Optimal Design of Private Litigation

Louis Kaplow*

Abstract

This article examines optimal legal system design in a model with private suits, evidence signals, court error, and two types of primary behavior: harmful acts that may be deterred and benign acts that may be chilled. The instruments examined are filing fees or subsidies that may be imposed on either party, damage awards and also payments by unsuccessful plaintiffs (each of which may be decoupled), and the stringency of the evidence threshold (burden of proof). With no constraints, results arbitrarily close to the first best can be implemented. Prior analyses of optimal damage awards, decoupling, and fee shifting are shown to involve special cases, with results changing qualitatively when implicit instrument restrictions are relaxed. The introduction of a simple filing fee can make it optimal to minimize what losing plaintiffs pay winning defendants and to reduce the evidence threshold as much as possible — even though the direct effect of these adjustments is to chill desirable behavior.

JEL Classes: D82, H23, K13, K41, K42

Keywords: litigation, law enforcement, adjudication, courts, instrument choice, burden of proof, filing fees, punitive damages, fee shifting, decoupling

© Louis Kaplow. All rights reserved.

*Harvard University and National Bureau of Economic Research. I am grateful to Jonathan Borowsky, Aaron Edlin, Steven Shavell, Holger Spamann, Kathryn Spier, and participants at Harvard, NBER, and the Law and Economics Theory conference for comments, and to the John M. Olin Center for Law, Economics, and Business at Harvard University for financial support.
1. Introduction

Much law enforcement is accomplished using private suits, usually filed by victims, rather than through public enforcement, the latter subject having received greater attention in the literature on optimal legal system design. With private enforcement, actors’ primary behavior and plaintiffs’ filing decisions are influenced by a variety of instruments. The social welfare problem is to maximize benefits from individuals’ primary activity — which here includes both harmful and benign acts — net of harms caused and system costs. It is important to explore optimal system design in this setting, particularly in light of the fundamental divergence between private and social welfare with regard to the incentive to sue (Shavell 1982). In addition, we wish to understand the extent to which social welfare may be sacrificed by the fact that actual legal systems restrict the instrument set in ways that have received little attention and to determine how such limitations change the optimal use of those instruments that remain.

In the model presented in section 2, three types of behavioral effects matter: the deterrence of harmful acts, the chilling of benign acts, and plaintiffs’ inclination to sue. These may be influenced by several instruments: a fee (or subsidy) imposed on each of the plaintiff and the defendant at the time of suit, a damage award paid by losing defendants to successful plaintiffs (which may be decoupled, allowing the two amounts to differ), and a transfer paid by losing plaintiffs to successful defendants (which may also be decoupled) — as well as an evidence threshold (akin to a burden of proof), indicating how strong the evidence must be for liability to be imposed.

When individuals commit an act of either type, a signal is generated, the distributions of these signals differing for the two types of acts. This signal is observed by a prospective plaintiff, who decides whether to sue. When there is a suit, each party pays a fee (which may be zero or negative) and incurs litigation costs. Then the tribunal itself receives its signal — which is a noisy observation of the original signal — and imposes liability depending on whether the observed evidence strength exceeds the legal threshold. Finally, transfers are paid, the amount and direction depending on whether liability is imposed. Individuals’ decisions whether to undertake their acts and plaintiffs’ decisions whether to sue reflect private expected benefits and the expected costs of the legal system just described. The legal system’s raison d’être is to deter harmful activity, but the planner also seeks to use unavoidably costly litigation as little as possible and to enhance the system’s diagnosticity, which promotes deterrence and reduces chilling.

Before proceeding to the analysis of how this is accomplished, section 2 also explores instrument redundancy. Two of the transfer instruments are redundant in a narrow accounting or mechanical sense, and an additional instrument is to a degree redundant in practice when others are sufficiently unconstrained. The relationships among these instruments help to illuminate prior literature wherein each paper considers only a couple particular instruments.

---

1Much of the literature surveyed in Polinsky and Shavell (2007), on public enforcement, focuses on optimal design, whereas the majority in Spier (2007), on private litigation, involves positive analysis of suit and settlement.

2On the latter, see note 11.
Section 3 demonstrates that if three instruments — a plaintiff filing fee, a standard (non-decoupled) damages award, and a (non-decoupled) transfer from losing plaintiffs to defendants — are unconstrained, we can to span nearly all relevant choices of the three types of behavior (regarding harmful acts, benign acts, and suit). Moreover, it is feasible to implement a result arbitrarily close to the first best: first-best deterrence (individuals commit harmful acts if and only if their benefit exceeds the harm), no chilling of benign acts, and negligible litigation costs. The analysis is not straightforward because of the need to provide differential incentives to those who may commit harmful and benign acts, although in this unconstrained setting such is possible even though the evidence threshold is taken to be fixed.

In practice, instruments are typically restricted, by limits on feasibility (parties’ payments cannot exceed their wealth) and by various legal institutions (perhaps for reasons outside this model). Accordingly, most of the analysis is devoted to optimal system design under various restrictions. Section 4 analyzes a setting in which the available instruments are a filing fee that may be imposed on plaintiffs (only), a (non-decoupled, that is, ordinary) damages award that losing defendants pay to plaintiffs (which may be subject to a maximum), a (non-decoupled) transfer that losing plaintiffs pay to defendants, and the evidence threshold. With these instruments and restrictions, the optimal sanction is maximal, the optimal transfer from losing plaintiffs to defendants is minimal (negative, if permitted), and the optimal evidence threshold is as low as possible.

These results hold despite the presence of court error involving the imposition of sanctions on individuals who committed benign acts that we wish to chill as little as possible. The core intuition behind these somewhat surprising conclusions is that plaintiffs’ self-interested filing decisions themselves reflect the initial signal and hence sort cases by quality to that extent: Plaintiffs file only those cases that are stronger than some critical level, which minimum serves as a de facto evidence threshold. By raising the filing fee — in conjunction with raising the sanction, lowering the transfer paid by losing plaintiffs, or reducing the de jure evidence threshold — one can, for example, maintain plaintiffs’ filing decisions, holding fixed this de facto evidence threshold. But there’s more: each of the stated adjustments actually allows for an even higher filing fee and thus a more stringent de facto evidence threshold, making it possible to simultaneously increase deterrence, hold chilling constant, and reduce filing rates and thus total litigation costs. Even the lower de jure evidence threshold enables a higher de facto threshold and thereby can better protect individuals who commit benign acts. (The actual experiment holds chilling constant, but we could instead have held deterrence constant, which would have yielded a reduction in chilling.) A parallel intuition for these results is that each policy experiment makes the system harsher on defendants in inframarginal cases while making it more generous by removing marginal cases; even when the former bears relatively more heavily on benign acts, the latter, because it involves the weakest cases from among those previously filed, relatively benefits benign acts to an even greater extent.

The present analysis casts prior literature on the optimal design of private litigation in a
different light.\textsuperscript{3} The most relevant articles consider notably simpler environments. No previous analyses of private litigation include the combination of explicitly modeled signals, two types of court error, both harmful and benign acts, and legal system costs.\textsuperscript{4} The omission of benign behavior is particularly important when one considers such matters as defensive medicine, securities lawsuits that might raise the cost of capital (e.g., in IPOs), and antitrust litigation that may discourage efficient behavior.\textsuperscript{5} Moreover, policies that might be best in models with only harmful acts, because they maximize deterrence, could be worst if one introduced benign acts that would be chilled. Yet Png (1986) is the only paper with two-way errors and benign as well as harmful acts; however, in his model all cases are filed and litigation is costless, so little of the current challenge arises.

In addition, attention in a given paper is usually confined to (at most) two instruments: damages combined with one other. Polinsky and Che (1991) examine decoupling, in a model with only harmful acts and thus one-way error. In a similar framework, Kaplow (1993) addresses fee shifting, and Polinsky and Rubinfeld (1996) introduce the possibility that losing plaintiffs pay defendants. Each setting is a special case in many respects. Moreover, as the foregoing explains, the optimal use of one instrument can change qualitatively when implicit restrictions on other instruments are relaxed — in some instances, by the mere introduction of a filing fee. Indeed, the sometimes surprising results in the present model are due in significant part to relaxing prior work’s implicit assumptions, whose significance in generating contrary conclusions was unappreciated.\textsuperscript{6}

2. Model

A. Setup

There are two types of acts that may be committed, a harmful one, $H$, and a benign one, $B$. The harmful type of act imposes an external social cost of $h$; the benign type of act involves no externality.\textsuperscript{7} A mass of individuals normalized to 1 may commit the harmful type of act. Those who may commit the benign type of act have a mass of $\gamma$; an interpretation is that $\gamma$ indicates the relative mass of benign acts that may be undertaken in situations in which they

---

\textsuperscript{3}There is also some prior positive analysis of some of the elements at play here. See Polinsky and Shavell (1989).

\textsuperscript{4}Kaplow (2011) includes this array but with public enforcement; also, fewer instruments are examined (some of which are moot when the government chooses enforcement effort by fiat).

\textsuperscript{5}For suggestive empirical evidence, see, for example, Studdert et al. (2005) on the avoidance of medical procedures and patients posing high litigation risks; Kessler, Sage, and Becker (2005) on medical malpractice and physician supply; and Lin, Pukthuanthong, and Walker (2013) on IPO underpricing and litigation risk.

\textsuperscript{6}This paper does not pursue an unrestricted mechanism design approach, under which the only limitations are derived from the information structure and underlying technology, because the purpose here is to explore the optimal design of private litigation of a form broadly employed and the subject of prior literature. The results in section 3, where it is possible to achieve an outcome arbitrarily close to the first best, obviously would not differ under the pure mechanism design approach. In section 4, if instrument constraints are taken to depend, for example, on limits on parties wealth, then any further limitations that make the outcome fall short of the first best might be mitigated if a broader class of instruments were allowed. In particular, the tribunal here is assumed to observe the parties’ signal with noise, but there exist information revelation games that are superior to litigation as ordinarily conducted.

\textsuperscript{7}One could readily allow a smaller negative externality or a positive externality.
might initially be confused with harmful acts (because other benign acts do not give rise to the possibility of suit and sanctions). One could instead imagine that the same individuals may commit both types of acts, and that \( \gamma \) indicates the relative frequency of opportunities to commit the benign type of act.\(^8\) An individual’s benefit from committing an act is \( b \), with density functions \( f^i(b) \) (which are positive for positive values of \( b \)) and cumulative distribution functions \( F^i(b) \), where \( i = H, B \). Individuals know what type of act they are able to commit and its benefits to them, but the tribunal knows neither acts’ types nor actors’ benefits. Individuals commit their acts if and only if their private benefit \( b \) exceeds expected legal costs, as specified below. (All actors are taken to be risk neutral.\(^9\))

When an act is committed, it generates a signal \( x \) that is observed by the actor and also by a prospective plaintiff. For each type of act, the densities and cumulative distribution functions for \( x \) are given by \( g^i(x) \) and \( G^i(x) \), respectively, which are assumed to satisfy the strict monotone likelihood ratio property: \( x_1 > x_0 \) implies that \( g^H(x_1)/g^B(x_1) > g^H(x_0)/g^B(x_0) \), which is to say that higher values of \( x \) are relatively more likely to be generated by acts of type \( H \) than those of type \( B \).

A prospective plaintiff (\( P \)) sues a prospective defendant (\( D \)) — usually referred to as a plaintiff and a defendant for simplicity — if and only if the expected gains from suit exceed the expected costs. The legal system operates as follows: First, if a suit is filed, each party pays the government a fee \( r^j \), where \( j = P, D \); these fees may be negative (subsidies). (The superscript \( i \) will always be used to refer to types of acts, \( H \) and \( B \), and the superscript \( j \) to refer to parties, \( P \) and \( D \).) After a suit is filed, all cases proceed to trial and parties incur litigation costs, \( c^j.10 \)

At trial, the tribunal assigns liability if and only if the evidence it observes, \( \chi \), exceeds the evidence threshold, \( x^T.11 \) The tribunal’s signal is related to the parties’ signal (which the tribunal does not observe): \( \chi = x + \varepsilon \), where \( \varepsilon \) varies according to the density function \( z(\varepsilon) \). (It might be natural to suppose that \( \varepsilon \) has a mean of zero and that its distribution satisfies other standard properties; however, such restrictions turn out to have little purchase and, in any event, the mean of \( \varepsilon \) in this setting is just a normalization because \( x^T \) may be translated accordingly.)

