NBER WORKING PAPERS SERIES

DECOUPLING LIABILITY:
OPTIMAL INCENTIVES FOR CARE AND LITIGATION

A. Mitchell Polinsky

Yeon-Koo Che

Working Paper No. 3634

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1991

Stanford University and National Bureau of Economic Research; and Stanford University, respectively. Both authors' research on this article was supported by the John M. Olin Program in Law and Economics at Stanford Law School. Helpful comments were provided by Keith Hylton, Avery Katz, Daniel Rubinfeld, Steven Shavell, Warren Schwartz, and participants at a seminar at Stanford. This paper is part of NBER's research program in Law and Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3634
February 1991

DECOUPLING LIABILITY: OPTIMAL INCENTIVES FOR CARE AND LITIGATION

ABSTRACT

A "decoupled" liability system is one in which the award to the plaintiff differs from the payment by the defendant. The optimal system of decoupling makes the defendant's payment as high as possible. Such a policy allows the award to the plaintiff to be lowered, thereby reducing the plaintiff's incentive to sue -- and hence litigation costs -- without sacrificing the defendant's incentive to exercise care. The optimal award to the plaintiff may be less than or greater than the optimal payment by the defendant. The possibility of an out-of-court settlement does not qualitatively affect these results. If the settlement can be monitored, it may be desirable to decouple it as well.

A. Mitchell Polinsky
Stanford Law School
Stanford University
Stanford, CA 94305
and
National Bureau
of Economic Research

Yeon-Koo Che
Stanford Law School
Stanford University
Stanford, CA 94305

1. <u>Introduction and Summary</u>

In suits between private individuals, liability usually is "coupled" in the sense that, aside from the parties' litigation costs, a successful plaintiff receives what the defendant pays. This article studies a system of "decoupled" liability -- in which the plaintiff is awarded an amount different from what the defendant is made to pay. If the plaintiff is awarded less than what the defendant pays, the government obtains the difference; if the plaintiff is awarded more, the government provides the difference.

Decoupled liability already occurs in certain circumstances. For example, in several states punitive damages are decoupled, with the plaintiff receiving 25% to 67% of the punitive damage amount paid by the defendant (the percentage depends on the state); the rest goes to the state treasury or to a public compensation fund.[1] Also, decoupled liability has been proposed in the context of private antitrust suits.[2]

The rationale for decoupling liability that will be investigated here is easily explained. Consider any level of liability when liability is coupled. This level of liability will determine the incentive of the victim to sue (the higher the award, the greater the incentive) and the incentive of the injurer to take care. The parties' behavior in turn will determine the level of social costs -- assumed to be the sum of the injurer's cost of taking care,

[1] The plaintiff receives 67% of the punitive damage amount in Colorado, 40% in Florida, and 25% in Iowa (under specified circumstances). The remainder goes to the State General Fund in Colorado, the General Revenue Fund or the Public Medical Assistance Trust Fund in Florida, and the Civil Reparations Trust Fund in Iowa. See Colo. Rev. Stat. sec. 13-21-102 (Supp. 1987); Fla. Stat. sec. 768.73(2)-(4); and Iowa Code se. 668A.1(1)-(2).

[2] See, for example, Schwartz (1980, pp. 1092-1096; 1981, pp. 10-15), Salop and White (1986, p. 1037), and Polinsky (1986); these articles are briefly commented upon in note 28 below. (The term "decoupling" apparently was first used by Salop and White.)

the victim's expected harm, and the parties' expected litigation costs.[3]

Now consider decoupling liability, starting at the specified level of coupled liability. First raise the amount paid by the injurer, which will cause him to take more care. Then lower the amount awarded to the victim -- which will reduce his incentive to sue and thereby cause the injurer to take less care -- until the injurer's care is back to its level under coupled liability. Since the level of care is the same under this decoupled system and the original coupled system, so is the injurer's cost of taking care and the victim's expected harm. But since the plaintiff is awarded less under the decoupled system, he will sue less often and, consequently, litigation costs will be lower. Thus, starting from any level of coupled liability, there always exists a decoupled system of liability that reduces social costs.

This logic also can be used to establish one of the main results of the article -- that in the optimal system of decoupled liability the defendant's payment is as high as possible. For if the payment by the defendant is not at its upper bound, it is possible to raise the defendant's payment and lower the plaintiff's award in such a way that the injurer's care is not affected but the parties' expected litigation costs are lowered.

With the payment by the defendant set at its upper bound, the optimal award to the plaintiff depends on how the plaintiff's award affects the injurer's care (through the plaintiff's incentive to sue) and the parties' litigation costs. The optimal award to the plaintiff minimizes the sum of the cost of the defendant's care, the expected harm to the victim, and the expected litigation costs of the parties.

---

[3] See generally Polinsky and Rubinfeld (1988).

It will be shown that the optimal award to the plaintiff may be less than or greater than the optimal payment by the defendant. To understand why either relationship is possible, consider two limiting cases involving the level of harm. As the level of harm approaches zero, the optimal award to the plaintiff must approach zero; it is not worthwhile to encourage the plaintiff to sue since the value of inducing the injurer to take care becomes small, and litigation is costly. In this case, the optimal award to the plaintiff will be less than the optimal payment by the defendant (which is at its upper bound). Conversely, as the level of harm becomes large, suits become more valuable, and it is optimal to continue to raise the award to the plaintiff. In this case, the optimal award to the plaintiff will exceed the optimal payment by the defendant.

