Should Plaintiffs Win What Defendants Lose? Litigation Stakes, Litigation Effort, and the Benefits of Decoupling

Albert Coi and Chris William Satchuriko

ABSTRACT
In a 1991 paper, Polinsky and Shavitz argue that lowering plaintiffs' recovery and raising defendants' damages can deliver the same level of deterrence with fewer legal suits. A subsequent paper by Kahan and Tushnet provisionally corroborates Polinsky and Shavitz's analysis in an extended model that accounts for the effect of litigation states on litigation effort levels. In contrast, we show that when litigation effort is exogenous, Polinsky and Shavitz's proposal to lower recovery and raise damages may no longer improve social welfare. We then characterize the kinds of suits when it is in fact suboptimal to set recovery below damages. Of significance for the current policy debate, we find that such suits play many of the empirical premises about litigation that guided conventional arguments in favor of making recovery less than damages. Our findings are robust to the possibility of out-of-court settlement, plaintiff employment of contingent fee lawyers, and alternative fee-sharing rules.

1. INTRODUCTION
Punitive damage awards have demonstrated remarkable staying power as a source of controversy in scholarship, legislation, and the media. Yet despite the persistent discord, most commentators agree that ef-
effective deterrence—a well-acknowledged purpose of punitive awards—generally requires that damages be something more than purely compensatory. Although compensatory damages would produce appropriate deterrence were injurers always called to task for the harm they caused, not all harms are litigated and not all deserving plaintiffs win. To deter ideally, therefore, damages must be more than compensatory to make up for their less-than-comprehensive imposition (Bensel 1931, p. 325; Becker 1968).  

What remains less clear is why plaintiffs should be the beneficiaries of this upward adjustment in damages. Plaintiffs may well be doing society’s work in bringing defendants to task for dangerous product designs that put many at risk or for commercial practices that threaten the smooth functioning of markets. But what is the logical relationship between what is needed to inspire plaintiffs appropriately to bring suit and what is needed to ensure that defendants fully internalize the costs that they impose on others? Why, in particular, should we assume that the appropriate bounty for plaintiffs is the same as the appropriate fine for the defendants?  

In a deservedly influential paper, Polinsky and Che (1991) make a powerful case that the award paid to plaintiffs should generally be less than the liability imposed on defendants. They argue that whenever the plaintiff’s recovery equals the defendant’s damages, the same level of deterrence can be produced with lower social cost by lowering recovery and simultaneously raising damages, thus “decoupling” the two transfers. As depicted schematically in Figure 1, lowering recovery reduces the number of cases filed, thus reducing both deterrence and the social

3. Several qualifications to this position appear in the literature. First, Calfee and Crosswell (1984, p. 993, n.69) point out that optimal punitive damages may be less than compensatory when (1) the probability of liability is significantly more elastic in the level of care than in the level of harm and (2) damages cannot be adjusted on a case-by-case basis according to the defendant’s actual probability of liability. See also Görs (1994) and Crosswell and Calfee (1986). Second, Polinsky and Rubinfeld (1988) argue that when the social costs of litigation are fully accounted for, compensatory damages may not be socially optimal even in a world in which defendants are always liable for the harm they cause. Lastly, Coase (1944) suggests that raising punitive damages above compensatory damages may be unnecessary under a fault standard, if the standard is clear enough to injurers and the percentage of harms litigated is not too small. For a more general discussion of the law and economics of punitive damages, see Polinsky and Shavell (1998).


3. Herein lies, what the defendant says will be called “damages,” and what the plaintiff receives, “recovery.”
costs of litigation. Raising damages alone by an appropriate amount 
estrengthens deterrence. But because plaintiffs do not receive the increase, 
suits remain at their now lower level. Thus, deterrence is restored with- 
out also restoring litigation costs.4

As Kahan and Tuckman (1995) have pointed out, however, Polinsky 
and Che's analysis ignores an important feature of litigation mechan- 
is.5 Polinsky and Che (1991) consider the effect of their policy prescription 
on the number of suits filed but not the manner in which filed suits 
proceed. Which lawyer a party retains, how many billable hours she 
authorizes, where she hires as an expert or investigator, how many wit- 
tnesses she deposes and for how long, how many documents and things 
she requests for inspection, how prepared she is to resist such requests 
from her opponent, how carefully she inspects what she does receive— 
all of these decisions affect both the social cost and the deterrent force 
of litigation, independently of the number of filed suits. Moreover, all 
of these decisions are likely to be sensitive to the parties' stakes in the 
case. According to one study, "[H]igher stakes are associated with sig- 
nificantly higher total lawyer work hours, significantly higher lawyer 
work hours on discovery, and significantly longer time to disposition" 
(Kaklik et al. 1998, p. 638). Specifically, median total lawyer work 

4. Importantly, this may mean lowering recovery while raising damages—always 
social welfare improving in Polinsky and Che’s model. Yet, as Polinsky and Che (1991) 
explain, this does not imply that optimal damages always exceed optimal recovery. This 
is because the narrower is not feasible when damages are already equal in all of the 
defendants’ weak (which, in Polinsky and Che’s framework, is always the case in the total 
optimum). For more on this point, see note 22. Given that actual damage awards to 
non-frivolous defendants, it may be fair to say that the practical message of Polinsky 
and Che’s analysis for litigation reform is that recovery should be decreased and damages 
increased.

5. Kahan and Tuckman also extend Polinsky and Che’s analysis in considering the 
effect of agency problem between lawyer and client. See, for example, Kahan and Tuckman 
(1991, proposition 2). We consider one such agency problem in Section 4.1.
hours were more than two and a half times larger for cases with monetary stakes over $500,000 than for cases with monetary stakes of $500,000 or less, while mean total lawyer work hours were almost four times larger (Kakalik et al. 1998, p. 648, table 2-7).

Kahan and Tuckman do much to advance our understanding of the benefits of decoupling simply by broadening the scope of analysis to include not just changes in the number of filings, but also changes in how filed suits proceed. In the end, however, Kahan and Tuckman provisionally conclude that such effects do not fundamentally alter Polinsky and Che's basic arguments for decoupling. This paper reaches a different conclusion. We show that when the effect of litigation stakes on litigation effort is fully accounted for, the policy maneuver by which Polinsky and Che proove the benefits of decoupling—lowering recovery and raising damages—may no longer improve social welfare. In addition, we characterize the kinds of suits in which the optimal level of recovery is no less than the optimal level of damages. Ironically—and of some theoretical significance in the current policy debate—we find that this class of cases bears a strong resemblance to the negative prototype of litigation that is invoked by some advocates of reduced recovery.

Our conclusions about the welfare implications of Polinsky and Che's decoupling maneuver differ from those of Kahan and Tuckman because we consider the effect on litigation effort of both decreasing recovery and increasing damages. Kahan and Tuckman (1995) consider only the

6. Their results concern only cases closing in 9 or more months after filing. Another empirical study that used a different data found that the "percentage of the monetary stakes in the case had a strong relationship to total litigation costs of any of the case characteristics we studied." (Wolfing et al. 1999, pp. 327, 332). This study also found that "the stakes in the litigation were positively correlated with the length of the case: the higher the stakes, the longer the case lasted." (p. 333). Note that this study has a rather special definition of "stake" (p. 349, n. 56).

7. We find that, in the absence of a single problem in the plaintiff-lawyer relationship, special losses (medical, reductions of recovery, reduced litigation costs and the expected award possible by the defendant in the case of a trial) to that extent, special losses combined with increased awards could be used to reduce litigation costs while maintaining deterrence, as suggested by Polinsky and Che" (Kahan and Tuckman 1995, p. 140).