If there is a finding of liability, \( D \) pays \( s^D \) and \( P \) receives \( s^P \); that is, damages may be decoupled, although (for reasons given in section 2.C) most of the analysis will restrict attention

\(^8\)Alternatively, one could allow individuals to choose one of the two acts or inaction, which would complicate the exposition but have only a modest effect on the qualitative results: deterrence of harmful acts would induce some individuals to switch to benign acts rather than inaction (making deterrence more valuable, and chilling of benign acts would cause some individuals to switch to harmful acts rather than inaction (making chilling more detrimental). If plaintiffs, defendants, or both were risk averse, then behavior would be influenced in familiar ways. Moreover, social welfare (expression 4 just below) would differ because transfers would no longer be neutral. See, for example, Polinsky and Shavell (1979), and also note 16.

\(^9\)The analysis throughout abstracts from settlements. With symmetric information and assuming that some costs are incurred before settlement (which is typical in reality), the qualitative results would be unchanged. Compare Polinsky and Che (1991). As mentioned in the conclusion, one could allow the parties’ initial signals to differ, thereby introducing asymmetric information and making for more interesting explicit analysis of settlement.

\(^10\)The evidence threshold is related to the notion of a burden of proof. The latter concept, as it is understood in the legal sphere, ordinarily refers in essence to a specified Bayesian posterior probability rather than to a cutoff on the strength of the evidence (a likelihood ratio); these turn out to differ in important ways. See Kaplow (2011).
to the case in which $s^D = s^P = s$. If there is a finding of no liability, $P$ pays $t^P$ and $D$ receives $t^D$, although often it will be assumed that $t^D = t^P = t$. (To repeat for the sake of clarity: for the non-decoupled case, $s$ corresponds to a familiar damages payment, what the defendant pays to the plaintiff in the event of liability, and $t$ is a transfer from the plaintiff to the defendant when the plaintiff loses at trial.) The values of the $s^j$ and $t^j$ are taken to be nonnegative unless otherwise indicated. Sometimes it will be supposed that $s$ or $s^D$ cannot exceed some maximum, $s^{Max}$, as is common in much of the literature on the economics of law enforcement.

It is useful to summarize much of the foregoing in a timeline:

1. The government sets all policy instruments: the $r^j$, $s^j$, $t^j$, and $x^T$.
2. (Prospective) defendants learn their type of act ($H$ or $B$) and private benefit $b$.
3. They decide whether to act.
4. For each act undertaken, the signal $x$ is observed by the parties.
5. (Prospective) plaintiffs decide whether to sue.
6. If a suit is brought, the fees $r^j$ are paid to the government and the costs $c^j$ are incurred by the parties.
7. When there is a suit, the tribunal observes $\chi$. If $\chi > x^T$, damages of $s^j$ are paid or received, as the case may be. If $\chi \leq x^T$, transfers of $t^j$ are paid or received.

**B. Behavior and Social Welfare**

We can now characterize parties’ actions. Starting with plaintiffs, it is useful to introduce the notation $\pi(x|x^T)$, sometimes abbreviated as $\pi(x)$ for simplicity, which refers to the probability that a tribunal will assign liability when the initial signal is $x$. Note that the liability rule, $\chi > x^T$, can be expressed as $x + \varepsilon > x^T$, which is equivalent to $\varepsilon > x^T - x$. Hence,

$$
(1) \quad \pi(x|x^T) = \int_{x^T-x}^{\infty} z(\varepsilon) d\varepsilon.
$$

As one would expect, for a given evidence threshold $x^T$, $\pi(x|x^T)$ is increasing in $x$ (the lower limit of integration falls). And, for any $x$, $\pi(x|x^T)$ is decreasing in $x^T$.

(Prospective) plaintiffs sue if and only if the expected transfers they would thereby receive, net of the expected transfers they would have to pay, exceed their costs of suit, that is, if and only if: $\pi(x|x^T)s^P - (1 - \pi(x|x^T))r^P - r^P > c^P$. This may conveniently be rewritten in terms of a critical probability of liability:

$$
(2) \quad \pi(x|x^T) > \frac{t^P + r^P + c^P}{s^P + t^P} \equiv \pi(x^*),
$$

where $x^*$ is the critical value of $x$ that generates the probability indicated by the fraction on the
right side of the inequality. Because, as remarked, \( \pi(x|x^T) \) is increasing in \( x \), this means that plaintiffs sue if and only if \( x > x^\ast \). Note that the value of \( x^\ast \) depends (implicitly) on four instruments — the three transfers to or from plaintiffs and the evidence threshold — and also on plaintiffs’ litigation costs. Observe that \( x^\ast \) does not depend on any of the defendants’ instruments (or on defendants’ litigation costs), although in cases without decoupling, e.g., when \( s^D = s^P = s \), there will be a connection with some of them.

Turn now to primary behavior. (Prospective) defendants will commit their harmful or benign acts, as the case may be, if and only if their private benefit from doing so exceeds the expected costs that arise in the event of suit, which are comprised of expected net transfers (those to be paid, minus the value of any to be received) as well as legal costs. This condition is:

\[
(3) \quad b > \lim_{x^\ast} \left[ \int_{x^\ast}^{\infty} \left( p^D + c^D + \pi(x|x^T) s^D - \left(1 - \pi(x|x^T) \right) t^D \right) g^i(x) dx \equiv b^i, \]

where \( b^i \) denotes the private benefit of the prospective actor who is just indifferent about whether to act. This integral is over all values of \( x \) for which a plaintiff would sue (for otherwise no costs are incurred). The term in brackets in the integrand reflects the net aggregate expected cost when a suit is filed. For convenience, individuals who refrain from committing acts of type \( H \) will be described as deterred, and those who refrain from acts of type \( B \) will be described as chilled. Observe that actors’ behavior is influenced by all of the plaintiffs’ instruments (as well as plaintiffs’ litigation costs) because, as explained with regard to expression (2), \( x^\ast \) implicitly depends on all of these.

Social welfare, \( W \), is taken to be the sum of individuals’ benefits from committing both types of acts minus the harm caused by acts of type \( H \) and litigation costs expected to be incurred when either type of act is committed. Accordingly, social welfare is given by:

\[
(4) \quad W = \lim_{b^\ast} \left[ \int_{b^\ast}^{\infty} \left( b - \left(1 - G^H(x^\ast)\right) (c^P + c^D) \right) f^H(b) db \right. \\
+ \left. \gamma \int_{b^\ast}^{\infty} \left( b - \left(1 - G^B(x^\ast)\right) (c^P + c^D) \right) f^B(b) db. \right]
\]

12The expression \( \pi(x^\ast) \) involves some abuse of notation because the pertinent \( x^\ast \) is conditional on \( x^T \) (which is relevant in expression 20, when differentiating with respect to \( x^T \)). Note that the right side of the inequality can equal or exceed one for some parameter values, meaning literally that suit arises only when probability of victory is in excess of one, which obviously holds for no value of \( x \), meaning that no one will sue. Also, without further restrictions on the transfers, the critical probability can be zero or negative, meaning that all plaintiffs would sue. For the most part, attention is confined to intermediate cases, that is, when \( \pi(x^\ast) \in (0,1) \).
The first term indicates all social benefits and costs with regard to individuals who commit harmful acts. These undeterred individuals, as indicated by expression (3), are those with benefits above the critical value, $b^I$. For each such act (now interpreting the expression in brackets in the integrand), there is the private benefit of the act, minus the harm it causes, and also minus the expected litigation costs associated with the act. To elaborate this latter expression, plaintiffs will sue when $x > x^*$ (implicitly determined by expression 2), which probability is given by one minus the cumulative distribution for the signal resulting from harmful acts, evaluated at $x^*$. And, when there is suit, both parties incur litigation costs. The second term, regarding benign acts, is analogous except that there is no direct harm associated with them. And this term is weighted by $\gamma$, the relative mass of benign acts.

Finally, although none of our instruments appear directly in expression (4) for social welfare, recall that all of them influence the values of the $b^I$ and that the plaintiffs’ instruments as well as the evidence threshold also determine $x^*$ and hence affect the expected litigation costs for each act committed.

C. Instrument Redundancy

To avoid a priori restrictions and to broadly encompass prior work, the model here admits six transfer instruments: $r^j, s^j, t^j, j = P, D$. This section explains that two of these are redundant in a mechanical, accounting sense, and that another may sometimes be dispensed with as a practical matter. (Note that the analysis in this section is largely independent of most of the model’s structure – such as primary behavior, parties’ signals, and court error – and hence is substantially more general.)

The former point can be illustrated in a number of ways. As a pure accounting matter, suppose, for example, that we increase $s^D$ by 1 and likewise reduce each of $r^D$ and $t^D$ by 1. A defendant pays 1 more (or, equivalently, receives 1 less) after trial, regardless of whether it loses or wins, and it pays 1 less (or, equivalently, receives 1 more) before trial, when the case is filed. Nothing really changes.

To understand this mechanical redundancy conceptually, observe that we have two parties, $P$ and $D$, and two outcomes, liability and no liability; hence, it is sufficient to specify four payoffs. Accordingly, we could take no liability as our baseline, define the two parties’ payoffs in that situation, and then further specify the difference in net outcomes when there is liability. For defendants, this depiction corresponds to rewriting the integrand in expression (3) as: $c^D + (r^D - t^D) + \pi(x|x^T)(s^D + t^D)$; the second term, $r^D - t^D$, could be reinterpreted as the defendant’s filing fee and the factor $s^D + t^D$ as its damages payment. Similarly, plaintiffs’ filing decision (the condition that generated expression 2) was $\pi(x|x^T)s^P - (1 - \pi(x|x^T))t^P - r^P > c^P$. Analogously, the left side can be written as $- (r^P + t^P) + \pi(x|x^T)(s^D + t^D)$, with the first term reinterpreted as the plaintiff’s filing fee and the second factor as its damages receipt. For each party, we have two

---

13An alternative way to view this is that, for each outcome with regard to liability, we need one instrument to indicate the transfer between the parties and another instrument to indicate the transfer between the parties as a group and the government.
payoffs determined by three parameters, rendering one of them redundant, meaning that any one could be fixed at any specified level (including zero) without limiting the social planner.

Furthermore, because this redundancy is mechanical, the additional instrument cannot relax real constraints. For example, assume that the maximum feasible \( s_D \) is given by \( s^{\text{max}} \), which is the level of prospective defendants’ wealth.\(^{14}\) Suppose that, instead of increasing \( s_D \), we raise \( r_D \) from zero (and increase \( t_D \) by a corresponding amount), which we can see has an identical effect on primary behavior. However, if a wealth-constrained defendant must now pay \( r_D \) up front, it has that much less wealth available to pay \( s_D \) in the event that it loses the case, so the wealth constraint is not effectively relaxed.\(^{15}\) Indeed, in the present model, there is no significance to whether any filing fee is paid at the time of filing or at the conclusion of the case.

Before proceeding, it is convenient to select as our four instruments the set \{\( r_P, s_P, s_D, t \)\}. To implement the four payoffs, one can set \( t \) to target \( D \)’s baseline payoff, then \( r_P \) to target \( P \)’s baseline payoff, and finally \( s_P \) and \( s_D \) to generate each party’s increment when there is liability.