Thus far, the discussion has assumed implicitly that all suits result in trials. In practice, however, most cases settle out of court. It will be demonstrated that the possibility of a settlement does not affect the result that the optimal payment by the defendant if the case goes to trial is as high as possible. This is for two reasons. First, as before, by making the defendant's payment as high as possible, the award to the plaintiff -- and his incentive to sue -- can be lowered. This will reduce either trial costs or settlement costs (which are assumed to be positive). Second, given a suit, by raising the defendant's trial payment and lowering the plaintiff's trial award, the likelihood of a settlement is enhanced because the defendant will be willing to pay more in settlement and the plaintiff will be willing to accept less. Increasing the likelihood of a settlement is beneficial because settlement costs are less than trial costs.

In addition to decoupling the trial outcome, a court sometimes may be able to monitor, and therefore decouple, the settlement.[4] (For instance, in class action suits, settlements often have to be approved by the court.) It will be shown that if suits otherwise would not settle, it is beneficial to be able to decouple settlements. This is because, by decoupling the settlement as well as the trial outcome, the likelihood of a settlement can be increased. For example, given any decoupled trial outcome, settling can be made more attractive if the settlement amount paid by the defendant to the plaintiff is supplemented by the government.

The article is organized as follows. Section 2 introduces the basic model, in which suits are assumed to result in trials, and Section 3 derives the optimal system of decoupling in this context. Section 4 extends the model to include the possibility of settlements. Section 5 derives the optimal system of decoupling in the extended model, first when only the trial outcome can be decoupled, and then when both the trial outcome and the settlement can be decoupled. Section 6 contains several concluding remarks, including about the prior literature that discusses decoupling liability, and the relationship between our decoupling analysis and the economic theory of public enforcement.

2. The Basic Model

There is one risk-neutral injurer and many risk-neutral potential victims. The injurer chooses a level of care that affects the probability of

---

[4] Although we will refer to a "court" as having control over the policy instruments, in practice liability may be decoupled by an administrative agency or the legislature. The decoupling of antitrust damages, for example, presumably would be implemented through Congressional legislation.

an accident.[5] If an accident occurs, one of the potential victims is harmed (hereafter frequently referred to as "the victim"). The level of harm is fixed and the same for all potential victims. Let

c = injurer's level of care,

$p(c)$ = probability of an accident ($p' < 0$; $p'' > 0$),

ℓ = loss if an accident occurs.

Care is measured in units that cost one dollar each, so c also represents the injurer's cost of care.

If the victim sues the injurer, each side bears its own legal costs.[6] The injurer's cost of litigation is fixed. Each potential victim's cost of litigation also is fixed, but is assumed to vary among victims.[7] This variation might be attributed, for example, to differences among individuals in the value of their time or to differences among lawyers in their fees. Let

a = potential victim's trial cost ($a > 0$),

$f(a)$ = probability density of a ($f(a) > 0$ for all $a > 0$),[8]

b = injurer's trial cost ($b > 0$).

---

[5] Although the present article discusses liability in the context of accidents, it will be clear that the analysis also generally applies to non-accidental harms, such as antitrust violations.

[6] The basic ideas in this article also would apply under other rules for allocating legal costs, such as the rule that the loser pays the winner's costs.

[7] An equivalent interpretation of this assumption is that there is one potential victim whose litigation cost is uncertain before an accident but known after an accident. (An alternative assumption that would generate similar results is that the level of harm varies among potential victims, rather than their cost of litigation.)

[8] The assumption that a has positive density for all positive values of a is made mainly for expositional convenience. See note 14 below for a discussion of how the results would be affected if this assumption were changed.

It is assumed without loss of generality that the victim will prevail at trial if he brings a suit.[9] The plaintiff then receives an award and the defendant makes a payment. There is some upper bound on the defendant's payment. Let

x = award to the plaintiff at trial,

y = payment by the defendant at trial,

m = maximum possible payment of defendant ($y \leq m$).

The defendant's payment may be bounded for any number of reasons -- his limited wealth, considerations of fairness, and so forth. For purposes of our analysis, it does not matter which reason applies or what the bound is.

If there is an accident, the victim will bring a suit if his trial cost is less than the award he will receive at trial -- that is, if $a < x$.[10] Thus, the probability that a suit will be brought is $F(x)$, where $F(.)$ is the cumulative distribution of a.

Given the victim's suit decision and the payment the injurer will have to make at trial, the injurer will choose his level of care to minimize the sum of his cost of care, his expected payment at trial, and his expected trial cost:

$$(1) \qquad \min_{c} \; c + p(c)F(x)[y + b].$$

The social problem is to choose the award to the victim and the payment by the injurer that minimize the sum of the injurer's cost of taking care, the

[9] If the victim were to win at trial with probability less than one, the qualitative results of our analysis would not be affected. The assumption in the text corresponds to the rule of strict liability; for a discussion of how the analysis would apply to a rule of negligence, see comment (a) in section 6 below.

[10] There is no loss of generality in assuming that the victim will not sue if $a = x$.

victim's expected harm, and the parties' expected trial costs.[11] This problem is solved subject to the constraints that there is some upper bound on the payment by the injurer, and that the victim and the injurer are each maximizing his own welfare.

Thus, the social problem can be written as

$$(2) \qquad \min_{x,y} \; c + p(c)[\ell + \int_0^x af(a)da + F(x)b],$$

where it is understood that c is determined by x and y according to (1), and that $y \leq m$.[12] The integral in (2) is the expected value of the potential victims' trial cost, conditional on a suit being brought (that is, for values of a between 0 and x).

The optimal values of x and y will be denoted x^* and y^*. It is assumed that x^* is positive and unique.

3. Optimal Decoupling in the Basic Model

It will first be shown that the optimal payment by the injurer at trial is as high as possible: $y^* = m$.

