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

Albert Choi and Chris William Sanchirico

USC Center for Law, Economics & Organization
Research Paper No. C02-7

CENTER FOR LAW, ECONOMICS AND ORGANIZATION RESEARCH PAPER SERIES

Sponsored by the John M. Olin Foundation

University of Southern California Law School
Los Angeles, CA 90089-0071

Should Plaintiffs Win What Defendants Lose?

Litigation Stakes, Litigation Effort, and the Benefits of Decoupling[§]

Albert Choi[*]
Chris William Sanchirico[**]

April 8, 2002[+]

ABSTRACT: Professors Polinsky and Che advocate "decoupling" what plaintiffs recover from what defendants pay in damages, specifically arguing that lowering recovery and raising damages (by appropriate amounts) delivers the same level of primary activity deterrence with fewer filed suits. Professors Kahan and Tuckman extend Polinsky and Che's analysis to account for the effect of parties' litigation stakes on the cost of each filed suit, provisionally concluding that Polinsky and Che's basic argument remains intact. This article reaches a different conclusion. We show that when the effect of litigation stakes on litigation effort is more fully taken into account, 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. Of rhetorical significance in the current policy debate, we find that such suits resemble the negative picture of modern litigation invoked by some advocates of reduced recovery. Our basic findings are robust to the possibility of out-of-court settlement, plaintiffs' employment of contingent fee lawyers, and alternative fee-shifting rules.

[§] For helpful comments and suggestions we thank Linda Cohen, Gillian Hadfield, and participants at USC Law School's Conference on Mechanism Design and the Law (February 2002) and Boston College Law School's Faculty Workshop (February 2002),

[*] UVA Economics Department, Rouss Hall 114, PO Box 400182, Charlottesville, VA 22904-4182; ahc4p@virginia.edu; (434) 924-7845.

[**] UVA Law School, 580 Massie Road, Charlottesville, VA 22903; csanchirico@virginia.edu; (434) 924-3229; www.cstone.net/~csanchir; http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=2.

[+] First draft: February 2002.

I. INTRODUCTION **1**

II. THE BASIC MODEL **9**

A. TRIAL 10
B. PLAINTIFF'S FILING DECISION 12
C. DEFENDANT'S PRIMARY ACTIVITY CHOICE 12
D. SOCIAL WELFARE PROBLEM 13

III. ANALYSIS OF THE BASIC MODEL **14**

A. SEPARATING THE NUMBER OF SUITS FROM THE COST AND DETERRENT FORCE OF EACH SUIT 14
B. THE MARGINAL DETERRENCE COST OF RECOVERY AND DAMAGES 15
C. UPPER BOUND ON THE MARGINAL DETERRENCE COST OF RECOVERY 17
D. DAMAGES' MARGINAL DETERRENCE COST IN HIGH-STAKES, DEEP-POCKETS SUITS WITH ERROR-PRONE JURIES. 18
E. OPTIMAL RECOVERY NO LESS THAN OPTIMAL DAMAGES IN HIGH-STAKES, DEEP-POCKETS SUITS WITH ERROR-PRONE JURIES 20

IV. NUMERICAL EXAMPLES **21**

A. FUNCTIONAL FORM ASSUMPTIONS 21
B. CENTRAL CASE 22
C. CHANGES IN PARAMETER VALUES 23

V. VARIATIONS IN THE BASIC MODEL **26**

A. CONTINGENT FEE PLAINTIFF LAWYERS 26
B. SETTLEMENT 28
C. THE BRITISH RULE 31
D. BUDGET BALANCE IN TRANSFERS 32
E. WHEN FILING FEES ARE NEGLIGIBLE AND NOT ADJUSTABLE, BUT INFRA-MARGINAL EFFECTS DOMINATE 33

VI. CONCLUSION **35**

VII. TECHNICAL APPENDIX **37**

A. SOME TECHNICAL REMARKS ON THE EVIDENCE PRODUCTION GAME AT TRIAL 37
B. PROPOSITION 1 37
C. PROPOSITION 2 37
D. PROPOSITION 3 38
E. PROPOSITION 4 39
F. PROPOSITION 6 41
G. PROPOSITION 7 42
H. PROPOSITION 8 43
I. PROPOSITION 9 44

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 would agree that effective deterrence—a well-acknowledged purpose of punitive awards—generally requires that damages be something more than purely compensatory. Were injurers always called to task for the harm they caused, compensatory damages would produce appropriate deterrence. But 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.[1]

What remains less clear is why plaintiffs should be the beneficiaries of this upward adjustment. 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

---

[1] Jeremy Bentham, The Theory of Legislation 325 (1931) and Gary Becker, Crime and Punishment: An Economic Approach, 76 J. Pol. Econ. 169 (1968).

The literature contains some important qualifications to this result. One well-known qualification emphasizes the fact that the probability of liability is not exogenous, but rather is affected by the defendant's primary activity action. In this case, grossing up compensatory damages by the probability of detection at the socially desired level of care (to take an example from torts) will actually encourage the defendant to take too much care. See, John E. Calfee & Richard Craswell, Some Effects of Uncertainty on Compliance with Legal Standards, 70 Va. L. Rev. 965, 995 (1984); C. Goetz, Cases and Materials on Law and Economics 299-303 (1984); Richard Craswell & John E. Calfee, Incentives to Comply with Uncertain Legal Standards, J. L. Econ. & Org. (1986). The optimal (fixed) damages multiplier will depend on the relative elasticities of the probability of liability and the amount of harm caused (both calculated with respect to the defendant's level of care). The optimal multiplier may indeed be less than one, but only when the probability of liability is significantly more elastic than the level of harm. (For example, if the probability that the defendant will be held liable even when he takes the socially desired level of care is 10%, then the optimal damages multiplier will be less than one, only if the probability of liability is nine times more elastic than the level of harm.) Furthermore, as Calfee and Craswell are careful to point out, "a *variable* multiplier that multiplied each defendant's damages by one divided by *that* defendant's chance of being found liable would actually give defendants the correct incentives" and the resulting damages multiplier would always be greater than one. [emphasis added] Calfee and Craswell at 995 n69.

Explicitly accounting for the costs of litigation also complicates the conventional analysis as described in this first paragraph of text. For example, compensatory damages may not, in fact, be socially optimal even in a world in which defendants are always liable for the harm they caused. On the effects of accounting for costs (when damages equal recovery), see, e.g., A. Mitchell Polinsky & Daniel Rubinfeld, The Welfare Implications of Costly Litigation for the Level of Liability, 17 J. Legal Stud., 151-164 (1988).

For a recent, general discussion of the law and economics of punitive damages, see A. Mitchell Polinsky & Steven Shavell, Punitive Damages: An Economic Analysis, 111 Harv. L. Rev. 869 (1998).

that they impose on others? Why should we assume that the appropriate "bounty" for plaintiffs is the same as the appropriate "fine" for the defendants?

In a deservedly influential paper, Professors Polinsky and Che make a powerful case for why the award paid to plaintiffs should generally be *less* than the liability imposed on defendants.[2] 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 simultaneously lowering recovery and increasing damages, thus "decoupling" the two transfers. As depicted schematically in Figure 1, lowering recovery reduces the number of cases filed, thus reducing both the social costs of litigation and deterrence. Increasing damages restores deterrence without increasing the number of filings, and so without also restoring litigation costs.[3]

$$\left.\begin{array}{c}\Uparrow\\ \text{Damages}\end{array}\right\}\ \text{more deterrence}$$

$$\left.\begin{array}{c}\text{Recovery}\\ \Downarrow\end{array}\right\}\ \text{less deterrence; fewer suits}$$

Figure 1: Schematic of Polinsky & Che Argument

As Professors Kahan and Tuckman have pointed out, however, Polinsky and Che's analysis does not address an important part of the problem.[4] Polinsky and Che's prescription to reduce

[2] A. Mitchell Polinsky and Yeon-Koo Che, "Decoupling liability: optimal incentives for care and litigation," 22 RAND J. Econ. 562 (1991). See also Hylton, Keith, "The influence of Litigation Costs on Deterrence Under Strict Liability and Under Negligence," 10 Int. Rev. L & Econ. 161 (1990); Katz, Avery, "The Effect of Frivolous Lawsuits on the Settlement of Litigation," 10 Int. Rev. L & Econ. 3 (1990); Polinsky, A. Mitchell, "Detrebling versus Decoupling Antitrust Damages: Lessons from the Theory of Enforcement," 74 Georgetown L. J. 1231 (1986); Polinsky, A. Mitchell & Steven Shavell, "Legal Error, Litigation, and the Incentive to Obey the Law," 5 J. L., Econ. & Org. 99 (1989). Steve Salop & Lawrence White, "Economic Analysis of Private Antitrust Enforcement," 74 Georgetown L. J. 1001 (1986). Polinsky and Che attribute the idea of decoupling liability to Warren F. Schwartz, "An Overview of the Economics of Antitrust Enforcement," 68 Georgetown L. J. 1075, 1093 (1980).

[3] Importantly, this maneuver—lowering recovery while raising damages—is always social welfare improving in Polinsky and Che's model. Yet, as Polinsky and Che explain, this does not imply that optimal damages always exceed optimal recovery. This is because the maneuver is not feasible when damages are already equal to all of the defendant's wealth (which, in Polinsky and Che's framework, is always the case at a social optimum). For more on this point, see note 23, infra. Given that actual damage levels do not often bankrupt 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 raised.

[4] Marcel Kahan & Bruce Tuckman, Special Levies on Pun itive Damages: Decoupling, Agency Problems, and Litigation Expenditures, 15 Int. Rev. L. & Econ. 175 (1995). Kahan and Tuckman also extend Polinsky and Che's analysis by considering the effect of agency problems between lawyer and client. See, e.g., Id. [Proposition 2]. We consider one such agency problem in Section V.A.

recovery and increase damages considers the effect of this policy change on the number of suits filed, but not the effect on how filed suits proceed. Which lawyer a party retains; how many billable hours are authorized; whom the party hires as an expert;[5] whom the party hires as an investigator; how many witnesses she deposes,[6] and for how long; how many documents and things she requests for inspection;[7] how prepared she is to resist such requests from the other side;[8] and 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, "higher stakes are associated with significantly higher total lawyer work hours, significantly higher lawyer work hours on discovery, and significantly longer time to disposition."[9] Specifically, median total lawyer work hours were more than two and a half times larger for cases with monetary stakes over $500,000 than for cases with monetary stakes $500,000 or less, while mean total lawyer work hours were almost four times larger.[10]

Thus, Kahan and Tuckman do much to advance our understanding of the benefits of decoupling simply by broadening the scope of the analysis to include not just changes in filings, but also changes in how "infra-marginal suits" proceed. In the end, however, Kahan and Tuckman provisionally conclude that such infra-marginal effects do not fundamentally alter Polinsky and

---

[5] See, e.g., Fed. R. Evid. 702 ["Testimony of Experts"]; Fed. R. Civ. Pro. 26(a)(2) ["General Provisions Governing Discovery; Duty of Disclosure," "Disclosure of Expert Testimony"], and 26(b)(4) ["Trial Preparation: Experts"].

[6] See, e.g., Fed. R. Civ. Pro. 30 ("Depositions upon Oral Questions"). Note that Fed. R. Civ. Pro. 30(a)(1)(A), 30(d)(2) limit the number and duration of depositions absent stipulation of the parties or express authorization by the court. Furthermore, Fed. R. Civ. Pro. 26(b)(2) grants the court broad authority to impose discovery limits.