Next, it is straightforward to demonstrate that another of the plaintiff’s instruments may be redundant in practice. Plaintiffs’ instruments are relevant only through their affect on plaintiffs’ filing decisions (they do not enter expression 3 except through their influence on \( x^* \)). These decisions, recall, are governed by

\[
(2) \quad \pi(x|x^T) > \frac{t_P + r_P + c^P}{s_P + t_P} \equiv \pi(x^*).
\]

Clearly, there is substantial freedom to target a given level of \( \pi(x^*) \) as long as there is a single unconstrained transfer instrument for plaintiffs. Specifically, we can eliminate the decoupling of damages payments in the following manner: raise or lower \( s_P \), as the case may be, until it equals \( s_D \) while, at the same time, raising or lowering \( r_P \) (in the same direction) to the degree that keeps \( x^* \) constant. When \( r_P \) may thus be adjusted, it is sufficient to analyze the three transfer instruments \{\( r_P, s, t \)\} — and the evidence threshold \( x^T \).\(^{16}\)

The foregoing discussion also casts new light on prior understandings of various instruments. Start with decoupling, which is to say allowing \( s_D \neq s_P \) (studied, for example, in

---

\(^{14}\)In similar spirit, one could consider a restriction on the amount a plaintiff could potentially pay, which would place a maximum on \( r^P \) and, as appropriate, on the sum of \( r^P \) and \( r^D \). See note 16 and section 4.C.

\(^{15}\)If there were some external norm or other factor that limited the sanction to a level strictly below defendants’ wealth but did not likewise constrain the functionally equivalent defendant filing fee, than this indirect technique could succeed.

\(^{16}\)When plaintiffs are risk neutral and there is no constraint on \( r^P \), which will be assumed in much of the analysis below, the statement in the text is complete. If there is a constraint on \( r^P \) (say, a limit to plaintiffs’ wealth), then permitting reductions in \( s^P \) offers an alternative manner of discouraging suits. If plaintiffs are risk averse, the analysis of behavior is largely the same (a different adjustment of \( r^P \) will be required when eliminating decoupling, but the same \( x^* \) can be maintained), but social welfare will differ due to risk-bearing costs: for example, raising \( s^P \) and \( r^P \) in a manner that holds \( x^* \) constant increases risk-bearing. Finally, note that this set of adjustments, while holding \( x^* \) constant, is favorable to plaintiffs with \( x > x^* \) (they have an expected gain, at the expense of the government’s budget).
Polinsky and Che (1991), or \( r^D \neq r^P \). Regarding the former, suppose that we start without decoupling, i.e., \( s^D = s^P = s \), and then contemplate raising or lowering just \( s^P \). The earlier discussion in this section indicates that, as an accounting matter, this is equivalent to instead introducing (allowing the adjustment of) a filing fee, \( r^P \) — and also adding and making a corresponding adjustment to \( t^P \). Moreover, the discussion just above shows how the latter is actually unnecessary since this way of eliminating decoupling only influences \( x^* \) (any effects on the \( b' \) are only through this channel) — and, as long as we are free to adjust \( r^P \), any level of \( x^* \) can be induced. Understanding decoupling as tantamount to allowing filing fees (or subsidies) on plaintiffs helps to unify analysis and also to make intuitions more transparent.

The present analysis can also usefully be related to fee shifting, which has received the most attention in various law and economics literatures (although usually without explicit analysis of primary behavior). Consider two-way fee shifting, such as under the English rule, and let us focus now on a basic setting without any decoupling. This rule is simply a special case of the present model in which \( s \) — ordinarily in private damages actions being taken to equal the harm suffered, \( h \) — is augmented by \( c^D \), and we also take \( t \) (relative to the ordinary benchmark of 0) to equal \( c^P \). Clearly, if \( s \) and \( t \) are not otherwise constrained to equal \( h \) and 0, respectively, there is no particular significance to fee shifting in the present sort of model. Only by chance would \( s = h + c^P \) and \( t = c^D \) be optimal. Moreover, if we started with \( s = h \) and \( t = 0 \), only by chance would moving toward fee shifting raise rather than lower social welfare. In addition, fee shifting is often thought to be problematic for two reasons: a party’s prospect of having its costs shifted to the other party reduces the marginal cost of expenditures, which makes litigation more expensive, and it can also be nontrivially costly in practice for the adjudicator to determine \( c^P \) and \( c^D \). See Katz (1987). Accordingly, if the main appeal of fee shifting is to improve plaintiffs’ filing incentives or primary behavior, then it may make more sense to adjust directly the outcome-contingent transfers, \( s \) and \( t \).

Finally, observe that, contrary to what is assumed in most prior work on the economics of litigation, it is natural to entertain the use of nontrivial filing fees (or equivalents, like decoupling) when assessing of a broad range of policies. Consider any change that reduces the costs of adjudication: because this may make suit more attractive, total costs could rise. However, if one can impose (or raise) \( r^P \) to keep filing behavior constant — substituting a transfer payment for a preexisting real resource cost — one might instead achieve an unambiguous improvement. As will now be seen, this degree of flexibility also has important implications for optimal system design in the present setting, which focuses on primary behavior as well as system costs.

---

\footnote{Also, fee shifting raises stakes, which increases incentives to spend on litigation, but so would adjusting \( s \) and \( t \) in an analogous fashion. In addition, in some models fee shifting will influence the likelihood of settlement (notably, the implicit increase in stakes makes a given degree of asymmetric information more likely to disrupt settlement), which also affects total system costs. But adjusting \( s \) and \( t \) directly would, again, have the same effects.}

\footnote{See also Kaplow (1993), where shifting victorious plaintiffs’ costs has the further disadvantage of (relatively) favoring filing by plaintiffs with higher litigation costs.}
3. Spanning and the First Best

The social welfare maximization problem — setting the instruments so as to maximize expression (4) — is concerned with controlling three margins of behavior, two concerning defendants’ primary activity (deterrence, \( b^H \), and chilling, \( b^B \)), and one concerning plaintiffs’ decisions whether to sue (\( x^* \)). We can view the planner’s problem as choosing individuals’ behaviors, subject to constraints, in order to maximize the objective function. In light of the preceding discussion, we will suppose that the available transfer instruments are \( r^P, s, \) and \( t \). In addition, the value of \( x^T \) is taken to be fixed: Because we have three targets, it is natural to consider whether three instruments are sufficient to implement any outcome.

To preview the construction, the first step will be to show that, for any given and admissible (see just below) values of \( s \) and \( t \), there exists an \( r^P \) such that any \( x^* \) can be supported. Next, given any value of \( x^* \), it will be shown that it is possible to find an \( s, t \) combination that supports any \( b^H \). As \( s \) or \( t \) is varied to target the designated \( b^H \), we will adjust \( r^P \) in the background so as to keep \( x^* \) fixed, recalling that, other than through \( x^* \), \( r^P \) has no affect on the \( b^H \). Finally, for a given \( \{ x^*, b^H \} \), we will vary \( t \) — adjusting \( s \) so as to keep \( b^H \) constant and, in light of the changes in both \( t \) and \( s \), adjusting \( r^P \) to keep \( x^* \) constant — to implement different levels of \( b^B \). As we will see, it is possible to drive \( b^B \) as low as we wish, but there will be a ceiling on how high a level of \( b^B \) can be implemented, for the given \( \{ x^*, b^H \} \). Hence, these three instruments allow substantial, although incomplete spanning of possible levels of the endogenous variables \( \{ b^H, b^B, x^* \} \). Despite the limitations, however, it is possible to implement an outcome arbitrarily close to the first best.

We begin with a choice of \( r^P \) to implement any given \( x^* \), taking as given a level of \( s \) and \( t \). From expression (1), any \( x^* \) (with a given \( x^T \)) is associated with a unique value of \( \pi(x^*) \), which, from expression (2) and our current instrument set is determined by

\[
\pi(x^*) = \frac{t + r^P + c^P}{s + t}.
\]

For the analysis to make sense in the present model, we will require the denominator, \( s + t \), to be positive. (Aside from this restriction, we will be allowing negative values of the instruments.) If the denominator were negative, this would mean that plaintiffs would prefer to lose. And, if this were so, then it would be low \( x \) cases that would be relatively attractive: Plaintiffs would sue if and only if \( x < x^* \). (If \( s + t = 0 \), they would be indifferent about the outcome and the value of \( x \) would be immaterial.) Accordingly, in the analysis to follow, we will impose the constraint \( s + t > 0 \).

For any admissible \( s \) and \( t \), we can induce \( \pi(x^*) = 0 \) by choosing \( r^P = -(t + c^P) \). In that event, plaintiffs would sue for any value of \( x \). Likewise, we can induce \( \pi(x^*) = 1 \) by choosing \( r^P = s - c^P \), in which event no one would sue. Because \( \pi(x^*) \) rises continuously with \( r^P \), we clearly can span the interval \([0,1]\) and thus induce any \( x^* \). In the remainder of the analysis it will simply be assumed that, whenever \( x^* \) is taken to be held constant but we are changing \( s \) or \( t \), \( r^P \) is
adjusted as implicitly determined by expression (5). As mentioned, because $r^p$ only influences the $b^p$ through its effect on $x^*$, there are no feedbacks on the $b^p$ that need to be taken into account. Also, for the remainder of the analysis, we will limit attention to finite values of $x^*$, that is, cases in which $\pi(x^*) \in (0,1)$. In particular, the arguments do not hold when $\pi(x^*) = 1$ because, when no one sues, both $b^i$ equal 0; adjusting $s$ or $t$ has no effect.

Next, consider our ability to implement any chosen level of $b^H$, taking $x^*$ as given in the manner just described. Using our current instrument set, and suppressing $x^T$ in the notation for convenience, we can use expression (3) to state

$$\begin{aligned}
(6) \quad b^i &= \int_{x^*}^{\infty} \left[ c^D + \pi(x)s - (1 - \pi(x))t \right] g^i(x) dx. \\
& \text{Starting from a given } r^p, s, \text{ and } t, \text{ expression (6) determines a value of } b^H. \text{ (It also implies a value of } b^B, \text{ which will be ignored in this segment of the analysis.) From that initial point, consider increasing } s. \text{ Clearly } b^H \text{ rises, and at a constant rate. (The pertinent derivative is the integral of } \pi(x)\text{ from } x^* \text{ to infinity, weighted by the density } g^H(x), \text{ which as a whole is positive and independent of all the instruments, given that } r^p \text{ is adjusted so as to keep } x^* \text{ constant.) Accordingly, we can implement as high a level of } b^H \text{ as we wish (any level above its initial value). (Keep in mind that in this section we have imposed no extrinsic constraints on } s \text{ or the other instruments.) To achieve lower values of } b^H, \text{ we can correspondingly reduce } s. \text{ However, in this instance we have our constraint that } s + t > 0, \text{ so it might not be possible to drive } b^H \text{ down to 0. Nevertheless, if we cannot, we can instead fix } s \text{ and increase } t: \text{ this will decrease } b^H \text{ at a constant rate, allowing us to span all values down to zero. Note that, because we are raising } t \text{ in this instance, we are not hampered by the requirement that } s + t > 0.}
\end{aligned}$$

The third part of the construction is more involved. We will start with a given $x^*$ and $b^H$, supported by some given $r^p$, $s$, and $t$, which, from expression (6), implies some value of $b^B$. We now will consider an experiment (indicated notationally at various points by “exp”) that attempts to manipulate $b^B$ by adjusting $t$. We will need, however, simultaneously to adjust $s$ so as to keep $b^H$ fixed because changes in $t$ also influence $b^H$. (As before, we will also adjust $r^p$ in the background to keep $x^*$ fixed; this can be set to the side because, as mentioned, this does not feed back on the $b^i$.) Because we need to know the properties of this adjustment in $s$, we derive it explicitly, as follows:

$$\begin{aligned}
(7) \quad \frac{db^H}{dt} \text{ }_{\text{exp}} &= \int_{x^*}^{\infty} \left[ \pi(x) \frac{ds}{dt} \text{ }_{\text{exp}} - (1 - \pi(x)) \right] g^H(x) dx = 0. \\
& \text{This implies:}
\end{aligned}$$
Note further that the value of this derivative is constant, given that we are holding $x^*$ constant.