8. Kumar (1999) reaches a similar conclusion for different reasons. He argues that the frequent enforcement and low-sanction regime created by indemnification of corporate directors can be beneficial to the extent that the resulting increase in litigation reduces legal uncertainty and to the extent that indemnification is seen as a `cost'. More recently, Andrew Shaughnessy and Jennifer Rongman (2003) have provided an extensive (though positive) analysis of the effect on asymmetric information of limiting litigation of forcing plaintiffs to "sue" part of their negative damages award with the other. They find that slip-and-fall statutes generally lead to more frequent suits and at lower amounts.
effect of decreasing recovery (see, for example, p. 179, proposition 1). As Kahan and Tuckman rightly argue, the effect on litigation effort of decreasing recovery, taken alone, merely amplifies the effects considered by Polinsky and Che. When recovery is reduced, both litigation cost and deterrence still fall—now, not just because there are fewer suits, but also because plaintiffs who still file pursue their cases less vigorously. This is shown in Figure 2, which adds to Figure 1 the effect on deterrence and litigation costs of changes in plaintiffs' litigation effort. (In this and subsequent figures, added effects are shown in darkened capitals, while the effects identified by Polinsky and Che [1991] are shown in lowercase grayscale letters.)

In contrast, the welfare implications of increasing damages change significantly when the effect on litigation effort is taken into account. In Polinsky and Che's (1991) framework, in which litigation effort is implicitly held constant, increasing damages costslessly raises deterrence. When litigation effort is allowed to vary, raising the stakes for defendants causes them to devote more resources to their defense, and this increases the cost of each filed suit. Although it is still true that increasing damages increases the deterrent force of each suit that is filed, this additional deterrence is no longer costless. Each filed suit is now more expensive. Moreover, defendants' increase in litigation effort will feed back into plaintiffs' filing decisions. Filing suit will now be less attractive for potential plaintiffs since they will now face more fervent opposition from...
Damages

\[ \therefore \text{LESS deterrence; FEWER, COSTLEER suits} \]

Recovery

\[ \therefore \text{less deterrence; CHEAPER, cheaper suits} \]

Figure 3. Adding the effect on litigation effort of increasing damages

defendants. The consequent reduction in filings will act to lower both deterrence and litigation costs. Figure 3 adds to Figure 2 both this indirect filing effect and the direct effect of defendants’ litigation effort response (underlined).

The crosscurrent of effects emanating from defendants’ increased effort draws into question whether the full maneuver advocated by Polinsky and Che—a reduction in recovery along with a deterrence-restoring increase in damages—is still welfare improving (let alone feasible). Indeed, under many plausible scenarios the maneuver will be welfare reducing. Of special interest, given the empirical findings reported above, is the situation wherein the effect of the maneuver on “inefficient suits”\(^{10}\) dominates the effect on the number of filings.

For expositional purposes, consider, in particular, the limiting case wherein the number of filed suits remains constant. In the absence of filing effects, Figure 3 reduces to Figure 4. Reducing recovery, as per the first step in the Polinsky and Che maneuver, decreases both deterrence and litigation costs because plaintiffs pursue their complaints with less intensity. Raising damages by an appropriate amount will restore deterrence but also increase litigation costs because defendants have more to lose and react by lodging a more vigorous defense. The Polinsky and Che maneuver reduces social welfare if the decrease in litigation costs from reducing recovery is outweighed by the increase in litigation costs from raising damages enough to restore deterrence.

Given indications in the data that the manner in which filed suits proceed is at least as important as the number of suits filed, the more

\(^{10}\) “Inefficient suits” are suits that would still be filed in response to small reductions in the net benefits of filing.
The possibility that inframarginal effects may net to a reduction in social welfare is by itself of practical importance. However, it would still be useful to identify the kinds of lawsuits in which this was the case. This article provides such a characterization. Significant for current policy debate, the argument against reducing recovery and raising damages turns out to be stronger in the kind of suit that seems to motivate litigation reform—including various forms of decoupling. Allowing plaintiffs to keep all of what defendants lose is most likely to be optimal in high-stakes suits with deep-pockets defendants and contingent-fee plaintiffs lawyers.

The result that raising damages and lowering recovery is unlikely to be welfare improving in high-stakes suits is fairly intuitive. Focus again on the scenario laid out in our discussion of Figure 4, wherein inframarginal effects dominate. The key to our analysis is the recognition that when litigation stakes are high, the increase in deterrence from a unit increase in damages is much smaller than the decrease in deterrence from a unit decrease in recovery.

11. One by-product of our analysis is to show that optimal damages will not in general equal all of defendant's wealth when we account for the effect of litigation stakes on litigation effort. See note 4. This finding can be added to the literature’s list of reasons why the (unacknowledged) interpretation of Gary Becker’s (1968) famous results—that increasing fines and lowering the probability of sanction is always welfare improving—does not hold in the general case. See, for example, Polinsky and Shavell (1979).

12. In Indiana, for example, punitive damages may be as high as four times compensatory damages, but plaintiffs receive only 25 percent of such damages, with the rest going to a fund for victims of violent crimes. For a summary of various state laws that reduce the plaintiff’s award, see Daughtery and Reinganum (2001, at 137–38) and Kahn and Tuckman (1995, p. 177, n. 1). For a discussion of the law governing the imposition of punitive damages—which, of course, increases the defendant’s sanctions—see generally Kocher, Wiemer, and Ghandi (2000).
Increasing damages by $1 affects the defendant's expected trial losses—and so deterrence—in three separate ways, the first two of which cancel out. These first two ways concern the effect on expected trial losses of the defendant's increase in litigation effort. First, greater litigation effort increases the defendant's litigation costs, thus increasing her expected trial losses. Second, greater litigation effort reduces the probability that the defendant will lose the case, which lowers her expected trial losses. These two effects cancel out at the margin because the defendant has already optimally set her litigation effort to balance marginal costs and benefits with respect to expected trial losses. This cancellation is a manifestation of the “envelope theorem.” The remaining third effect is the simplest. Increasing damages by $1 increases the defendant's trial losses by the chance that the defendant will have to pay out that additional dollar. If, for example, the chance that the defendant loses the suit is 50 percent, then each dollar increase in damages increases deterrence by $0.50.

Reducing recovery, on the other hand, reduces the defendant's expected trial losses solely by reducing the plaintiff's litigation effort and thereby also reducing the probability that the defendant will be held liable. The impact of this change on the defendant's expected trial losses depends on what is at stake for the defendant. If the defendant stands to lose only $50, then a percentage point decrease in the chance of liability decreases the defendant's expected losses by only $0.50. On the other hand, if the stakes are $5,000,000, then a percentage point decrease in the chance of liability decreases the defendant's expected trial loss by $25,000. (Note that the envelope theorem does not operate here because the change in the probability of liability is induced by the plaintiff's change in litigation effort and the plaintiff's litigation costs are not borne by the defendant.)

Thus, the reduction in deterrence from reducing recovery by $1 is leveraged by the defendant's stakes in the case and will tend to be large when those stakes are high. In contrast, the increase in deterrence from increasing damages by $1 is essentially independent of the stakes of the case and will be a fraction of that dollar that corresponds to the defendant's chance of being held liable. As a consequence, executing Polinsky and Che's (1991) deterrence-maintaining maneuvers in a high-stakes suit requires raising damages by much more than recovery is reduced. The result is that the additional litigation cost from raising damages—incurrd by virtue of the defendant's stepped-up defense—is likely to over-
whenever the cost savings from reducing recovery—borne from the plaintiff’s stepped-down prosecution.

As we show in the formal analysis to follow, this basic point is strikingly robust. Indeed, the argument against raising damages and lowering recovery is even stronger in a setting in which plaintiffs’ lawyers are paid on a contingent-free basis and in which cases settle out of court. Moreover, the point holds under both the American rule—whereby parties pay their own costs—and the British rule—whereby the loser pays both sides’ costs.

The rest of the paper is organized as follows. Sections 2 and 3 present the basic model and results, while important variations on the basic model are considered in Section 4. A technical Appendix hoists the formal statements of results, where necessary, along with all mathematical proofs.

2. THE BASIC MODEL

The model consists of a population of (potential) defendants, who make primary-activity choices; a population of (potential) plaintiffs, who decide whether to sue; and three sequential and contingent phases: the primary activity, the plaintiff’s filing decision, and the trial.15 We describe the three phases in reverse order with the aid of Figure 5.