[7] See, e.g., Fed. R. Civ. Pro. 34 ("Production of Documents and Things and Entry Upon Land for Inspection and Other Purposes;" for parties) and 45 ("Subpoena;" for nonparties).

[8] See, e.g., Fed. R. Civ. Pro. 26(c) ["General Provisions Governing Discovery; Duty of Disclosure," "Protective orders"].

[9] James S. Kakalik, Deborah R. Hensler, Daniel McCaffrey, Marian Oshiro, Nicholas M. Pace, and Mary E. Vaiana, Discovery Management: Further Analysis of the Civil Justice Reform Act Evaluation Data, 39 B.C. L. Rev 613, 638 (1998) ("Second study") ; James S. Kakalik, Terence Dunworth, Laural A. Hill, Daniel McCaffrey, Marian Oshiro, Nicholas M. Pace, and Mary E. Vaiana, An Evaluation of Judicial Case Management Under the Civil Justice Reform Act, RAND, MR-802-ICJ (1996).

[10] Katalik et al., Second Study, at 648 (Table 2.8). These results concern only cases closing in nine or more months after filing.

Another empirical study using independent data found that that "the size of the monetary stakes in the case had the strongest relationship to total litigation costs of any of the case characteristics we studied." Thomas E. Willging, Donna Stienstra, John Shapard & Dean Miletich, An Empirical Study of Discovery and Disclosure Practice Under the 1993 Federal Rule Amendments, 39 BC L Rev 525, 527, 532 (1998). This study also found

Che's basic argument for decoupling.[11] This article reaches a different conclusion. We show that when infra-marginal suit effects are fully accounted for, the policy maneuver by which Polinsky and Che prove the benefits of decoupling—lowering recovery and raising damages—may no longer be welfare-improving.[12] 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 rhetorical significance in the current policy debate—we find that this class of cases bears a strong resemblance to the negative prototype of litigation invoked by some advocates of reduced recovery.

Our conclusions about the welfare implications of Polinsky and Che's maneuver differ from those of Kahan and Tuckman because we consider the infra-marginal effects of both decreasing recovery *and* increasing damages, while Kahan and Tuckman focus only on the infra-marginal effect of decreasing recovery.[13] As Kahan and Tuckman rightly argue, the infra-marginal effect of decreasing recovery, taken alone, merely amplifies the effects considered by Polinsky and Che. When recovery is reduced, both litigation costs and deterrence still fall—now, not just because there are fewer suits, but also because plaintiffs who still file suit pursue their cases less vigorously.[14] This is shown in Figure 2, which adds to Figure 1 the infra-marginal effects of decreasing recovery (in capital letters).

---

that "the stakes in the litigation were positively correlated with the length of the case: the higher the stakes, the longer the case lasted." Id. at 533. Note that this study has a rather special definition of "stakes." See Id. n36.

[11] Id. at 180 ("We find that in the absence of agency problems in the plaintiff-lawyer relationship, special levies [i.e., the reduction of recovery] reduce litigation costs and the expected award payable by the defendant in the case of trial. To that extent, special levies combined with increased awards could be used to reduce litigation costs while maintaining deterrence, as suggested by Polinsky and Che.").

[12] Professor Kamar offers another reason why decoupling may not be welfare improving. Ehud Kamar, "Shareholder Litigation Under Indeterminate Corporate Law," 66 U Chi L Rev 887 (1999). (Arguing that the frequent enforcement and low sanctions regime created by indemnification of corporate fiduciaries by the shareholders that sue them can actually be beneficial in that it increases litigation and so reduces legal uncertainty and is preferable to high sanctions and low enforcement for risk-averse fiduciaries).

More recently, Professors Daughety and Reinganum have provided an extensive (mostly) positive analysis of the effect on asymmetric-information settlement bargaining of forcing plaintiffs to "split" part of their punitive damages award with the state. Daughety and Reinganum find that split award statutes generally lead to more frequent settlement at lower amounts. Andrew Daughety and Jennifer Reinganum, "Found Money? Split Award Statutes and Settlement of Punitive Damages Cases," (2001), available at www.ssrn.com.

[13] See, e.g., Id. at 179 [Proposition 1].

[14] The discussion in this introduction, like Kahan and Tuckman's formal analysis, considers only the direct effects of changes in litigation stakes. Thus, the direct effect of decreasing recovery on the plaintiff's litigation effort is analyzed. But the second-order "cross effect" on the defendant's

$\Uparrow$ Damages $\}$ more deterrence

Recovery $\Downarrow$ $\}$ less/*LESS* deterrence; fewer, *CHEAPER* suits

Figure 2: Adding the infra-marginal effect of decreasing recovery

In contrast to the effect of decreasing recovery, the effect of increasing damages changes significantly when infra-marginal effects are taken into account. Recall that in Polinsky and Che's framework, increasing damages costlessly restores the deterrence lost from decreasing recovery, as depicted in the top half of Figure 1. But when infra-marginal effects are accounted for, raising the stakes for defendants will cause them to devote more resources to their defense, and this will increase the cost of each filed suit. Thus, although it will still be true that increasing damages increases the deterrent force of infra-marginal suits, such suits will now also be more expensive. Moreover, defendants' response to increased stakes 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 defendants. The consequent reduction in filings will act to lower both deterrence and litigation costs. Figure 3 adds both this filing effect and the direct infra-marginal effect of increasing damages to Figure 2.

litigation effort of this change in the plaintiff's litigation effort is not considered—and so is implicitly assumed not to be decisive in the analysis. Our formal analysis does allow for cross-effects. For more on this see note 22 and Section III.D.

$\Uparrow$
Damages $\Big\}$ more / *LESS* deterrence; *fewer, COSTLIER suits*

Recovery $\Big\}$ less / less deterrence; fewer, cheaper suits
$\Downarrow$

Figure 3: Adding the infra-marginal effect of increasing damages

As Figure 3 indicates, if we account for the impact on both filings and infra-marginal suits, increasing damages has an ambiguous effect on both deterrence and social cost. This ambiguity, in turn, makes it impossible to conclude that the full maneuver advocated by Polinsky and Che—the reduction in recovery along with the increase in damages—is welfare improving.

In fact, it is easy to devise plausible examples in which decreasing recovery and increasing damages, so as to keep deterrence constant, actually increases social costs, thus decreasing overall social welfare. Suppose, for instance, that the effect on infra-marginal suits dominates the effect on the number of filings in the social welfare equation. In that case, Figure 3 effectively reduces to Figure 4. Raising damages increases both deterrence and litigation costs, as infra-marginal defendants have more to lose and react by lodging a more vigorous defense. Inversely, reducing recovery decreases both deterrence and litigation costs, as infra-marginal plaintiffs pursue their complaints with less intensity. Imagine, then, that the increase in litigation costs from raising damages (by enough to maintain deterrence) overwhelms the litigation cost savings from the initial reduction in recovery. If so, the system will be producing the same amount of deterrence at greater litigation cost, not less, and Polinsky and Che's maneuver will be welfare reducing, rather than welfare improving.

$\Uparrow$ Damages	more deterrence;	costlier suits
Recovery $\Downarrow$	less deterrence;	cheaper suits

Figure 4: When infra-marginal effects dominate

Thus, lowering recovery and raising damages need not be welfare improving when the full effect of litigation stakes on litigation effort is explicitly taken into account. Given indications in the data that the manner in which filed suits proceed is at least as important as how many suits are filed, this finding by itself is of some practical import. But the article goes beyond this cautionary "possibility result" to characterize the *kinds* of cases in which optimal recovery will be no less than optimal damages.[15] This characterization turns out to be particularly relevant to the current policy debate, because the argument *against* reducing recovery and raising damages is strongest in precisely the kind of case that seems to motivate litigation reform—including various forms of decoupling.[16] In particular, the argument for allowing plaintiffs to keep all of what defendants lose is most compelling in high-stakes suits (e.g., those involving product liability, toxic torts, or deceptive commercial practices) with deep-pockets defendants (e.g., large corporations), unpredictable juries, and contingent-fee plaintiff lawyers.

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

[15] One byproduct 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 3, supra. This finding can be added to the literature's list of reasons why the (unnuanced) interpretation of Becker's famous result (supra note 1)—that increasing fines and lowering the probability of detection is always welfare improving—does not hold in the general case. See, e.g., A. Mitchell Polinsky & Steven Shavell, The Optimal Tradeoff between the Probability and Magnitude of Fines, 69 Am. Econ. Rev. 880, 883 (1979).

[16] In Indiana, for example, punitive damages may be as high as three times compensatory damages, but plaintiffs receive only 25% of such damages, the rest going to a fund for violent crime victims. For a summary of various state laws that reduce the plaintiff's award, see Daughety and

damages are increased to restore the deterrence lost by decreasing recovery, and wherein infra-marginal effects dominate. The key to our analysis is the recognition that when litigation stakes are high, the positive impact on deterrence of increasing damages is much smaller than the negative impact on deterrence of reducing recovery. Increasing damages by one dollar increases the defendant's expected trial losses—and so deterrence—by the chance that the defendant will have to pay that additional dollar. If, for example, the chance that the defendant loses the suit is 50%, then each dollar increase in damages increases deterrence by fifty cents. Reducing recovery, on the other hand, reduces the defendant's expected trial losses via 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 fifty cents. On the other hand, if the stakes are high for the defendant, say $5,000,000, then a percentage point decrease in the chance of liability decreases the defendant's expected trial loss by $50,000. Thus, the reduction in deterrence from reducing recovery by one dollar 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 one dollar is essentially independent of the stakes of the case and will be a fraction of that dollar corresponding to the defendant's chance of being held liable. As a consequence, executing Polinsky and Che's deterrence-maintaining maneuver 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—due to the defendant's stepped-up defense—is likely to overwhelm the cost savings from reducing recovery—due to the plaintiff's

---

Reinganum, supra note __ at __; Kahan and Tuckman, supra note 4 at 175 n1. For a discussion of the law governing the imposition of punitive stepped-down prosecution. In the end, reducing recovery, and then raising damages to restore the lost deterrence, will increase, not decrease system costs.

In general terms, when the stakes are high for defendants, recovery is a more efficient provider of infra-marginal deterrence than is damages. To reduce recovery and increase damages so as to keep deterrence constant is to substitute the less-efficient producer of infra-marginal deterrence for the more-efficient producer. Where infra-marginal effects are important, this is unlikely to improve social welfare.

As we show in the formal analysis to follow, this basic point—that recovery tends to be a more efficient "deterrer" in high-stakes suits—is strikingly robust. The argument against raising damages and lowering recovery is even stronger in a setting in which plaintiffs' lawyers are paid on a contingent fee basis and in which most 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 II and III present the basic model and results, while Section III provides some illustrative examples. Important variations on the basic model are considered in Section V. A technical appendix houses the formal statements of all results along with their mathematical proofs.

II. 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 sequentially contingent phases: the primary activity, the plaintiff's filing decision, and the trial.[17] At trial, both parties

---

damages—which, of course, increases the defendant's sanction—see generally Ghiardi & Kircher, Punitive Damages Law and Practice (1996).