This can be proved by contradiction. Suppose that the optimal value of y were $y_0 < m$ and that the optimal value of x were some $x_0 > 0$. Then, if an accident occurs, the sum of the injurer's expected payment at trial and his expected trial cost is $F(x_0)[y_0 + b]$. Now raise y_0 to m and lower x_0 to x_1 such that

---

[11] The injurer's payment and the victim's award are not added to or subtracted from this sum because they are transfers of income rather than social costs or benefits.

[12] Both (1) and (2) already incorporate the constraint that the victim will bring a suit if $a < x$.

(3) $$F(x_0)[y_0 + b] = F(x_1)[m + b].$$

By construction, the expected costs borne by the injurer if there is an accident are unchanged. Therefore, the injurer will choose the same level of care as before; see (1). Now observe from the social problem (2) that the sum of the injurer's cost of care and the victim's expected loss is unaffected, but that, because x is lower, both the victim's and the injurer's expected trial cost are reduced. Thus, the original y_0 and x_0 could not have been optimal.

Given $y^* = m$, the optimal award to the victim, x^*, can be determined by minimizing the objective function in (2) just over x, where c now is determined by x from

(4) $$\underset{c}{\text{MIN}} \quad c + p(c)F(x)[m + b].$$

It is clear from (4) that the injurer's care c is increasing in the victim's award x (since raising x raises F(x), the probability that a suit will be brought). Let c(x) be the solution to (4), with $c' > 0$.

The first-order condition (recall that a unique interior solution is assumed) that determines the optimal award to the victim, x^*, can be written as

(5) $$-p'c'[\ell + \int_0^x af(a)da + F(x)b] = c' + pf(x)[x + b].$$

The left-hand side of (5) is the marginal benefit of raising x. As x goes up, the injurer's care rises and the probability of an accident therefore falls. The fall in the accident probability reduces the expected harm to the victim and the expected trial cost of both parties (trials can only occur if an accident occurs). The right-hand side of (5) is the marginal cost of raising

x. The marginal cost consists of the increased care that the injurer is induced to take, and the increase in the parties' expected trial costs caused by the greater likelihood that a suit will be brought (the increase in trial costs is $x + b$ because, for the "marginal" suit, $a = x$).

It should be clear at this point that there is no simple relationship between the optimal award to the victim, x^*, and the optimal payment by the injurer, y^*. The factors that determine x^* -- such as the responsiveness of the accident probability to the injurer's choice of care, and the magnitude of the parties' trial costs -- may have nothing per se to do with the factors that determine y^* -- such as the injurer's wealth or considerations of fairness.

In general, x^* may be less than or greater than y^*. To illustrate the former possibility, consider what happens to x^* and y^* as the victim's loss, ℓ, approaches zero. For the usual reason, y^* remains at m. But x^* must approach zero as ℓ approaches zero. This can be demonstrated by contradiction. Suppose that x^* is bounded away from 0, say by $\bar{x} > 0$. This implies some minimum level of care, say $\bar{c} > 0$.[13] Thus, social costs at x^* are at least $\bar{c}$. Compare this to social costs when $x = 0$; then, since there are no suits, c is zero and social costs are simply $p(0)\ell$. But as ℓ approaches 0, these social costs approach 0, and become less than $\bar{c}$. Therefore, x^* must also approach 0 as ℓ approaches zero, showing that for ℓ

[13] That $\bar{c}$ must be positive can be proved by contradiction. Suppose $x^* > 0$ and $\bar{c} = 0$. Then social welfare could be improved by setting $x = 0$ (since the level of care would be the same and expected trial costs would be lower). So it must be that $c > 0$ when $x^* > 0$.

low enough, $x^* < y^*$.[14]

By similar reasoning, it can be demonstrated that as ℓ tends to infinity, x^* tends to infinity. The details are omitted here, but the intuition is straightforward. As ℓ tends to infinity, the value of taking additional care to reduce the probability of an accident increases without bound. The only way to induce the defendant to take more care is by raising the award to the plaintiff, x, so that the defendant will be sued with a higher probability if an accident occurs. (The defendant's payment cannot be raised because $y^* = m$.) Therefore, x^* also must tend to infinity as ℓ tends to infinity, showing that for ℓ sufficiently large, $x^* > y^*$.[15]

In general, as suggested by the preceding discussion, x^* is an increasing function of the magnitude of the loss, ℓ.[16] Thus, for accidents with relatively low losses, x^* will be less than y^*, and for accidents with relatively high losses, x^* will be greater than y^*.

---

[14] If the trial costs of potential victims have a positive lower bound, then x^* would tend to that lower bound as ℓ tends to zero. Assuming this lower bound is less than $y^* = m$, then for ℓ low enough, $x^* < y^*$, as claimed. However, if the lower bound exceeds m, then x^* always would exceed y^*. (Analogous observations apply to the discussion in the next paragraph if the trial costs of victims have an upper bound.)

[15] A potential problem with setting x greater than y is that this creates an incentive for individuals to "fabricate" harms -- to claim that an accident has occurred when one has not -- in order to obtain the implicit government subsidy equal to x - y. If individuals cannot be deterred from fabricating harms (say by criminal penalties), it may be desirable to narrow the "gap" between x and y.

[16] It can be seen from (5) that raising ℓ does not affect the marginal cost of raising x, but it does increase the marginal benefit of raising x.

4. The Extended Model

The model of section 2 will be modified to allow for the possibility of a settlement in the following way. After an accident occurs, the plaintiff makes a "take it or leave it" settlement demand.[17] If it is accepted by the defendant, the case is settled for this amount. During the settlement process, both parties incur settlement costs (which are assumed to be less than their respective trial costs).[18] If the plaintiff's settlement demand is rejected by the defendant, the plaintiff then decides whether to go to trial or to drop the suit. For simplicity, both parties are assumed to have perfect information (including about each other's litigation costs). Let

s = plaintiff's settlement demand,

$\alpha(a)$ = plaintiff's settlement cost ($0 < \alpha(a) < a$),[19]

β = defendant's settlement cost ($0 < \beta < b$).