2.1. Trial

Each trial matches a particular plaintiff with a particular defendant. The plaintiff produces the quantity \( x \geq 0 \) of evidence, and the defendant, \( y \geq 0 \). The net weight of the evidence is \( x - y \). The fact finder’s perception of the net weight of the evidence is \( x - y + \epsilon \), where \( \epsilon \) is an error term, uniformly distributed between \(-b \) and \( b \).16 Plaintiff wins if the fact finder’s perception of the net weight of the evidence favors her case. Thus, the probability of plaintiff victory is

\[
    p(x, y) = \Pr(x - y + \epsilon \geq 0) = \Pr(x \geq y + \epsilon) = \frac{b - (y - x)}{2b}.
\]

15. We consider arbitration in Section 4.2.

16. The working paper version of the article (Chen and Sanchirico 2003) considers general probability distributions. See note 21 for the significance of this assumption.

17. The following expression assumes that \(-b \leq y - x \leq b \). If \( y - x < -b \), then \( p(x, y) = 1 \). If \( y - x > b \), then \( p(x, y) = 0 \).
If the plaintiff wins, the defendant pays damages of $D$ to the court and the plaintiff receives recovery of $R$. Recovery and damages are independent policy variables in the social welfare problem. Damages cannot exceed the defendant’s wealth $W$.

The plaintiff’s evidence costs are $k + c(e)$, where $k \geq 0$ represents a separately notated fixed-cost component. (As in Polinsky and Che’s model, this fixed-cost component varies across potential plaintiffs, as described below.) We assume that $c' > 0$ and $c^2 > 0$. Similarly, the defendant bears evidence production costs of $q(y)$, where $q' > 0$ and $q^2 > 0$. According to our assumptions on evidence cost derivatives, the cost to each party of further tilting the evidentiary balance in her favor is increasing in the amount of evidence that that party already has on the scale. The first few units of evidentiary weight are low-hanging fruit, and the cost of additional evidentiary weight is ever greater as the party must look ever higher in the tree.

16. The defendant’s evidence costs may also have a fixed-cost component, but this is not separately notated.
We assume that both parties are risk neutral. Thus, the plaintiff chooses evidence \( x \geq 0 \) to maximize expected litigation payoffs \( p(x, y)R - c(x) - k \), while the defendant chooses evidence \( y \geq 0 \) to minimize expected litigation losses \( p(x, y)D + c(y) \). We will assume that the parameters of the model are such that in the relevant range of policy variables both plaintiff and defendant choose a level of effort that is strictly greater than zero and that the two effort levels in combination produce a probability of plaintiff victory that is strictly between zero and one. This allows us to conclude that the parties choose effort levels that satisfy the following first- and second-order conditions:

\[
\begin{align*}
 p_x R - c' &= 0 \quad \text{and} \quad p_x D + c' &= 0, \quad (2a) \\
p_y R - c'' &\leq 0 \quad \text{and} \quad p_y D + c'' &\geq 0. \quad (2b)
\end{align*}
\]

Given equation (1), the second-order conditions reduce to \( c', c'' \geq 0 \), which we have assumed. Furthermore, our assumptions imply that the first-order conditions will have a unique solution in \( x \) and \( y \).

### 2.2. The Plaintiff’s Filing Decision

Each plaintiff decides whether to file suit on the basis of whether she expects litigation to be a profitable venture. In deciding whether to file suit, each plaintiff knows her own evidence costs, including her fixed cost \( k \), and she correctly anticipates the (unique) trial equilibrium that will obtain should she file. Further, the plaintiff may be required to pay a “filing fee” \( K \), which is a third social policy variable in addition to \( R \) and \( D \). Thus, the plaintiff files suit if and only if

\[
pR - c - K \geq 0 \quad \text{or} \quad k + pR - c - K,
\]

where \( p \) and \( c \) are determined by the expected equilibrium at trial, which is in turn determined by \( R \) and \( D \). Thus, writing \( k(D, R, K) = pR - c - K \) for the “marginal files,” the plaintiff files suit whenever \( k \leq k \).

### 2.3. The Defendant’s Primary Activity Choice

Each defendant in a population of defendants decides whether to engage in a particular activity that may cause harm to others. A defendant’s net private benefit from engaging in the activity (‘‘acting’’) is \( \beta \geq 0 \), where \( \beta \) is distributed among the population of defendants according to the cumulative distribution \( J \) with density \( j \). If the defendant acts, harm of \( \delta > 0 \) is inflicted on one plaintiff drawn at random from the plaintiff population. Plaintiffs in this population have different fixed costs \( k \).
cumulative distribution of the fixed cost of the injured plaintiff is $G$. The density is $g$. The plaintiff then decides whether to file suit against the defendant, as described above. If the defendant refrains from engaging in the harmful activity ("refrains"), there is no harm and no litigation. Each defendant makes his primary activity decision knowing his own private benefits and the distribution $G$ of a potential plaintiff’s costs and predicting each potential plaintiff’s filing decision as well as the expected equilibrium in evidence production should the plaintiff choose to file. The defendant chooses to act if and only if the private benefits from acting exceed the expected loss from litigation, including evidence costs:

$$\beta \geq G(\hat{k})(\mu D + q),$$

where $G(\hat{k})$ is the defendant’s chance of being sued if he acts and $(\mu D + q)$ is the expected trial loss if the defendant is sued. It will be convenient to define the variables “per-suit deterrence” $\Delta = \mu D + q$ and “all-in deterrence” $\Omega = G(\hat{k})\Delta$. In this notation, the defendant acts if and only if $\beta \geq \Omega$. Note that per-suit deterrence depends on $D$ and $R$, whereas all-in deterrence depends on $D$, $R$, and the filing fee $K$.17

2.4. Social Welfare Problem

The social cost arising from the primary activity is the integration of net harm over acting defendants

$$\int_{x=0}^{x=\Omega} (\bar{b} - b)(\bar{dx}) \text{d}x.$$ (4)

17. Some notes on the generality of this structure. First, the model applies to any primary activity choice that generates emissions. For example, in a tort setting, we can think of “acting” as “acting negligently” and “refraining” as acting with due care. Or we can think of “acting” as “acting negligently” and “refraining” as engaging in a given activity at a high level rather than at a low level. Second, our dual assumption that there is no litigation in the absence of harm and no harm when the defendant refrains simplifies the modeling without changing the basic results. Our main conclusions hold as long as defendants are not likely to file when they act than when they do not. Third, our assumption that recovery and damages are scaled follows Polinsky and Che’s (1991) model and, in any, strictly simplifying. Were the level of damages and recovery a function of the evidence, our basic results would still persist. Fourth, our modeling task is also greatly simplified by having plaintiffs differ by only the fixed component in evidence costs. As a result of this assumption, all plaintiffs will face the same evidence choice problem because their fixed costs differences do not affect their marginal evidence costs. Thus, all filed costs will be the same in terms of evidence production and the probability of plaintiff activity.
The expected social cost of litigation is

$$1 - \phi(0) \int_{0}^{\infty} (k + c + e + \phi(0)k) \, d\phi,$$

where $[1 - \phi(D)]$ represents the proportion of acting defendants, $(k + c + e)$ is the cost of each suit, and the integration is over filed suits. Litigation effort being independent of $k$, we may write $\phi = k + c + e$ for per-suit costs. Expected social costs are then $[1 - \phi(D)] G(\phi)^2$. The socially optimal configuration of $R$, $D$, and $K$ is that which minimizes all-in social cost, the sum of equations (4) and (5).

