 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.

Let U_0 and U_1 be defined as above. A type-θ injurer chooses not to compensate his victim if $U_1 < U_0$, that is if

$$\beta\left(\overline{\theta}_1 - \overline{\theta}_0\right) < (1 - \theta)L. \tag{5}$$

Beliefs about one's type belong to the interval $[0, \theta_h]$, so that $\overline{\theta}_1 - \overline{\theta}_0 \leq \theta_h$. Assumption 1 can be rewritten as $\beta\theta_h < (1 - \theta_h)L$, implying that (5) always holds. Thus, we have:

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

When the support of types is an interval of the form $[0, \theta_h]$, assumption 1 is necessary to rule out voluntary compensation in the absence of further restrictions on the probability distribution of types. Although our injurers differ from standard *homo economicus*, they therefore behave the same way following the occurrence of harm. By contrast, if they were to voluntarily compensate their victims, they would impose upon themselves the same penalties as under the strict liability legal regime.[9]

Posterior information. Throughout the paper, with a qualification in the case of negligence based liability, an injurer's precautions are assumed to be private information. Precautions can affect esteem only indirectly 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.[10] When the tort regime is negligence, we will suppose that some evidence about the injurer's precautions becomes available to the court and that the evidence is sufficiently informative to assess whether the injurer complied with due care. However, the general public is assumed to be informed only of the court's ruling, not of the detailed evidence disclosed at trial. Social esteem depends on information available at large in the general public.

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

about an injurer's predisposition would not reduce to events such as B or G if, at equilibrium, injurers attempted to signal their type by voluntarily compensating victims. From lemma 1, because no one compensates, nothing can be inferred from the fact that a particular injurer did not offer compensation.

The probability of the random events depends on the injurer's precautions; we denote with $\varphi(p)$ the probability of B. Inferences about an injurer's predisposition depend on whether B or G occurred. A general formulation for the expected utility of a type-θ injurer is then

$$\overline{U} = \overline{y}(p) - \theta\overline{x}(p) + \beta\left[\varphi(p)\overline{\theta}_B + (1-\varphi(p))\overline{\theta}_G\right], \tag{6}$$

where $\overline{y}(p)$ is the injurer's expected material payoff, $\overline{x}(p)$ is the expected uncompensated harm for which he will feel 'morally responsible', and $\overline{\theta}_B$ and $\overline{\theta}_G$ are society's posterior beliefs about one's type. The expected values $\overline{y}(p)$ and $\overline{x}(p)$ depend on the legal regime because it determines whether the victim or the injurer ultimately bears the accidental loss. Society's posterior beliefs are part of an equilibrium.

DEFINITION 1: *A perfect Bayesian equilibrium consists of strategies $\widehat{p}(\theta)$ and of beliefs $\overline{\theta}_B$ and $\overline{\theta}_G$ such that*

(i) $\widehat{p}(\theta)$ solves $\max_p \overline{y}(p) - \theta\overline{x}(p) + \beta[\varphi(p)\overline{\theta}_B + (1-\varphi(p))\overline{\theta}_G]$, $\theta \in [0, \theta_h]$;

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

$$\overline{\theta}_B = \frac{\int_0^{\theta_h} \theta\varphi(\widehat{p}(\theta)) f(\theta)\, d\theta}{\overline{\varphi}_B}. \tag{7}$$

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

$$\overline{\theta}_G = \frac{\int_0^{\theta_h} \theta[1-\varphi(\widehat{p}(\theta))] f(\theta)\, d\theta}{1-\overline{\varphi}_B}; \tag{8}$$

Beliefs satisfy Bayesian up-dating when the conditioning events have positive probability over the population of injurers. When a conditioning event

has zero probability, the posterior belief is to some extent arbitrary. In some settings, bad news will never occur at equilibrium. This arises when all injurers conform to a standard of behavior precluding bad news. In that case, $\overline{\theta}_G = \overline{\theta}$ and the equilibrium will be 'supported' by a range of out-of-equilibrium beliefs $\overline{\theta}_B$. It will often be convenient to think of these beliefs as $\overline{\theta}_B = 0$. The reason is that the worst moral type will then be the one who would lose the least from a deviation to an action generating a positive probability of bad news.[11]

Of particular interest is $\Delta \equiv \overline{\theta}_G - \overline{\theta}_B$, the gap in social esteem between good and bad news, which we will refer to as the *reputational penalty*. When both B and G have positive probability, (7) and (8) can be combined to yield

$$\Delta = -\frac{\int_0^{\theta_h} \theta[\varphi(\widehat{p}(\theta)) - \overline{\varphi}_B] f(\theta)\, d\theta}{\overline{\varphi}_B(1-\overline{\varphi}_B)}. \tag{9}$$

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

3 No Liability vs Perfectly Enforced Liability

While injurers are not willing to compensate their victims *ex post*, they will want to take some precautions to prevent the occurrence of harm. We consider two benchmarks. The first is the case where harm is not subject to legal liability. Incentives to take precautions then rely solely on moral and esteem concerns. In particular, individuals will take some precautions because the marginal precaution cost is nil at the zero precaution level and they anticipate the guilt of causing uncompensated harm (given that they will not want to compensate *ex post*). Next, we introduce perfectly enforced legal liability and examine how this combines with normative motivations.

No liability. Causing harm often does not trigger legal liability. It may be that the harm is not subject to judicial sanction because it is trivial or considered part of the usual risks of life. Even when legal liability applies in principle, victims often do not file suit because judicial procedures are too expensive compared to the stakes.

Let k denote the probability that the injurers' involvement in causing harm becomes publicly known. Consider first the no-publicity case where an individual causing harm is never detected (e.g., damages to one's car in a parking lot). A type θ injurer then chooses his precaution level p to maximize

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

Because no information about the injurer is made public, his perceived type is the prior mean. Using the function defined in the previous section, the type-θ injurer chooses $p = P(\theta L)$. Given Assumption 1, this is greater than the efficient p^* but less than unity, i.e., some precautions are taken out of moral concerns. More morally concerned individuals exert greater care.

Suppose now that k is positive. 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 $\varphi(p) = pk$. 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

$$1 - p + p(1 - k) = 1 - pk.$$

Either harm has not occurred or it has occurred but has not been observed by society at large or has been observed but not causally related to the particular injurer (i.e., the injurer is not detected).

