

ISSN 1936-5349 (print)  
ISSN 1936-5357 (online)

# HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

## ON THE OPTIMAL BURDEN OF PROOF

Louis Kaplow

Discussion Paper No. 712

01/2012

Harvard Law School  
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:  
[http://www.law.harvard.edu/programs/olin\\_center/](http://www.law.harvard.edu/programs/olin_center/)

The Social Science Research Network Electronic Paper Collection:  
<http://ssrn.com/>

# On the Optimal Burden of Proof

Louis Kaplow\*

## *Abstract*

The burden of proof is a central feature of adjudication, and analogues exist in many other settings. It constitutes an important but largely unappreciated policy instrument that interacts with the level of enforcement effort and magnitude of sanctions in controlling harmful activity. Models are examined in which the prospect of sanctions affects not only harmful acts but also benign ones, on account of the prospect of mistaken application of sanctions. Accordingly, determination of the optimal strength of the burden of proof, as well as optimal enforcement effort and sanctions, involves trading off deterrence and the chilling of desirable behavior, the latter being absent in previous work. The character of the optimum differs markedly from prior results and from conventional understandings of proof burdens, which can be understood as involving Bayesian posterior probabilities. Additionally, there are important divergences across models in which enforcement involves monitoring (posting officials to be on the lookout for harmful acts), investigation (inquiry triggered by the costless observation of particular harmful acts), and auditing (scrutiny of a random selection of acts). A number of extensions are analyzed, in one instance nullifying key results in prior work.

JEL Classes: D81, K14, K41, K42

Keywords: burden of proof, law enforcement, type 1 and type 2 errors

*Forthcoming, Journal of Political Economy*

---

\*Harvard University and National Bureau of Economic Research. I am grateful to Steven Shavell, Christopher Snyder, Kathryn Spier, the referees, and presentation participants at the 2011 AEA meetings, Harvard, NYU, and the NBER for comments, to the John M. Olin Center for Law, Economics, and Business at Harvard University for financial support, and to Michael Atamas, Sarah Beckerman, and Noam Lerer for research assistance.

## 1. Introduction

All systems of adjudication must determine how high to set the burden of proof: the strength of evidence required for the imposition of sanctions, award of damages, or provision of other forms of relief. Proof burdens are set in other contexts as well: internal organizational decision-making (standards for promotion, employee discipline, or product launches), contractual requirements (defect rates for rejection of a supplier's shipment), and medical treatment.

The familiar tradeoff of type 1 and type 2 errors, such as in medical treatment, focuses on the costs and benefits of various post-decision outcomes. In many settings, however, including for the legal system, a (or the) primary purpose and effect of decisions and resulting actions is the creation of ex ante incentives. The prospect of correct imposition of sanctions deters harmful activity; mistaken exoneration dilutes it. Moreover, the expectation that sanctions will sometimes be mistakenly applied to benign acts tends to chill desirable behavior. Consider, for example, antitrust, where misapplication of sanctions may discourage efficient, pro-competitive behavior (for example, promotional product pricing may look like predation); securities regulation, where the prospect of erroneous sanctions may increase the cost of capital (IPO's may result in liability even in the absence of any misrepresentation); medical malpractice, where worries about false positives may discourage cost-effective care (including denial of care to high-risk patients and reduced physician supply in certain fields); and contract breach generally, where concern for misassessment of performance may deter efficient contracting or misdirect behavior governed by contracts.<sup>1</sup> In these settings and others, choosing the burden of proof poses an error tradeoff, but a more complicated one than in standard problems like medical decision-making because the focus is on ex ante behavior, which is endogenous.

This article extends models of law enforcement by treating the proof burden as a policy instrument alongside the traditional ones, enforcement effort and the level of sanctions. Because the results differ in important ways depending on the enforcement technology, three types of models are analyzed. Monitoring – posting lookouts for harmful acts, such as police patrols or private security guards – is considered first because its features are elements of the other cases. Investigation – inquiry triggered by the costless observation of particular harmful acts, such as with many crimes, accidents, and contract breaches – differs because benign acts can be mistakenly sanctioned only as a consequence of legal action triggered initially by the commission of a harmful act. Auditing – scrutiny of a random selection of acts, such as tax examinations, inspections for regulatory compliance, and intellectual property owners' surveys of competitors' products – is analytically identical to monitoring in terms of behavior but differs in administrative costs because greater activity (harmful and benign) requires more audits for a given detection rate.