3. Analysis of the Basic Model

3.1. Separating the Number of Suits from the Cost and Deterrent Force of Each Suit

When changes in litigation effort are drawn into the analysis of decoupling, it becomes important whether any reduction in plaintiffs' expected litigation winnings is imposed up front—like a taker price for playing the "litigation lottery"—or on the back end, as a tax on whatever recovery is obtained. While up-front charges discourage some plaintiffs from filing, they act like "lump-sum taxes" with respect to how plaintiffs behave in suits that are filed. As such, up-front fees will not significantly dampen plaintiffs' favor in prosecuting the suits they choose to file. On the other hand, reducing back-end winnings in an effort to reduce the number of filings will also significantly affect how filed suits proceed as plaintiffs dedicate fewer resources to the case. A corollary to this point is that when up-front fees are a policy variable, as $K$ is here, the job of controlling the number of suits should be fully delegated to such fees, while damages and recovery (that is, the plaintiff's trial outcome-dependent winnings) should be fully determined by their effect on the cost and deterrent force of inframarginal suits. To see why, imagine setting recovery and damages in a way that did not produce the resulting level of per-suit deterrence $\Delta$ at the lowest possible

18. The important aspect of these fees is not their timing per se but the fact that they are not contingent on the outcome of the suit. Under current law, plaintiffs pay "filing fees," but these are usually negligible.

19. Note, however, that wealth effects from up-front fees—via wealth constraints or changes in the marginal utility of "income"—may affect on plaintiffs litigation effort.
per-suit cost $T$. Consider adjusting recovery and damages to change this. From the perspective of total social costs, equation (4) plus (5), we might be concerned that altering recovery and damages to provide the same level of per-suit deterrence at lower per-suit cost would alter the number of filed suits $(1 - f(D)) G(R)$ in such a way that the net effect on social welfare would be negative. But when up-front fees are a policy instrument, this concern is unwarranted. As we adjust recovery and damages to make each filed suit a more efficient provider of deterrent force $\Delta$, we can simultaneously adjust plaintiffs' up-front fee $K$ to cancel out any effect that changes in recovery and damages might have on the number of suits filed. This idea is captured formally in the following proposition:

**Proposition 1.** At the socially optimal levels of recovery, damages, and filing fee, recovery and damages provide their level of per-suit deterrence at the lowest possible per-suit cost.

We will, therefore, focus our attention in the remainder of this section on the problem of providing per-suit deterrence at lowest per-suit cost. Nevertheless, as shown in Section 4.4, our analysis accommodates not only the case where filing fees can be suitably adjusted, as assumed in proposition 1, but also the case in which filing fees are fixed and imputable effects are dominant.

### 3.2. The Marginal Deterrence Cost of Recovery and Damages

Central to the problem of providing per-suit deterrence at minimal per-suit cost is the concept of "marginal (per-suit deterrence) cost" as applied to both recovery and damages. Given current levels of recovery and damages as well as the parties' implied evidence production choices, we may ask, how much more in per-suit costs would we have to invest in order to produce one more unit of per-suit deterrence by changing $R$? By changing $D$? We can answer these questions by differentiating both per-suit costs and per-suit deterrence by $R$ and $D$, respectively, and then considering the ratio of the former derivative over the latter:

$$\text{MDC}_R = \frac{\delta R}{\Delta} \quad \text{and} \quad \text{MDC}_D = \frac{\delta D}{\Delta}.$$

The next proposition establishes that the marginal deterrence costs of recovery and damages must be equal at an interior social optimum. To see why, imagine that each additional unit of per-suit deterrence costs $\$1$ when additional deterrence is provided by increasing $R$ and $\$2$ when
additional deterrence is provided by increasing \( D \). Then decreasing \( D \) by an amount that reduces deterrence by one unit saves $2 in litigation costs per suit, and the lost unit of deterrence can be made up by increasing \( R \) at a cost of only $1. The result is that per-suit deterrence remains constant while per-suit costs fall by $1. (And in the spirit of proposition 1, any effect on the number of suits filed can then be "sterilized" by adjusting the filing fee.)

Proposition 1. At the optimal levels of recovery, damages, and the filing fee, the marginal deterrence cost of recovery must equal that of damages.

Proposition 2 will enable us to make statements about the relative sizes of optimal \( R \) and \( D \). In particular, our strategy will be to identify a region of the parameter space on which damages' marginal deterrence cost \( MDC_D \) strictly exceeds that of recovery \( MDC_R \) whenever it is the case that \( R \leq D \). Given proposition 2, this will imply that the social optimum over this region of the parameter space cannot entail \( R \leq D \).

3.3. The Marginal Deterrence Cost of Recovery

When recovery is no more than damages, each additional dollar of per-suit deterrence generated by increasing recovery increases per-suit costs by no more than $1. In particular,

Proposition 2A. The marginal deterrence cost of recovery \( MDC_R \) is always \( R/D \).

The intuition for why the deterrence cost of recovery is less than one whenever recovery is less than damages is straightforward. When the plaintiff is producing his privately optimal amount of evidence, increasing \( R \) has the same marginal effect on the plaintiff's costs as on her expected winnings (\( pR = c \))—otherwise, she could improve her situation by producing more or less evidence. Furthermore, when \( R \leq D \), the marginal effect on the plaintiff's expected winnings is less than the marginal effect on the defendant's expected loss (\( pD \geq pR \)). By transitivity, therefore, increasing recovery increases the plaintiff's evidence costs by less than it increases the defendant's expected loss.

20. This assumes that this tripartite adjustment does not violate the court system's budget constraint. In fact, it will not, as we explain in our working paper (Chot and Sanchirico 2003).

21. What about the possibility that the plaintiff's recovery affects the defendant's evidence production and vice versa? When trial error is uniformly distributed and not weight
3.4. Marginal Deterrence Cost of Damages in High-Stakes, Deep-Pockets Suits

In this section, we show that if defendants have sufficiently deep pockets and damages are sufficiently large, then the marginal deterrence cost of damages will also be large. First, we must provide an expression for this marginal deterrence cost.

Proposition 2B. The marginal deterrence cost of damages $MD_{DC}$ is $\frac{\partial}{\partial \theta}$.

The derivation of the ratio here is straightforward. Consider first the denominator. As described informally in Section 1, a marginal in-
crease in damages increases the defendant's expected loss in any given suit simply by the probability \( p \) that he will have to pay it. The impact of the defendant's own responsive increase in litigation effort \( y \) drops out because of the "envelope theorem": it must be that the defendant has set his effort level \( y \) so that it has no marginal effect on his expected losses; otherwise, he would not be at an optimum before the change in \( D \).

With regard to the marginal social cost of litigation, the numerator, a marginal increase in damages causes the defendant to increase litigation effort by \( y_n \), which in turn increases the defendant's litigation costs by the marginal cost of the defendant's effort \( c' \). We are now ready to state the main result of the section.

**Proposition 3.** Suppose that \( c' \) is bounded from above. Then the marginal deterrence cost of damages can be made arbitrarily large (whatever the size of recovery) by ensuring that damages and the defendant's wealth are sufficiently large.

The intuition for this result follows easily from the ratio in proposition 2b. A unit increase in damages increases the defendant's expected loss in any given suit by the probability \( p \) that he will have to pay it, which will always be less than one, no matter what the starting level of damages. In contrast, the defendant's litigation costs will be quite responsive to a further increase in damages if the level of damages is already high. To see why, assume for the moment that the defendant's marginal increase in litigation effort in response to an additional unit of damages \( y_n \) is constant in \( D \). In that case, the increase in costs is exactly proportional to the marginal costs of the defendant's litigation effort \( c' \), which the defendant's second-order condition requires to be increasing. Under this scenario, each additional unit of damages insures a fixed amount of additional effort from the defendant, which in turn costs more and more as damages continue to increase. Now remove the assumption that the defendant's responsiveness \( y_n \) is constant. This raises the possibility that the defendant's responsiveness to additional damages declines so rapidly that the effect just identified is dissipated and the ratio falls. The assumption that \( c' \) is bounded from above is sufficient to ensure that such changes in \( y_n \) do not dominate. Many common cost functions satisfy this condition. In any event, we prove proposition 3 in the Appendix for the weaker condition that \( c' \) does not grow exponentially ad infinitum. (See assumption 1 in the Appendix for a formal statement of this condition.)
3.5. Optimal Recovery Is No Less Than Optimal Damages In
High-Stakes, Deep-Pockets Suits

Proposition 2A tells us that the marginal deterrence cost of recovery is less than one whenever recovery is less than damages. Proposition 3 implies that the marginal deterrence cost of damages will strictly exceed one when the defendant has deep pockets and damages are large. Proposition 2 tells us that we cannot be at a social optimum if the marginal deterrence cost of damages strictly exceeds the marginal deterrence cost of recovery. Combining these findings yields the conclusion that if it is socially optimal (and feasible) for the defendant to pay large damages, then it is also socially optimal for the plaintiff to recover at least as much.