The expected utility of a type-θ injurer is now

$$\overline{U} = b - c(p) - \theta pL + \beta[pk\overline{\theta}_B + (1 - pk)\overline{\theta}_G]. \tag{10}$$

This can be rewritten as

$$\overline{U} = b + \beta\overline{\theta}_G - c(p) - p(\theta L + \beta k\Delta), \text{ where } \Delta \equiv \overline{\theta}_G - \overline{\theta}_B. \tag{11}$$

In equilibrium, society's beliefs will satisfy $\overline{\theta}_B < \overline{\theta} < \overline{\theta}_G$; that is, event B inflicts disesteem while event G provides esteem. Given the reputational penalty, the best-response function of a type-θ individual is then

$$p_0(\theta, \Delta) = P(\theta L + \beta k\Delta), \tag{12}$$

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

The equilibrium reputational penalty is obtained by substituting the injurers' best response functions (12) for $\widehat{p}(\theta)$ in the right-hand side of (9) and solving the resulting equation for Δ. The solution satisfies $0 < \Delta < \theta_h$.[12] Less morally concerned individuals exert less care and therefore are more often involved in causing accidents; hence, a below average moral type is inferred from the occurrence of harm. One can show that injurers exert greater care the greater the publicity about involvement in causing harm. Nevertheless, moral and esteem concerns are never strong enough to induce efficient care, even when causing harm is always detected.[13]

Perfectly enforced legal rules. Perfect enforcement corresponds to the elementary version of the economic model of torts. Suppose that, following the occurrence of harm, victims always have access to perfect evidence and can therefore always prove causation. This suffices under strict liability. Under the negligence rule, victims receive evidence showing whether the defendant was negligent or not. A lawsuit imposes a small cost on plaintiffs, so that they sue only if they have a valid claim, but litigation costs are otherwise negligible. Suppose also that out-of-court settlements are infeasible, hence all suits go to trial. Finally, suppose that trial outcomes are the

only information that becomes available to the general public. We derive the implications of these assumptions and then show that some can be relaxed.

Under *strict liability*, victims always sue. Injurers therefore know that, should they cause harm, they will have to pay compensatory damages and that their involvement in causing harm will be public information. The bad news event B is "injurer was 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. Albeit unwillingly, they therefore comply with the underlying moral prescription. As a result, they suffer no moral disutility but bear the victims' losses. The expected utility of a type-θ injurer is then

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

Expected utility does not depend on the injurers' type and best responses are therefore type independent. It follows that the events G and B provide no information about type, i.e., at equilibrium $\overline{\theta}_B = \overline{\theta}_G = \overline{\theta}$. All injurers choose the efficient precaution level p^* and there is no reputational penalty from causing harm.

Under the *negligence* rule, injurers who comply with due care are never sued. Those who do not comply and cause harm are found negligent and this becomes public information. Thus, the bad news event B is "injurer was 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, hence was not sued. A type-θ injurer now has the expected utility

$$\overline{U} = \begin{cases} b - c(p) - \theta pL + \beta\overline{\theta}_G \text{ if } p \leq p^*, \\ b - c(p) - pL + \beta[p\overline{\theta}_B + (1-p)\overline{\theta}_G] \text{ if } p > p^*. \end{cases} \tag{13}$$

The upper branch is the expected utility when the precaution level satisfies due care. With probability one, the injurer's perceived type will then be $\overline{\theta}_G$.

With probability p, the injurer will nevertheless cause uncompensated harm, which yields moral disutility. The lower branch is for an injurer who does not comply with due care. With probability $1-p$, harm will not occur and the belief will again be $\overline{\theta}_G$. With probability p, the injurer will be sued and will pay damages. He then suffers no moral disutility from having caused harm but his perceived type is $\overline{\theta}_B$.

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 $\overline{\theta}_B \leq \overline{\theta}_G$,

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

Combining both results (and noting that complying with due care does strictly better than barely not complying), the utility maximizing precaution level is p^*. Because all injurers exercise due care, $\overline{\theta}_G = \overline{\theta}$ and $\overline{\theta}_B$ is an out-of-equilibrium belief. From the above argument, any $\overline{\theta}_B \leq \overline{\theta}$ supports the equilibrium. Figure 1 illustrates the discontinuity in the expected utility of a type-θ individual.

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

Our assumptions about the litigation subgame can be modified in many respects without affecting the outcome. For instance, it would not matter if out-of-court settlements were allowed and their content remained secret. Victims would obtain the same damages as they would at trial, which is all that matters. The outcome would also remain the same if involvement in causing harm is assumed to be publicly observable independently of court

rulings or lawsuits. Under strict liability, causing harm and being held liable are the same events. Under the negligence rule, individuals causing harm are never found negligent.

The following proposition summarizes our results for both no liability and perfectly enforced liability.

Proposition 1 *Under no liability, individuals known to have caused harm suffer disesteem; moral and image concerns mitigate carelessness but injurers exert sub-optimal care. Under perfectly enforced liability rules, individuals known to have caused harm suffer no diseesteem; injurers exert efficient care and normative motivations play no role in providing incentives.*

When formal legal sanctions are introduced and enforcement is perfect, moral concerns either disappear (under strict liability) or are superfluous for the provision of incentives (under the negligence rule). Under strict liability, esteem concerns have no bite because an adverse court ruling inflicts no disesteem. Under the negligence rule, a ruling of negligence would inflict disesteem (given, say, $\overline{\theta}_B = 0$), but the threat of disesteem is superfluous in inducing efficient care. Legal liability causes 'motivational crowding-out' in the sense that formal incentives are not simply added to informal ones. However, there is no net crowding out effect because all injurers exert greater care than in the absence of formal sanctions.

4 Imperfectly Enforced Legal Rules

We now assume that victims have access to perfect evidence only with probability q. Specifically, 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. Given a small cost of filing suit, worthy victims therefore litigate only with probability q. By contrast with perfect enforcement,

esteem concerns will now be shown to matter. What information becomes public therefore also matters. For instance, even though litigation costs are negligible, injurers would favor confidential settlements — possibly allowing victims to extract 'hush money' — if disesteem can thereby be avoided.[14] For simplicity, we consider a litigation subgame where *ex post* public information effectively reduces to a binary outcome, as in the previous section. First, the filing of a lawsuit is public information and of course so would be the outcome at trial. Secondly, 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. Finally, an injurer's involvement in causing harm is not publicly observable independently of lawsuits.

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 public information 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 want to settle for more than the damages he would have paid if the case had gone to trial (and the victim would not accept less). On the other hand, if the victim does not have a viable case, filing suit is not worth the small filing cost.

Strict liability. Because victims file suit only with probability q, the probability of legal damages (and the probability of bad news) is $\varphi_S(p) = pq$, where the subscript stands for strict liability. 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 expected utility of a type-θ injurer is

$$\overline{U} = b - c(p) - pqL - \theta p(1-q)L + \beta[pq\overline{\theta}_B + (1-pq)\overline{\theta}_G],$$

When he is not sued, the injurer suffers moral disutility from causing uncompensated harm. The best response function is

$$p_S(\theta, \Delta) = P[(q + \theta(1 - q))L + \beta q \Delta], \tag{14}$$

An injurer's precautions depend on his type whenever the probability of enforcement 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 penalty solves equation (9) with $p_S(\theta, \Delta)$ substituted for $\widehat{p}(\theta)$ in the right-hand side.