Like $b^H$, $b^B$ also rises in $s$ and falls in $t$. Accordingly, our experiment — which involves, let us say, increasing $t$, which implies from expression (8) that $s$ also increases — will have conflicting effects on $b^B$. We will now demonstrate that increasing $t$ in this experiment causes $b^B$ to fall. The basic intuition is that increasing $t$ is relatively favorable to type $B$ (benign) acts whereas increasing $s$ is relatively unfavorable to type $H$ (harmful) acts. And, because our experiment is designed to keep $b^H$ constant, both of these considerations indicate that raising $t$ is net favorable to $b^B$, meaning that it (chilling) falls. Before proceeding to prove this claim, note that raising $t$, which also entails raising $s$, poses no difficulty with our constraint that $s + t > 0$. Hence, our result implies that, from the initial value of $b^B$, we will be able to drive it down as much as we wish. But, as we shall see, to raise $b^B$ requires reducing $t$, and also $s$, so at some point the constraint will bind; hence, there will be an upper limit on how high of a $b^B$ we can implement for our given $x^*$ and $b^H$.

To further preview the analysis, when comparing relevant terms below, it will be useful to normalize the two density functions, the $g_i(x)$, so that these normalized densities each have the same mass above $x^*$. Then we will take advantage of the fact that the density associated with $H$ is an upward shift of that associated with $B$ (the monotone likelihood ratio property implies first-order stochastic dominance).

Accordingly, let us first explain why it is true that

$$\left. \frac{ds}{dt} \right|_{\text{exp}} = \frac{\int_{x^*}^{\infty} (1 - \pi(x)) g^H(x) dx}{\int_{x^*}^{\infty} \pi(x) g^H(x) dx} > 0.$$ 

The leading derivative on each side, as indicated by expression (8), is positive. With the stated normalizations, note that on both sides we are integrating, with regard to the stated fractions, over the same total mass. Furthermore, as mentioned just above, the normalized density on the left side, $g^H(x)/(1 - G^H(x^*))$, is an upward shift relative to that on the right side, $g^B(x)/(1 - G^B(x^*))$. Finally, $\pi(x)$ is positive and increasing in $x$. Therefore, the inequality holds.

Extending this logic, we can also show that
(10) \[ \int_{x^*}^{\infty} (\pi(x) - 1) \frac{g^H(x)}{1 - G^H(x^*)} \, dx > \int_{x^*}^{\infty} (\pi(x) - 1) \frac{g^B(x)}{1 - G^B(x^*)} \, dx. \]

Here, the term \( \pi(x) - 1 \) is negative, but it is also increasing in \( x \). Because the normalized density on the left is an upward shift, it places more weight on the less negative values, so we again have the same inequality. \(^{19}\)

Next, let us sum the two terms on the left sides of expressions (9) and (10) and compare that total to the sum of the two terms on the right sides:

\[ \frac{ds}{dt} \bigg|_{\exp,x^*} \int_{x^*}^{\infty} \pi(x) \frac{g^H(x)}{1 - G^H(x^*)} \, dx + \int_{x^*}^{\infty} (\pi(x) - 1) \frac{g^H(x)}{1 - G^H(x^*)} \, dx \]

\[ > \frac{ds}{dt} \bigg|_{\exp,x^*} \int_{x^*}^{\infty} \pi(x) \frac{g^B(x)}{1 - G^B(x^*)} \, dx + \int_{x^*}^{\infty} (\pi(x) - 1) \frac{g^B(x)}{1 - G^B(x^*)} \, dx. \]

Comparison of the left side of inequality (11) with expression (7) indicates that the former simply equals \( (db^H/dt)_{\exp}(1 - G^H(x^*)) \). Likewise, the right side of inequality (11) equals \( (db^B/dt)_{\exp}(1 - G^B(x^*)) \). Therefore, we can write:

\[ (12) \frac{1 - G^B(x^*)}{1 - G^H(x^*)} \frac{db^H}{dt} \bigg|_{\exp} > \frac{db^B}{dt} \bigg|_{\exp}. \]

Finally, recall that our experiment adjusts \( s \) as we increase \( t \) so as to hold \( b^H \) constant. In other words, the left side of expression (12) equals 0. Hence, the right side is negative, which is to say that \( b^B \) falls as \( t \) rises (and \( s \) and \( r^P \) are adjusted as specified). Furthermore, from inspection of the right side of expression (11), it is apparent that the magnitude of this derivative is constant (\( x^* \) is held constant; neither \( s \) nor \( t \) appears anywhere; and, from expression 8, \( (ds/dt)_{\exp} \) is constant). Therefore, as asserted above, we can raise \( t \) to achieve any level of \( b^B \) from its initial level to as low as we would like (keeping in mind that, because we are raising both \( t \) and \( s \), starting from initial, feasible levels, the constraint that \( s + t > 0 \) does not bind).

And, as mentioned before, we can also lower \( t \) (and \( s \), etc.), to achieve higher levels of \( b^B \), but in this instance, the constraint that \( s + t > 0 \) will at some point bind. (In addition, reflection on this analysis should be sufficient to indicate that, if we could also vary the instrument \( x^T \), we

\(^{19}\)Alternatively, note that the “-1” terms on each side, when integrated, are equal and thus cancel, leaving the \( \pi(x) \) terms that we had before.
To offer a concrete illustration of these points, suppose that we have $x^* = 0$, that initially $t = 0$, and we take a positive value of $s$ that is arbitrarily close to 0. (For the moment, hold $x^T$ fixed.) This implies that $b_H > b_D (1 - G_{H0}(0))$. Likewise, we have an initial value of $b_B = c_B (1 - G_{B0}(0))$, which, note, is lower than $b_H$. How much can we raise $b_B$, keeping in mind that we must keep both $x$ and $t$ constant? To raise $b_B$, we need to lower $t$, and as per our experiment, $s$ as well. But we are bound by the constraint that $s + t > 0$, which from our initial conditions is almost binding already. Hence, this initial value of $b_B$ is the approximate upper bound. Finally, suppose that we could also adjust $x^T$ (making other adjustments that hold $b_H$ and $x^*$ fixed). Changing $x^T$ in one direction or the other would, in general, enable a further increase in $b_B$. But since both $s$ and $t$ are so close to zero, this would make little difference. (Suppose that, instead of starting with $s$ and $t$ at or near zero, we instead started with any $x^T$, which might be chosen as high or as low as we might like. That, actually, was the original construction employed in this section, where $x^T$ was taken as given but no restriction was imposed. We saw that, for a designated (finite) $b_H$, there is a limit on how high a $b_B$ can be implemented. To say something further about this ceiling, analysis of expression (6) for the $b_i$ indicates that $t < c_B$ is a sufficient condition for $b_B < b_H$. And, as we have seen, raising $t$ in our experiment, which also involves raising $s$ (and so forth) entails lowering $b_B$.)

Combining these conclusions, we can appreciate the limits to spanning. Given $b_H$ and $x^*$, there is a limit to how much we can raise $b_B$. Note that, the higher is our initial $b_H$, the higher will be the ceiling on $b_B$. Likewise, given a $b_B$ and $x^*$, there may be a limit to how much we can reduce $b_H$, and the lower is our initial $b_B$, the lower will be the floor on $b_H$. These limitations derive from two assumptions: we are restricting attention to situations in which stronger cases are those plaintiffs prefer to win — hence the constraint that $s + t > 0$ — and we are maintaining the strict monotone likelihood ratio property, which means that stronger evidence is relatively more often generated by harmful acts than by benign acts. As a consequence, we cannot simultaneously implement an arbitrarily low $b_H$ and an arbitrarily high $b_B$. To summarize:

**Proposition 1:** If there are no extrinsic constraints on plaintiff filing fees, damage awards, and payments from losing plaintiffs to winning defendants — the instruments $r^p, s, \text{and } t$ — then:

a. It is feasible to implement any value of $x^*$ (a critical filing threshold) and any nonnegative value of $b_H$ (level of deterrence) and, for those values, any nonnegative value of $b_B$ (degree of chilling) up to some upper limit.

b. It is feasible to implement any value of $x^*$ and any nonnegative value of $b_B$ and, for those values, any nonnegative value of $b_H$ down to some lower limit (which might be 0).

---

20To offer a concrete illustration of these points, suppose that we have $x^* = 0$, that initially $t = 0$, and we take a positive value of $s$ that is arbitrarily close to 0. (For the moment, hold $x^T$ fixed.) This implies that $b_H = c_B (1 - G_{H0}(0))$. Likewise, we have an initial value of $b_B = c_B (1 - G_{B0}(0))$, which, note, is lower than $b_H$. How much can we raise $b_B$, keeping in mind that we must keep both $x^*$ and $b_H$ constant? To raise $b_B$, we need to lower $t$, and as per our experiment, $s$ as well. But we are bound by the constraint that $s + t > 0$, which from our initial conditions is almost binding already. Hence, this initial value of $b_B$ is the approximate upper bound. Finally, suppose that we could also adjust $x^T$ (making other adjustments that hold $b_H$ and $x^*$ fixed). Changing $x^T$ in one direction or the other would, in general, enable a further increase in $b_B$. But since both $s$ and $t$ are so close to zero, this would make little difference. (Suppose that, instead of starting with $s$ and $t$ at or near zero, we instead started with any $x^T$, which might be chosen as high or as low as we might like. That, actually, was the original construction employed in this section, where $x^T$ was taken as given but no restriction was imposed. We saw that, for a designated (finite) $b_H$, there is a limit on how high a $b_B$ can be implemented. To say something further about this ceiling, analysis of expression (6) for the $b_i$ indicates that $t < c_B$ is a sufficient condition for $b_B < b_H$. And, as we have seen, raising $t$ in our experiment, which also involves raising $s$ (and so forth) entails lowering $b_B$.)

21Consider a simple variation of the example in the preceding footnote.
In this world in which the instruments \( r^p \), \( s \), and \( t \) are unconstrained, except for the requirement that \( s + t > 0 \) (which is necessary for plaintiffs to prefer to win and thus prefer to bring stronger rather than weaker cases), we can achieve substantial spanning. Moreover, with regard to the maximization of social welfare, it would seem that we may be in good shape because our only limits were on making \( b^g \) (chilling) high or \( b^H \) (deterrence) low. But since, in cases of greatest interest, chilling reduces welfare and deterrence raises welfare, these limits may not be of concern. Indeed, it is straightforward from the foregoing to demonstrate:

**Proposition 2:** If there are no extrinsic constraints on plaintiff filing fees, damage awards, and payments from losing plaintiffs to winning defendants — the instruments \( r^p \), \( s \), and \( t \) — it is feasible to implement a result arbitrarily close to the first best. Specifically, one can induce \( b^H = h \) (first-best deterrence), \( b^g = 0 \) (first-best chilling, which is to say, none), and \( x^* \) arbitrarily high (implying negligible suits and hence negligible litigation costs).

Note that, in an ideal world, three things would be true: (1) There would be no litigation, so no litigation costs would be incurred. (2) Individuals with acts of type \( H \) would commit them if and only if \( b > h \). (Note that, when there is no litigation, we need not add the expected litigation costs associated with an act on the right side of this inequality, which had been included in expression (4) for social welfare.) (3) Individuals with acts of type \( B \) would commit them if and only if \( b > 0 \). (In the ideal world, we can ignore litigation costs here as well.)

The proof of Proposition 2 is sufficiently straightforward that a few remarks should suffice. First, it has been demonstrated that one can implement any \( x^* \) and \( b^H \), so we can pick \( x^* \) as high as we like and, at the same time, implement \( b^H = h \) by choosing a sufficiently large \( s \). Moreover, this can be accomplished even if we restrict \( t \) to equal 0. Although previously it was possible that we might need to raise \( t \) above 0 to reduce \( b^H \), once we may choose \( x^* \) as high as we wish, this is unnecessary because, for finite \( s \) and \( t = 0 \), \( b^H \) approaches 0 as we increase \( x^* \).