To state this result in terms of the primitives of the model, we need only connect the size of optimal damages to an exogenous variable of interest. The size of the potential harm h caused by the defendant’s activity is one such variable. When harm is large (as when the defendant is designing a product for mass distribution or in handling large quantities of toxic waste), the marginal benefit of additional deterrence is also large, and so then is the optimal level of damages. Thus, our main result:

Proposition 4. Suppose that both the defendant’s wealth and the potential harm from the defendant’s activity are sufficiently large. Then, optimal recovery is no less than optimal damages.13

22. Proposition 4 should be carefully contrasted with a similar possibility that arises in Polinsky and Ché (1991) model. Even though recovering recovery while raising damages is always socially welfare improving in Polinsky and Ché’s framework, this does not imply that optimal damages always exceed optimal recovery in their model. "[A]s 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 may exceed the optimal payment by the defendant” (p. 562).

However, in Polinsky and Ché’s model, the possibility that optimal recovery will exceed optimal damages is entirely a result of the fact that the defendant’s wealth constraint always binds at a social optimum. See Polinsky and Ché (1991, p. 563): “If the optimal recovery of discounted liability the defendant’s payment is as high as possible.” And "As harm tends to infinity, the value of taking additional care to reduce the probability of an accident increases without bound. Since [optimal damages] equal [the defendant’s wealth], the only way to induce the defendant to take more care is by suit-for-recovery so that he will be used with a higher probability if an accident occurs. Therefore, as harm tends to infinity, [optimal recovery] must also tend to infinity, even though there is [harm] sufficiently large, [optimal recovery] is strictly greater than [optimal damages]” (p. 366, emphasis added).

The fact that the defendant’s wealth constraint always binds at an optimum in Polinsky and Ché’s framework is, in turn, a result of the fact that Polinsky and Ché do not account...
4. VARIATIONS IN THE BASIC MODEL

4.3. Contingent-Fee Plaintiff Lawyers

In a majority of tort cases, plaintiffs hire lawyers on a contingent-fee basis. In such cases, plaintiffs’ lawyers have greater de facto control over evidence production decisions than their clients. In this section, we show that incorporating these features into our model only reinforces our basic findings.

Suppose the plaintiff’s lawyer receives \( \theta \leq 1 \) fraction of the litigation return but bears the entire cost of evidence production. Assume that the lawyer chooses the amount of evidence to be presented and does so in order to maximize his own return \( \theta p R - c \) rather than that of his client. Thus, the plaintiff’s evidence production and the equilibrium per-suit deterrence is determined by the plaintiff’s lawyer’s first-order condition \( \theta p R - c = m \).

Substituting the lawyer’s first-order condition into the marginal deterrence cost of recovery \( \text{MDC}_R \) yields \( \theta (R/D) \), which is smaller than \( R/D \). At the same time, the marginal deterrence cost of damages \( \text{MDC}_d \) is still \( c/\theta p \). Hence, we have the following corollary:

**Corollary 1.** Suppose the plaintiff’s lawyer receives a \( \theta \) fraction of the return from litigation while bearing the entire cost of evidence production, where \( 0 < \theta \leq 1 \). Suppose also that the plaintiff’s lawyer makes the plaintiff’s evidence production decisions to maximize her own return. Then proposition 4 still holds.

4.2. Settlement

In this section, we incorporate into our model the conventional treatment of settlement, which focuses on the existence and position of the settle-
ment range, and show that this modification does not affect our basic result.23

Suppose the parties can settle the lawsuit after the plaintiff has filed but before proceeding to the evidence production stage. Foremost, in order for them to settle, there must be some surplus from not proceeding to trial. Conven­
tionally, when recovery is less than damages, the most that the defendant would be willing to pay in settlement is always greater than the least the plaintiff would accept: (pD + q) - (pR - c) = p(D - R) - q + c > 0.25 In other words, there is always a positive settlement surplus in this case.

As is well recognized, however, the existence of a positive settlement range is best regarded as a necessary, rather than sufficient, condition for settlement to occur. Even when there is a surplus to be split, not all negotiating partners will be able to agree on precisely what that split should be. Therefore, we assume that some (possibly large) fraction of all cases with positive settlement ranges actually do settle.26 We also

23. Polinsky and Che (1991, pp. 566-68) consider settlement in the context of a perfect-information model in which the plaintiff makes a "take-it-or-leave-it" settlement demand to the defendant and settlement itself imparts costs. Polinsky and Che also study the case in which the court can observe the settlement amount and make additional awards or incur additional costs based on the settlement amount. Such court monitoring of settlement is necessary since otherwise the benefits of decoupling will be undone through settlement. Kelton and Tuckman (1995, pp. 178-79, 181-82) work with a model of settlement that is similar to that in which the parties may have incomplete (although commonly known) beliefs about what will happen at trial. See Daugherty and Reinganum (2003) for a mostly positive analysis of decoupling in the case of asymmetric information settlement bargaining. (Daugherty and Reinganum also study the revenue-maximizing tax on plaintiffs' recovery.)

24. This result is partly due to the fact that the parties agree on the probability of plaintiff's victory. However, even when the parties do not agree, there is no settlement range when recovery is significantly greater than damages. See Polinsky and Che (1991, p. 567 n.13).

25. This note discusses the significance of changes in the size of a monopoly settlement range. In many models with monopoly settlement ranges, there would appear to be no systematic relationship between the size of the settlement range and the likelihood that a settlement will be reached. Thus, there is no reason to believe that a suit whose settlement range is from $100 to $1,000 is any more likely to settle than a suit whose settlement range is from $100 to $10. On the one hand, the benefits of decoupling are greater when there is more territory within the settlement range to be captured, and this acts to make settlement less likely the larger the range. But to what extent this effect dominates is an empirical question that remains to be answered by the literature. In this paper, we treat the phenomenon of nonsentlement in the presence of a settlement range as random and
assert that when the case settles, each party gets a fixed fraction of the settlement surplus. Thus, letting \( \gamma \in [0, 1] \) be the plaintiff's fraction, the settlement amount is a weighted average of the parties' trial payoffs:

\[
S = \gamma p + (1 - \gamma) c = \gamma (pR - c) + (1 - \gamma) (pD + q) = \gamma (pR - c) + (1 - \gamma) (pD + q) - (pR - c).
\]

The defendant's ex ante expected loss from a D-R suit—that is, per-suit deterrence—is also a weighted average of the parties' trial payoffs:

\[
A = (1 - \gamma) (pD + q) + \gamma c = (1 - \gamma) (pD + q) + \gamma (pR - c), \quad \text{where} \quad \gamma = \frac{1 - \delta}{1 - \gamma}.
\]

The problem of providing per-suit deterrence \( A \) at minimal per-suit cost then becomes \( \max_{x, \gamma} \min_{x} (1 - \gamma) (x + c) + (1 - \gamma) (pD + q) + \gamma (pR - c) = 1 \), or equivalently, \( \min_{x, \gamma} c + (1 - \gamma) (pD + q) + \gamma (pR - c) = \delta \).