In the preceding section no liability was compared with perfectly enforced strict liability, a rule known to induce efficient care when informal motivations are non-existent. Does the introduction of imperfectly enforced strict liability also increase incentives to exert care? A natural comparison is with no liability for the same probability of publicity about involvement in causing harm. From (12), a type-θ injurer's incentives to exert care under no liability are given by $\theta L + \beta k \Delta_0$, where Δ_0 is the equilibrium reputational penalty under no liability. From (14), the injurer's incentives under strict liability are $(q+\theta(1-q))L+\beta q \Delta_S$, 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 the lower moral incentives are more than compensated by the greater expected legal damages. However, reputational incentives are likely to be weaker, i.e., $\Delta_S < \Delta_0$. The intuition is that the introduction of legal liability has a relatively greater effect on the incentives of injurers with low moral concerns. When precautions become more alike between types, the bad news-good news signal will be less informative about moral type; hence esteem concerns will have less bite because the reputational penalty will be smaller.[15]

Proposition 2 *When strict liability is imperfectly enforced, injurers known to have caused harm suffer disesteem, both moral and esteem concerns miti-*

gate carelessness, and greater care is exerted than under no liability with the same (or a smaller) probability of observing involvement in causing harm.

The proposition is a "no net crowding-out" result: even when they are imperfectly enforced, formal sanctions increase incentives to exert care compared to no liability, although there is some crowding-out of informal motivations. Note that in practice a move from no liability to a strict liability regime with negligible litigation costs may well increase publicity, i.e., $q > k$, because victims now have monetary incentives to reveal the occurrence of harm by filing suit.

Negligence rule. The probability of being sued and found negligent is $\varphi_N(p) = 0$ if $p \leq p^*$ and $\varphi_N(p) = pq$ otherwise, where the subscript refers to the negligence rule. 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 is

$$\overline{U} = \begin{cases} b - c(p) - \theta pL + \beta\overline{\theta}_G \text{ if } p \leq p^*, \\ b - c(p) - p(q + \theta(1-q))L + \beta[pq\overline{\theta}_B + (1-pq)\overline{\theta}_G] \text{ if } p > p^*. \end{cases} \tag{15}$$

There will now be situations where all injurers exert efficient care even though enforcement is imperfect. This is not surprising considering that in the standard model injurers comply with due care when the probability of enforcement is above the critical q_N defined in (3). In the present case, however, compliance will also obtain even with enforcement below q_N. The pattern of compliance is straightforward: if some moral type complies with due care, so will higher types (see Lemma 2 in the Appendix). Hence, either everyone complies, no one does or high moral types do and low ones do not.

The smallest probability of enforcement consistent with a full compliance equilibrium is now q_1 solving

$$b - c(p^*) + \beta\overline{\theta} = \max_p b - c(p) - qpL + \beta(1-pq)\overline{\theta}. \tag{16}$$

The right-hand side is the expected utility of the non-complying no-moral-concern type (i.e., $\theta = 0$) who anticipates $\overline{\theta}_G = \overline{\theta}$ if he is not found negligent and $\overline{\theta}_B = 0$ if he is. Note that this yields the largest reputational penalty consistent with a full compliance equilibrium. When the probability of enforcement is q_1, the worst moral type is just indifferent between complying and not. Obviously $q_1 < q_N$.

Proposition 3 *Under the negligence rule enforced with probability q, all injurers exert more care than under no liability with the same (or a smaller) probability of publicity.*
(i) When $q \geq q_N$, the equilibrium is the same as under the perfectly enforced negligence rule and normative motivations play no role.
(ii) When $q_1 \leq q < q_N$, all injurers comply with due care. Moral concerns play no role, but esteem concerns provide incentives to comply.
(iii) There exists $q_0 < q_1$ such that, when $q_0 \leq q < q_1$, higher moral types comply with due care, lower ones do not. Moral and esteem concerns influence the decision to comply and, for non compliers, they mitigate carelessness.
(iv) When $q < q_0$, no one complies and the outcome is the same as under strict liability with the same probability of enforcement.

The proof is in the Appendix. Figure 2 provides an illustration. The curves $\overline{p}_N$ and $\overline{p}_S$ depict the average probability of harm under the negligence and strict liability regimes as a function of the probability of enforcement. The curve $\overline{p}_0$ is the average probability of accident under no liability as a function of the probability of publicity.

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

Under the negligence rule, when enforcement is above q_1, all injurers comply with due care. Because even type $\theta = 0$ complies, moral incentives

play no role in inducing compliance. When $q \geq q_N$, esteem concerns play no role either, as indeed in the standard model. However, when $q \in [q_1, q_N)$ the full compliance equilibrium is supported only with beliefs satisfying $\overline{\theta}_B < \overline{\theta} = \overline{\theta}_G$, i.e., lower moral types conform because of the risk of 'looking bad' if they were to be found negligent. In particular, at the threshold q_1, $\overline{\theta}_B = 0$ is the unique out-of-equilibrium belief consistent with full compliance. Injurers with $\theta = 0$ are then induced to comply with due care because of the threat of being perceived as the worst moral type (which they are) if found negligent. Thus, esteem concerns provide useful incentives when $q < q_N$.

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

Figure 3 illustrates the expected utility of type $\theta = 0$ in a full compliance equilibrium with $q > q_1$. The utility from complying with due care is $\overline{U}_c$. The expected utility from not complying is $\overline{U}_{nc}(q)$; if he were not to comply, the best this injurer could do is to choose the precaution level denoted by $p_{nc}(q)$. The utility from not complying is drawn given the out-of-equilibrium belief $\overline{\theta}_B = 0$, so that the reputational penalty from being found negligent is $\Delta = \overline{\theta}$ (when $q > q_1$ a smaller reputational penalty would obviously also induce compliance). The utility difference between complying and not complying is then

$$\overline{U}_c - \overline{U}_{nc}(q) = qp_{nc}(q)[L + \beta\overline{\theta}] - [c(p^*) - c(p_{nc}(q))].$$

By not complying the injurer would save on precaution costs, but this is more than compensated by the risk of legal damages and of disesteem. When q is reduced, $\overline{U}_c$ does not change but $\overline{U}_{nc}(q)$ increases. At $q = q_1$ both are equal.