Second, for this instrument setting, \( b^g \) will be positive. Hence, we can now raise \( t \) (and correspondingly raise \( s \) so as to keep \( b^H \) fixed at \( h \)) to reduce \( b^g \) as much as we like, specifically, until \( b^g = 0 \). Therefore, we can indeed implement the stated levels for the \( b^i \).

Third, by taking the initial \( x^* \) ever higher, we can drive litigation costs arbitrarily close to zero — now taking as given the specified primary behavior, which can be maintained by appropriately increasing \( s \) and \( t \). (Note that our constraint that \( s + t > 0 \) does not interfere with this construction.)

It is interesting that this ability to approximate the first best holds in what is in some respects a challenging environment. There are errors in adjudication, of both types. And we care not just about controlling harmful acts (deterrence), as in nearly all prior work on optimal legal system design, but also about avoiding the chilling of benign acts. In addition, enforcement effort is endogenous, so we must control plaintiffs’ filing decisions. Finally, note that our solution can be implemented even while holding the evidence threshold, \( x^T \), fixed at an arbitrary initial level.
4. Optimal System Design with Instrument Restrictions

Section 3 indicates that the instrument set \{r^*, s, t\} is sufficient to span much of the space of the endogenous behavioral variables, \(x^*, b^H, \) and \(b^B\), and to achieve a result arbitrarily close to the first best. However, this result was derived assuming that there were no constraints on any of the instruments (but for the supposition that \(s + t > 0\), which implies that plaintiffs prefer to win their cases). This section is concerned with the implications of instrument restrictions and, in particular, how the level of damages paid by losing defendants, \(s\), the transfer paid by losing plaintiffs, \(t\), and the evidence threshold, \(x^T\), are each optimally set when the other two are taken to be fixed (at levels that need not be optimal).

Most of the discussion (until section C) will be conducted under the assumption that it is possible to control \(x^*\) directly — that is, without influencing actors’ behavior with regard to the two types of acts except through \(x^*\) itself. As discussed in section 2.C, this can be done in various ways. Here, continuing to focus on the subset of transfer instruments \{\(r^*, s, t\)\}, we will assume that \(r^*\) may be set freely. Accordingly, when we vary other instruments below, we will also be adjusting \(r^*\) so as to move \(x^*\) in the prescribed manner. As we shall see, a number of surprising results arise.

A. Analysis

The three propositions proved here are that extreme values of each of the three instruments are optimal; specifically, \(s\) should be as high as possible, and both \(t\) and \(x^T\) should be as low as possible. Each result is demonstrated with the same sort of experiment. First, it will be supposed that the instrument under investigation is not at its maximum or minimum feasible value, as the case may be. Second, the instrument will be moved slightly in the indicated direction while simultaneously adjusting \(x^*\) (by the requisite adjustment of \(r^*\)) so as to keep \(b^B\) constant. That is, the overall expected burden on benign acts will be held constant, so that chilling is unchanged. Third, in each case it will be straightforward to show that this experiment involves a higher \(x^*\) and thus, on that account, lower litigation costs, which raises social welfare. Fourth, through a more elaborate argument, it will be demonstrated that the experiment also necessarily increases deterrence, which is to say, raises \(b^H\).

---

22In practice, negative values of \(r^*\) may be problematic due to frivolous suits or collusion. A partial remedy would be to pay any filing subsidy only at the end of trial (i.e., to alter the post-trial payoffs in the manner described in section 2.C). Whether an unconstrained optimum will involve a filing subsidy depends, of course, on all the parameters of the model. Note, however, that the results in Propositions 3–5 (and also Proposition 2 in section 3) involve raising \(r^*\) to values that may be extremely high, in which case the present concern would be moot (although one might instead face a plaintiff wealth constraint, as discussed in note 16 and in section 4.C, below).

23The reader may wonder why a seemingly easier-to-analyze experiment, which holds \(x^*\) fixed, was not undertaken. To be sure, that experiment holds the direct contribution to litigation costs constant and, as will become clear, also zeros out one of the channels by which changing the other instruments influences both deterrence and chilling (because, from expressions 14, 19, and 20 below, one of the two channels by which instruments otherwise influence the commission of acts in this setting is through their effect on \(x^*\), the lower limit of integration). However, this value of \(x^*\) — which is lower than that implied by the experiment undertaken in the text — implies that chilling rises, and one cannot typically derive unambiguous results regarding social welfare even if it can be demonstrated that deterrence increases to a greater extent.
assumption that will be conditionally maintained here that, at the optimum, greater deterrence, ceteris paribus, would raise social welfare, this increase in deterrence implies that social welfare rises. (An increase in social welfare can also be demonstrated in a similar fashion under the alternative assumption that, at the optimum, greater chilling is welfare reducing.) Finally, it should be noted that attention is confined to optima in which some suits are filed (otherwise marginal adjustments to the instruments have no effects) and where defendants prefer that the marginal suit not be filed (see the discussion following expression 14). These assumptions are revisited in section B.

Proposition 3: If the optimum involves some suits \( \pi(x^*) < 1 \), defendants prefer that the marginal suit not be filed, and deterrence is welfare increasing (or chilling is welfare reducing) at the margin, then the optimal level of damages, \( s \), cannot be interior. Specifically, it would be possible to raise \( s \) and raise \( r^p \) such that welfare increases, so that the optimal \( s \) equals \( s^{Max} \), if such a maximum exists.

Proof: Suppose, in contradiction to the claim in the Proposition, that we have \( s < s^{Max} \), and, moreover, there are some suits, defendants disprefer marginal suits, and deterrence is welfare increasing. Our experiment is to raise \( s \) while adjusting \( r^p \) so as to keep \( b^B \) fixed. The effect of this experiment (denoted again by “exp”) on social welfare (4) is given by:

\[
\frac{dW}{ds}\bigg|_{exp} = \frac{db^H}{ds}\bigg|_{exp} \left( h + (1 - G^H(x^*)) (c^p + c^D) - b^H \right) f^H(b^H) \\
+ \int_{b^H}^{\infty} g^H(x^*) \left( \frac{dx^*}{ds}\bigg|_{exp} \right) (c^p + c^D) f^H(b) db \\
- \gamma \frac{db^B}{ds}\bigg|_{exp} \left( b^B - (1 - G^B(x^*)) (c^p + c^D) \right) f^B(b^B) \\
+ \gamma \int_{b^B}^{\infty} g^B(x^*) \left( \frac{dx^*}{ds}\bigg|_{exp} \right) (c^p + c^D) f^B(b) db.
\]

The first term on the right side shows the effect on welfare due to the induced change in deterrence. For a unit increase in deterrence, there is a net social gain for each act deterred that has three components (corresponding to the three terms in square brackets): the harm avoided, the savings in expected adjudication costs (which would have been incurred had the act instead been committed), and a forgone benefit (which, note, has a value equal to \( b^H \) because it is the marginal act that is deterred). This net effect is weighted by the value of the density function of
benefits from harmful acts, evaluated for the marginal act, indicating how many such acts are deterred for a unit increase in the expected burden on such acts. We will return to the question of the sign of the key derivative, \( (db^H/ds)_{\text{exp}} \).

The third term, the effect on welfare due to increased chilling, is analogous to the first term, just described. However, by construction, \( (db^B/ds)_{\text{exp}} = 0 \), so this term equals zero.

The second term indicates the savings in social costs relating to suits involving harmful acts due to an increase in \( x^* \) (if such occurs; if \( x^* \) falls, this will be a social cost). For all harmful acts committed (hence the integral), the change in suits is given by the product of three factors: the frequency with which harmful acts emit the signal \( x^* \) (indicating how much suits change for a given change in \( x^* \)), the change in \( x^* \) (which we will sign momentarily), and the total costs per suit. The fourth term, regarding changes in social costs relating to suits involving benign acts, is analogous. Note in particular that its sign is the same as that for the second term, both depending on the sign of \( (dx*/ds)_{\text{exp}} \).

Next, we can inquire into how this experiment influences behavior, specifically, the \( b^i \). To be sure, we know that it does not affect chilling, but by examining the chilling effect and setting it equal to zero, we can proceed to sign both \( (dx*/ds)_{\text{exp}} \) and \( (db^H/ds)_{\text{exp}} \). This behavioral effect, determined by differentiating expression (6), is given by:

\[
(14) \quad \frac{db^i}{ds}_{\text{exp}} = -\frac{dx^*}{ds}_{\text{exp}} \left[ c^D + \pi(x^*)s - (1 - \pi(x^*))t \right] g^i(x^*) + \int_{x^*}^{\infty} \pi(x) g^i(x) dx.
\]

The first term indicates that, for a given increase in \( x^* \) due to the experiment, \( (dx*/ds)_{\text{exp}} \), the \( b^i \) will fall, which is to say that behavior will be encouraged. To the extent that certain suits (those near the threshold \( x^* \)) are no longer filed, actors will face lower overall expected costs: they will no longer incur litigation costs, \( c^D \), in those cases, and they will no longer be subject to the prospect of paying \( s \) in those cases, which has probability \( \pi(x^*) \); they also forgo the prospect of receiving the transfer \( t \), but the Proposition’s assumption that defendants disprefer the marginal suit guarantees that this factor is not dominant. In addition, the number of suits that no longer are brought for a unit increase in \( x^* \) is given by the density for the signal, evaluated at \( x^* \). The second term indicates that, for the postulated increase in \( s \), all suits that are filed in any event (corresponding to values of \( x \) above \( x^* \)) now have a higher sanction being applied, which sanction is imposed, for a given \( x \), with probability \( \pi(x) \).

Now, if the experiment had simply increased \( s \) and left \( r^o \) unchanged, we know that the \( b^i \), and, in particular, \( b^B \), would rise on both accounts: from expression (2), a higher \( s \), ceteris paribus, reduces \( \pi(x^*) \) and thus the value of \( x^* \), making the first term in expression (14) positive, and the second term is positive, as just explained. Therefore, if our experiment is to hold \( b^B \) constant, it must raise \( x^* \). Stated more fully, it must be that \( r^o \) is elevated by enough not only to offset the reduction in \( x^* \) caused by the increase in \( s \) (which would make the first term zero, leaving the positive second term), but also to actually raise \( x^* \), by just enough to generate a negative first term with a magnitude equal to that of the (positive) second term. Accordingly, we
have proved that the experiment results in positive values for the second and fourth terms in expression (13) for the change in social welfare.

Finally, we need to determine the effect of the experiment on $b^H$. Specifically, we will show that $(db^H/ds)_{\text{exp}} > 0$. The argument makes use of the monotone likelihood ratio property for the densities of the signals. The core idea is that, as a consequence of this property, the influence of raising $s$ on deterrence due to its direct effect on expected sanctions will be relatively greater for harmful acts than for benign acts; hence, deterrence rises relative to chilling; and, since chilling is unchanged, deterrence therefore rises absolutely. The subtlety concerns the emphasized term “relatively,” which can be made explicit through a particular normalization, specifically, with respect to the values of the densities, $g'(x)$, when each are evaluated at $x^*$.

This suggestion can be made concrete by returning to the mathematics, working backward in a sense. To begin, consider the validity of the following inequality:

$$\int_{x^*}^{\infty} \left[ \frac{g^H(x)}{g^B(x)} \right] \pi(x) g^B(x) dx > \int_{x^*}^{\infty} \pi(x) g^B(x) dx.$$

The term in brackets on the left side is the likelihood ratio evaluated at $x$ divided by the likelihood ratio evaluated at $x^*$. When this term is evaluated at the lower limit of integration, $x^*$, it equals 1. For all higher values of $x$ — that is, for the values over which we are integrating — the strict monotone likelihood ratio property indicates that this term exceeds 1. Furthermore, $\pi(x)$ and $g^B(x)$ are both positive. Hence, the integrand on the left side exceeds that on the right side for all values of $x$ (except at the lower limit of integration), which establishes the inequality.