[17] We consider settlement in Section V.B.

choose how much costly evidence to produce. We describe the three phases of the model in reverse order with the aid of Figure 5

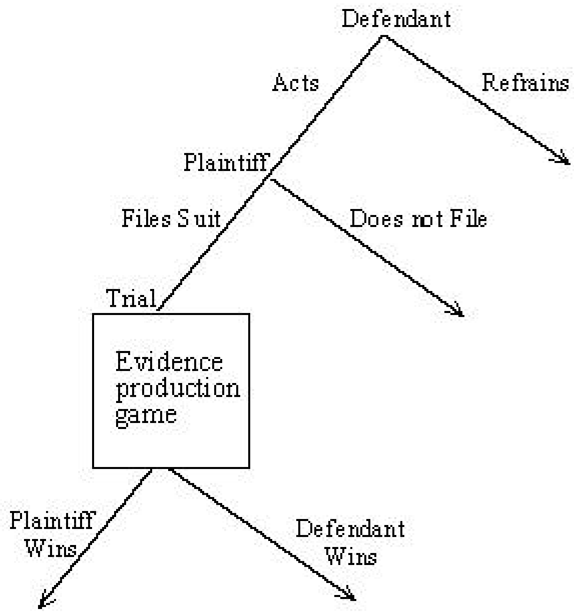

Figure 5: The phases of the model

A. *Trial*

Each trial matches a particular plaintiff with a particular defendant. The plaintiff produces the quantity $x \geq 0$ of evidence, the defendant, $y \geq 0$. The net weight of the evidence is $x - y$. The fact-finder observes the evidence with error. Its perception of the evidence is $x - y + \boldsymbol{e}$, where the error term $\boldsymbol{e}$ has cumulative distribution F with density f. Plaintiff wins if the fact-finder's perception of the evidence favors her case. Thus, the probability of plaintiff victory is:

$$p(x,y) \equiv \Pr\left(\underbrace{x-y}_{\substack{\text{pro-plaint.} \\ \text{net weight} \\ \text{of evidence}}} + \underbrace{e}_{\substack{\text{pro-plaint.} \\ \text{net error}}} \geq 0 \right) = \Pr(e \geq y-x) = 1-F(y-x) \,.$$

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 $k + c(x)$, 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 $c'' > 0$. Similarly, the defendant bears evidence production costs of $V(y)$, where $V' > 0$ and $V'' > 0$. (Any fixed-cost component in the defendant's evidence cost function is not separately notated.) According to our assumptions on evidence cost derivatives, the cost to each party of tilting the evidentiary balance one tick in that party's 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.

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 $y \geq 0$ to minimize expected litigation losses $p(x,y)D + V(y)$. In equilibrium, the following first and second order conditions must be satisfied:

$$p_x R - c' = 0 \text{ and } p_y D + V' = 0$$

$$p_{xx} R - c'' \leq 0 \text{ and } p_{yy} D + V'' \geq 0 \,.$$

B. *Plaintiff's Filing Decision*

Each plaintiff decides whether to file suit based on 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. The plaintiff may be required to pay a *filing fee K*, a third social policy variable in addition to R and D. Thus, the plaintiff files suit if and only if

$$pR - c - k - K \geq 0 \text{ or } k \leq pR - c - K$$

where p and c are determined by the expected equilibrium at trial, which is in turn determined by R and D. Thus, writing $\hat{k}(D, R, K) \equiv pR - c - K$ for the *marginal filer*, the plaintiff files suit whenever $k \leq \hat{k}$.

C. *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 $b \geq 0$, where b is distributed among the population of defendants according to the cumulative distribution J with density j. If the defendant acts, harm of $h > 0$ is inflicted on one plaintiff drawn at random from the plaintiff population. In particular, the fixed evidence cost component k of the plaintiff is drawn from the cumulative distribution G and density g. This 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 potential plaintiffs' costs, and predicting the expected equilibrium in evidence production at trial. Thus, the defendant chooses to act, if and only if the private benefits from acting exceed the expected loss from litigation, including evidence costs:

$$b \geq \underbrace{G\left(\hat{k}\right)}_{\substack{\text{chance of} \\ \text{being sued,} \\ \text{if act}}} \underbrace{\left(pD+V\right)}_{\substack{\text{expected trial} \\ \text{loss, if sued}}} .$$

It will be convenient to define *per suit deterrence* $\Delta \equiv pD+V$ and *all-in deterrence* $\Omega \equiv G\left(\hat{k}\right)\Delta$. In this notation, the defendant acts if and only if $b \geq \Omega$. Note that per suit deterrence depends on *D* and *R* whereas all-in deterrence depends on *D, R,* and *K.*[18]

D. *Social Welfare Problem*

The social cost arising from the primary activity is

$$\underbrace{\int_{b=\Omega}^{\infty} \left(h-b\right) j\left(b\right) db}_{\text{integration over acting defendants}} . \tag{1}$$

The expected social cost of litigation is

$$\underbrace{\left(1-J(\Omega)\right)}_{\substack{\text{acting} \\ \text{defendants}}} \underbrace{\int_{k=0}^{\hat{k}} \underbrace{\left(k+c+V\right)}_{\text{cost of each suit}} g\left(k\right) dk}_{\text{integration over filed suits}} . \tag{2}$$

The socially optimal configuration of *R*, *D*, and *K* is that which minimizes *all-in social cost*, the sum of (1) and (2).

[18] Some notes on the generality of this structure: First, the model applies to any primary activity choice that generates externalities. For example, in a torts setting, we can think of "acting" as "acting negligently," and "refraining" as acting with due care. Or we could think of "acting," rather than "refraining," as engaging in a given activity at a high level, rather than a low level. Secondly, 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 more likely to be sued when they act than when they do not. Thirdly, our assumptio n that recovery and damages are scalars follows Polinsky and Che's model, and is, again, merely simplifying. Were the level of damages and recovery a function of the evidence, our basic results would still pertain. Fourthly, our modeling task is also greatly simplified by having plaintiffs differ by only the fixed component of evidence costs. As a result of this assumption, all plaintiffs will face the same evidence choice problem because their fixed cost differences do not affect their marginal evidence costs. Thus, all filed cases will be the same in terms of evidence production and the probability of plaintiff victory. Furthermore, even with significant cross-effects, the set of filing plaintiffs will have a threshold structure: all plaintiffs with cost parameters below some level will file, all with cost parameters above this level will not. Using an alternative specification, even one as simple as $kc(x)$, significantly complicates the analysis. For example, with significant cross-effects, it is not necessarily true that the set of filing plaintiffs has a threshold structure when plaintiffs'

III. ANALYSIS OF THE BASIC MODEL

A. *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 upfront—like a ticket price for playing the "litigation lottery"—or on the backend, as a tax on recovery.[19] While upfront 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, upfront fees will not significantly dampen plaintiffs' fervor in prosecuting the suits they choose to file.[20] On the other hand, reducing backend 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 upfront fees, like K, also are a policy variable, the job of controlling the number of suits should be fully delegated to such fees, while damages and recovery (i.e., the plaintiff's trial outcome-dependent winnings) should be fully determined by their effect on the cost and deterrent force of infra-marginal suits. To see why, imagine setting recovery and damages in a way that did not produce the resulting per suit deterrence at lowest possible per suit cost. We might be concerned that altering recovery and damages to provide per suit deterrence at lower per suit cost would alter the number of filed suits in such a way that the net effect on social welfare was negative. But when upfront fees are a policy instrument, this concern is not justified. As we adjust recovery and damages to make each filed suit a more efficient provider of deterrent force, we can simultaneously adjust plaintiffs' upfront fee to cancel out any negative effect that

---

costs have the form $kc(x)$. In other words, it is possible with this form of costs that a higher k plaintiff will file, while a lower k will not. This irregularity would, nevertheless, be eliminated were cross-effects not dominant.

[19] 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 do pay "filing fees," but these are usually negligible.

changes in recovery and damages might have on the number of suits filed. This idea is captured formally in the following result:

> PROPOSITION 1: *At the socially optimal levels of recovery, damages and the 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 on the problem of providing per suit deterrence at lowest per suit cost. However, as shown in Section V.C, our analysis accommodates not only the case where filing fees can be suitably adjusted, as in Proposition 1, but also the case in which filing fees are not suitably adjustable and infra-marginal effects dominate.

B. *The marginal deterrence cost of recovery and damages*

Crucial 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 incur in order to produce one more unit of per suit deterrence by changing R? We can answer this question by differentiating both per suit costs $c+V$ and per suit deterrence $pD+V$ by R and then considering the ratio of the former derivative over the latter. A marginal increase in R will increase per suit cost by $c'x_R + V'y_R$,[21] and per suit deterrence by $p_x D x_R + \left(p_y D + V'\right) y_R$. Using the defendant's first order condition ($p_y D + V' = 0$) to simplify the derivative of per suit deterrence, and taking the ratio of the two derivatives yields

$$MDC_R \equiv \frac{c'x_R + V'y_R}{p_x D x_R}.$$

---

[20] Note, however, that wealth effects from upfront fees—via wealth constraints or changes in the marginal utility of "income"—may have an impact on the plaintiff's litigation effort.

Similarly, increasing D will increase per suit cost by $c'x_D + V'y_D$, and per suit deterrence by $p_x D x_D + \left(p_y D + V'\right) y_D + p$. Notice the additional term p in the derivate of per suit deterrence. This term reflects the direct effect of increasing D on the defendant's expected losses. Again using the defendant's first order condition ($p_y D + V' = 0$) to simplify the derivative of per suit deterrence, we obtain the ratio

$$MDC_D \equiv \frac{c'x_D + V'y_D}{p + p_y D x_D}.$$

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* saves \$2 in litigation costs per suit and the lost unit of deterrence can be made up by increasing *R* at a cost of \$1, with the result that per suit deterrence remains constant while per suit costs fall by one dollar. (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 2: *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

---

[21] The notation "x_R," for example, denotes the derivative of the plaintiff's choice of evidence production in recovery.

that $R \leq D$. Given Proposition 2, this will imply that the social optimum over this region of the parameter space cannot entail $R \leq D$.

C. *Upper bound on the marginal deterrence cost of recovery*

The next proposition shows that the marginal deterrence cost of recovery is never more than $1 above the value of the ratio $\frac{R}{D}$. This holds for all error distributions *f* and without regard to the size of the stakes, R and *D*. The result implies that the marginal deterrence cost of recovery is never more than $2 when $R \leq D \Leftrightarrow \frac{R}{D} \leq 1$.

The intuition for this result can be gleaned by assuming that *cross-effects* in the parties' evidence choices are negligible (i.e., $y_R, x_D \approx 0$). (The result itself does not depend on this assumption.) In this case, the marginal deterrence cost of recovery $\frac{c' x_R + V' y_R}{p_x D x_R}$ reduces to

$$MDC_R = \frac{c' x_R}{p_x D x_R} = \frac{p_x R x_R}{p_x D x_R} = \frac{R}{D},$$

where we have substituted from the plaintiff's first order condition $p_x R - c' = 0$. Therefore, when recovery is no less than damages, each additional dollar of per suit deterrence generated by increasing recovery increases per suit costs by no more than one dollar.

Why is this so? When the plaintiff is producing her privately optimal amount of evidence, increasing *R* has the same marginal effect on the plaintiff's costs as on her expected winnings ($p_x R = c'$)—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 of the plaintiff's additional evidence production by the plaintiff is less than the marginal effect on the defendant's expected loss ($p_x D \geq p_x R$). By transitivity, therefore, increasing recovery increases the plaintiff's evidence costs by less than it increases the defendant's expected loss. The increase in the defendant's expected loss is, in turn, the increase in per suit deterrence.

PROPOSITION 3: *The marginal deterrence cost of recovery is never more than* $\frac{R}{D}+1$.

D. *Damages' marginal deterrence cost in high-stakes, deep-pockets suits with error-prone juries*

Definitive statements about the marginal deterrence cost of damages require additional assumptions. In this section we show that if the fact-finder is sufficiently error prone, defendants have sufficiently deep pockets, and damages are sufficiently large, then the marginal deterrence cost of damages will also be large.