When enforcement is smaller than q_1, both moral and reputational incentives play a role. When $q \geq q_0$, the more morally concerned types — those with θ above a threshold that depends on q — comply with due care. Intuitively, it is 'less costly' to comply for more morally concerned individuals.

Less morally concerned injurers do not comply, but their carelessness is mitigated by both moral and esteem concerns as under an imperfectly enforced strict liability regime.

The intuition for the result is that a ruling of negligence provides sharper information about the injurer's behavior than the mere occurrence of harm. Compared with strict liability or no liability, the negligence regime may 'crowd in' normative motivations through the greater role of esteem concerns. When the probability of enforcement is not too low, a small departure from due care generates the risk of a large reputational penalty. As a result, injurers or at least some of them will choose to 'conform', i.e., to pool on the due care precaution level.

The outcome would remain the same qualitatively if involvement in causing harm were public information independently of lawsuits. The event G would then be partitioned 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 $\overline{\theta}_{G_1} = \overline{\theta}_{G_2} = \overline{\theta}$. When not all injurers comply, it is straightforward to see that $\overline{\theta}_{G_1} > \overline{\theta}$, but whether G_2 would be good or bad news depends on the probability of enforcement.

5 Discussion

The different legal regimes (no liability, strict liability and the negligence rule) were compared in terms of how close the injurers' precautions were to the efficient precaution level of the standard model. When injurers have moral and esteem concerns, however, it is not clear that p^* is still the appropriate target. We now compare the different regimes in an explicit utilitarian framework.

We also discuss a related but more intricate issue. In our analysis, the

moral prescription was exogenously given. We then examined how legal incentives interact with normative concerns to deter careless behavior. We did not consider the degree to which the 'legal norm' was consistent with or differed from the moral norm, nor the possibility that the 'legal norm' could influence the individuals' moral preferences.

Comparison of regimes. The costs of enforcing formal sanctions would naturally bear on the comparison of legal regimes. Nevertheless, we will continue to abstract from such costs.

Suppose that injurers can also be victims; that is, they can themselves suffer harm caused by other agents. For instance, the risk generating activity under consideration is an activity such as driving which everyone engages in.[16] Let $\overline{U}_j(\theta)$ be the equilibrium expected utility of a type-θ injurer as defined in the previous sections, where j denotes the legal regime. Because individuals are now potentially both injurer and victim, the pecuniary part of one's utility is modified to take into account the expected loss that the individual faces due to the actions of others. Under no liability, the expected wealth of the type-θ individual is then $b - c(\widehat{p}_0(\theta)) - \overline{p}_0 L$. The term $\overline{p}_0 L$ is the expected loss the individual faces as potential victim, given the average probability $\overline{p}_0$ that another individual will cause harm in the no-liability equilibrium. Under the strict liability and negligence regimes, the expected wealth of the type-θ individual is

$$b - c(\widehat{p}_j(\theta)) - \varphi_j(\widehat{p}_j(\theta))L - \int_0^{\theta_h} \left[\widehat{p}_j(\theta) - \varphi_j(\widehat{p}_j(\theta))\right] L f(\theta)\, d\theta, \quad j = S, N.$$

The last term is the expected loss as victim net of the legal damages that may eventually be awarded.

Per-capita welfare is[17]

$$W_j = \int_0^{\theta_h} \overline{U}_j(\theta) f(\theta)\, d\theta, \quad j = 0, S, N.$$

Under no liability, this is easily seen to yield

$$W_0 = (b - \overline{c}_0 - \overline{p}_0 L) - \int_0^{\theta_h} \theta \widehat{p}_0(\theta) L f(\theta)\, d\theta + \beta\overline{\theta},$$

where $\overline{c}_0$ is the average precaution cost. The expression inside the first parentheses is the average net wealth. The middle term is the average moral disutility from causing uncompensated harm. The last term is the average utility from social esteem; reputational benefits and penalties cancel out. Similarly, under the strict liability and negligence regimes, total welfare is

$$W_j = (b - \overline{c}_j - \overline{p}_j L) - \int_0^{\theta_h} \theta \left[\widehat{p}_j(\theta) - \varphi_j(\widehat{p}_j(\theta))\right] L f(\theta)\, d\theta + \beta\overline{\theta}, \quad j = S, N.$$

Consider the perfectly enforced strict liability regime. All individuals then take efficient care and full compensatory damages are always paid. Hence

$$W_S = b - p^* L - c(p^*) + \beta\overline{\theta} \equiv W^*, \tag{17}$$

Clearly, welfare cannot be greater than W^*: wealth is maximized and there is no moral disutility from having caused harm. The perfectly enforced negligence rule also maximizes wealth. However, total welfare is now

$$W_N = b - p^* L - c(p^*) - \overline{\theta} p^* L + \beta\overline{\theta} = W^* - \overline{\theta} p^* L. \tag{18}$$

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 as injurers suffer moral disutility from inflicting uncompensated harm.

Proposition 4 *When liability rules are perfectly enforced,* $W_0 < W_N < W_S$.

Strict liability yields greater welfare than the negligence rule 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 (given Assumption 1). In turn, welfare is greater under negligence than under no liability because average wealth is larger and because the moral cost of imposing uncompensated harm is smaller; both results follow from the fact that $\widehat{p}_0(\theta) > p^*$ for all types.

When enforcement is imperfect, the welfare comparison is not as straightforward. For instance, it is not clear how negligence compares with strict liability with the same probability of enforcement. Suppose $q \geq q_1$ 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 moral norms. In the above analysis, legal liability reduced to a pure system of monetary penalties contingent on some evidence. We now inquire about the values underlying legal liability.

The moral prescription we postulated was that harming others should be avoided and that one should compensate for the harm that one does cause. Compliance with the prescription would arise spontaneously if all individuals were sufficiently 'morally concerned', say with $\theta \geq 1$. In a less than ideal world, such individuals rarely exist. However, strict legal liability can in principle (when enforcement is perfect) induce individuals to behave in perfect conformity with the moral prescription. Thus, one could say that strict liability 'expresses' perfectly the underlying moral norm.

This is not so with the negligence rule we considered. Under this rule, there is no 'legal wrongdoing' when harm occurs and the injurer's actions complied with the legal due care standard. In our analysis, moral values were taken to be unaffected by the legal norm; that is, the individuals' conception of 'moral wrongdoing' remained based solely on the prescription 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'. For instance, a driver may feel bad from hitting a pedestrian even though he is not found negligent.