We can now rearrange expression (15), multiplying both sides by $1/g^B(x^*)$ and cancelling the $g^B(x)$'s on the left side to yield:

$$\int_{x^*}^{\infty} \pi(x) \left[ \frac{g^H(x)}{g^H(x^*)} \right] dx > \int_{x^*}^{\infty} \pi(x) \left[ \frac{g^B(x)}{g^B(x^*)} \right] dx.$$
Finally, multiply both sides by $g^H(x^*)$ and, on the right side, factor out $1/g^B(x^*)$, to yield:

\[
- \frac{dx^*}{ds} \bigg|_{\text{exp}} \left[ c^D + \pi(x^*)s - (1 - \pi(x^*))t \right] \pi(x) \left[ \frac{g^H(x)}{g^H(x^*)} \right] dx > \left( \frac{g^H(x^*)}{g^B(x^*)} \right) \left( - \frac{dx^*}{ds} \bigg|_{\text{exp}} \left[ c^D + \pi(x^*)s - (1 - \pi(x^*))t \right] g^B(x^*) + \int_{x^*}^{\infty} \pi(x) g^B(x) dx \right).
\]

Expression (18) establishes our result, which can be seen as follows: First, the left side is $(db^H/ds)|_{\text{exp}}$, as can be seen from expression (14). Likewise, the term in large parentheses on the right side of expression (18) is $(db^B/ds)|_{\text{exp}}$. Moreover, our experiment sets this value equal to zero, so the right side of expression (18) equals zero (so we can ignore the scaling ratio that precedes the term in large parentheses). Therefore, $(db^H/ds)|_{\text{exp}} > 0$.

(The argument is complete if deterrence is welfare increasing at the margin. If it is not, then Proposition 3’s alternative condition must hold: chilling is welfare reducing at the margin. In this case, simply alter the experiment such that, instead of having $(db^B/ds)|_{\text{exp}} = 0$, we adjust $r^p$ such that $(db^H/ds)|_{\text{exp}} = 0$. In that case, everything follows until the last sequence of analysis, which, when we now make the appropriate substitutions from expression (14) into expression (18), we have the left side of expression (18) equal to zero and hence the right side is negative. Therefore, $(db^B/ds)|_{\text{exp}} < 0$, which completes the argument for this alternative premise.)

To summarize, we have shown that the experiment of raising $s$ while adjusting $r^p$ so as to keep $b^B$ constant raises $b^H$ (deterrence) along with reducing the level of suits (because holding $b^B$ fixed when $s$ is increased requires increasing $r^p$ sufficiently to increase $x^*$, which means that plaintiffs need a stronger signal for suit to be profitable).

Proposition 3, indicating that it is optimal to raise damages (akin to the sanction in cases with public enforcement) to its maximum feasible level is reminiscent of the result based on

---

24When proving Propositions 4 and 5 below, this final argument will not be repeated.
Becker (1968) that one can raise the sanction and reduce enforcement effort so as to maintain deterrence and save enforcement costs. The present model differs in having private enforcement (making enforcement effort endogenous) but also in a number of other respects, the most important of which is the introduction of two types of error (specifically, including the erroneous imposition of sanctions) and, relatedly, of benign activity that may be chilled as a consequence. Nevertheless, the same result is obtained, although the just-mentioned extensions make the argument much more involved.

We next turn to the other instruments, \( t \) and \( xT \), which have received very little attention, and none in this sort of setting. Here, one might suppose that the raising the values of these instruments would be beneficial — or at least have a distinctive beneficial effect — as a consequence of lightening the burden on benign acts. Of course, reducing \( s \) also lightens that burden. The difference is that, given differential strengths of different cases (indicated by the magnitude of \( x \)), a plausible conjecture would be that raising \( t \) and \( xT \) can be advantageous because each would seem to favor benign acts relative to harmful acts. As we shall see, however, in the present framework such is not the case. Indeed, the opposite is true.

**Proposition 4:** If the optimum involves some suits (\( \pi(x^*) < 1 \)), defendants prefer that the marginal suit not be filed, and deterrence is welfare increasing (or chilling is welfare reducing) at the margin, then the optimal payment from losing plaintiffs to winning defendants, \( t \), cannot be interior. Specifically, if it were, it would be possible to reduce \( t \) and raise \( rP \) such that welfare increases, so that the optimal \( t \) equals 0.

**Proof:** The argument follows closely that given for Proposition 3, so the exposition will be abbreviated. Suppose, in contradiction to the claim in Proposition 4, that we have \( t > 0 \). Our experiment is to lower \( t \) while adjusting \( rP \) so as to keep \( bB \) fixed. However, because derivatives are understood to relate to increases in a variable (and repetitive reminders to reverse signs would be tedious), let us instead analyze an experiment that increases our variable of interest, \( t \), while adjusting \( rP \) so as to keep \( bB \) fixed. We will show that this experiment unambiguously reduces social welfare, implying that the reverse experiment, in which we indeed lower \( t \), raises welfare.

The effect of this experiment on social welfare (4) is given by an expression identical to expression (13) above, except that each derivative with respect to \( s \) is replaced by one with respect to \( t \). Accordingly, we can proceed directly to analyze how the experiment influences behavior.

Differentiating expression (6), we have:

\[
\left( \frac{db^I}{dt} \right)_{|exp} = -\left. \frac{dx^*}{dt} \right|_{exp} \left[ e^D + \pi(x^*) s - (1 - \pi(x^*)) t \right] g^I(x^*) - \int_{x^*}^\infty (1 - \pi(x)) g^I(x) dx.
\]

Expression (19) differs from expression (14) in a number of respects: The derivatives here are with respect to \( t \), of course. The second term is subtracted rather than added because raising \( t \)
rewards actors whereas raising $s$ punishes them. Finally, the integrand in the second term is $1 - \pi(x)$ rather than $\pi(x)$. There is, however, another key difference in substance: the direct effect of raising $t$ on $x^*$ is to increase it (whereas when raising $s$, the direct effect is to decrease $x^*$).

That is, raising the amount that losing plaintiffs pay to defendants makes suit less attractive, so the marginally viable plaintiff suit will need to have a stronger signal $x$.

Combining these points, the direct effect of raising $t$ is to reduce $b^H$ on account of both terms in expression (19). Therefore, if we wish to hold $b^H$ constant, we need more suits: a lower $x^*$, which is implemented by reducing $r^P$. As a consequence, this experiment raises litigation costs (which is to say, the terms analogous to the second and fourth ones in expression 13 are negative).

To determine the effect on $b^H$, we proceed as before. Because the mathematical manipulations are virtually identical, they are omitted. The key difference concerns the analogue to inequality (15). The corresponding inequality indeed holds. (It suffices to observe that, just as $\pi(x)$ being positive was used to establish the previous inequality, the fact that $1 - \pi(x)$ is positive can be used to establish the analogous inequality here.) However, here we wish to be characterizing the second term in expression (19), which is subtracted (whereas the corresponding term in expression 14 was added). Hence, to implement the preceding series of manipulations (that ultimately must mimic, on each side of the inequality, what is now expression 19), we need first to multiply both sides by negative one. This reverses the inequality. Therefore, when the corresponding analysis is complete, it shows that the experiment of increasing $t$ while holding $b^H$ constant causes deterrence ($b^H$) to fall.

Taken together, the experiment that increases $t$ raises litigation costs and reduces deterrence, both of which are welfare decreasing. Hence, the reverse experiment, involving a reduction in $t$, raises social welfare.$\blacksquare$

Proposition 4’s indication that it is optimal to reduce the amount that losing plaintiffs pay defendants as much as possible may seem surprising. Indeed, had we not restricted $t$ to be nonnegative, the foregoing argument indicates that $t$ should be as negative as possible: specifically, it should approach $-s$ (recall the constraint in section 3 that $s + t > 0$). In other words, it would be optimal for winning defendants to pay losing plaintiffs. Such a result is even more astonishing when one considers the fact that the present model, unlike most prior work (including what little allows for transfers when defendants win), introduces benign conduct that tends to be chilled as a side-effect of the attempt to regulate harmful acts. This problem would seem to make attractive a requirement that losing plaintiffs pay winning defendants, for two reasons: such payments occur more often in weak cases, which are more likely to involve benign acts, and the prospect of having to make such payments might be imagined to relatively discourage the filing of weak cases. How can the contrary conclusion be explained?

The core intuition actually is quite simple: Plaintiffs’ filing decisions, out of their own self-interest, result in their selection of the strongest cases. As will be elaborated when discussing Proposition 5, just below, the filing cutoff $x^*$ serves as a de facto evidence threshold (a.k.a., burden of proof; see note 11). Hence, any sort of transfer, including $t$, which is a transfer
when the defendant wins, occurs only in suits with high values of $x$, specifically, ones in excess of $x^*$. Note further that, because our experiment, when reducing $t$, also holds $b^p$ constant, it necessarily involves a higher $x^*$ (which is implemented by increasing $r^p$), as the proof indicates.

The consequences of these two features for benign acts are as follows: First, although our experiment has defendants receive less when they win — or, in the seemingly whimsical extension, paying when they win — this occurs only in filed cases, which are the stronger ones, those for which $x > x^*$. Second, the experiment simultaneously benefits defendants by having fewer suits filed. Moreover, this reduction in filings, due to the increase in $x^*$, involves eliminating the weakest cases from among those previously filed. Taken together, therefore, the experiment is relatively favorable to defendants who committed benign acts rather than harmful acts. Our preliminary intuition that a lower $t$ has the opposite relative effect ignores that the second component of the experiment operates not merely in the reverse direction, but to an even greater degree. Defendant’ gains from the experiment arise in cases ($x = x^*$) that are weaker than all of those ($x > x^*$) in which they suffer a detriment.

Some of these points also help explain the error in the conjecture that a higher $t$ improves the quality of plaintiffs’ choices of which cases to file. As long as $t > -s$ (which is satisfied if $t$ is nonnegative), plaintiffs prefer to win rather than to lose. Therefore, they prefer cases with a higher $\pi(x)$ and hence with a higher value of $x$. Although reducing $t$ diminishes the strength of this preference, it does not eliminate it. Filing decisions can still be summarized by a threshold value, $x^*$, where cases are filed if and only if $x > x^*$. And, as just explained, plaintiffs’ selections are actually sharpened by this experiment because the increase in $r^p$ that holds $b^p$ constant entails a higher $x^*$, thereby inducing plaintiffs to refrain from filing the weakest cases that otherwise would have been filed.

**Proposition 5:** If the optimum involves some suits ($\pi(x^*) < 1$), defendants prefer that the marginal suit not be filed, and deterrence is welfare increasing (or chilling is welfare reducing) at the margin, then the optimal evidence threshold, $x^*$, cannot be interior. Specifically, if it were, it would be possible to reduce $x^*$ and raise $r^p$ such that welfare increases.

**Proof:** The argument establishing Proposition 5 follows very closely that given for Proposition 4. Our experiment is to lower $x^*$ while adjusting $r^p$ so as to keep $b^p$ fixed. Again, it will ease the exposition to consider, until the end, an experiment that instead increases our variable of interest, $x^*$, while adjusting $r^p$ so as to keep $b^p$ fixed. We will show that this experiment reduces social welfare, implying that the reverse experiment, in which we lower $x^*$, raises welfare.

The effect of this experiment on social welfare (4) is given by an expression identical to expression (13) above, except that each derivative with respect to $s$ is replaced by one with respect to $x^*$. Accordingly, we can proceed to analyze how the experiment influences behavior.