At optimal values of \( D \) and \( R \), the settlement analogy to proposition 2 must hold. That is, with the uniform distribution of trial errors, the marginal deterrence cost of recovery

\[
MDC_{D} = \frac{c\gamma}{(1 - \gamma) pR + \gamma p^2D + p^2R}
\]

must equal the marginal deterrence cost for damages

\[
MDC_{C} = \frac{c^2\gamma}{(1 - \gamma) pR + \gamma p^2D + p^2R R^2}.
\]

The numerators of these ratios (derived after eliminating \((1 - \gamma)\) from the minimand) are the same as the numerators in the no-settlement case. The presence of settling cases has no effect on the relative per-suit cost of increasing the instruments; the fact that cost increases affect only a

unrelated to the width of the settlement range. Thus, we assume this is enormous.

In any event, when we account for inframarginal suit effects, increasing recovery and decreasing damages does not necessarily decrease the size of the settlement range, as one might imagine. Khan and Terrman (1993, p. 160, proposition 3 and discussion) note a similar ambiguity.

For example, decreasing damages by one unit and increasing recovery by one unit increases the defendant's expected trial loss by \( pDx_{2} = p \). When damages are large, therefore, the effect of increasing recovery will swamp the effect of decreasing damages and the defendant will actually be willing to pay more in settlement on 

For the plaintiff, decreasing damages by one unit and increasing recovery by one unit increases expected trial payoffs by \( p - F_{p}\), where \( p < c \). Therefore, for high- stakes cases, both bounds on the settlement range increase and the issue becomes whether the more the defendant will pay is increasing faster or slower than the least the plaintiff will accept. If \( R \) is significantly less than \( D \), the stakes for both parties are large, and their evidence costs are similar so that the defendant's expected trial losses will increase faster than the plaintiff's expected trial gains, and the settlement range will increase on net.

26. Our results go through as long as the settlement amount is increasing in both the plaintiff's expected trial payoffs and the defendant's expected trial losses.
fraction of filed suits makes both tools more efficient in precisely the same proportion.

Therefore, the difference in these marginal deterrence cost ratios is wholly located in the ratios’ denominators, that is, their effects on per-suit deterrence. In particular, with settlement, the effect on per-suit deterrence of increasing either instrument depends not only on the change in the defendant’s expected trial loss, but also on the change in the plaintiff’s expected trial winnings. The plaintiff’s expected trial winnings affect the size of the settlement amount and so the deterrence force of the suit.

Nevertheless, our results on optimal recovery and damages for the no-settlement case will still pertain. To see why, when we compare the denominators of these two ratios, \((1 - \eta) x_y D + \eta p\) versus \((1 - \eta) + \eta y_y R\), where \(p\) is negative and \(y_y\) is positive, we see that each is a weighted average of two expressions with the same weights. The expressions weighted by \((1 - \eta)\) in each denominator are the same as for the no-settlement case. The new expressions, those weighted by \(\eta\), are exactly and unambiguously compared. For the marginal deterrence cost of recovery, we are averaging in a positive number \(p\), whereas for the marginal deterrence cost of damages, we are averaging in a negative number \(p x_y R\). Thus, incorporating settlement into the model makes damages relatively less deterrence efficient.

Intuitively, while a unit increase in damages increases the defendant’s expected loss from trial as before (by \(p\)), it also reduces the size of the settlement amount by depressing (by \(-p x_y R > 0\)) the plaintiff’s expected trial payoffs and thus her threat point in bargaining. The plaintiff’s expected trial payoffs decline because greater damage would induce the defendant to lodge a more vigorous defense should the case proceed to trial. On the other hand, recovery remains attractive as an instrument for creating per-suit deterrence. A unit increase in recovery increases both the plaintiff’s expected return (which in turn increases the size of settlement by \(p\)) and the defendant’s expected loss from trial (by \(p x_y D\)).

**Proposition 5.** Proposition 4 still holds when the possibility of settlement is incorporated into the model.

### 4.3. The British Rule

Thus far, we have assumed that each party bears its own litigation cost. Although this is the most prevalent form of litigation cost allocation in the United States, other countries, such as Britain, have adopted a “loser-
pay-all" rule. An impressive literature catalogues the pluses and minuses of each approach, but for the purposes of our paper, whether the cost allocation rule is American or British makes no difference.

Indeed, the marginal deterrence cost of recovery is smaller under the British rule than under the American. In particular, any given increase in the plaintiff's evidence production has a larger negative impact on the defendant's expected trial payoffs. As under the American rule, increasing recovery causes the plaintiff to produce more evidence, and this, in turn, increases the probability of a plaintiff victory. This increase will increase the defendant's expected trial losses by a larger amount under the British rule because a losing defendant must now pay the plaintiff's costs as well as his own. Moreover, under the British rule, the increase in the plaintiff's evidence costs will also directly increase the size of the losing defendant's pay-off.

On the other hand, the marginal deterrence cost for damages is the same under the British rule as under the American. While it is true that a given increase in the defendant's evidence reduces the defendant's expected trial losses by a larger amount under the British rule than under the American, because the defendant has already accounted for this in setting his level of evidence production, the impact of the defendant's own adjustment in evidence production has no marginal effect on his expected trial losses.

Proposition 6. Proposition 4 still holds if litigation costs are allocated according to the British rule.

4.4. When Filing Fees Are Negligible and Not Adjustable but Inframarginal Effects Dominate

Our analysis has focused on the per-suit effects of changes in recovery and damages. In Section 3.1, we justified this focus by assuming that the policymaker can impose fees on the plaintiff—we have called them "filing fees"—that do not depend on the outcome of the suit.28 Another reason to focus on per-suit effects would be that they are relatively important empirically. In this section, we investigate the theoretical im-

27. This rule is also used in the United States in certain limited circumstances. For example, under 42 U.S. Code, sec. 1983 and 1988, losing defendants must pay plaintiffs' attorneys' fees in certain civil rights actions that meet very high standards. Costs other than for the attorneys' fees are also routinely shared under rules such as Federal Rule of Civil Procedure 36(d)(1).
28. This was discussed in Section 3.1. Recall that the filing fee is just a stand-in for charges that are not dependent on the outcome of the suit.
applications of this second approach. We examine the relative size of optimal damages and recovery under the restriction that filing fees are negligible (more generally, not adjustable). We then find a region of the parameter space over which per-suit effects dominate so that our prior results on the relative size of optimal recovery and damages continue to hold.

When filing fees are restricted to be negligible, we must expand our analysis of recovery and damages to encompass the effect of these instruments on all-in social costs via the number of suits filed. Increasing recovery will increase the number of suits filed. More filings will mean both more all-in deterrence and greater litigation costs. Increasing damages, on the other hand, will decrease the number of suits filed, and this will act to decrease both all-in deterrence and litigation costs.

Filing effects operate through changes in the marginal filer $k$. The impact on social costs for any given charge in the marginal filer depends on how many additional plaintiffs are affected by that change, which depends, in turn, on the height of the density $g(k)$. When this density is small, the impact on all-in social costs of a given change in the marginal files will be small. At the same time, per-suit effects are independent of the density $g$ of plaintiffs’ fixed evidence costs. Therefore, per-suit effects will dominate in comparing the all-in deterrence efficiency of recovery and damages when the distribution of plaintiffs’ costs is spread thinly along the number line. Such an attribute is present in cases where there is a large degree of heterogeneity among potential plaintiffs, such as in large product liability suits wherein numerous injured consumers have widely varying access costs to the legal system. It is again important to note that these cases are reminiscent of those that typically inspire calls for tort reform, including proposals to decouple damages from recovery. 29

29. The derivatives of the marginal file in $R$ and $D$ are $p = p_{\text{d}}$ and $p$, respectively. Aggregation of the discussion of complications arising from cross effects is made in note 21. Increasing recovery may usually decrease filings when cross effects are strong. The defendant may respond to additional plaintiff evidence with more evidence of his own, and this will act to decrease the plaintiff’s trial payoffs. This indirect effect on the plaintiff’s expected trial payoffs may outweigh the direct effect of increasing recovery.