It has been argued that legal rules have normative power in that they affect behavior not only by shaping the material payoffs but also by directly influencing motives. If law expresses values, it could change the individuals' perception of the moral prescription.[18] In the present context, one possibility is that the legal norm of due care modifies the interpretation of wrongdoing. Individuals who comply with the legal rule of conduct experience no moral disutility from causing harm. Because of the legal norm defined by due care, the individual's utility in (4) is now

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

where $\delta(p) = 0$ if $p \leq p^*$ and $\delta(p) = 1$ otherwise. Welfare under the negligence rule is then

$$W_N = (b - \overline{c}_N - \overline{p}_N L) - \int_0^{\theta_h} \theta\delta(\widehat{p}_N(\theta)) \left[\widehat{p}_N(\theta) - \varphi_N(\widehat{p}_N(\theta))\right] Lf(\theta)\, d\theta + \beta\overline{\theta}$$

where $\widehat{p}_N(\theta)$ now denotes the equilibrium strategy under the new moral preferences.

Proposition 5 *Suppose individuals have no moral concerns about causing uncompensated harm when they comply with due care under the negligence rule. If the probability of enforcement satisfies $q \geq q_1$ as defined in Proposition 3, $W_N = W^* \geq W_S$ with strict inequality when $q < 1$.*

Everything else equal, compared to the situation where moral preferences are unaffected by the legal norm, complying with due care is as desirable when $\theta > 0$ and it remains the same when $\theta = 0$. Hence, the threshold q_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 now dominates strict liability.

Causation requirement under the negligence rule. Grady (1983) remarked that the standard accident model disregards causation requirements for a ruling of negligence. The issue of causation yields interesting insights in our analysis. As a matter of doctrine, in both common law and civil law systems, injurers are liable for damages if they did not comply with due care and if their negligence can be shown to have 'caused' the accident, i.e., the accident would not have occurred had they not been negligent. To deal with causation requirements, a more elaborate description of the stochastic environment is needed.[19]

We consider the simple model proposed by Kahan (1989); see also Schweizer (2009). In this model, when an accident occurs and the injurer was negligent (i.e., $p > p^*$), the probability that negligence will be found to be the cause of the accident is equal to $(p - p^*)/p$. Given a probability of enforcement q, the expected pecuniary payoff as a function of the precaution level is then

$$\overline{y} = \begin{cases} b - c(p) \text{ if } p \leq p^*, \\ b - c(p) - q(p - p^*)L. \end{cases}$$

By contrast with the 'full liability' formulation discussed until now, there is no discontinuity in the expected payoff function at the due care level. It follows that, in the model without normative motivations, an imperfectly enforced negligence rule yields the same outcome as strict liability with the same probability of enforcement.

We now introduce normative motivations into this model. Suppose the injurer feels no guilt when he complied with due care or when he did not comply but his negligence was not the cause of harm (i.e., $\delta = 0$ in the

notation of the previous section). The expected utility is then

$$\overline{U} = \begin{cases} b - c(p) + \beta\overline{\theta}_G \text{ if } p \leq p^*, \\ b - c(p) - (p - p^*)(q + \theta(1-q))L + \beta[pq\overline{\theta}_B + (1-pq)\overline{\theta}_G] \text{ if } p > p^*. \end{cases}$$

The difference in expected utility between complying and barely not complying is now equal to

$$p^* q \beta(\overline{\theta}_G - \overline{\theta}_B).$$

There is a now a discontinuity in the expected payoff, but it is due solely to the reputational penalty imposed by a ruling of negligence. It is easily seen that the results of Proposition 3 would remain qualitatively the same, except that we now have $q_N = 1$ and that q_0 and q_1 would be larger. When enforcement is above q_1, there is bunching on the legal due care standard only because of image concerns.

6 Concluding remarks

This paper incorporates normative motivations into the economic model of tort rules. We showed that informal motivations and legal sanctions complement one another when each on its own would yield suboptimal precautions. If the substantive laws are well designed but imperfectly enforced, a consideration for the moral duty not to cause uncompensated harm or the attempt to signal such concerns improves efficiency. On the other hand, if compliance with the moral duty is imperfect, appropriate liability rules improve efficiency even if the law is imperfectly enforced. There is no net crowding-out of overall incentives following the introduction of formal sanctions. In particular, when enforcement is imperfect, the negligence rule has the potential to do better than strict liability because image concerns have more bite when potential injurers face the threat of being found negligent.

In the unilateral accident model, potential victims are passive and can do nothing to avoid harm. Who 'caused' harm is then straightforward, although the issue of 'causation' may still arise with respect to legal liability or moral responsibility, as discussed above. How would our analysis extend to the so-called bilateral model where both injurers and victims can take precautions? Whether the injurer 'caused' harm or should feel 'responsible' may then be problematic. If the victim acted foolishly, will the injurer feel bad if an accident occurs? Our results can be extended to the bilateral case through the concept of causation in the manner discussed in the previous section, but now taking into account the behavior of both injurer and victim, or through the δ function describing when moral responsibility is triggered. To illustrate, consider the simple negligence rule (as opposed, say, to contributory negligence) and suppose that the injurer feels no guilt when he has complied with due care but would feel guilty otherwise. Then we would get the same results concerning the role of informal incentives under an imperfectly enforced negligence rule. The same holds *mutatis mutandis* if a non-complying injurer feels guilt only when the victim did not act inefficiently.

One limitation of our analysis is that social interactions arise only through reputational effects, as in Bénabou and Tirole (2006). These effects are endogenous, i.e., they depend on what can be inferred from various occurrences. However, it may well be that 'intrinsic motivation' also depends on one's perception of the extent to which others adhere to the moral norm, a form of reciprocity or conditional cooperation (see Sugden 1984, Fischbacher et al. 2001, and Frey and Torgler 2007). Similarly, the desire to be perceived as virtuous may presumably also depend on the importance of virtue to others. An interesting extension of the present analysis would be to add these channels of social interactions to the accident model.

Appendix

Proof of proposition 1. We complete the argument in the text by proving that, in the no-liability case, the equilibrium Δ belongs to $(0, \theta_h)$. Define

$$\psi(\Delta) \equiv - \frac{\int_0^{\theta_h} \theta[p_0(\theta, \Delta) - \overline{p}_0(\Delta)] f(\theta)\, d\theta}{\overline{p}_0(\Delta)\,(1 - k\overline{p}_0(\Delta))}, \tag{A1}$$

where $\overline{p}_0(\Delta)$ is the average best response over all types. At equilibrium, $\Delta = \psi(\Delta)$. From (12), $p_0(\theta, \Delta)$ is non zero and is strictly decreasing in θ. Hence, $\overline{p}_0(\Delta) \in (0, 1)$ and therefore $\psi(\Delta) \in (0, \theta_h)$ for all $\Delta \in [0, \theta_h]$, implying that any solution belongs to the open interval as well. Existence of a solution follows from the continuity of $\psi(\Delta)$. QED

For the no-liability case, a solution to (A1) is a stable equilibrium if $\psi'(\Delta) < 1$. Clearly, there is at least one solution (e.g., the one with the largest Δ) at which the $\psi(\Delta)$ curve cuts the forty-five degree line from above, implying $\psi'(\Delta) < 1$. The equilibrium penalty depends on k and can be written as $\Delta(k)$. It can be shown that at any stable solution $k\Delta(k)$ is increasing in k, which proves the claim in the text that greater publicity increases incentives to exert care.