Differentiating expression (6) — keeping in mind that $\pi(x)$ is implicitly a function of $x^*$, as indicated by expression (1) — we have:
Expression (20) is quite similar to expression (19), but a few comments are in order. Regarding the first term, we again have a situation in which the direct effect of increasing our variable of interest, here $x^T$, is to increase $x^*$. Raising the hurdle for victory at trial makes suit less attractive. Interestingly, this direct effect is one-to-one, as can be seen from expression (1): that is, to maintain the preexisting probability of success, which from expression (5) is unchanged (that is, until we adjust $r^p$), $x^*$ rises by the same amount as does $x^T$, which is as we would have expected. Regarding the second term, the $s + t$ in the integrand reflects that the stakes in a case — the amount parties’ prospects change depending on whether liability is found — are the sum of the two transfers, as discussed in section 2.C. From expression (1), we can see that raising $x^T$ changes the probability of liability, at a given $x$, by $-z(x^T - x)$, which is to say, the probability falls. This explains why the second term is also negative.

As a result, the direct effect of raising $x^T$ is to reduce $b^H$ on account of both terms in expression (20). Therefore, if we wish to hold $b^H$ constant, we need more suits: a lower $x^*$, which is implemented by reducing $r^p$. Therefore, this experiment raises litigation costs.

To determine the effect on $b^H$, we perform the same sort of manipulations as before. The pertinent integrand, that in the second term of expression (20), is again positive. As with the proof of Proposition 4 (expression 19) but unlike in the case of Proposition 3 (expression 14), this term is subtracted, so a slight modification of our previous analysis indicates that the experiment of increasing $x^T$ while holding $b^H$ constant causes deterrence ($b^H$) to fall.

Taken together, the experiment that increases $x^T$ raises litigation costs and reduces deterrence, both of which are welfare decreasing. Hence, the reverse experiment, involving a reduction in $x^T$, raises social welfare. ■

In principle, therefore, it is optimal to lower $x^T$ as much as possible, tending to negative infinity. In doing so, one should keep in mind that the experiment also raises $x^*$, through an increase in $r^p$. In the process, however, $x^*$ does not increase without bound, for even if all suits that are filed result in a plaintiff victory and thus a defendant payment of $s$, this imposes only a finite cost on prospective actors. And further increases in $x^*$ reduce that cost.  

Proposition 5, like Proposition 4, presents what may seem to be a surprising result, one that appears counterintuitive for similar reasons. In a setting such as the present one, when there are benign acts, so chilling is a concern, a higher evidence threshold may seem appealing on two grounds: it is relatively generous with regard to weaker cases that are before the tribunal, which is favorable to benign acts, and it tends to discourage the filing of weaker cases. But the
apparent implication does not follow, for the same reasons as before.

Specifically, our present instrument, \(x^T\), can be thought of as the legal, or de jure, evidence threshold (a relative of the burden of proof; see note 11). As mentioned before, \(x^*\) is the practical, or de facto, evidence threshold. Plaintiffs only file stronger cases, those for which \(x > x^*\). And if one wishes to increase (decrease) this de facto evidence threshold, one can do so directly, by raising (lowering) \(r^p\). It turns out that adjusting the de jure threshold, \(x^T\), is unnecessary if one wishes merely to adjust the de facto threshold, \(x^*\). And, in this model, using an instrument like \(r^p\) is every bit as direct of a way of doing so.

But there is more: Our experiment, when adjusting the de jure evidence threshold, \(x^T\), also adjusts \(r^p\) so as to keep \(b^P\) constant. That adjustment, in turn, directly influences the de facto threshold, \(x^*\). And here’s the kicker: This second adjustment moves the de facto evidence threshold, \(x^*\), in the opposite direction from that of our manipulation of \(x^T\), and the total adjustment does so to a greater extent. The latter, recall from the proof, is true because the change in \(x^*\) needs to wash out, with regard to actors’ expected costs of committing acts of type \(B\), both the entire change in \(x^*\) caused directly by the change in \(x^T\) — that alone would produce an exact offset — and also the change in expected transfers in inframarginal cases (those filed regardless) — which requires a further opposing movement in \(x^*\). (These two components correspond to the analysis of the first and second terms, respectively, in expression 20.)

To summarize, the net effect of our experiment, as a whole, on the de facto evidence threshold, \(x^*\), is opposite in direction to our movement of the de jure evidence threshold, \(x^T\). In light of this fact, we can see why the results are reversed from those we might have anticipated.

The practical lesson is that, in a setting such as the present one, if one wants to focus the legal system’s burden most heavily on acts that generate strong signals — which are relatively more often generated by harmful acts than by benign ones — the best way to do so is to raise the de facto evidence threshold, \(x^*\), not the de jure evidence threshold, \(x^T\). Moreover, once one does this, the latter is actually counterproductive: when the system, through plaintiffs’ filing decisions, is already confronting a smaller but stronger set of cases, it is attractive to require defendants to pay damages more often (and receive transfers from plaintiffs less often). And this is accomplished by reducing, not increasing, the de jure evidence threshold, \(x^T\).26

Put another way, in a model such as this one in which the cases that come before the tribunal are endogenously determined (selected) by self-interested plaintiffs, the best way to increase the legal system’s overall diagnosticity — which determines the precision with which sanctions are targeted at harmful rather than benign acts — is to make plaintiffs’ filing decisions

---

26As an aside, note that, by analogy to the point made when discussing Proposition 4 that a reduction in \(t\) decreases the extent by which plaintiffs gain from bringing stronger versus weaker cases but does not eliminate it, reducing \(x^T\) has the same effect of decreasing the relative gain from stronger cases but not eliminating it.

27This point is complementary to the argument that private enforcement tends to be superior to public enforcement precisely where potential plaintiffs possess superior information to the government about defendants’ acts. See Shavell (1993).
even more selective. Because marginal cases are weaker than inframarginal cases, it is better to be harsher toward plaintiffs with regard to the former and more lenient with regard to the latter. Reducing $x^T$, which enables a net increase in $x^*$ (while keeping chilling constant, a point elaborated in section B), makes just this sort of tradeoff. Proposition 4’s demonstration that it is optimal to reduce $t$, enabling an increase in $x^*$, follows the same logic. Likewise for Proposition 3, showing the optimality of an increase in $s$. All three experiments are generous to plaintiffs inframarginally — on cases they file regardless — which enables the system to be tougher at the filing margin.

B. Discussion

Initially, let us revisit the three conditions in these propositions. The first is that the optimum involves some suits. Clearly, if it does not, then marginal adjustments to $s$, $t$, or $x^T$ will be inconsequential. Moreover, it is possible for an optimum to involve no suits. Specifically, suppose that $h$ is very small whereas the $c_j$ are large. In addition, assume that $\gamma$ is large and the distributions of individuals’ benefits from acts are such that few have low benefits. Then the deterrence gain from some enforcement would be very small, whereas aggregate litigation costs would be relatively high (because so few acts are deterred or chilled). When litigation — or, more generally, enforcement — is costly and deterrence benefits are small, the optimum may involve no action. Accordingly, this article (like essentially all the pertinent literature) addresses optimal legal system design when some enforcement is desirable.

Second, it is assumed that defendants prefer that the marginal suit not be filed. Obviously this will be so when litigation costs $c'$ are positive (as assumed here), the sanction $s$ is positive (it is assumed here to be nonnegative), and the transfer $t$ that defendants receive when they win is not too large. If $t = 0$, which is shown here to be optimal (although employing the stated assumption), there is no problem. If, however, $x^*$ is low, implying that $\pi(x^*)$ is low, so $1 - \pi(x^*)$ is high, and $t$ is high, and $s$ and $c'$ are not very high, it is possible that defendants would welcome the marginal suit due to the significant possibility of receiving a large transfer. Although one might suppose that plaintiffs would never bring such a suit, keep in mind that $r''$ can be negative (a subsidy, and perhaps a large one), so this possibility cannot be ruled out a priori. As a practical matter (outside the model), such a configuration may not be sustainable because it might encourage collusive suits since both parties gain, at the expense of the government. (See note 22.)

Third, the propositions assume (for the first branch of our condition) that deterrence is welfare increasing at the margin. For this to be so, the first term of expression (13) (and corresponding expressions for the latter two propositions) tells us that the sum of the harm avoided from deterring an act and the expected litigation costs associated with an act exceeds the private benefit of the marginal harmful act. In simple models of law enforcement, deterrence is

---

28Moreover, even if $\pi(x^*)$, as given by expression (5), precisely equals 1, raising $s$ or reducing $t$ would have some of the effects described in the proofs of Propositions 3 and 4 but still be insufficient to establish the results. Specifically, initial changes in the instruments would influence the commission of acts only through the effect on $x^*$. Hence, the adjustment to $r''$ that keeps $b''$ constant would simply undo this effect on $x^*$, leaving all behavior, and thus social welfare, unchanged.
always beneficial at the margin at an optimum because, if it were detrimental or even neutral, reducing enforcement effort would be desirable or neutral with regard to behavior and also save enforcement resources. In models such as this one, however, less enforcement (say, induced by increasing \( r^p \)) has an additional behavioral effect, a reduction in the chilling of benign acts. For this reason, one might have supposed that it must likewise be true that deterrence is necessarily welfare increasing at the margin. (In addition, in all of our experiments, as explained at the conclusion of the proof of Proposition 3, we could instead have held deterrence, that is to say, \( b^D \), constant, and the same sort of analysis shows that chilling, \( b^B \), would fall, which on this account would be beneficial.)

However, in the present setup, incremental chilling is not necessarily detrimental at an optimum. From the third term of expression (13), the social cost of chilling the marginal act is the forgone benefit minus the expected litigation cost it would have generated. Consider, then, a set of parameters and a setting of the instruments such that the chilling effect is very small, so that the forgone benefit \( b^B \) from the marginal chilled act is likewise small. Moreover, assume further that \( x^* \) is low and the distribution function \( c^B(x) \) is such that benign acts often generate signals (values of \( x \)) that are above \( x^* \). Finally, suppose that litigation costs, the \( c^l \), are high. In that event, chilling a benign act could readily be beneficial: little private benefit is lost, but the reduction in expected litigation costs is significant.

Taken together, we cannot a priori rule out the case in which, at an optimum, deterrence is welfare reducing at the margin and chilling is welfare increasing.\(^{29}\) Note, however, that there is an important sense in which such an optimum is odd. Specifically, it would raise social welfare to reverse the outcome in some cases, as follows: for a small mass in which \( \chi \) is extremely high, find no liability rather than liability, and for a small mass in which \( \chi \) is extremely low, find liability rather than no liability.\(^{30}\) This seemingly perverse adjustment to the decision rule would slightly lessen the expected burden on harmful acts (which is desirable since deterrence is welfare reducing at the margin) and slightly increase the expected burden on benign acts (which is desirable since chilling is welfare increasing at the margin).

Having remarked on the propositions’ three conditions, let us step back briefly to consider the implications of these results when they do hold, which the foregoing suggests may include most cases of practical interest. All of the propositions in this section involve extreme prescriptions, and the latter two seem particularly jarring. In this regard, however, we should recall the original implication from Becker (1968) that, in a very simple setting with public enforcement, the optimal sanction is extreme, even when the level of harm (\( h \) here) is quite low. The reaction of scholars in the field was to labor long and hard to identify factors outside the basic model — which one might then bring in — that might generate less extreme implications.\(^{31}\) Much was learned from these efforts. In particular, making models more realistic in various ways brought additional, important dimensions of behavior and possible instruments to the fore.

\(^{29}\)Even then, the results hold if the direct savings in litigation costs (from raising \( x^* \)) are sufficiently large so as to outweigh welfare reductions from changes in deterrence or chilling.

\(^{30}\)And, if necessary, adjust \( r^p \) so as to hold \( x^* \) constant.

\(^{31}\)See the survey by Polinsky and Shavell (2007).
Moreover, the extensions often had additional implications, sometimes for the setting of other instruments.