To gain intuition, consider again the case in which cross-effects are negligible ($x_D, y_R \approx 0$). In this case, the marginal deterrence costs of damages $\frac{c' x_D + V' y_D}{p + p_y D x_D}$ reduces to

$$MDC_D = \frac{V' y_D}{p}.$$

Given that $p \leq 1$, the marginal deterrence cost for damages will be greater than $V' y_D$. As *D* grows ever larger, the defendant presents more and more evidence *y*. The marginal cost of evidence for the defendant, V', grows accordingly as each additional unit of evidentiary weight from defendant—inspired by his increased stakes—costs more than the last. Therefore, assuming that the defendant's responsiveness to greater stakes y_D does not decay too quickly as *D* increases (it may, in fact, grow), $V' y_D$ will grow without bound.

Absent the hypothesis that cross-effects are small, definitive results on the size of damages' marginal deterrence cost are elusive. Indeed, when one parties' evidence production has a strong impact on the other's optimal evidence choice (as opposed to his expected trial payoffs), almost anything can happen—a common phenomenon in the analysis of Nash equilibrium comparative statics.[22]

Cross-effects are relatively insignificant when the fact-finder is prone to error in interpreting the evidence. One party's evidence production affects the other's optimal choice via changes in the *marginal* probability of plaintiff victory (as opposed to the probability of victory itself). For example, more evidence from the plaintiff would inspire the defendant to present less evidence, if the plaintiff's additional evidence dampened the marginal (negative) impact of the defendant's evidence on the probability of plaintiff victory. Thus, cross-effects operate through the cross-derivative of the probability of plaintiff victory

$$p_{xy} = \frac{\partial^2 \left(1 - F\left(y - x\right)\right)}{\partial x \partial y},$$

which, in turn, equals the derivative f' of the density of fact-finder error e. When the fact-finder tends to correctly perceive the evidence, the density *f* of error e will peak relatively sharply at zero, and its slope f' will be relatively steep on either side of zero. On the other hand, when the fact-finder often misperceives the true weight of the error, the error density will be relatively flat and its slope will remain relatively close to zero. When the density's slope remains close to zero, so do cross-effects: a change in one party's evidence production has little effect on the slope of the error density, and so little effect on the marginal benefits of additional evidence production for the other side.

PROPOSITION 4: *The marginal deterrence cost of damages can be made arbitrarily large by ensuring that damages and the defendant's wealth are sufficiently large and the fact-finder's perception of evidentiary weight is sufficiently error-prone.*

[22] Kahan and Tuckman assume that cross-effects are zero. Polinsky and Che do not encounter the issue of cross-effects in evidence production.

E. *Optimal recovery no less than optimal damages in high-stakes, deep-pockets suits with error-prone juries*

Proposition 4 implies that the marginal deterrence cost of damages will strictly exceed 2 if the defendant has deep pockets, the stakes are high, and the fact finder is error-prone. Proposition 3 implies that the marginal deterrence cost of recovery is never more than 2 when $R \leq D$. And Proposition 2 tells us that we cannot be a social optimum if the marginal deterrence cost of damages is strictly greater than the marginal deterrence cost of recovery. Combining these findings yields our main characterization result.

> PROPOSITION 5: *Suppose that the defendant's wealth is sufficiently large and the fact-finder's perception of evidentiary weight is sufficiently error prone. Then, if optimal damages are sufficiently large, optimal recovery is no less than optimal damages.*[23]

Finally, we provide conditions on primitives under which optimal recovery would be no less than optimal damages. In particular, we show that when harm h is large (as when defendant is designing a product for mass distribution, or is handling large quantities of toxic waste), optimal recovery will not be less than optimal damages.

---

[23] Proposition 5 and Proposition 6 below should be carefully contrasted with a similar possibility that arises in Polinsky and Che's model. Even though lowering recovery while raising damages is always social welfare improving in Polinsky and Che's framework, this does not imply that optimal damages always exceed optimal recovery in their model. Polinsky and Che, supra note 2 at 563 ("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 may exceed the optimal payment by the defendant.")

However, in Polinsky and Che'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. Id. at 563 ("In the optimal system of decoupled liability the defendant's payment is as high as possible."), 566 ("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 raising [recovery] so that he will be sued with a higher probability if an accident occurs. Therefore, as [harm] tends to infinity, [optimal recovery] must also tend to infinity, showing that for [harm] sufficiently large, [optimal recovery is strictly greater than optimal damages]." [emphasis added])

The fact that the defendant's wealth constraint always binds at an optimum in Polinsky and Che's framework is, in turn, a result of the fact that Polinsky and Che do not account for the per suit cost impact of increasing damages. In their model, increasing damages is always a cost-free means of increasing deterrence that can be profitably substituted for the costly production of deterrence via recovery. In our model, in contrast, optimal recovery may be more than damages even when damages do not equal all of the defendant's wealth. Indeed, optimal damages typically will not equal all of the defendant's wealth.

This distinction is important for legal policy, because damages rarely equal all of the defendant's wealth in practice. Rather, the policy debate takes place in a range where damages could be feasibly increased, and the debate concerns whether various incremental changes would be welfare improving. In this range, our model's prescriptions are very different from those of the Polinsky and Che model.

PROPOSITION 6: *Suppose that the defendant's wealth is sufficiently large and the fact-finder's perception of evidentiary weight is sufficiently error-prone. Suppose also that the harm from defendant's primary activity is sufficiently large. Then optimal recovery is no less than optimal damages.*

IV. NUMERICAL EXAMPLES

This section serves the dual purpose of numerically illustrating the operation of our model and also establishing that the characterization results recited above obtain over an interesting and policy-relevant region of the model's parameter space.

A. *Functional form assumptions*

We impose several functional form assumptions on the model presented above so that we can numerically calculate optimal damages and recovery. First, we assume that fact-finder error e is distributed according to the truncated normal distribution with mean zero, variance parameter s^2, and *range* $[\underline{e}, \overline{e}]$. Thus,

$$f(e) = \frac{e^{-\frac{e^2}{s^2}}}{\int_{-\overline{e}}^{\overline{e}} e^{-\frac{\tilde{e}^2}{s^2}} d\tilde{e}}.$$

Next, we assume that the parties' evidence cost functions are $c(x) = a x^2$ and $V(y) = b y^2$, where the scalars a and b allow us to separately vary the parties' marginal costs of evidence production. Note that the defendant's evidence costs satisfy the assumption imposed in the last section.

As is justified by Proposition 1, we examine the problem of providing a given level of per suit deterrence at the least per suit cost. However, unlike the analysis so far, we impose an additional budget constraint on transfers. Recovery paid out from the court to a winning plaintiff may not exceed damages paid in by the losing defendant: $R \le D$. (Section V.D discusses budget constraints in more detail.)

B. *Central Case*

Figure 1 represents our central case wherein evidence costs are symmetric ($a = b = 1/2$) and variance is relatively high. The horizontal axes in all the panels measure the required level of deterrence (Δ). The upper left panel presents the graphs of optimal damages and recovery. We can clearly see that when the level of deterrence is sufficiently high, optimal recovery is no less than optimal damages. The upper right panel shows the respective amounts of evidence presented by both sides. Due to the symmetry of evidence cost, when damages equals recovery, as it does for large required levels of deterrence, both sides present the same amount of evidence to the court. This implies that the probability of plaintiff's winning will be at 50%, as shown in the lower left panel. Lastly, the lower right panel shows the defendant's expected loss and the plaintiff's expected gain. The defendant's expected loss is a 45 degree line, signifying that the required level of deterrence (the x-axis) is being precisely provided. The plaintiff's expected trial gain at first increases more slowly than the defendant's expected trial loss, and eventually decreases, though at a slower rate than the defendant's expected trial losses increase. The plaintiff's trial payoffs begin to decrease after the point at which, simultaneously, recovery and damages become the same, the parties' levels of evidence production become the same, and the probability of plaintiff victory fixes on 50%. At this point, the detriment of the plaintiff's additional litigation expenses begins to dominate the benefit of increased recovery.

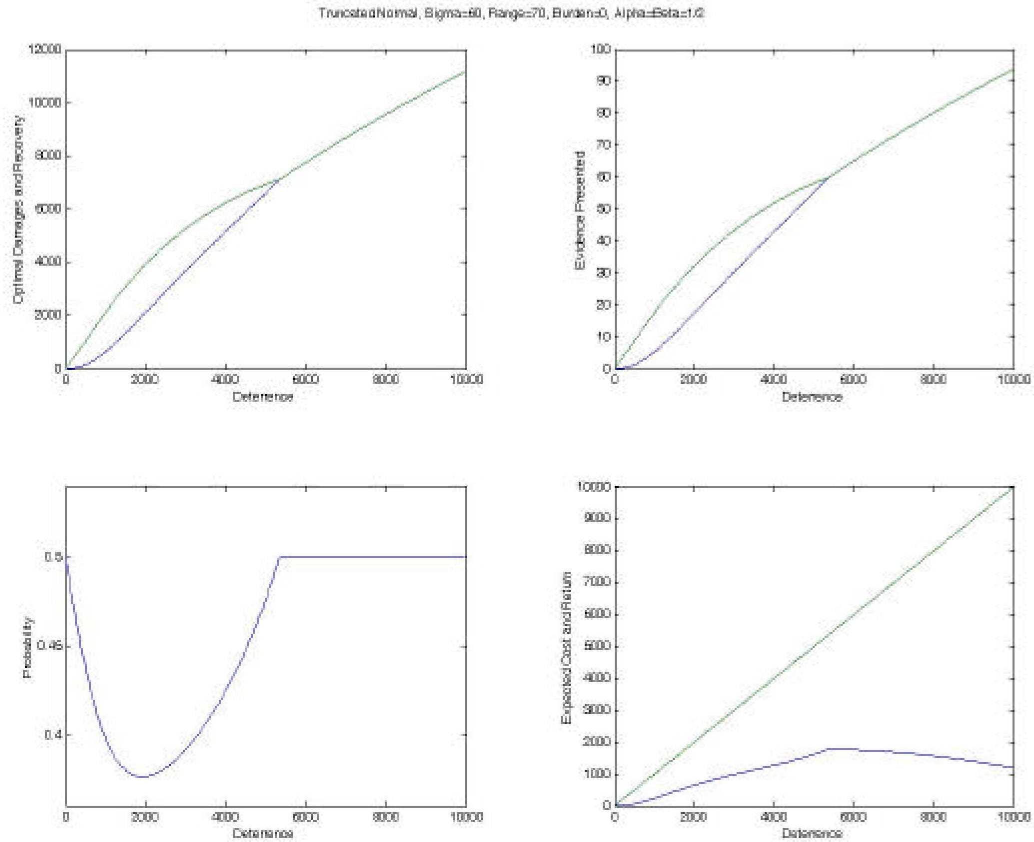

FIGURE 1: $s = 60,\ a = b = 1/2,\ \bar{e} = -\underline{e} = 70$

C. *Changes in Parameter Values*

The general results of our prototype case are not sensitive to small changes in parameters. Figure 2 shows the result of increasing the variance in fact-finder error. We see that optimal recovery is still no less than optimal damages when large amounts of deterrence are required.