Among the more important findings applicable to all three models is that the

---

<sup>1</sup>For empirical evidence suggestive of the importance of these phenomena, see, for example, Walker et al. (2010) on IPO litigation risk; Studdert et al. (2005) on the avoidance of medical procedures and patients posing high litigation risks; and Kessler, Sage, and Becker (2005) on medical malpractice and physician supply.

determinants of the optimal evidence threshold depart qualitatively and substantially from familiar understandings regarding the burden of proof – which involve consideration of the likelihood that an individual before the tribunal committed a harmful act rather than a benign one. For example, it is demonstrated that raising the evidence threshold can increase the frequency with which individuals who commit benign acts are subject to sanctions and that raising this threshold can result in a lower ex post likelihood that an individual must have committed the harmful type of act in order to be subject to sanctions.

Section 2 presents the first model, involving monitoring, which is employed in the majority of the analysis because it is somewhat more streamlined and, as mentioned, nearly every effect also appears in the other models. There are two groups of individuals who may commit acts that yield private benefits. For one group, the act causes external harm; for the other, the act is socially benign. The government chooses a level of enforcement that determines the fraction of acts that are scrutinized. As mentioned, the posting of monitors (police patrols) has this feature. At this first level, enforcement does not distinguish the two types of acts, a difference that is initially unobservable to the authority. But the process of scrutinizing acts ultimately gives rise to evidence, which is then assessed to determine whether to impose sanctions. The evidence is modeled as a signal. The distribution of signals is higher for harmful acts than for benign ones. The choice of a burden of proof can be thought of as selecting a cutoff, where sanctions are applied if and only if the signal exceeds the chosen threshold. Any (nonextreme) rule entails both true and false positives and negatives.

Section 3 analyzes optimal policy, focusing on the level of enforcement effort and the setting of the evidence threshold. As one would expect, the prospect of chilling desirable behavior is central to both choices. Greater enforcement effort, in addition to involving the usual tradeoff of deterrence and enforcement costs, also raises the expected sanction on benign acts, so optimal enforcement effort is lower on this account. The evidence threshold is set to equate the marginal effects of deterring harmful acts and chilling benign acts. The benefit of deterring a harmful act falls at the margin whereas the cost of chilling a benign act rises, although nonuniform distributions of benefits for each type of act can produce countervailing effects.

Additional attention is devoted to the respects in which greater enforcement effort and a lower evidence threshold are substitute means of achieving deterrence. The former involves direct resource costs whereas the latter does not. Accordingly, for there to be an optimal mix, it must be that a lower evidence threshold has some relative disadvantage. Under standard assumptions, a lower evidence threshold (compared to greater enforcement effort) causes relatively more chilling per unit of deterrence enhancement – and, if it did not, one could not be in the neighborhood of an optimum. The analysis also suggests a new, complementary version of the so-called Becker (1968) argument favoring maximal sanctions: it is generally optimal to raise sanctions (if they are not maximal) and simultaneously raise the evidence threshold, holding deterrence constant; the only effect of this combination is a reduction in chilling costs. Perhaps surprisingly, therefore, the prospect of mistakes makes it harder rather than easier to rationalize the widespread practice of employing only moderate sanctions for less harmful acts. Moreover, this finding overturns key results in prior literature that abstracts from errors and, accordingly, does not feature the evidence threshold as an instrument.

Section 3 also relates the present analysis, which focuses on the choice of an evidence threshold, to conventional notions about the burden of proof, which may be understood in terms of Bayesian posterior probabilities. Consider, for example, the preponderance of the evidence standard, often taken to mean that it must be more probable than not that the defendant before the tribunal committed a harmful act (which may include failing to take due care, breaching a contract, and so forth). The determinants of the optimal evidence threshold and of the preponderance standard (or other ex post probability-based standards) are quite different. (They do have two common factors, but the other seven terms in the two expressions differ.) Comparative statics can also be counterintuitive. Raising the evidence threshold can result in a lower ex post likelihood that an individual before the tribunal must have committed a harmful act in order to be subject to sanctions. Also, raising the evidence threshold can increase the absolute frequency with which individuals who commit benign acts are mistakenly subject to sanctions.