30. Indeed, a somewhat more general version of our model, for condition on plaintiffs’ costs also corresponds to another of the complaints about litigation that has inspired reforms such as decoupling. Specifically, in a model where all defendants might be sued even if they are not at fault in the primary activity, there will be two denials for plaintiffs’ costs—one for when the defendant acts and one for when he refrains. (The difference between these two denials will be the source of the defendant’s primary activity incentives: passive incentives require that the defendants be less likely to be sued than is plaintiffs’
Proposition 7. Consider the case where filing fees are restricted to be negligible (more generally, not adjustable). Suppose that the defendant's wealth is sufficiently large and the distribution of plaintiffs' costs is sufficiently diffuse. Then proposition 4 still holds.

5. CONCLUSION

Should plaintiffs win what defendants lose? Answering this question requires examining how litigation stakes influence not only the number of suits filed, but also the manner in which filed suits proceed. Existing research has uncovered important lessons about the effect of litigation stakes on filings and about the effect of plaintiffs' stakes on per-suit costs. The primary contribution of this paper has been to expand the analysis to include the effect of both parties' stakes on both filings and inframarginal suits. This expansion has itself uncovered some lessons. In particular, the literature's provisional conclusion that plaintiffs' recovery should be less than defendants' damages no longer holds in all cases. Moreover, among the cases in which the conclusion does not hold, we find precisely the negative paradigm of modern litigation that has inspired some policy commentators to advocate awarding plaintiffs less than what defendants pay.

Our paper may also offer a more general lesson about the law and economics of litigation. With some exceptions, most of this literature focuses on the incentive to file and to settle, leaving discovery and evidence production relatively undermodeled. While this has certainly been a successful research strategy to date, the analysis in this paper indicates that adding to the model even the broadest outlines of how filed suits proceed may have a significant effect on what conclusions can be drawn from the analysis.

notes (endnotes added here, see, for example, Grig and McAller (2000), Baye, Kovash, and De Vries (2000), Sanchisco (2000, p. 362, 2001, p. 320), and Bernardo, Talley, and Welch (2000)).
APPENDIX: FORMAL STATEMENTS AND PROOFS OF PROPOSITIONS

1. Proposition 1

Formal Statement of Proposition 1. If \( R^*, D^*, \) and \( K^* \) minimize all-in social cost while generating marginal files \( \hat{e}^* \), per-suit deterrence \( \Delta^* \), and per-suit evidence costs \( e^* + c^* \), then \( R^* \) and \( D^* \) also solve the problem:

\[
\min_{x \geq 0} c + qD + c = \Delta^*.
\]

Proof. Suppose, on the contrary, that \( \hat{R} \) and \( \hat{D} \) yield per-suit deterrence \( \Delta^* \) at lower per-suit cost \( \hat{c} = c + \hat{c}^* = c^* \). Set the filing for \( K^* \) so that \( \hat{p} \hat{K} - \hat{c} \hat{K} = \hat{p} \hat{K}^* - c \hat{K}^* \). Since \( \hat{R}, \hat{D}, \) and \( \hat{K} \) yield the same set of filing plaintiffs \([0, K^*]\) and the same per-suit deterrence \( \Delta^* \) as \( R^*, D^*, \) and \( K^* \), they also yield the same all-in deterrence \( \Delta^* \). Therefore, they yield the same primary activity costs, equation (4). However, each suit is strictly less costly. Therefore, litigation costs, equation (5), are strictly lower under \( \hat{R}, \hat{D}, \) and \( \hat{K} \). This contradicts the statement that \( R^*, D^*, \) and \( K^* \) minimize all-in social costs. Q.E.D.

2. Proposition 2

Formal Statement of Proposition 2. At any interior social optimum at which marginal deterrence costs for \( R \) and \( D \) are finite and positive, \( \text{MDC}_0 = \text{MDC}_0 \).

Proof. The problem of providing a given level of per-suit deterrence at minimal per-suit cost is:

\[
\min_{x \geq 0} c(x) + c(y) \text{ subject to } p(x) + c(y) = \Delta^*.
\]

First-order conditions for an interior solution to this problem can be written as:

\[
c'x + c'y = \lambda x p_x D \text{ and } c'x + c'y = \lambda y p_y D_x D_y,
\]

where \( \lambda \) is the Lagrange multiplier. If \( p_x p_y D_x D_y \neq 0 \) and \( p_x p_y D_x D_y = 0 \), then we can isolate \( \lambda \) on the right-hand side of both equations. The resulting ratios on the left-hand side will both be equal to \( \lambda \) and so equal to each other. On the other hand, if \( p_x p_y D_x D_y = 0 \), then the first-order condition implies that \( c'x + c'y = 0 \) and \( \text{MDC}_0 \) is not well defined, having zeros in both the numerator and denominator. Similarly, if \( p_x p_y D_x D_y = 0 \), then \( \text{MDC}_0 \) is not well defined. Q.E.D.

3. Proposition 2A

A marginal increase in \( R \) will increase per-suit cost by \( c'x + c'y \) and per-suit deterrence by \( p_x D_x + p_y D_y + c'y \). Using the defendant's first-order condition \( (p_x D_x + c'y = 0) \) to simplify the derivative of per-suit deterrence and taking the ratio of the two derivatives yields:

\[
\frac{\text{MDC}_0}{\text{MDC}_0} = \frac{c'x + c'y}{p_x D_x}.
\]

From equation (1), \( p_x = -1/2c'y \), and therefore, \( p_x = 0 \). By the implicit func-
tion theorem applied to equations (1a), \( y_1 = -\left(p_r \partial D|_{p_r} D\right) = 0 \). By similar reasoning, \( c_{r} = 0 \).

Therefore, the marginal deterrence cost of recovery \( c' x_D + c'y_D|_{p_r} D x_D \) reduces to

\[
MDC_r = \frac{c' x_D}{p_r D x_D} = \frac{p_r R x_D}{p_r D x_D} = \frac{R}{D}
\]

where we have substituted from the plaintiff's first order condition \( p_r R - c' = 0 \). Q.E.D.

4. Proposition 28

Increasing \( D \) will increase per unit cost by \( c' x_D + c'y_D \) and per unit deterrence by \( p_r D x_D + p_r D y_D + p \). Again, using the defendant's first-order condition \( p_r D + c' = 0 \) to simplify the derivative of per unit deterrence, we obtain

\[
MDC_D = \frac{c' x_D + c'y_D}{p_r D x_D + p}
\]

Given zero cross effects, as noted in the proof of proposition 2A,

\[
MDC_D = \frac{c'y_D}{p}
\]

Q.E.D.

5. Proposition 3

We prove something stronger than proposition 3. In particular, rather than assume that the second derivative \( c'' \) of defendant's cost function is bounded, we assume only that it does not grow exponentially ad infinitum.

Assumption 1. \( \lim_{y \to \infty} c(y) = 0 \).

This condition holds for a great variety of functional forms, including all polynomial functions \( c(y) = \alpha y^\nu + \cdots + \alpha y + \alpha_0 \) on \( y \geq 0 \). For example, it holds for \( c(y) = y^\epsilon \), whereas the boundedness condition in the main text does not. [Still, not all cost functions satisfy the assumption; consider, for example, the exponential cost function \( c = e^\nu \).] The assumption plays a role in ensuring that the defendant's evidentiary response to changes in damages does not decay too quickly.