Comparing Figures 1 and 2, the first point at which optimal recovery is equal to optimal damages is farther along the x-axis with a higher variance.[24] The reason relates to the symmetry in

[24] At first glance, this may seem to contradict our formal results in which a large variance (more precisely, a small error density derivative) plays a role in defining the cases where optimal recovery is no less than optimal recovery. In fact, it is not inconsistent for two reasons. First, our result on the effect of variance was not a monotonicity result, but a limiting result. Second, our identification of one class of cases wherein optimal recovery is less than optimal damages does not preclude the existence of cases outside this class that yield the same result.

the parties' evidence costs and the burden of proof. With similar marginal costs of evidence and neutral burden of proof, when damages is roughly equal to recovery, the litigants produce about equal amounts of evidence. This will keep the probability of the defendant's liability at 50% and the marginal probability at $f(0)$. Then, as the truncated normal distribution takes on a larger variance, the mode of the distribution decreases, and looking at the marginal deterrence cost ratio for damages of $\frac{V' y_D}{p} = \frac{f D y_D}{p}$, fall in *f* signifies that damages becomes more deterrence efficient.[25]

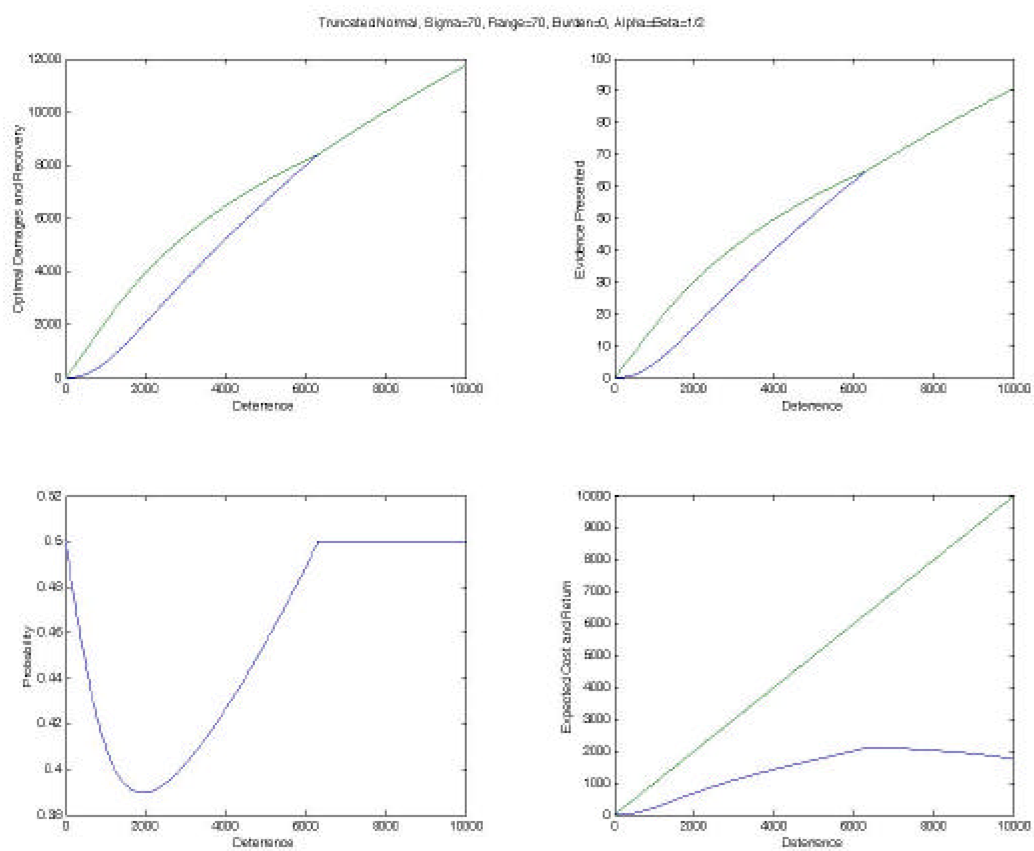

FIGURE 2: $s = 70$, $a = b = 1/2$, $\bar{e} = -\underline{e} = 70$

[25] If evidence costs were sufficiently asymmetric and/or the burden is sufficiently one-sided, the relevant portion of the error distribution might be *increasing* in variance at these same recovery-damages pairs. In this case, increasing variance would increase damages' marginal deterrence cost and cause optimal recovery to approach optimal damages more quickly. More generally, the comparison of Figures 1 and 2 illustrates the fact that very little can be said definitively about the relative sizes of optimal recovery and optimal damages in the middling case, and this is in turn what motivates our focus in this paper on limiting arguments.

Figure 3 represents the case in which the defendant's marginal evidence costs are twice as high as the plaintiff's. The upper right graph verifies that as the size of the case becomes bigger—i.e., as requisite level of deterrence gets larger—the defendant increases evidence production at a slower pace than the plaintiff. After a certain point, the defendant produces much less evidence than the plaintiff. Because of the relative bias in evidence production, the probability of the plaintiff's prevailing (lower left graph) and the plaintiff's expected return (lower right graph) increase to a much higher level as well.

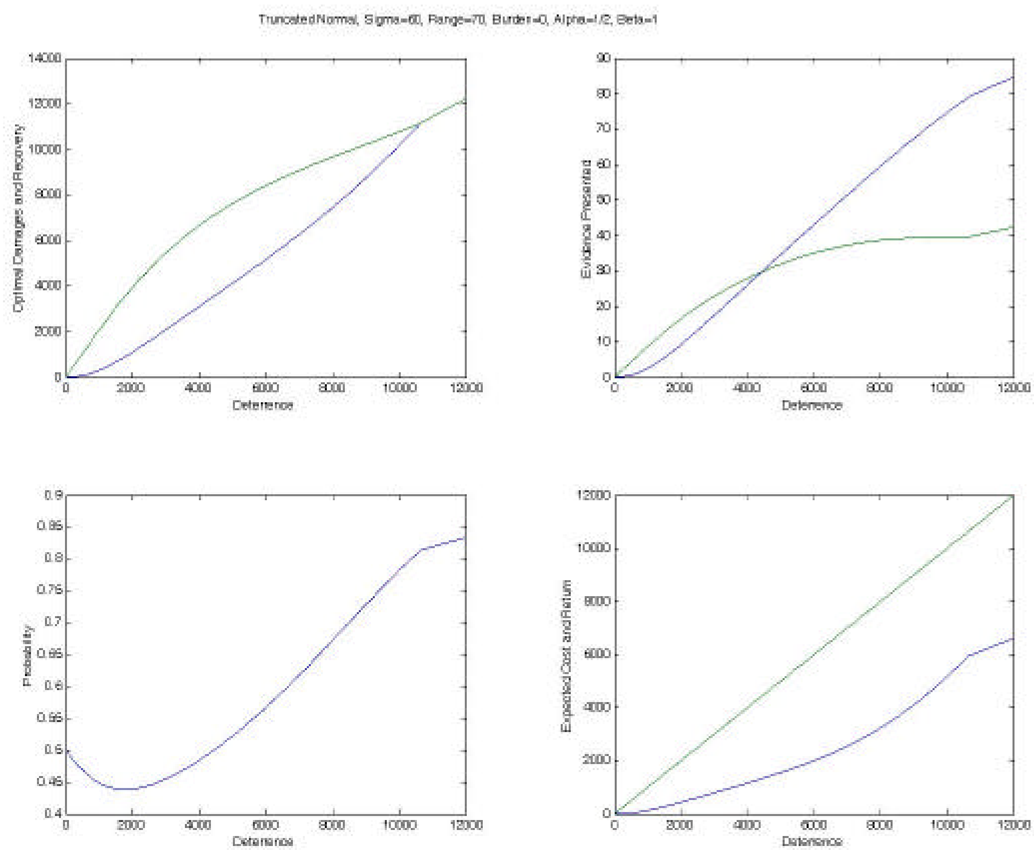

FIGURE 3: $s = 60$, $a=1/2$, $b=1$, $\bar{e}=-\underline{e}=70$

Interestingly, increasing the defendant's marginal cost of evidence has made decoupling optimal over a larger range, as is clear from the upper left panel. To understand this, consider again the marginal deterrence cost of damages when cross-effects are small: $\frac{V' y_D}{p}$. When the defendant faces a steeper marginal cost curve, (V' is larger), an increase in damages will increase the amount of evidence presented by the defendant at a much slower pace than before (i.e., y_D becomes smaller). Of course, at the same time V' is larger. While the model leaves it uncertain which effect will dominate (the increase in V' versus the decrease in y_D), in this simulation case, the former effect seems to be stronger. The reason is that with quadratic cost, marginal cost increases with respect to more evidence, but the concavity of the cost (which determines the responsiveness, y_D) is fixed at a constant.

V. VARIATIONS IN THE BASIC MODEL

A. *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. Incorporating these features into our model only reinforces our basic findings.

Suppose the plaintiff's lawyer receives $q \leq 1$ 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 rather than that of his client. Thus, the plaintiff's evidence production is determined by the plaintiffs' lawyer's first-order condition, $q\, p_x R - c' = 0$.

Consider again the case of negligible cross-effects, wherein the marginal deterrence costs of recovery and damages are well approximated by $\frac{V' y_D}{p}$ and $\frac{c' x_R}{p_1 x_R D}$, respectively. Substituting the plaintiff's *lawyer's* first-order condition into the marginal deterrence cost of recovery yields $q \frac{R}{D}$. Given that $q \leq 1$, this is smaller than $\frac{R}{D}$, which, as we have seen, is the marginal deterrence cost of recovery in the absence of cross-effects. Thus, contingent fee arrangements for plaintiff's lawyers make recovery more deterrence efficient. At the privately optimal level of evidence production for the plaintiff, the change in evidence costs for a given increase in evidence is now only a fraction of the change in expected recovery ($q\, p_x R = c'$)—since the lawyer, who chooses the plaintiff's evidence, pays all the costs and receives only a fraction of the recovery. In addition, when $R \leq D$, the increase in the plaintiff's expected recovery for a given change in the plaintiff's evidence production is just a fraction of the increase in the defendant's expected loss ($p_x R \leq p_x D$). Therefore, by transitivity, the increase in the plaintiff's evidence costs is now but a fraction of a fraction of the increase in the defendant's expected trial loss, which is per suit deterrence. Accordingly, recovery is even more deterrence efficient. In contrast, damages' marginal deterrence cost is not (directly) affected by the nature of the plaintiff's arrangement with her lawyer.

COROLLARY 1: *Suppose the plaintiff's lawyer receives a è fraction of the return from litigation while bearing the entire cost of evidence production, where* $0 < q \leq 1$. *Suppose, also, that the plaintiff's lawyer makes the plaintiff's evidence production decisions to maximize her own return. Then, Propositions 5 and 6 still hold.*

B. *Settlement*

In this section, we incorporate into our model the conventional treatment of settlement, which focuses on the existence and position of the settlement range, showing that this modification does not affect our basic result.[26]

Recall that our argument that optimal recovery is no less than optimal damages is based on a comparison of marginal deterrence costs when recovery is less than damages. Conveniently, 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: $\left(pD+V\right)-\left(pR-c\right)$ $=p\left(D-R\right)+V+c>0$.[27] In other words, there is always a positive settlement range. 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 k of all cases with positive settlement ranges actually do settle.[28] If the case settles, each party gets a fixed fraction of the settlement surplus.[29]

---

[26] Polinsky and Che, supra note 2 at 566-568, 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 imposes costs. Polinsky and Che also study the case in which the court can observe the settlement amount and make additional awards or impose additional charges based on the settlement amount. Kahan and Tuckman, supra note 4 at 178-179, 180, work with a model of settlement that is similar to ours, but in which the parties may have disparate (though commonly known) beliefs about what will happen at trial. See Daughety and Reinganum, supra note 12, for a mostly positive analysis of decoupling in the case of asymmetric information settlement bargaining. (Daughety and Reinganum also study the revenue-maximizing tax on plaintiffs' recovery.)