It is assumed that α is increasing in a (a plaintiff's trial cost), and that the gap between a and α also is increasing in a. These assumptions would be satisfied, for example, if a plaintiff's settlement cost is a constant fraction of his trial cost.

First observe that, as before, a plaintiff will bring a suit if and only

[17] This assumption is not as special as it may appear; results qualitatively similar to those discussed in this article generally would occur if the injurer made a take-it-or-leave-it settlement offer (but see note 23 below).

[18] It will be seen that decoupling liability always is socially valuable when settlement costs are positive, whereas if settlement costs are zero, there are some circumstances in which decoupling liability is socially valuable and other circumstances in which it is not needed. The assumption that settlement costs are positive is made both to avoid the additional complexity of having to distinguish between these two sets of circumstances and because it is the more realistic assumption.

[19] When there can be no confusion, $\alpha(a)$ will be written simply as α.

if his trial cost is less than his trial award: $a < x$. This result can be explained as follows. If a is less than x, then clearly the plaintiff would bring a suit even if it were to result in a trial; since a settlement makes him at least as well off as the trial outcome (otherwise he would not agree to the settlement), his incentive to sue is at least as great if the suit might result in a settlement. If a is greater than x, then the plaintiff would not bring a suit if it were to result in a trial. The defendant, knowing this, will reject any settlement demand; and the plaintiff will then drop the suit. In other words, if the plaintiff's trial cost exceeds his award at trial, he cannot obtain a settlement from the defendant because his threat to sue is not credible. Thus, the plaintiff will bring a suit if and only if $a < x$.

Given a suit, next consider whether a settlement is feasible. If the plaintiff goes to trial his net gain is $x - a$, the award less his trial cost. If he settles, his net gain is $s - \alpha$, the settlement amount less his settlement cost. Thus, the plaintiff will prefer to settle if $s - \alpha \geq x - a$ or, equivalently, $s \geq x - (a - \alpha)$.[20] If the defendant goes to trial his total payment is $y + b$, the payment to the government plus his trial cost, whereas if he settles his total payment is $s + \beta$, the settlement amount plus his settlement cost. Thus, the defendant will prefer to settle provided $s + \beta \leq y + b$ or $s \leq y + (b - \beta)$. Consequently, a settlement will be feasible if

$$x - (a - \alpha) \leq y + (b - \beta). \tag{6}$$

If the plaintiff's award is less than or equal to the defendant's payment ($x \leq y$), then (6) always will be satisfied. But if the plaintiff's

[20] There is no loss of generality in assuming that the plaintiff prefers to settle when $s = x - (a - \alpha)$. An analogous statement applies below to the defendant.

award exceeds the defendant's payment ($x > y$), a settlement might not be feasible.

When a settlement is feasible, a settlement will occur and will equal

(7) $$s = y + (b - \beta).$$

This is because it is in the plaintiff's interest to make his take-it-or-leave-it settlement demand as high as possible, provided it will be acceptable to the defendant. If a settlement is not feasible -- that is, if (6) does not hold -- then a suit will result in a trial.[21]

Now consider the injurer's choice of care. If a suit results in a trial, the injurer's total payment is $y + b$. If a suit results in a settlement, the injurer's total payment is $s + \beta$ or, using (7), $y + b$. In other words, the injurer pays $y + b$ whenever he is sued, regardless of whether the suit goes to trial or settles. This is because the plaintiff's settlement demand makes the injurer indifferent between going to trial and settling. Thus, since the plaintiff's suit decision and the injurer's total payment is the same as in the basic model, the injurer's choice of care is determined as in the basic model, according to (1).

Before describing the social problem, it is necessary to define a threshold level of the plaintiff's trial cost -- denoted $\hat{a}$ -- such that a trial will occur if the plaintiff's trial cost is below $\hat{a}$, and a settlement will occur if the plaintiff's trial cost is above $\hat{a}$.[22] Intuitively, one

---

[21] In most economic models of litigation, a trial can occur only if there is asymmetric information (or a difference of opinion) about the plaintiff's probability of prevailing or the magnitude of the harm. Here, even though the parties' have perfect information, a trial might occur because of the decoupling of liability (when the plaintiff's award sufficiently exceeds the defendant's payment).

[22] The precise definition of $\hat{a}$ is contained in the appendix.

would expect such an $\hat{a}$ to exist since the higher the plaintiff's trial cost, the greater the benefit to the plaintiff of settling. In general, $\hat{a}$ may be zero -- in which case a suit always will result in a settlement -- or $\hat{a}$ may exceed the plaintiff's award, x -- in which case a suit always will result in a trial. For purposes of discussion here, $\hat{a}$ is assumed to be positive and less than x; then $\hat{a}$ is defined by the level of the plaintiff's trial cost that satisfies with equality the condition for a settlement to be feasible (6):

$$(8) \qquad x - (\hat{a} - \alpha(\hat{a})) = y + (b - \beta).$$

The social problem now can be written as

$$(9) \qquad \underset{x,y}{\text{MIN}} \quad c + p(c)[\ell + \int_{0}^{\hat{a}} (a + b)f(a)da + \int_{\hat{a}}^{x} (\alpha(a) + \beta)f(a)da],$$

where c is determined by x and y according to (1), $\hat{a}$ is determined by x and y according to (8), and $y \leq m$. The first integral in (9) is the parties' expected trial costs and the second integral is their expected settlement costs.

5. Optimal Decoupling in the Extended Model

There are two natural cases to consider, depending on whether the settlement can be observed and decoupled by the court (in addition to the trial outcome being decoupled).