It should be remarked that, relative to the handful of prior models involving optimal system design with private litigation, the results derived here were generated not by adding restrictions but by expanding the scope in a number of ways. First, benign acts were introduced, the chilling of which is an important concern when attempting to design an optimal legal system. Second and related, error was introduced, allowing for errors of both types (which little prior literature considers, particularly in connection with benign acts); relatedly, the probability of a plaintiff victory is modeled explicitly in terms of signals (evidence), allowing for differences in what the parties and the tribunal observe. Third, this model (including in this section) allows more instruments than most prior work, and in a setting in which enforcement effort is endogenous, through plaintiffs’ decisions whether to sue.

It is particularly interesting that many of these extensions actually contributed to the extreme results derived here. The presence of benign acts and error might, as discussed previously, have led us to suspect that rewarding victorious defendants and raising the evidence threshold would be attractive. In addition, introducing additional instruments — importantly in this section, merely allowing a filing fee — was central in demonstrating these unexpected outcomes. Therefore, to the extent that limited prior work does not obtain such extreme results, we can now see that any seemingly more intuitive conclusions actually are fragile, in a sense artifacts of restrictions on instruments that were implicit in the setup (and which have not previously been discussed in the literature). Therefore, although the present demonstrations should be regarded primarily as exploratory rather than prescriptive, they are more informative than are perhaps more appealing conclusions in prior work that are not robust to the introduction of realistic phenomena such as the possibility of error involving innocent behavior and the availability of additional and actually quite simple instruments.

Finally, it is useful to compare this section’s results to Proposition 2 in the preceding section, where it was shown how, without restrictions on the magnitudes of the instruments, it is feasible to implement a result arbitrarily close to the first best. Regarding Proposition 3, that it is optimal to raise \( s \) ever higher (until limited by any maximum, \( s^{Max} \)), the construction showing how to implement Proposition 2’s result was much the same. The connection with Proposition 4, that it is optimal to reduce \( t \), is less apparent. The supposition for this Proposition is that deterrence is welfare increasing or chilling is welfare reducing at the margin, but neither is true with respect to Proposition 2, for, if there are no restrictions on instruments, both deterrence and chilling are at first-best levels (\( b^{H} = h \) and \( b^{B} = 0 \)). Additionally, the proof of Proposition 4 entails that reducing \( t \) enables a reduction in litigation costs, but when there is no constraint on \( s \), we can already reduce them arbitrarily close to zero. Proposition 5 indicates that, with an interior solution, it is always optimal to reduce \( x^{T} \) (on the same grounds that it is optimal to reduce \( t \)). For the reasons just stated, this too is unnecessary in demonstrating Proposition 2.

Combining the results of all four propositions regarding optimal system design, we can see that the contours of an optimal regime will be determined importantly by instrument restrictions. These may include limitations on the values instruments may take: for example, if
we require that \( s \leq s_{\text{Max}} \) or that \( t \geq 0 \). In addition to any outside, perhaps institutional restriction, section 3 also noted the further implicit constraint that \( s + t > 0 \), so if there is a maximum on \( s \), this implies a (negative) minimum on \( t \). A qualitatively different sort of limitation would arise if some instrument may not be used at all, as we now explore.

**C. Additional Instrument Restrictions: When Suits Cannot Be Regulated Directly**

Propositions 3–5 tell us how to set the instruments \( s \), \( t \), and \( x^T \) when we are able to regulate suits (the level of \( x^* \)) directly and the only other restrictions we face are a possible maximum on \( s \) or minimum on \( t \). Note that these results hold even if only the single instrument in question (\( s \), \( t \), or \( x^T \), as the case may be) can be adjusted along with \( r^P \), with institutional constraints tying down the levels of other instruments. Finally, section 2.C explained how suits (the level of \( x^* \)) can alternatively be regulated — without any feedback effects on the \( b^i \) through \( r^P \), the decoupling of \( s \) (the ability to set \( s^p \) independently of \( s^D \)), or the decoupling of \( t \) (the ability to set \( t^P \) independently of \( t^D \)). However, just as we may have a constraint like \( s \leq s_{\text{Max}} \), due perhaps to limits on defendants’ wealth, so it may be that there is an upper limit on \( r^P \) (and, if relevant, on \( r^P + t^P \)) due to limits on plaintiffs’ wealth (see note 16). In that event there may be additional bounds on the ability to raise social welfare in the manner dictated by Propositions 3–5. Moreover, many actual legal systems do not impose significant filing fees (or subsidies) on plaintiffs or employ either form of decoupling.

For these reasons, this section briefly considers more restricted systems: specifically, those in which it is not possible to regulate plaintiffs’ filing decisions (\( x^* \)) directly. Consider a system in which \( r^P \) is fixed — for convenience, at 0 — and in which there is no decoupling of either sort. Furthermore, we will continue to assume that \( r^D = 0 \).

In such a regime, do any results like those in Propositions 3–5 continue to hold? The answer is no. Sparing the reader extensive derivations (some tedious but straightforward, others subtle), a number of observations can be offered. To begin, raising (or lowering) each of our remaining instruments, \( s \), \( t \), or \( x^T \), one at a time, presents the welfare tradeoffs one would have expected in this type of model: Raising \( s \) hurts both types of defendants and helps plaintiffs, so it increases deterrence and chilling as well as the level of lawsuits. That is, \( b^H \) and \( b^B \) rise and \( x^* \) falls. The latter is always, ceteris paribus, welfare reducing. As discussed just above, enhanced deterrence tends to be good and additional chilling bad, although either or both could be the opposite. The only thing that we can say with confidence regarding the optimum in the present setting is that it cannot be true that both deterrence and chilling are undesirable: If they were, then reducing \( s \) would unambiguously increase social welfare.

The effects of raising \( t \) are, of course, the opposite in all respects from those of raising \( s \). Likewise for raising the evidence threshold, \( x^T \).\(^{32}\)

---

\(^{32}\)It is interesting to inquire further about the effects of raising \( x^T \). As explained previously, effects that we might intuitively have associated with a higher de jure evidence threshold are produced in this model by the higher de facto evidence threshold — the filing cutoff \( x^* \) — which is influenced by all of the other instruments as well. The experiment underlying Proposition 5 allowed us to adjust \( r^P \) so as to move \( x^* \) as we wished. Here we are assuming that this is not
We can also ask whether further characterizations might be possible with the sorts of experiments we have examined earlier in this section wherein two instruments are simultaneously adjusted in particular ways. For example, we might increase \( s \) while simultaneously increasing \( t \) by just the amount that keeps chilling, the level of \( b^H \), fixed. When this is done, it is possible to show that \( x^* \) rises. So far, chilling is unchanged and litigation costs are saved with regard to both types of acts. However, the sign of the effect on deterrence, the level of \( b^H \), is ambiguous. (And no simple parameter restrictions remove the ambiguity.) For the other two pairings of instruments — raising \( s \) and \( x^T \) in a manner that holds \( b^H \) constant, and raising \( t \) while reducing \( x^T \) in a manner that holds \( b^H \) constant — it is not possible to sign the effect on either \( x^* \) or \( b^H \). In summary, even when we consider a broader array of experiments, no additional, sharp characterizations can be offered.

5. Conclusion

In a model with endogenous filing decisions, court error, and two types of behavior (harmful and benign), this article analyzes the optimal levels of filing fees, damages, and payments by losing plaintiffs — each of which may differ for each party — as well as the stringency of the evidence threshold. If there are no instrument constraints, three instruments suffice to span most possible outcomes for the three endogenous behaviors: the commission of harmful and benign acts, and filing decisions. In particular, one can achieve an outcome arbitrarily close to the first best. Additional analysis examines instrument redundancy and the manner in which the present model incorporates prior work on decoupling and fee shifting as special cases.

Further analysis explores the optimal choice of particular instruments, including when there are restrictions of various sorts. As long as there is some free instrument that can regulate suits directly, such as a filing fee charged to plaintiffs, each other instrument considered is optimally set at an extreme level: the damage award as high as possible, the transfer paid by losing plaintiffs as low as possible, and the evidence threshold as low as possible. The core explanation for the latter two (more surprising) results is that plaintiffs’ self-interest already entails selection of the strongest cases. As a consequence, a filing fee can target a cutoff for suit that serves as a de facto evidence threshold, and it is demonstrated that greater generosity toward plaintiffs in filed cases, which are the strongest — even through reducing the de jure evidence threshold, at trial — enables a tougher de facto threshold up front, the combination of which is relatively favorable to benign acts versus harmful ones. All such results, however, may change qualitatively when additional external constraints are imposed on the availability of instruments, possible in any direct manner. Suppose instead that we increase \( x^T \) and simultaneously increase \( s \) so as to keep \( x^* \) fixed. The remaining effects on the \( b^* \) stem from the fact that, for those cases filed (for \( x > x^* \)), the expected sanctions change, upward because a higher \( s \) is imposed and downward because it is imposed less often. How much less often? The increase in \( x^* \) means, through expression (1), that, for every value of \( x \), there is now a different critical value of the error, \( \varepsilon \), above which there is liability. Further analysis indicates that (for a single-peaked density function \( z \) that has its maximum at \( \varepsilon = 0 \)), if \( x^* > x^T \), both deterrence and chilling rise in this experiment (because the effect of raising \( x^T \) on inframarginal cases is less than the effect on marginal cases). However, if \( x^* \) is sufficiently below \( x^T \), deterrence may rise while chilling falls, and for even lower \( x^* \) relative to \( x^T \), both deterrence and chilling may fall. These, in a sense, are the distinctive effects of adjusting \( x^T \).
which seems quite common in actual legal systems although the reasons for this are often not apparent.

As explained, many of the present conclusions differ qualitatively (and in some cases radically) from those in the small body of prior literature on optimal legal system design with private enforcement, with the divergences attributable to implicit instrument restrictions in previous models. It is also notable that the explicit modeling of signals — which allows for two types of court error that influence two types of behavior, including, importantly, the possible chilling of benign conduct — fundamentally changes much of the analysis but need not generate the sorts of results that one might have anticipated, such as the general desirability of making losing plaintiffs pay or utilizing tougher evidence thresholds.

Although the present framework is more general and encompassing than those in prior work, the results are best viewed as outlining a partial and preliminary conceptual understanding of optimal legal system design with private enforcement. First, a number of familiar complications considered in the literature on public enforcement, such as risk aversion and socially costly sanctions, would influence the conclusions, perhaps in familiar ways. See notes 9 and 16, and Polinsky and Shavell (2007). Second, it is essential to explicitly model signals (evidence) — with attention to what is observed by the parties, plaintiffs and defendants; how that may differ from what is observed by the tribunal; and how these signals are generated by different types of behavior — but their central role also suggests that different modeling choices might generate different results. Of particular interest would be variations where parties’ initial signals differed from each other’s, which would introduce asymmetric information that, in turn, could allow for more interesting analysis of settlement and pre-trial discovery. Also, the signal observed by the tribunal could be made a function not only of parties’ signals but also of their endogenously chosen litigation expenditures. Another sort of extension would introduce additional dimensions of heterogeneity, such as in parties’ litigation costs (which makes selection at the time of filing contingent on particular plaintiffs’ costs in addition to the signal of case strength) or in types of underlying behavior wherein the tribunal (or perhaps plaintiffs as well) cannot observe pertinent differences.

---

33One avenue would be to explore the value of accuracy of the tribunal’s noisy observation χ of the parties’ signal x. Compare Kaplow and Shavell (1994). Preliminary analysis in the present model did not yield sharp results. Another variation would be to model the tribunal as observing x and the parties observing χ (perhaps the process of trial yields greater clarity, although the opposite assumption employed here, that the tribunal never learns as much as the parties observe directly, may be more realistic in many settings). Some reflection on this case leads to a conjecture that some results will differ because the tribunals’ final decisions are more diagnostic than plaintiffs’ filing decisions. Moreover, the analysis will be more complex because plaintiffs who observe χ need to form beliefs about the distribution of x that the tribunal will subsequently observe, and these beliefs involve Bayesian inferences reflecting underlying primary behavior, which itself is endogenous.

34On settlement in the present, symmetric information setting, see note 10.
References