[27] This result is partly due to the fact that the parties agree on the probability of plaintiff victory. However, even when the parties so agree, there is no settlement range when recovery is significantly greater than damages. See, Polinsky and Che, supra note 2 at 567 n 15.

[28] Among suits that have 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 \$1000 is any more likely to settle than a suit whose settlement range is from \$100 to \$150. On the one hand, the gains from settlement are larger for both parties in the suit with the larger range, midrange settlements being far in excess of the best each could do in the absence of a deal. This acts to make settlement more likely the larger the range. Yet on the other hand, the benefits of bargaining more aggressively are greater where there is more territory within the settlement range to be captured. And this acts to make settlement less likely the larger the range. Whether and to what extent either effect dominates is an empirical question that remains yet to be answered in the literature. In this paper, we treat the phenomenon of non-settlement in the presence of a settlement range as random and unrelated to the width of the settlement range. Thus, we assume that k is exogenous.

In any event, when we account for infra-marginal suit effects, increasing recovery and decreasing damages does not necessarily decrease the size of the settlement range. Kahan and Tuckman, supra note 4, notice a similar ambiguity at 180 [Proposition 3 and discussion].

For example, if cross-effects are zero, then decreasing damages by one unit and increasing recovery by one unit increases the defendant's expected trial loss by $p_1 Dx_R - p$. When damages are large, therefore, the effect of increasing recovery will swamp the effect of decreasing damages and the defendant will actually be worse off on net. For the plaintiff, decreasing damages by one unit and increasing recovery by one unit increase plaintiffs

Thus, letting $g \in [0,1]$ be the plaintiff's fraction, the settlement amount is a weighted average of the parties' trial payoffs: $S \equiv (pR - c) + g\left((pD+V) - (pR-c)\right) = g\,(pD+V) + (1-g)(pR-c)$. The defendant's ex ante expected loss from a filed suit—i.e., per suit deterrence—is also a weighted average of the parties' trial payoffs: $\Delta \equiv (1-k)(pD+V) + k\,S = (1-h)(pD+V) + h\,(pR-c)$, where $h \equiv k(1-g)$. The problem of providing per suit deterrence Δ at minimal per suit cost then becomes[30] $\min_{\mathrm{R,D}} (1-k)(V+c) : (1-h)(pD+V) + h\,(pR-c) = \Delta$, or equivalently, $\min_{\mathrm{R,D}} c + V : (1-h)(pD+V) + h(pR-c) = \Delta$.

At optimal D and R, the settlement analogy to Proposition 2 must hold. That is, the marginal deterrence cost of recovery

$$MDC_R = \frac{V' y_R + c' x_R}{(1-h)\,p_x x_R D + h\,(p + p_y y_R R)}$$

must equal the marginal deterrence cost for damages

$$MDC_D = \frac{V' y_D + c' x_D}{(1-h)(p + p_x x_D D) + h\,p_y y_D R}.$$

The numerators of these ratios 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 fraction of filed suits makes both tools more efficient in precisely the same proportion.

---

expected trial payoffs by $p - p_x R y_D$. Therefore, for high-stakes cases, both parties' threat points improve and the issue becomes whether the most 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, then the defendant's expected trial losses will go up faster than the plaintiff's expected trial gains and the settlement range will increase on net.

[29] Our results go through as long as the settlement amount is increasing in both the plaintiffs expected trial payoffs and the defendant's expected trial losses.

[30] Given that ê is strictly bounded away from zero, we can ignore the multiplier $(1-k)$ in the objective function. We will carry it through the analysis nonetheless so as not to create confusion.

Therefore, the difference in these marginal deterrence cost ratios, as compared to the case without settlement, is wholly located in the ratios' denominators: i.e., 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, examine the marginal deterrence cost ratios when cross-effects are negligible:

$$\frac{c' x_R}{(1-h)\,p_x x_R D + h\,p} \text{ versus } \frac{V' y_D}{(1-h)\,p + h\,\underbrace{p_y}_{-}\,\underbrace{y_D}_{+}\,R}.$$

Comparing the denominators of these two ratios, we see that each is a weighted average of two expressions with the same weights. The expressions weighted by $(1-h)$ in each denominator are the same as for the no settlement case. The new expressions, those weighted by h, are easily and unambiguously compared. For recovery, on the left, we are averaging in a positive number p. For damages, on the right, we are averaging in a negative number $p_y y_D R$. 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 $-p_y y_D R > 0$) the plaintiff's expected trial payoffs, and thus her threat point in bargaining. The plaintiff's expected trial payoffs decline, because greater damages 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, even as scale increases. 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 $p_x x_R D$).

> PROPOSITION 7: *Propositions 4 and 5 still hold when the possibility of settlement is incorporated into the model.*

C. *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-pays-all" rule.[31] An impressive literature catalogues the plusses and minuses of each approach. But for the purposes of our main point, whether the cost allocation rule is American or British makes no difference. Recovery is more deterrence efficient than damages under both rules when the stakes are high and the fact-finder is error prone.

Indeed, when cross-effects are small, 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 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 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 plaintiffs' 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 payout.

[31] This rule is also used in the United States in certain limited circumstances. For example, under 42 U.S.C. §1983 and §1988, losing defendants must pay plaintiffs' attorney's fees in certain civil rights actions (but not vice versa). Costs other than for the attorney are also routinely shifted under rules such as Fed. R. Civ. Pro. 54(d)(1).

On the other hand, when cross-effects are small, the expression for 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. But the defendant has already accounted for this in setting his level of evidence production under the British rule, and, as under the American rule, the impact of the defendant's own adjustment in evidence production has no marginal effect on his expected trial losses.

> PROPOSITION 8: *Propositions 4 and 5 still hold, if litigation costs are allocated according to the British rule.*

D. *Budget Balance in Transfers*

Several notions of budget balance are possible. Most simply, we might require that, in the case the plaintiff wins, transfers to the court in the form of damages do not exceed transfers from the court in the form of recovery: $R \le D$. Alternatively, we might allow the court to use any monies collected as filing fees in the current case to offset any excess of recovery over damages: $K \ge (R - D)$. Or, given that filing fees are certainly paid, while damages and recovery are probabilistically imposed, we might require only that filing fees cover the *expected* shortfall in backend transfers: $K \ge p(R - D)$. This would, in turn, be equivalent to system wide budget balance: $G(\hat{k})K \ge G(\hat{k})p(R - D)$.[32]

In analyzing the effect of adopting any of these concepts of budget balance, we begin with the proposition that the filing fee will not be strictly negative in practice. In reality, plaintiffs always have the option of filing a hastily worded complaint, thus qualifying for the filing subsidy, and then failing to pursue the case. Therefore, were K strictly negative, virtually all plaintiffs would file suit, most merely to obtain the subsidy, and it is difficult to believe that this could be socially optimal.[33]

Given that the filing fee is nonnegative, the only potential problem for our analysis arises when the filing fee is precisely zero. At that point, squeezing recovery and damages back together would not be feasible were it the case that the sterilizing adjustment in the filing fee (designed to hold the number of filed suits constant) required decreasing that fee. In fact, however, when the fact-finder is sufficiently prone to error, the proper compensating adjustment to the filing fee will always be to *increase* it. In particular, when cross-effects are relatively small, increasing recovery and decreasing damages will always improve plaintiffs' prospective trial payoffs: the plaintiff gets more when he wins and is more likely to win for the fact that the defendant is defending less vigorously. Therefore, keeping constant the number of suits as we bring recovery and damages together will require raising the fee to cancel out the growing attractiveness of litigation for plaintiffs.

> COROLLARY 2: *Suppose that the social planner minimizes all-in social costs subject to any one of the budget constraints mentioned above. Suppose also that the optimal filing fee is nonnegative. Then Propositions 4 and 5 still hold.*

E. *When filing fees are negligible and not adjustable, but infra-marginal effects dominate*

Our analysis has focused on the per suit effects of changes in recovery and damages. In Section III.A, we justified this focus by assuming that the policy maker can impose fees on the plaintiff—

---

[32] Without loss of generality, our budget constraint and social welfare objective might also include the fixed public costs of the system. Another alternative would be to add to the social welfare function the social cost of raising $G(\hat{k})\,K - G(\hat{k})\,p(\,R - D)$ in public funds through tax receipts. Given Kuhn-Tucker techniques, this would differ from imposing a budget constraint only when the budget constrain was non-binding.

we have called them "filing fees"—that do not depend on the outcome of the suit.[34] Another reason to focus on per suit effects would be that they are relatively important empirically. In the current section we investigate the theoretical implications 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. Lastly, we argue that this region of the parameter space is rhetorically significant in the current policy debate.

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. In addition to per suit effects, increasing recovery will increase the number of suits filed—at least when cross-effects are not significant.[35] 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 $\hat{k}$. The impact on social costs for any given change 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, $g(\hat{k})$. When this density is small, the impact on all-in social costs of a given change in the marginal filer will be small. At the same time, per suit effects are independent of the density g of plaintiff's fixed evidence costs.

---

[33] This result could be established formally within our model, if we added public clerking costs and made the realistic assumption regarding plaintiff costs to ensure that only a small fraction of the population would ever find suit worthwhile in the absence of a filing subsidy if they planned to present zero evidence.

[34] This was discussed in Section III.A. Recall that the filing fee is just a stand-in for charges that are not dependent on the outcome of the suit.

[35] The derivatives of the marginal filer in R and D are $p + p_y y_R R$ and p, respectively. Apropos of the discussion of complications arising from cross-effects in the text surrounding note 22, increasing recovery may actually 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.

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.

It is worth noting that, in a somewhat more general version of our model, this condition on plaintiffs' costs corresponds to one of the complaints about litigation that has inspired reforms such as decoupling. Specifically, in the model wherein defendants might be sued even if they are not at fault in the primary activity, there will be two densities for plaintiffs' costs—one for when the defendant acts, and one for when he refrains. (The difference between these two densities will be the source of the defendant's primary activity incentive: positive incentives require that the defendant is less likely to be sued (i.e., plaintiffs' costs tend to be higher), if he is not at fault.) In this more general model, the condition for small filing effects is that *both* densities be small. This requirement, in turn, limits the extent to which the densities can differ from one another. And this corresponds to a world in which there is only a very loose association between what defendants actually do in the primary activity and whether or not they are sued.[36]

> PROPOSITION 9: *Consider the case where filing fees are restricted to be negligible (more generally, not adjustable). Suppose that the defendant's wealth is sufficiently large, the fact-finder's perception of evidentiary weight is sufficiently error prone, and the distribution of plaintiffs' costs is sufficiently diffuse. Then Propositions 4 and 5 still hold.*

VI. 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'

[36] A proof of the analogy to Proposition 9 when the model is expanded to include "false suits" is available from the authors.

stakes on *both* filings and infra-marginal suits. This expansion has also 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 notable exceptions, most of this literature focuses on the incentive to file and to settle, leaving discovery and evidence production relatively under-modeled. 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.

VII. TECHNICAL APPENDIX

A. *Some technical remarks on the evidence production game at trial*

To ensure an interior solution to each party's evidence choice problem we assume that $c'(0)$ and $V'(0)$ are arbitrarily small, while the density *f* is positive in the relevant range for $y-x$. We conduct comparative statics on a single equilibrium of the evidence production game described in the text. There may be several equilibria depending on the shape of the trial noise distribution *F*.