5.1 Settlements Cannot be Decoupled

The principal result to be demonstrated in this case is that the optimal payment by the defendant at trial remains as high as possible: $y^* = m$. A general proof of this proposition (without any assumptions about $\hat{a}$) is contained in the appendix. An informal argument will be presented here.

The structure of the argument is similar to that used to prove the corresponding result in the basic model. Suppose that the optimal value of y were less than m and that the optimal value of x is positive. This would lead to some level of care, chosen by the injurer according to (1). Now raise y and lower x so that $F(x)[y + b]$, and therefore the injurer's care, is held constant. Since in the social problem (9) the sum of the injurer's cost of care and the victim's expected loss is unaffected, the only question concerns what happens to the parties' expected trial costs and expected settlement costs.

With x lower, the range of the plaintiff's trial cost, a, over which a suit will be brought becomes smaller. Because the suits that are discouraged are those that would have settled (since $\hat{a}$ is assumed to be less than x), this effect of lowering x reduces expected settlement costs. Also, for those suits that are still brought, raising y and lowering x lowers $\hat{a}$, the threshold level of the plaintiff's trial cost below which a suit results in a trial (see (8)). So expected trial costs fall too. Since raising y and lowering x reduces both settlement costs and trial costs, the original y less than m could not have been optimal.

The proof that $y^* = m$ in the extended model may be compared with the corresponding proof in the basic model. In both models, lowering the plaintiff's award x is beneficial because it discourages costly suits. But in the extended model there is the additional effect that raising y and lowering x has on the trial-versus-settlement decision. This effect is beneficial too -- for by raising the defendant's trial payment and lowering the plaintiff's trial award, the likelihood of a settlement is enhanced (the defendant will be willing to pay more in settlement and the plaintiff will be willing to accept less).

Given $y^* = m$, the optimal choice of x is determined in a way similar to that discussed in the basic model. However, the marginal benefit of raising x now is lower because the resulting fall in the probability of an accident reduces expected settlement costs rather than expected trial costs (for suits in which $a \geq \hat{a}$). The marginal cost of raising x may be lower or higher. It tends to be lower because, as x rises, the "marginal" plaintiff (one for whom $a = x$) will settle rather than go to trial. But the marginal cost of raising x tends to be higher because raising x makes settling less attractive for "inframarginal" plaintiffs (see (6)). Thus, x^* in the extended model may be greater or less than x^* in the basic model.

The observations made in the basic model about the relationship between x^* and y^* carry over essentially unchanged to the extended model. Thus, for accidents with relatively low losses, x^* will be less than y^*, and for accidents with relatively high losses, x^* will be greater than y^*.

Finally, it might be wondered whether it is socially desirable to discourage settlements when they cannot be decoupled, since they might undermine the beneficial effects of decoupling the trial outcome. Given the assumption that the plaintiff makes a take-it-or-leave-it settlement demand -- resulting in a settlement at the upper end of the settlement range -- it is not desirable to discourage settlements. In essence, this is because the injurer's incentive to take care then is not diminished by a settlement; and since settlements are less costly than trials, they are desirable.[23]

---

[23] However, if the defendant were to make a take-it-or-leave-it settlement offer in our model, it might be socially desirable to discourage settlements if they cannot be decoupled. For then settlements would occur at the lower end of the settlement range; the injurer consequently would take less care; and the lesser cost of settlements would have to be balanced against their reduced deterrent effect. The potential undesirability of settlements in the context of decoupling was first noted by Schwartz (1980, p.

5.2 Settlements Can be Decoupled

It is assumed in this case that the court can observe the settlement amount, s, and then award an additional amount to the plaintiff and/or make the defendant pay an additional amount. Let

x' = additional award to the victim if there is a settlement,

y' = additional payment by the injurer if there is a settlement.

The main point to be developed in this case is that if suits otherwise would not settle, it is desirable to decouple settlements because suits can thereby be encouraged to settle. It also will be shown that when decoupling settlements is beneficial, the additional award to the victim, x', exceeds the additional payment by the injurer, y' -- resulting in, on balance, a transfer to the parties from the government. As before, intuitive arguments will be emphasized here, with formal proofs provided in the appendix.

First note that even if settlements cannot be decoupled, all suits might settle. In particular this will be true if $x \leq y + (b - \beta)$, since then a settlement will be feasible regardless of a; see (6). Thus, decoupling settlements can be valuable only when some suits otherwise would result in trials, that is when

(10) $$x - (a - \alpha) > y + (b - \beta)$$

for some a, which implies that

(11) $$x > y + (b - \beta).$$

If (11) holds, there will be a range of a, starting at zero, over which a suit will result in a trial. Condition (11) will be assumed to hold for the remainder of this section.[24]

1095).

[24] Note that this is equivalent to assuming that $\hat{a} > 0$.

To see how decoupling settlements can encourage suits to settle, consider the feasibility of a settlement when both the trial outcome and the settlement are decoupled. If the plaintiff goes to trial his net gain is $x - a$, whereas if he settles, his net gain is $s - \alpha + x'$. The plaintiff will prefer to settle if $s \geq x - (a - \alpha) - x'$. If the defendant goes to trial his total payment is $y + b$, whereas if he settles his total payment is $s + \beta + y'$. The defendant will prefer to settle if $s \leq y + (b - \beta) - y'$. Therefore, a settlement will be feasible if

(12) $$x - (a - \alpha) - x' \leq y + (b - \beta) - y'.$$

Given any x and y, it is straightforward to show that x' and y' can be chosen so that (12) is satisfied for all values of a. If $a = 0$ (the value of a least favorable to a settlement), any combination of x' and y' such that

(13) $$x' - y' \geq x - [y + (b - \beta)]$$

will satisfy (12). If a is positive, any combination of x' and y' satisfying (13) will satisfy (12) as well.