Proof of proposition 2. We prove only the last claim. Let Δ_0 and Δ_S be the equilibrium reputational penalties under no liability and strict liability respectively. Incentives are greater under strict liability if

$$[q + \theta(1 - q)]\, L + q\beta\Delta_S > \theta L + k\beta\Delta_0$$

or equivalently

$$q(1 - \theta)L + \beta(q\Delta_S - k\Delta_0) > 0.$$

When $q \geq k$,

$$\begin{aligned} q(1 - \theta)L + \beta(q\Delta_S - k\Delta_0) &\geq q\,[(1 - \theta)L + \beta(\Delta_S - \Delta_0)] \\ &> q\,[(1 - \theta_h)L - \beta\theta_h] > 0. \end{aligned}$$

The last inequality follows from Assumption 1, the second-to-last from $\theta \leq \theta_h$ and the fact that Δ_0, $\Delta_S \in (0, \theta_h)$. QED.

Before proving proposition 3, we first characterize the pattern of compliance under an imperfectly enforced negligence rule.

Lemma 2 *If a type θ' injurer complies with due care, so does a type $\theta'' > \theta'$.*

Proof. Let $\Delta = \overline{\theta}_G - \overline{\theta}_B > 0$. Write the expected utility in (15) as $\overline{U}(p, \theta)$ and let

$$\overline{U}_c(\theta) \equiv \max_{p \leq p^*} \overline{U}(p, \theta) = b - c(p^*) - \theta p^* L + \beta \overline{\theta}_G. \tag{A2}$$

This is the utility reached by a type-θ injurer who complies with due care. Let

$$\overline{U}_{nc}(\theta) \equiv \sup_{p > p^*} \overline{U}(p, \theta). \tag{A3}$$

This is the most an injurer can obtain when he does not comply. Then

$$\overline{U}_{nc}(\theta) = b - c(p_{nc}(\theta)) - p_{nc}(\theta)[(q + \theta(1-q))L + \beta q \Delta] + \beta \overline{\theta}_G \tag{A4}$$

where

$$p_{nc}(\theta) \equiv \begin{cases} P[(q + \theta(1-q))L + \beta q \Delta] \text{ if this is larger than } p^*, \\ p^* \text{ otherwise.} \end{cases} \tag{A5}$$

Let

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

We show that $\zeta(\theta) = 0$ has at most one solution in $[0, \theta_h]$, say $\widetilde{\theta}$, with $\zeta(\theta) > 0$ if $\theta > \widetilde{\theta}$ and $\zeta(\theta) < 0$ if $\theta < \widetilde{\theta}$. Applying the envelope theorem,

$$\zeta'(\theta) = p_{nc}(\theta)(1-q))L - p^* L$$

and therefore

$$\zeta''(\theta) = p'_{nc}(\theta)(1-q))L \leq 0.$$

Let us extend the range of possible values for θ to $[0,1]$ and look for solutions to $\zeta(\theta) = 0$ in this interval. From (A5), $p_{nc}(1) = p^*$. Hence $\zeta(1) = p^*q\beta\Delta > 0$ and $\zeta'(1) = -p^*qL < 0$. Because $\zeta(\theta)$ is concave and $\zeta'(1) < 0$, $\zeta(\theta)$ is (i) either everywhere decreasing in the interval $[0,1]$ or (ii) first increasing and then decreasing. Because $\zeta(1) > 0$, a solution to $\zeta(\theta) = 0$ does not exist in case (i) and at most one exists in case (ii). If one does, it is at some $\widetilde{\theta} < 1$ such that $\zeta'(\widetilde{\theta}) > 0$. Concavity then implies $\zeta(\theta) \leq 0$ for $\theta \leq \widetilde{\theta}$; concavity together with $\zeta(1) > 0$ imply $\zeta(\theta) > 0$ for $\theta \in (\widetilde{\theta}, 1]$. If $\widetilde{\theta} \geq \theta_h$, $\zeta(\theta)$ does not change sign in the interval $[0, \theta_h]$. If $\widetilde{\theta} < \theta_h$, lower types in $[0, \theta_h]$ do not comply, higher types do. QED

Proof of proposition 3. Part (i) follows from lemma 2 and the argument in section 2. To show (ii), let $q \in [q_1, q_N]$ and define

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

This is the expected utility of type $\theta = 0$ if found liable with probability q when bad news yields $\overline{\theta}_B$ and good news yields $\overline{\theta}$. The function is increasing in $\overline{\theta}_B$ and decreasing in q. Let $\widetilde{\theta}_B(q)$ solve

$$v(\widetilde{\theta}_B, k) = b - c(p^*) + \beta\overline{\theta}, \tag{A6}$$

where the right-hand side is the utility of complying when good news yields $\overline{\theta}$. Type $\theta = 0$ complies if $\overline{\theta}_B \leq \widetilde{\theta}_B(q)$. From (3), $\widetilde{\theta}_B(q) = \overline{\theta}$ when $q = q_N$; from (16), $\widetilde{\theta}_B(q) = 0$ when $q = q_1$. Moreover, $\widetilde{\theta}_B(q)$ is increasing and continuous. Thus, when $q \in [q_1, q_N)$, there is an upper bound $\widetilde{\theta}_B(q) < \overline{\theta}$ such that beliefs satisfying $\overline{\theta}_B \leq \widetilde{\theta}_B(q)$ induce compliance when $\theta = 0$; by lemma 2, such beliefs induce overall compliance.