Apropos of these assumptions, the litigation game we have defined will not under *any* parametric assumptions exhibit strategic complementarity (even if the ordering of either or both players' strategy spaces are reversed). This means that the techniques of "monotone comparative statics" are not available to us. Nor are results on equilibrium uniqueness.[37]

This lack of strategic complementarity is a deep structural characteristic of litigation models in which player's payoff functions are interdependent through the probability of plaintiff victory. If more defendant evidence inspires the plaintiff to produce more evidence, this must be because additional defendant evidence increases the marginal impact of net evidentiary weight in favor of the plaintiff. That means that the marginal impact of net evidentiary weight is decreasing in net evidentiary weight. This, in turn, implies that more plaintiff evidence, which increases net evidentiary weight, lowers the marginal impact of evidence production for defendant.

B. *Proposition 1*

If R, D*, and K* minimize all-in social cost, while generating marginal filer $\hat{k}^*$, per suit deterrence Δ^*, and per suit evidence costs $c^* + V^*$, then R* and D* also solve the problem:* $\min_{R,D} c+V: \; pD+V=\Delta^*$.

Proof: Suppose, on the contrary, that $\hat{R}$ and $\hat{D}$ yield per suit deterrence Δ^* at lower per suit cost $\hat{c}+\hat{V} < c^* + V^*$. Set the filing fee $\hat{K}$ so that $p'\hat{R}-\hat{c}-\hat{K} = p^*R^* - c^* - K^*$. Since $\hat{R}$, $\hat{D}$, and $\hat{K}$ yield the same set of filing plaintiffs $\left[0, k^*\right]$ and the same per suit deterrence Δ^* as *R**, *D*,* and *K**, they also yield the same all-in deterrence Ω^*. Therefore, they yield the same primary activity costs (1). However, each filed suit is strictly less costly. Therefore, litigation costs (2) 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. QED.

C. *Proposition 2*

At any interior social optimum at which marginal deterrence costs for R and D are finite and positive, $MDC_R = MDC_D$.

---

[37] See Vives, Xavier "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19(3): 305-321, (1990) and Milgrom, Paul and Christine Shannon, "Monotone Comparative Statics," *Econometrica*, 157-180 (January 1994).

Proof: The problem of providing a given level of per suit deterrence at minimal per suit cost is: $\min_{D,R} c(x)+V(y)$ subject to $p(x,y)D+V(y)=\Delta *$. First-order conditions for an interior solution to this problem can be written as $c'x_R+V'y_R = l\,p_x x_R D$ and $c'x_D+V'y_D = l\,(p+p_x D x_D)$, where l is the Lagrange multiplier. If $p_x x_R D \neq 0$ and $p+p_x D x_D \neq 0$, then we can isolate l on the right-hand side of both equations. The resulting ratios on the left-hand sides will both be equal to l and so equal to each other. On the other hand, if $p_x x_R D = 0$, then the first first-order condition implies that $c'x_R+V'y_R=0$ and MDC_R is not well defined, having zeros in both numerator and denominator. Similarly, if $p+p_x D x_D = 0$, then MDC_D is not well defined. QED.

D. *Proposition 3*

We will prove something more than the statement of the result in the text. First, some preliminary results:

LEMMA A1: *The plaintiff's first- and second-order conditions for evidence production are* $fR-c'=0$ *and* $-fR-c''\leq 0$. *The defendant's are* $-f'D+V'=0$ *and* $-f'D+V''\geq 0$. *The derivatives of the parties' evidence production in* R *and* D *are given by the matrix equation*:

$$\begin{bmatrix} x_R & x_D \\ y_R & y_D \end{bmatrix} = -\frac{1}{\Lambda}\begin{bmatrix} (-f'D+V'')f & f'Rf \\ -f'Df & -(-f'R-c'')f \end{bmatrix},$$

where $\Lambda \equiv (-f'R-c'')(-f'D+V'')-(f')^2 RD<0$.

Proof: We obtain the parties' first- and second-order conditions from the conditions given in the text (Section II.A) and the fact that $f=p_x=-p_y$ and $f'=-p_{xx}=-p_{yy}$. To obtain evidence choice derivatives, we apply the multivariate implicit function theorem. This theorem applies because, first, Λ is the determinant of the derivative (a 2x2 matrix) of the parties' first order conditions with respect to (x,y). Secondly, $\Lambda \neq 0$, indeed $\Lambda<0$. For if $f'=0$, then $f'=-p_{xx}=-p_{yy}=p_{xy}=0$, and the Λ expression becomes $-c''V''<0$. And if $f'\neq 0$, then Λ is less than $-(f')^2 RD<0$, because $(-f'R-c'')(-f'D+V'')\leq 0$. The implicit function then yields:

$$\begin{bmatrix} x_R & x_D \\ y_R & y_D \end{bmatrix} = -\begin{bmatrix} p_{xx}R-c'' & p_{xy}R \\ p_{xy}D & p_{yy}D+V'' \end{bmatrix}^{-1}\begin{bmatrix} p_x & 0 \\ 0 & p_y \end{bmatrix} = -\frac{1}{\Lambda}\begin{bmatrix} p_{yy}D+V'' & -p_{xy}R \\ -p_{xy}D & p_{xx}R-c'' \end{bmatrix}\begin{bmatrix} p_x & 0 \\ 0 & p_y \end{bmatrix}$$

$$= -\frac{1}{\Lambda}\begin{bmatrix} (p_{yy}D+V'')p_x & -p_{xy}Rp_y \\ -p_{xy}Dp_x & (p_{xx}R-c'')p_y \end{bmatrix} = -\frac{1}{\Lambda}\begin{bmatrix} (-f'D+V'')f & f'Rf \\ -f'Df & -(-f'R-c'')f \end{bmatrix}.$$

LEMMA A2: *If the marginal deterrence cost of recovery is well defined, then it reduces to*

$$MDC_R = \frac{R}{D}+\frac{y_R}{x_R} = \frac{R}{D}+\frac{f'D}{f'D-V''}.$$

Proof: Substituting the parties' first-order conditions from Lemma A1 into MDC_R (as given in the text) gives:

$$MDC_R = \frac{fRx_R + fDy_R}{\frac{D}{R}fRx_R} = \frac{R}{D} + \frac{y_R}{x_R}.$$

Using Lemma A1 to substitute for y_R and x_R, and noting that $\Lambda < 0$, we obtain

$$MDC_R = \frac{R}{D} + \frac{-f'Df}{\left(-f'D + V''\right)f} = \frac{R}{D} + \frac{-f'D}{-f'D + V''}.$$

PROPOSITION 3: *If, at the parties' privately optimal levels of evidence production,* $f'\left(y - x\right) > 0$*, then* $MDC_R < \frac{R}{D}$*. If* $f' = 0$*,* $MDC_R = \frac{R}{D}$*. And if* $f' < 0$*, then* $MDC_R < \frac{R}{D} + 1$*.*

Proof: From Lemma A1, the defendant's second-order condition is $-f'D + V'' \geq 0$, or $f'D - V'' \leq 0$. Therefore, if $f' > 0$, then $\frac{f'D}{f'D - V''} \leq 0$, and, from Lemma A2, $MDC_R = \frac{R}{D} + \frac{f'D}{f'D + V''} \leq \frac{R}{D}$. If , on the other hand, $f' < 0$, then $-f'D + V'' > -f'D > 0$, or $f'D - V'' < f'D < 0$. That implies that $\left|f'D - V''\right| > \left|f'D\right| > 0$ and, therefore, $0 < \frac{f'D}{f'D - V''} < 1$. We conclude from Lemma A2 that $MDC_R = \frac{R}{D} + \frac{f'D}{f'D + V''} < \frac{R}{D} + 1$, and if $f' = 0$, $\frac{f'D}{f'D + V''} = 0$, given $V'' > 0$, and $MDC_R = \frac{R}{D}$. QED.

E. *Proposition 4*

First, some preliminary results and assumptions.

LEMMA A3: $\lim_{\sup f' \to 0} MDC_D = \dfrac{f^2 D}{V'' p}$.[38]

Proof: From Lemma A1:

$$MDC_D = \frac{fRx_D + fDy_D}{fDx_D + p} = \frac{fR\left(\frac{f'Rf}{-\Lambda}\right) + fD\left(\frac{-\left(-f'R - c''\right)f}{-\Lambda}\right)}{fD\left(\frac{f'Rf}{-\Lambda}\right) + p} = \frac{f'R^2 f^2 + f^2 Df'R + f^2 Dc''}{f^2 Df'R - \Lambda p}.$$

Noting that

$$\lim_{\sup f' \to 0} \Lambda = -c''V''$$

we have,

$$\lim_{\sup f' \to 0} \frac{f'R^2 f^2 + f^2 Df'R + f^2 Dc''}{f^2 Df'R - \Lambda p} = \frac{f^2 Dc''}{c''V'' p} = \frac{f^2 D}{V'' p}.$$

We assume that the second derivative $V'' > 0$ of the defendant's marginal evidence cost does not grow exponentially ad infinitum:

ASSUMPTION 1: $\lim_{y \to \infty} \dfrac{V'''}{V''} = 0$.

[38] This notation means that for any sequence f_n of error densities with $\sup f_n' \to 0$, the corresponding sequence of MDC_D has this limit.

This condition holds for a great variety of functional forms, including all polynomial functions $V(y) = a_n y^n + \ldots + a_1 y + a_0$ on $y \geq 0$. For example, it holds for $V(y) = y^2$, wherein $V' = 2y$, $V'' = 2$, and $V''' = 0$. (Not all cost functions satisfy the assumption: consider, for example, the exponential cost function $V = e^y$.) The assumption plays a role in ensuring that the defendant's evidentiary response to changes in damages does not decay too quickly.

LEMMA A4: *Define the inverse $f(z)$ of the defendant's marginal cost function, $V'(y)$, so that $z = V'(f(z))$. Then, Assumption 1 implies that $\lim_{z\to\infty} z / V''(f(z)) = \infty$.*

Proof: Suppose, on the contrary, that $\lim_{z\to\infty} \frac{z}{V''(f(z))} \neq \infty$. Then, because the numerator of this fraction goes to infinity, the denominator must also. L'Hopital's rule then applies. Therefore, $\lim_{z\to\infty} \frac{z}{V''(f(z))} = \lim_{z\to\infty} \frac{1}{V'''f'} = \lim_{z\to\infty} \frac{1}{V'''/V''}$, where we have used the fact that, by the inverse function theorem, $f' = \frac{1}{V''}$. The right-hand limit is infinite by assumption, which contradicts our supposition.

PROPOSITION 4: *The following holds for any lower-bound M on the marginal deterrence cost of damages, however large this bound. Choose any $j > 0$, and consider the set of all error distributions whose densities f never fall below j on their supports. For this subset of error densities, there exists lower-bound wealth level $\overline{W}$ such that if the defendant's wealth exceeds this bound ($W \geq \overline{W}$), then there exists an upper bound $x > 0$ on the error density's derivative f' and a lower bound $\overline{D} \leq W$ on damages such that if these two bounds are satisfied (i.e., if $\sup f' < x$ and $\overline{D} \leq D \leq W$), then $MDC_D > M$.*

Proof: Given j, Lemma A4 implies that we can choose W large enough so that for all $D \geq \frac{1}{2}W$, $j \frac{jD}{V''(f(jD))} > M + 1$. By Lemma A3, for any given D and R pair, we can find $x > 0$ small enough so that, if $\sup f' < x$, then $MDC_D \geq \frac{f^2 D}{V'' p} - \frac{1}{2}$. By a standard result from real analysis, we can, in fact, choose a single $x > 0$ small enough, so that if $\sup f' < x$, then $MDC_D \geq \frac{f^2 D}{V'' p} - \frac{1}{2}$ for all (D, R) pairs on the compact set $[0, W]^2$. Therefore, if $\sup f' < x$, then for all $D \geq \frac{1}{2}W$,

$$MDC_D \geq \frac{f^2 D}{V'' p} - \frac{1}{2} \geq j \frac{jD}{V''(f(jD))} \frac{1}{p} - \frac{1}{2} \geq (M+1)\frac{1}{p} - \frac{1}{2} > M.$$

REMARK: Notice that the bound on the error density's derivative x is not chosen uniformly across all minimal density heights j. The smaller *f*, the greater *D* must be to ensure that MDC_D exceeds *M*, ignoring cross-effects. But the greater *D*, the greater cross-effects for any given $\sup f'$. Cross-effects do not necessarily work against the result. But they may. Thus we choose $\inf f$ first, then choose D large enough so that $MDC_D > M$, ignoring cross-effects, and then choose $\sup f'$ small enough so that adding cross-effects cannot defeat this inequality. More precisely, the bound on $\sup f'$ *can* be chosen uniformly over all $D \le W$.