Observe that, using (11), (13) implies that $x' - y'$ must be positive and sufficiently large. In other words, to encourage settlements when they otherwise would not occur, the government must decouple settlements in such a way as to provide a net transfer to the parties. This result should not be surprising since a net transfer increases the attractiveness to the parties of a settlement relative to a trial.

Clearly, the optimal values of x' and y' are not uniquely determined. Without loss of generality one can assume that (13) is satisfied with equality.[25] Even then, there are many combinations of x' and y' that will

[25] This assumption could be justified on the grounds that raising government revenue generally results in a deadweight burden, so it is better to raise the smallest amount necessary.

assure a settlement. One solution is to set the additional payment by the injurer equal to zero and then to supplement the settlement payment to the victim by a positive amount (equal to $x - [y + (b - \beta)]$).

Assuming x' and y' are chosen so as to guarantee a settlement, it can be shown, once again, that $y^* = m$. Only the first steps of this argument will be presented here.

Because the plaintiff's settlement demand will be as high as possible subject to its acceptability by the defendant, the settlement amount will equal

(14) $$s = y + (b - \beta) - y'.$$

The injurer's total payment as a result of the settlement is $s + \beta + y'$ or, using (14), $y + b$. (The reason the injurer's total payment does not depend on y' is that the payment of y' to the government reduces the settlement amount by y'.) For now familiar reasons, it is optimal to set y as high as possible so that x can be lowered, thereby reducing the incentive to sue. Since all suits result in settlements (given x' and y'), this lowers the parties' expected settlement costs.

The discussion thus far has assumed that the purpose of decoupling settlements is to encourage settlements when they otherwise would not occur. It is now straightforward to see that decoupling settlements for this reason is socially desirable. When settlements are decoupled, the injurer's total payment as a result of a settlement was shown in the previous paragraph to be $y + b$; this is the same as the total payment if the case goes to trial. Thus, the decoupling of settlements does not affect the injurer's incentive to take care, but it does lower litigation costs.[26]

Given $y^* = m$, the determination of x^* and the comparison of x^* to y^* are essentially as described in the basic model.

And since it was not desirable to discourage settlements when settlements cannot be decoupled, it clearly is not desirable to discourage them when they can be decoupled.

6. Concluding Remarks

This section contains observations about: the applicability of our analysis to a negligence rule; the relationship of the analysis to the economic theory of public enforcement; the prior literature that discusses decoupling liability; and the informational requirements of a system of decoupling.

(a) Applicability to a negligence rule. In our model of decoupling, it was assumed implicitly that the injurer's choice of care did not affect whether he was liable. This is equivalent to assuming that the rule of liability is strict liability. A natural question is whether the analysis of decoupling also applies to a negligence rule -- under which the injurer is liable only if he does not take some minimum level of care, referred to as the standard of care.

---

[26] There is a potential detrimental effect from decoupling settlements. A victim whose trial cost exceeds the trial award might nonetheless sue and then settle with the injurer in order for the parties to obtain the net transfer $(x' - y')$ from the government. However, since the injurer knows that such a victim would drop the suit if the injurer rejects the victim's settlement demand, the injurer would have to be paid to settle the suit -- either by the victim through a "negative" settlement $(s < 0)$ or by the government through a "negative" additional payment $(y' < 0)$. Consequently, such suits can be forestalled by a policy of decoupling settlements only if the settlement amount is positive and by restricting the additional payment by the injurer to be non-negative.

In theory, decoupling liability would not be necessary under a negligence rule for the following reason. If the standard of care corresponds to the first-best level of care (the level of care that minimizes the sum of the cost of care and the expected harm) and the level of liability for violating the standard is high enough, the injurer will meet the standard. Then the victim will not sue since he would not prevail. Thus, in principle, the first-best level of care could be attained without any litigation costs being incurred. There would be no reason to decouple liability in these circumstances.

In practice, however, a negligence rule is likely to lead to some suits for the following reasons. The injurer may be uncertain about what the standard of care is, and therefore may choose a level of care that leads with some probability to his being found negligent. Conversely, a court or a jury may be uncertain about what level of care was chosen by the injurer, and consequently may find him negligent with some probability. In general, therefore, a negligence rule can be characterized as a schedule of the probability of liability as a function of the injurer's care; the higher the injurer's care, the lower the chance of his being found liable.

Given this more realistic view of the negligence rule, it is straightforward to see that the analysis of decoupling in this article applies under a negligence rule as well. Whatever level of care results from the best choice of coupled liability under a negligence rule, the same level of care can be achieved with lower litigation costs by decoupling liability in the way described here.

(b) <u>Relationship to the economic theory of public enforcement</u>. The analysis of decoupling liability in this article closely parallels the economic theory of public enforcement associated with Becker (1968). Becker observed that the same amount of deterrence can be achieved by catching an injurer infrequently and fining him severely as by catching him more often and fining him less. He concluded that the best system of public enforcement involves using the highest possible fine and a correspondingly low probability of detection, since such a combination can achieve any given amount of deterrence with the lowest investment in detection resources.

The rationale for making the defendant's payment as high as possible in a system of decoupled liability is, in essence, the same as Becker's rationale for high fines in a system of public enforcement. Under public enforcement, a high fine allows the probability of detection to be lowered; under decoupled liability, a high payment allows the probability of a suit to be lowered. In both cases, enforcement costs are saved. Thus, one could interpret our analysis of decoupling as providing a private litigation analogue to Becker's theory of public enforcement.