When $q < q_1$, equation (A6) has no solution. In equilibrium some injurers will therefore not comply and both B and G will have positive probability,

implying $\Delta \in (0, \theta_h)$. If some injurers comply, lemma 2 implies the existence of a type threshold $\theta(q) < \theta_h$ above which injurers comply. For q sufficiently close to q_1, such a threshold necessarily exists. For q sufficiently small, however, it does not. To see this, define

$$u(q) \equiv \max_p b - c(p) - p(q + \theta_h(1-q))L + \beta(1-ph)\theta_h.$$

This is the expected utility of the high type $\theta = \theta_h$ when he is found liable with probability q with bad news yielding $\overline{\theta}_B = 0$ and good news yielding $\overline{\theta}_G = \theta_h$, i.e., the anticipated reputational penalty is $\Delta = \theta_h$. The function is strictly decreasing in q. Let q_c be the solution to

$$u(q_c) = b - c(p^*) - \theta_h p^* L + \beta\theta_h. \tag{A7}$$

The right-hand side is the expected utility of the same injurer if he complies. It is easily verified that (A7) has a solution satisfying $0 < q_c < q_1$. Because at equilibrium we must have $\Delta < \theta_h$, even the high type $\theta = \theta_h$ would not comply when $q \leq q_c$. Thus, there exists some $q_0 \in (q_c, q_1)$ as stated in (iii) and (iv). QED

Notes

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

[2]The distinction between guilt and shame as emotional sanctions has been made in many contexts, e.g., Kandel and Lazear (1992). We follow Elster (2007) in describing a prescription as a 'moral rule' when a violator experiences guilt; see also Shavell (2002), Kaplow and Shavell (2007) or for that

matter Adam Smith (1790). In the law and economics literature, such prescriptions are more often referred to as 'social norms', e.g., Posner (1997), Posner and Rasmusen (1999), and Posner (2000). They are of course social norms to the extent that violation elicits disapprobation or contempt.

[3]See Rasmusen (1996), Bohnet et al. (2001), Brekke et al. (2003), Harel and Klement (2007), Lazzarini et al. (2004), Zasu (2007), and for a recent survey McAdams and Rasmusen (2007).

[4]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 mitigate harm.

[5]In the next section, a large b also implies that individuals will not be deterred from engaging in the activity out of moral or image concerns.

[6]Equivalently one could have written $e = \overline{\theta}_I - \overline{\theta}$, which amounts to adding a constant to the utility function. The utility component βe could also be interpreted as the economic value of one's reputation; as such it would then be part of the 'material payoff'. Tort law can obviously also be supplemented by the standard reputation mechanism (see for instance Jin and Leslie 2003). However, we will stick to the interpretation that individuals have a preference for social esteem for its own sake.

[7]The action of not compensating is out-of-equilibrium. Any belief $\overline{\theta}_0 \in [\theta_l, \overline{\theta}]$ supports the equilibrium.

[8]When compensation is paid, the reputational utility is $\beta\overline{\theta}$. When no compensation is paid (an out-of-equilibrium action), it is $\beta\overline{\theta}_0$, say with $\overline{\theta}_0 = \theta_l$. Hence, compensation is the preferred action when β is sufficiently large.

[9]To illustrate in a different context, replace 'causing uncompensated harm' with 'stealing'. Standard *homo economicus* is deterred from stealing only because of the threat of criminal sanctions (or of retortion by others). If the threat is nonexistent, he will steal. When normative concerns are at work, the fear of 'feeling bad' and 'looking bad' may in principle deter stealing.

Under a preference restriction such as assumption 1, however, the individual will nevertheless steal even though he knows he will 'feel bad' and 'look bad'.

[10]See Shavell (1993). In a related context, Daughety and Reinganum (2010) analyze the effect of privacy versus publicity about one's actions.

[11]This is consistent with the D1 criterion (Cho and Kreps 1987) as used in Bernheim's (1994) theory of conformity.

[12]It need not be unique. We will nevertheless loosely refer to "the" equilibrium. Should there be multiple equilibria, we focus on the one with the largest reputational penalty (see the Appendix for details). Social interaction models often exhibit multiple equilibria; e.g., Rasmusen (1996), Glaeser et al. (1996) or Bénabou and Tirole (2006).

[13]Suppose that one's engagement in the activity becomes public knowledge only when involvement in causing harm is observed. If type θ does not engage, his utility is therefore $\beta\overline{\theta}_G$; if he does, it is given by (11) with $p = \widehat{p}(\theta)$. Since $\Delta < \theta_h$, it is then easily seen that Assumption 1 and $b > c(p^*) + p^*L$ imply that all types engage in the activity.

[14]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 analyzed by Daughety and Reinganum (1999, 2002).

[15]In the limit, Δ_S tends to zero as q approaches unity, but Δ_0 always remains strictly positive.

[16]Alternatively, one could have two classes of agents, potential victims and potential injurers, and sum utility over both classes.

[17]For the purpose of computing welfare, we follow Bénabou and Tirole (2006) in adding all three components of the utility function. See also Kaplow and Shavell (2007).

[18]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).

[19]Shavell (1985) makes the point that causation is often highly uncertain and that courts often find the injurer liable if one cannot rule out that the accident was due to the injurer's negligence. The usual 'full liability' model is then an appropriate description.

References

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

Bernheim, B.D. 1994. "A Theory of Conformity," 102 *Journal of Political Economy* 905-953.

Bohnet I., B. Frey and S. Huck. 2001. "More Order with Less Law: On Contract Enforcement, Trust and Crowding," 95 *American Political Science Review* 131-144.

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

Brown, J. P. 1973. "Toward an Economic Theory of Liability," 2 *Journal of Legal Studies* 323-349.

Cho I-K and D.M. Kreps. 1990. "Strategic Stability and Uniqueness in Signaling Games," 50 *Journal of Economic Theory* 381-413.

Cooter, R. 2000. "Expressive Law and Economics," 27 *Journal of Legal Studies* 585-608.

Cooter, R. and A. Porat. 2001. "Should Courts Deduct Nonlegal Sanctions from Damages?" 40 *The Journal of Legal Studies* 401–422.

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

_____. 2002. "Informational Externalities in Settlement Bargaining: Confidentiality and Correlated Culpability," 33 *RAND Journal of Economics* 587-604.

_____. 2010. "Public Goods, Social Pressure, and the Choice Between Privacy and Publicity," 2 *American Economic Journal: Microeconomics* 191-222.

Elster, J. 2007. *Explaining Social Behavior: More Nuts and Bolts for the Social Science*. Cambridge: Cambridge University Press.

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

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

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

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

Gneezy, U. and A. Rustichini. 2000. "Pay Enough or Don't Pay at All," 115 *Quarterly Journal of Economics* 791-810.

Grady, M. F. 1983. "A New Positive Economic Theory of Negligence," 92 *Yale Law Journal* 799-829.

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

Jin, G.Z. and P. Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards", 118 *Quarterly Journal of Economics* 409-451.

Kahan M. 1989. "Causation and Incentives to Take Care Under the Negligence Rule," 18 *Journal of Legal Studies* 427-447.

_____. 1997. "Social Influence, Social Meaning, and Deterrence," 83 *Virginia Law Review* 349-395.

Kandel, E. and E.P. Lazear. 1992. "Peer Pressure and Partnerships," 100 *Journal of Political Economy* 801-817.

Kaplow, L. and S. Shavell. 2002. *Fairness versus Welfare*. Cambridge: Harvard University Press.