REMARK: What is the precise relationship between the error distribution's variance, on the one hand, and the supremum of its density's derivative, on the other. 1) As the variance of error increases to infinity, the supremum of the derivative of the density converges "in probability" to 0. However, it does not necessarily converge to zero "with probability one." That is, while it is not necessarily true that $\lim_{s^2 \to 0} \sup f' = 0$, it is true that for all $d > 0$, $\lim_{s^2 \to 0} \Pr\left(\sup f' > d\right) = 0$.[39] 2) This limiting discrepancy between the supremum of the density's derivative and the variance is indicative of a drawback of using the variance as a measure of dispersion. The variance is a summary measure of dispersion: a sequence of distributions can be ever more concentrated on an ever smaller interval even as the variance goes to infinity. 3) However, if we take the variance to infinity within certain classes of distributions, such as the class of normal distributions or the class of truncated normal distributions, then it will indeed be true that $\lim_{s^2 \to 0} \sup f' = 0$.

F. *Proposition 6*[40]

Proof: First, we establish that per suit deterrence at a social optimum grows without bound with the level of harm. That is, letting Δ^* represent the per suit deterrence created by optimal R and D, we show that $\lim_{h \to \infty} \Delta^* = \infty$. Suppose, on the contrary, that Δ^* remains bounded. Then all-in deterrence $\Omega = G\left(\hat{k}\right)\Delta$ is also bounded, given that $G\left(\hat{k}\right) \le 1$. 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 R, D, and K) will be strictly negative at all levels of all-in deterrence below B, which contradicts the social optimality of Δ^*.

To see precisely why, take any $\Omega \le B$ and suppose that this all-in deterrence is created by some R, D, and K. Now consider changing both R and D in such manner that per suit 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; the change in R 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.) Note that this change in R and D may also change per suit costs. Consider, also, adjusting K to keep the number of filings constant. The net effect of these changes in R, D, and K will be an increase in all-in deterrence Ω and some change, positive or negative, in per suit costs. For large enough h, the positive social cost effect of the former will outweigh the latter per suit cost effect because

$$\frac{\partial SC}{\partial \Omega} = -(h - \Omega)\, j(\Omega) - j(\Omega)\int_0^{\hat{k}} (k + c + V)\, g(k)\, dk$$

[39] For a discussion of these convergence concepts, see Patrick Billingsley, Probability and Measure, pp. 330-331, New York, John Wiley & Sons (3rd ed 1995).

[40] The proof of Proposition 5 is omitted because it follows immediately from Propositions 2-4, as described in the text. From hereon, we will include a formal statement of results that are obvious variations on the formal statement of Proposition 4.

goes to negative infinity as h approaches infinity, where the per suit cost effect is finite and independent of h. Therefore, these levels of R, D, and K could not be socially optimal for large enough h.

Next we show that for large enough optimal per suit deterrence Δ^*, optimal recovery is no less than optimal damages. Suppose, on the contrary, that optimal recovery is infinitely often strictly less than optimal damages, as Δ^* grows without bound. By Proposition 4, this implies that infinitely often optimal damages—and so optimal recovery, by hypothesis—are bounded below $\bar{D}$, as that variable is used in Proposition 4. But this contradicts that Δ^*, a continuous function of R and D, grows without bound.

G. *Proposition 7*

First, some preliminary results.

LEMMA A5: *With settlement*, $\lim_{\sup f' \to 0} MDC_R \leq \dfrac{1-k}{1-h}\dfrac{R}{D}$.

Proof:

$$
\begin{aligned}
(1-k)\left(MDE_R\right)^{-1} &= \left(\frac{V' y_R + c' x_R}{(1-h)\, p_x x_R D + h\,(p + p_y y_R R)}\right)^{-1} \\
&= (1-h)\frac{p_x x_R D}{V' y_R + c' x_R} + h\,\frac{p + p_y y_R R}{V' y_R + c' x_R} \\
&= (1-h)\frac{p_x x_R D}{c' x_R} + h\,\frac{p}{c' x_R} && \text{[pass to limit using Lemma A1]} \\
&= (1-h)\frac{D}{R} + h\,\frac{p}{c' x_R} && \text{[plaintiff's first-order condition]} \\
&\geq (1-h)\frac{D}{R} && [\,x_R \geq 0 \text{ from Lemma A1}]
\end{aligned}
$$

REMARK: Without settlement, MDC_R always has a scale-independent upper bound whenever $R \leq D$ (Proposition 3). With settlement, we find this upper bound when cross-effects are small.

LEMMA A6: *With settlement*, $\lim_{\sup f' \to 0} MDC_D \geq \dfrac{1-k}{1-h}\dfrac{f^2 D}{V'' p}$.

Proof: First, we track the steps in the proof of Lemma A5:

$$
\begin{aligned}
(1-k)\left(MDE_D\right)^{-1} &= \left(\frac{V' y_D + c' x_D}{(1-h)(p + p_x x_D D) + h\, p_y y_D R} \right)^{-1} \\
&= (1-h)\frac{p + p_x x_D D}{V' y_D + c' x_D} + h\,\frac{p_y y_D R}{V' y_D + c' x_D} \\
&= (1-h)\frac{p}{V' y_D} + h\,\frac{\overbrace{p_y}^{-} y_D R}{V' y_D} \\
&\le (1-h)\frac{p}{V' y_D}
\end{aligned}
$$

The result then follows from Lemmas A1 and A3.

Proof of PROPOSITION 7: We apply the logic of the proof of Proposition 4 to the results in Lemmas A5 and A6. However, now we choose the upper bound $x > 0$ on $\sup f'$ so that not only

$$
MDC_D \ge \frac{1-k}{1-h}\frac{f^2 D}{V'' p} - \frac{1}{2},
$$

but also

$$
MDC_R \le \frac{1-k}{1-h}\frac{R}{D} + \frac{1}{2}.
$$

H. *Proposition 8*

Proof: Under the British rule, the plaintiff maximizes $pR - (1-p)(c+V)$, while the defendant minimizes $p(D+c+V)$, which is also per suit deterrence Δ. The parties' first-order conditions are $p_x(R+c+V) - (1-p)c' = 0$ and $p_y(D+c+V) + pV' = 0$. The marginal deterrence cost of damages is the same with small cross-effects ($\sup f' \to 0$) as under the American rule:

$$
\begin{aligned}
MDC_D &= \frac{c' x_D + V' y_D}{p + \left(p_y\left(D+c+V\right) + pV'\right) y_D + \left(p_x\left(D+c+V\right) + pc'\right) x_D} \\
&= \frac{c' x_D + V' y_D}{p + \left(p_x\left(D+c+V\right) + pc'\right) x_D} \\
&\to \frac{V' y_D}{p} = \frac{f^2 D}{V'' p}
\end{aligned}
$$

The marginal deterrence cost of recovery has a scale-independent bound when cross-effects are small and $R \le D$:

$$
\begin{aligned}
MDC_R &= \frac{c' x_R + V' y_R}{\left(p_y\left(D+c+V\right)+pV'\right)y_R + \left(p_x\left(D+c+V\right)+pc'\right)x_R} \\
&= \frac{c' x_R + V' y_R}{\left(p_x\left(D+c+V\right)+pc'\right)x_R} \\
&\to \frac{c' x_R}{\left(p_x\left(D+c+V\right)+pc'\right)x_R} \\
&= \frac{p_x\left(R+c+V\right)+pc'}{p_x\left(D+c+V\right)+pc'}
\end{aligned}
$$

which is less than 1 when $R \le D$. We then apply the logic of the proof of Propositions 4 and 7.

I. *Proposition 9*

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

$$
\begin{aligned}
\frac{\partial SC}{\partial R} &= -\Omega_R\, j(\Omega)\left(h-\Omega+G(\hat{k})\left(m(\hat{k})+c+V\right)\right)+(1-J(\Omega))\left(\left(\hat{k}+c+V\right)g(\hat{k})\hat{k}_R+\left(c'x_R+V'y_R\right)G(\hat{k})\right) \\
\frac{\partial SC}{\partial D} &= -\Omega_D\, j(\Omega)\left(h-\Omega+G(\hat{k})\left(m(\hat{k})+c+V\right)\right)+(1-J(\Omega))\left(\left(\hat{k}+c+V\right)g(\hat{k})\hat{k}_D+\left(c'x_D+V'y_D\right)G(\hat{k})\right)
\end{aligned}
$$

where $m\left(\hat{k}\right)=E\left[k \middle| k\le\hat{k}\right]$, $\Omega_R = g(\hat{k})\hat{k}_R\Delta + G(\hat{k})p_x x_R D$, $\Omega_D = g(\hat{k})\hat{k}_D\Delta + G(\hat{k})(p+p_x x_D D)$, $\hat{k}_R = p + p_y y_R R$, and $\hat{k}_D = p_y y_D R$. When $\sup g \to 0$, we have $\Omega_R \to G(\hat{k})p_x x_R D$ and $\Omega_D \to G(\hat{k})(p+p_x x_D D)$. Furthermore, $\hat{k}_R$ and $\hat{k}_D$ are bounded on $\left(R,D\right)\in\left[0,W\right]^2$. Therefore, in the limit

$$
\begin{aligned}
\frac{\partial SC}{\partial R} &= -\Omega_R\, j(\Omega)\left(h-\Omega+G(\hat{k})\left(m(\hat{k})+c+V\right)\right)+(1-J(\Omega))\left(c'x_R+V'y_R\right)G(\hat{k}) \\
\frac{\partial SC}{\partial D} &= -G(\hat{k})(p+p_x x_D D)\, j(\Omega)\left\{(h-\Omega)+\int_{k=0}^{\hat{k}}(k+c+V)\,dG\right\}+(1-J(\Omega))\left(c'x_D+V'y_D\right)G\left(\hat{k}\right).
\end{aligned}
$$

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

$$
\frac{(1-J(\Omega))\left(c'x_R+V'y_R\right)G\left(\hat{k}\right)}{G(\hat{k})(p_x x_R D)\, j(\Omega)\left\{(h-\Omega)+\int_{k=0}^{\hat{k}}(k+c+V)\,dG\right\}} = \frac{(1-J(\Omega))\left(c'x_D+V'y_D\right)G\left(\hat{k}\right)}{G(\hat{k})(p+p_x x_D D)\, j(\Omega)\left\{(h-\Omega)+\int_{k=0}^{\hat{k}}(k+c+V)\,dG\right\}}
$$

Simplifying yields the result of Proposition 2, $\frac{c'x_R+V'y_R}{p_x x_R D} = \frac{c'x_D+V'y_D}{p+p_x x_D D}$.