Becker's theory of public enforcement has been criticized on the grounds that severe fines -- potentially as high as an individual's wealth -- hardly ever are imposed. An analogous criticism could be leveled against the theory of decoupling liability proposed here. In both contexts, however, there are additional considerations that could be taken into account in the analysis that would lead to the conclusion that the optimal fine or the optimal payment by the defendant is not as high as possible. For example, if injurers are risk averse it generally is desirable to reduce the sanction and to increase the probability of its imposition in order to lower risk-bearing costs.[27]

---

[27] For a review of other reasons why optimal sanctions are not at their upper bound, see, for example, Carr-Hill and Stern (1979, pp. 281-295).

(c) Literature relevant to decoupling. The literature most relevant to our analysis is concerned with the private enforcement of public fines -- whereby the first private enforcer to discover and report the injurer receives the fine. See, for example, Becker and Stigler (1974, pp. 13-16), Landes and Posner (1975), and Polinsky (1980). Landes and Posner emphasized that private enforcement would lead to socially excessive enforcement, and that a tax on private enforcers therefore might be desirable. Polinsky showed that private enforcement also could lead to too little enforcement and that a subsidy to private enforcers might be needed. In other words, to use the language of the present article, it may be desirable to "decouple" the injurer's payment and the enforcer's award, with the enforcer receiving either less than or more than the fine.[28]

The results just summarized were developed in a framework in which anyone could become a private enforcer, not just the victim of the harm. Thus, one could view our contribution as extending this analysis to a setting that is more descriptive of private litigation systems -- in which the victim of the harm normally is the sole initiator of a suit.[29]

Several other reasons have been advanced for decoupling liability, although they are unrelated to the rationale presented here. For example, if

---

[28] The literature referred to in this paragraph has been applied to private antitrust enforcement in the articles cited in note 2 above. Although those articles advocated decoupling liability and anticipated some of the results demonstrated here, they did not analyze a formal model of decoupling or systematically consider the possibility of settlements.

[29] Other recent articles concerned with private litigation have mentioned a decoupling-type solution. The discussion that is closest in spirit to the present analysis is by Hylton (1990, pp. 164-165 & 169-170). He assumes that the award to the victim equals the victim's loss, and shows that the optimal payment by the injurer exceeds this amount. See also Katz (1990, pp. 19-20) and Polinsky and Shavell (1989, pp. 105-107).

harms are nonmonetary (such as pain and suffering), there is a conflict between optimally compensating victims and optimally deterring injurers. In terms of risk-allocation, the victim should be awarded relatively little (assuming his marginal utility of money is not much affected). In terms of deterrence, however, the injurer should be made to pay an amount reflecting how much the victim values preventing the harm from occurring in the first place. Thus, only by decoupling liability is it possible to achieve both optimal compensation and optimal deterrence.[30]

(d) Informational requirements. If a government agency were to attempt to optimally design a decoupled system of liability, a substantial amount of information would be required. The agency would need to know how changes in the plaintiff's award would affect his incentive to sue; how changes in the probability of suit would affect the injurer's choice of care; how this choice would affect the sum of the cost of care and the expected harm; and whether a suit would result in a trial or a settlement.

Although the information required to implement an ideal decoupling system is substantial, two points should be kept in mind. First, the agency does not have to know with great precision everything referred to in the previous paragraph. It simply has to have some estimate -- possibly quite imperfect -- of the required information. Second, essentially the same information would be needed to determine the optimal level of coupled liability. Thus, if it is worthwhile to select the level of coupled liability on the basis of how liability affects care and litigation, then it is better

[30] This point is widely recognized in the literature on liability. See, for example, Spence (1977) and Shavell (1987, pp. 228-235). Additional arguments for decoupling liability have been discussed, for example, by Shavell (1987, pp. 29-30 & 142-144).

to use that information to design a system of decoupled liability.

In practice, rather than attempting to determine the ideal decoupling system, it may be preferable to develop relatively simple rules of thumb. For example, one rule might be to treat cases differently on the basis of the size of the loss. For low-loss cases, the plaintiff could be awarded some fraction of what the defendant pays (in order to discourage excessive litigation), and for high-loss cases the plaintiff could be awarded some multiple of what the defendant pays.[31]

[31] The rationale for this suggestion is based on the observation in section 2 that $x^* < y^*$ if the loss is sufficiently low, and $x^* > y^*$ if the loss is sufficiently high.

Appendix

Proposition 1: When settlements are possible but cannot be decoupled, the optimal payment by the defendant at trial is as high as possible: $y^* = m$.

Proof: First define $\bar{a}(x, y) \equiv \min(\hat{a}(x, y), x)$, where $\hat{a}(x, y)$ solves for a in

$$x - (a - \alpha(a)) = y + (b - \beta) \tag{A1}$$

if $x > y + (b - \beta)$, and $\hat{a}(x, y) = 0$ otherwise. Since $\hat{a}(x, y)$ is increasing in x and decreasing in y (in the weak sense), so is $\bar{a}$. For a plaintiff with $a < \bar{a}$, a suit resulting in a trial occurs and litigation costs $a + b$ are incurred; for a plaintiff with $\bar{a} \leq a < x$, a suit resulting in a settlement occurs and settlement costs $\alpha + \beta$ are incurred; and for a plaintiff with $a \geq x$, a suit does not occur.