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

Landes, W. and R.A. Posner. 1987. *The Economic Structure of Tort Law*. Cambridge: Harvard University Press.

Lazzarini S., G. Miller and T. Zenger. 2004. "Order with Some Law: Complementarity versus Substitution of Formal and Informal Arrangements," 20 *Journal of Law, Economics and Organization* 261- 281.

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," 2 *Journal of Empirical Legal Studies* 87-113.

McAdams, R.H. and E. B. Rasmusen. 2005. "Norms in Law and Economics", in A.M. Polinsky and S. Shavell, eds., *Handbook of Law and Economics.* Amsterdam: North Holland.

Posner, R.A. 1997. "Social Norms and the Law: An Economic Approach," 87 *American Economic Review* 365-369.

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

Posner, E. 2000. *Law and Social Norms.* Cambridge: Harvard University Press.

Rasmusen, E.B. 1996. "Stigma and Self-Fulfilling Expectations of Criminality," 39 *Journal of Law and Economics* 519-544.

Schweizer, U. 2009. "Legal Damages for Losses of Chances," 29 *International Review of Law and Economics* 153-160.

Shavell, S. 1985. "Uncertainty over Causation and the Determination of Civil Liability," 28 *Journal of Law and Economics* 587-609.

_____. 1987. *Economic Analysis of Accident Law.* Harvard: Harvard University Press.

_____. 1993. "The Optimal Structure of Law Enforcement," 36 *Journal of Law and Economics* 255-287.

_____. 2004. "Law versus Morality as Regulators of Conduct," 4 *American Law and Economics Review* 227-257.

_____. 2007. "Liability for Accidents," in A.M. Polinsky and S. Shavell, eds., *Handbook of Law and Economics*. Amsterdam: North Holland.

Smith, A. 1790 (1976). *The Theory of Moral Sentiments*, 2nd ed. Oxford: Clarendon.

Sugden, R. 1984. "Reciprocity: The Supply of Public Goods through Voluntary Contributions," 94 *Economic Journal* 772-787.

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

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

Deffains and Fluet, Figure 1 of 3

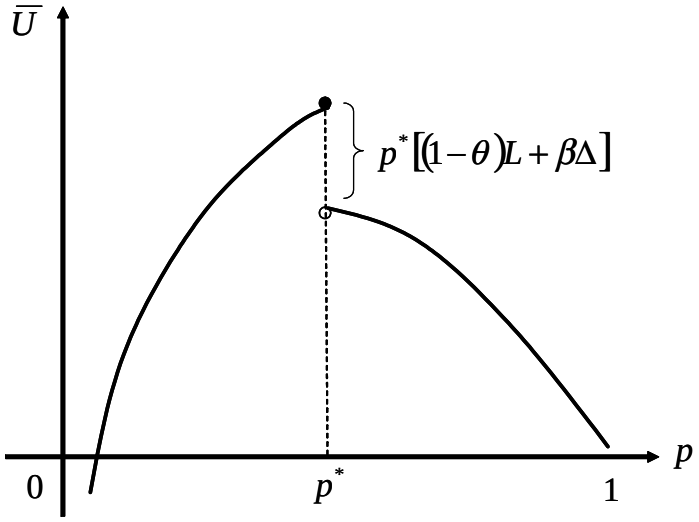

Figure 1: Expected utility under the negligence rule

Deffains and Fluet, Figure 2 of 3

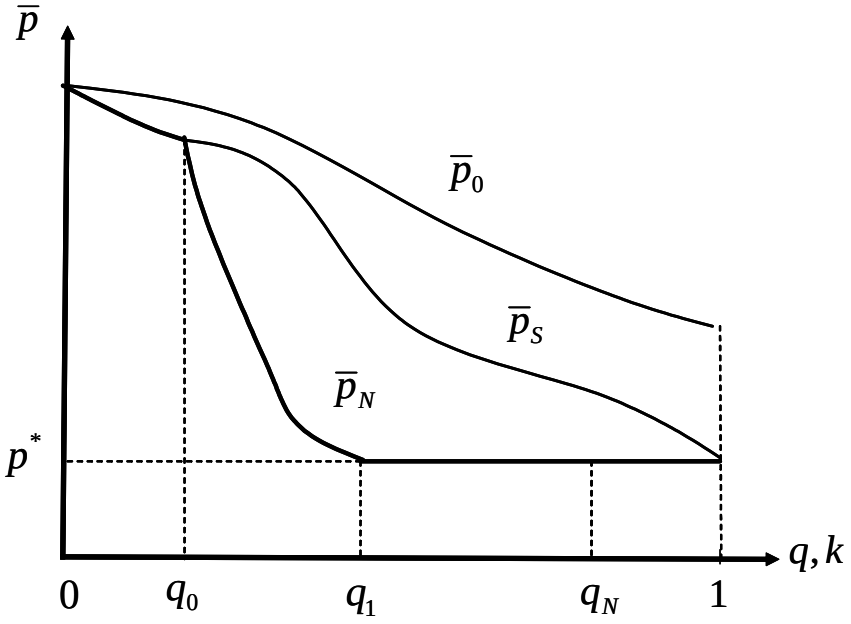

Figure 2: Liability regimes

Deffains and Fluet, Figure 3 of 3

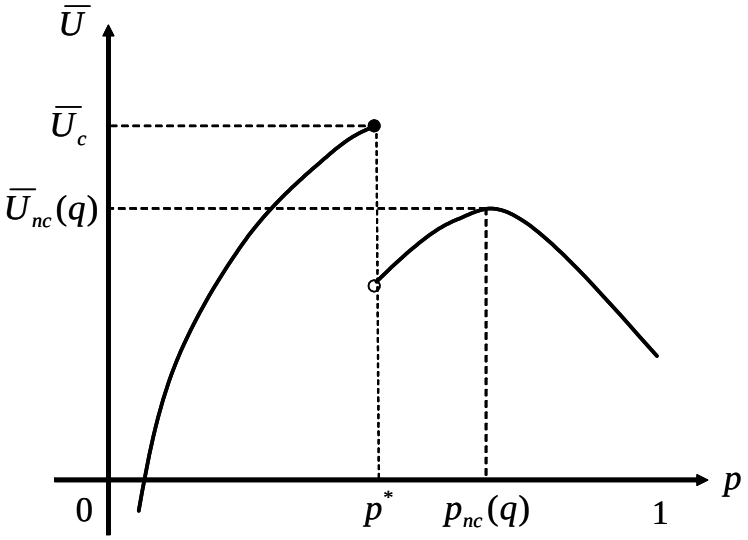

Figure 3: Expected utility for type $\theta = 0$ when $q > q_1$