Now on to the proof of (strengthened) proposition 3. First, because \( p \leq 1 \), the marginal deterrence cost for damages is always greater than \( c'y_D \). By the implicit
function theorem applied to equations (2a), \( y_0 = -\frac{p_2}{\partial_0}D = -\ddot{c}' \). By equation (3), \( p = \{ \overline{\gamma} - \overline{\alpha}/\overline{\beta}, \overline{x} = -1/\overline{\beta}, \) and \( p_2 = 0 \). Therefore,

\[
y_0 = -\frac{1}{2\overline{\beta}c''}. \tag{6}
\]

Furthermore, equations (2a) imply that \( \dot{c}' = -p_2D = (1/2\overline{\alpha})D \). Therefore,

\[
\text{MDC}_c = \frac{D}{2\overline{\alpha}c'} \tag{7}
\]

Noting that \( \dot{c}' \) is a function of \( \gamma \), which is in turn a function of \( D \), we now use assumption 1 to show that the factor \( D^2/\gamma \) on the right-hand side of this inequality never becomes infinite in \( D \). Suppose, on the contrary, that \( \lim_{c\to\infty}(D^2/\gamma) = \infty \). Then, because the numerator of this fraction does go to infinity, the denominator must be zero. (L'Hôpital's rule then applies. Therefore, using equation (6), \( \lim_{c\to\infty}(D^2/\gamma) = \lim_{c\to\infty}(1/\gamma) = 2(\overline{\alpha}m_{c'})^2 \). The right-hand limit is infinite by assumption, which contradicts our supposition. Q.E.D.

6. Proposition 4

Proof. First, we establish that per-case deterrence at a social optimum grows without bound with the level of harm. That is, letting \( \Delta^c \) represent the per-case deterrence created by optimal values \( h \) and \( D \), we show that \( \lim_{c\to\infty}\Delta^c = \infty \). Suppose, on the contrary, that \( \Delta^c \) remains bounded. Then all-in deterrence \( \Delta = G(\Delta^c) \) is also bounded, given that \( G(\Delta) \leq 1 \). Let \( B \) be the bound on all-in deterrence. However, if \( h \) is large enough, the marginal all-in social cost of increasing all-in deterrence (via changes in \( h, D, \) and \( R \)) will be strictly negative; at all levels of all in deterrence below \( B \), which contradicts the social optimality of \( \Delta^c \).

To see precisely why, take any \( h \leq B \) and suppose that this all-in deterrence is caused by some \( h, D, \) and \( R \). Now consider changing both \( h \) and \( D \) in such a manner that per-case deterrence increases. This will always be possible: an increase in \( D \), holding the plaintiff's evidence constant, will always increase the defendant's expected trial losses by \( p_2 \) the change in \( h \) can then ensure that the plaintiff does not change her evidence production; lastly, \( R \) must be strictly positive if we are at any positive level of deterrence.) Now that this change in \( h \) and \( D \) may also change per-case costs. Consider also adjusting \( R \) to keep the number of filings constant. The net effect of these changes in \( h, D, \) and \( R \) will be an increase in all-in deterrence \( \Delta \). and some change, positive or negative, in per-case costs. For large enough values of \( h \), the positive social cost effect of the former will outweigh the latter per-case cost effect because

\[
\int_0^{a+h} \frac{(\alpha + c' + \gamma b)\gamma}{\gamma + c'} \, \frac{1}{\gamma} \, d\gamma = -\left( a + D(x(x)) - \int_0^\infty (\alpha + c' + \gamma b)\gamma \right)
\]
goes to negative infinity as \( h \) approaches infinity, where the pre-suit cost effect is finite and independent of \( h \). Therefore, these levels of \( E, D, \) and \( R \) could not be socially optimal for large enough values of \( h \).

Next we show that for large enough optimal per suit deterrence \( D^* \), optimal recovery is no less than optimal damages. Suppose, on the contrary, that there exists a sequence of \( D^* \), \( R^* \), and \( D^* \) wherein \( \frac{R}{D} = m \) and \( R^* \leq D^* \). By propositions 2 and 2.4, this implies that MDC1 ≤ 1 along the sequence as well. Then by proposition 3, \( D^* \), and so \( R^* \) by hypothesis, are bounded below some \( B^* \) along this sequence. But this contradicts that \( D^* \), a continuous function of \( R \) and \( D \), grows without bound along the sequence.

7. **Proposition 5**

\[
(MDC_0)^{-1} = \left[ \frac{c}{1 - \psi(p + c + q)} \right] = \left[ 1 - \psi \right] \frac{D}{R} \left[ 1 - \psi \right] \frac{p}{C_k} \geq \left[ 1 - \psi \right] \frac{D}{R}.
\]

\[
(MDC_0)^{-1} = \left[ \frac{c}{1 - \psi(p + c + q)} \right] = \left[ 1 - \psi \right] \frac{p}{C_k} \left[ 1 - \psi \right] \frac{D}{R} \leq \left[ 1 - \psi \right] \frac{p}{C_k}.
\]

We then apply the logic of the proof of propositions 3 and 4 to these results.

8. **Proposition 6**

**Proof.** Under the British rule, the plaintiff maximizes \( pR - (1 - p)c + q_1 \), while the defendant minimizes \( pD \). Under the American rule, the parties' first-order conditions are \( pR = c + p = 1 - p(c + q) = 0 \) and \( pD = c + p = 0 \). The marginal cost of damages is the same under both rules.

\[
MDC_0 = \frac{\psi_k}{p} \left[ b(pD + c + q) + pcD \right] = \psi_k.
\]

The marginal cost of recovery has a scale-independent bound when \( R \leq D \):

\[
MDC_0 = \left[ b(pD + c + q) + pcD \right] = \psi_k.
\]

which is less than one when \( R \leq D \). We then apply the logic of the proof of propositions 3 and 4.

9. **Proposition 7**

**Proof.** It suffices to prove that proposition 2 (regarding the equality of pre-suit marginal deterrence costs at a social optimum) holds in the limit as
Suppose $g = 0$. First, holding $K$ constant, the partial derivatives of all-in social cost with respect to $R$ and $D$ are

$$
\frac{\partial SC}{\partial R} = -Q_2(0)|\hat{\theta} - \theta + G(\hat{k})|\hat{\alpha}(\hat{k}) + c + \omega|
$$

$$
+ (1 - f_0(\theta)|\hat{\alpha} + c + \omega|G(\hat{k}) + \phi \alpha G(\hat{k}),
$$

$$
\frac{\partial SC}{\partial D} = -Q_2(0)|\hat{\theta} - \theta + G(\hat{k})|\hat{\alpha}(\hat{k}) + c + \omega|
$$

$$
+ (1 - f_0(\theta)|\hat{\alpha} + c + \omega|G(\hat{k}) + \phi \alpha G(\hat{k}),
$$

where $\mu(k) = E(\theta | x_k = \hat{k})$, $Q_2 = g(\hat{k}, \hat{x}_D), Q_3 = g(\hat{k}, \hat{x}_D) + G(\hat{k})$, $\alpha_2 = \alpha(x_D)$, and $\alpha_3 = \alpha(x_D)$. When $sup g = 0$, we have $Q_2 = G(\hat{k}, \hat{x}_D)$ and $Q_3 = G(\hat{k}, \hat{x}_D)$. Furthermore, $\alpha_2$ and $\alpha_3$ are bounded on $[0, D) \in [0, W)$. Therefore, in the limit

$$
\frac{\partial SC}{\partial R} = -Q_2(0)|\hat{\theta} - \theta + G(\hat{k})|\hat{\alpha}(\hat{k}) + c + \omega| (1 - f_0(\theta)|\hat{\alpha} + c + \omega|G(\hat{k}),
$$

$$
\frac{\partial SC}{\partial D} = -Q_2(0)|\hat{\theta} - \theta + G(\hat{k})|\hat{\alpha}(\hat{k}) + c + \omega| (1 - f_0(\theta)|\hat{\alpha} + c + \omega|G(\hat{k}).
$$

These expressions, evaluated at the (possibly changing) social optimum, grow arbitrarily close to zero as $sup g = 0$. In the limit, therefore, the social optimum satisfies

$$
\frac{(1 - f_0(\theta)|\hat{\alpha} + c + \omega|G(\hat{k})}{G(\hat{k}, \hat{x}_D)|\hat{\theta} - \theta + G(\hat{k})|\hat{\alpha}(\hat{k}) + c + \omega|G(\hat{k})}
$$

This equation reduces to $R/D = \rho(\alpha|\hat{x}_D)$, which is the equality in proposition 2 with substitutions from propositions 2A and 2B. Q.E.D.

REFERENCES