The social problem can be written as:

$$\min_{c,x,y} \; c + p(c)\left[\ell + \int_0^x [\alpha(a) + \beta] f(a)da + \int_0^{\bar{a}(x,y)} [a - \alpha(a) + b - \beta] f(a)da\right] \tag{A2}$$

subject to

$$c \in \operatorname{argmin} \; c + p(c)F(x)[y + b], \tag{A3}$$

$$y \leq m. \tag{A4}$$

To prove that the second constraint (A4) is binding, suppose to the contrary that at the optimal choice (x_0, y_0), $y_0 < m$. Now consider an alternative choice (x_1, y_1) with $y_0 < y_1 \leq m$ and $F(x_1)[y_1 + b] = F(x_0)[y_0 + b]$. Thus, $x_1 < x_0$. It follows from the construction of the new pair (x_1, y_1) that the choice of the level of care c remains unchanged. But the value of the objective function (expected social cost) is lower under the new pair

(x_1, y_1), since the first term involving an integral in (A2) is strictly increasing in x, and the second term involving an integral is weakly increasing in x and weakly decreasing in y (since $\tilde{a}$ is weakly increasing in x and weakly decreasing in y).

Remark: It was assumed in section 4 of the text that $\hat{a}$ was positive and less than x. In that case, $\tilde{a} = \hat{a}$. The proof here does not impose any assumptions on $\hat{a}$.

Proposition 2: When settlements are possible and can be decoupled: (i) Decoupling settlements is valuable only if some suits otherwise would not settle (i.e., only if $x > y + (b - \beta)$); (ii) When decoupling settlements is valuable, optimal decoupling of settlements requires a minimum net transfer to the parties from the government (at least equal to $x - [y + (b - \beta)] > 0$); (iii) The optimal payment by the defendant at trial is $y^* = m$.

Proof: Let z denote the settlement subsidy to the two parties. (In terms of the notation of section 5.2, $z \equiv x' - y'$; the allocation of the subsidy does not matter.) To prove (i) and (ii), it suffices to show that the optimal policy involves a subsidy $z = 0$ if $x \leq y + (b - \beta)$, and $z \geq x - [y + (b - \beta)]$ if $x > y + (b - \beta)$.

Analogously to the proof of Proposition 1, define $\tilde{a}(x, y, z) \equiv \min(\hat{a}(x, y, z), x)$, where $\hat{a}(x, y, z)$ solves for a in

(A5) $$x - (a - \alpha(a)) = y + (b - \beta) + z$$

if $x > y + (b - \beta)$, and $\hat{a}(x, y, z) = 0$ otherwise. The social problem can be written as:

$$\text{(A6)} \quad \underset{c,x,y,z}{\text{MIN}} \quad c + p(c)[\ell + \int_0^x [\alpha(a) + \beta]f(a)da + \int_0^{\bar{a}(x,y,z)} [a - \alpha(a) + b - \beta]f(a)da]$$

subject to (A3) and (A4). Since $\bar{a}$ is weakly increasing in x and weakly decreasing in y, $y^* = m$ for the same reason as before, proving (iii).

Observe that the choice of the subsidy z does not affect the level of care c chosen by the injurer. Thus, if $x > y + (b - \beta)$, any $z \geq x - [y + (b - \beta)] > 0$ minimizes expected social cost by letting $\bar{a}(x, y, z) = 0$; and if $x \leq y + (b - \beta)$, since $\bar{a} = 0$ anyway, $z = 0$ is optimal (but not uniquely).

References

Becker, Gary S., "Crime and Punishment: An Economic Approach," Journal of Political Economy, Vol. 76, No. 2 (March-April 1968), pp. 169-217.

Becker, Gary S., and George J. Stigler, "Law Enforcement, Malfeasance, and Compensation of Enforcers," Journal of Legal Studies, Vol. 3, No. 1 (January 1974), pp. 1-18.

Carr-Hill, R. A., and N. H. Stern, Crime, The Police and Criminal Statistics (London: Academic Press, 1979).

Hylton, Keith N., "The Influence of Litigation Costs on Deterrence Under Strict Liability and Under Negligence," International Review of Law and Economics, Vol. 10, No. 2 (September 1990), pp. 161-171.

Katz, Avery, "The Effect of Frivolous Lawsuits on the Settlement of Litigation," International Review of Law and Economics, Vol. 10, No. 1 (May 1990), pp. 3-27.

Landes, William M., and Richard A. Posner, "The Private Enforcement of Law," Journal of Legal Studies, Vol. 4, No. 1 (January 1975), pp. 1-46.

Polinsky, A. Mitchell, "Private Versus Public Enforcement of Fines," Journal of Legal Studies, Vol. 9, No. 1 (January 1980), pp. 105-127.

Polinsky, A. Mitchell, "Detrebling versus Decoupling Antitrust Damages: Lessons from the Theory of Enforcement," Georgetown Law Journal, Vol. 74, No. 4 (April 1986), pp. 1231-1236.

Polinsky, A. Mitchell, and Daniel L. Rubinfeld, "The Welfare Implications of Costly Litigation for the Level of Liability," Journal of Legal Studies, Vol. 17, No. 1 (January 1988), pp. 151-164.

Polinsky, A. Mitchell, and Steven Shavell, "Legal Error, Litigation, and the Incentive to Obey the Law," Journal of Law, Economics, and Organization, Vol. 5, No. 1 (Spring 1989), pp. 99-108.

Salop, Steven C., and Lawrence J. White, "Economic Analysis of Private Antitrust Litigation," Georgetown Law Journal, Vol. 74, No. 4 (April 1986), pp. 1001-1064.

Schwartz, Warren F., "An Overview of the Economics of Antitrust Enforcement," Georgetown Law Journal, Vol. 68, No. 5 (June 1980), pp. 1075-1102.

Schwartz, Warren F., Private Enforcement of the Antitrust Laws: An Economic Critique (Washington D.C.: American Enterprise Institute, 1981).

Shavell, Steven, Economic Analysis of Accident Law (Cambridge: Harvard University Press, 1987).

Spence, A. Michael, "Consumer Misperceptions, Product Failure and Producer Liability," Review of Economic Studies, Vol. 94, No. 3 (October 1977), pp. 561-572.