The preceding points raise the question of how some prior work could have found the preponderance of the evidence standard to be optimal. To match more closely those models, a final subsection presents a modified, simpler version of the present model in which there are no benign acts, so individuals only decide whether or not to commit a harmful act. In this variant, the main result in previous papers is immediate. Nevertheless, this result is not the preponderance of the evidence standard but rather a statement about ex ante probabilities conditional on individuals' behavior. Moreover, comparison of the single factor that determines the optimum in this modified model with the first-order condition for the optimal evidence threshold in the central model here reveals that there is little relationship between the two (the latter condition including six additional factors). Hence, intuitions gleaned from the sort of model used in past work provide little guidance regarding the optimal evidence threshold in the settings examined in this article.

Section 4 presents a second model, for the case in which enforcement is through investigation rather than monitoring. Specifically, it is supposed that investigations are triggered by the observation of harmful acts, the sighting of which does not identify the perpetrator. For example, for murder, auto theft, or groundwater contamination, it is often clear that a violation occurred, and the authorities engage in investigation to determine who committed the harmful act. In this setting, the nature of the optimum for both enforcement effort and the evidence threshold depends on additional factors that were absent in the monitoring model. First, enforcement effort only needs to be expended when harmful acts occur, so deterrence reduces enforcement costs; this factor favors greater enforcement effort and a lower evidence threshold (the latter, by raising deterrence, similarly reduces enforcement costs). As a consequence, there need not be underdeterrence at the optimum, unlike when enforcement is by monitoring. Second, because deterrence reduces the frequency of investigations, it also reduces the likelihood that benign acts are mistakenly subject to sanctions and thus mitigates chilling costs. As well, chilling costs are in a sense lower to begin with since scrutiny is only triggered when harmful acts are committed. A number of these considerations are favorable to greater enforcement effort and a lower evidence threshold. Relatedly, when enforcement is by investigation, the factors determining the optimal evidence threshold differ in yet additional ways (by comparison to the model of monitoring) from the preponderance of the evidence standard or other qualitatively similar rules. Also, as in the model with monitoring, raising the evidence threshold can increase the absolute frequency with which individuals who commit benign acts are

mistakenly subject to sanctions, but the circumstances in which this is so differ qualitatively.

Section 5 offers four extensions. Auditing (a third model) is treated briefly because its differences from the monitoring model are modest. Behavioral effects are the same, so many results from the monitoring model are applicable. However, enforcement costs are qualitatively different, which increases the divergence between the determination of the optimal evidence threshold and that which satisfies conventional burden of proof notions.

Costly sanctions are considered next. Their introduction might seem to favor a heightened evidence threshold because this reduces the frequency with which sanctions are imposed on both the harmful and benign acts that are committed. There is, however, the competing effect that raising the evidence threshold reduces the deterrence of harmful acts and the chilling of benign acts, both of which increase the pool of individuals potentially subject to sanctions. Accordingly, the common view that the existence of costly sanctions obviously justifies a tougher proof requirement in the criminal context is undermined. Incapacitation benefits of sanctions (notably, imprisonment) as well as any social costs associated with the very act of mistaken imposition of sanctions are each analyzed as variations in social sanction costs, with the implication that the effects of these considerations on the optimal evidence threshold are similarly ambiguous, again contrary to conventional wisdom.

The third extension considers the accuracy of evidence. First explored is how an increase in accuracy affects the optimal evidence threshold. It is explained that there is serious ambiguity in posing such a question and, moreover, that there are a number of competing influences, so no clear answer can be offered. Second, it is explained how one can use the present analysis to identify the value of increased accuracy. Finally, a fourth extension addresses the relevance of whether there is a small, highly concentrated group of benign acts that might mistakenly enter adjudication versus a broad, more dispersed group, each less likely to be swept up in the legal system but with the same overall frequency of misidentification of benign acts. The former situation tends to involve greater chilling costs and thus to favor a higher evidence threshold, although this conclusion need not hold due to nonconstancy of the density functions for individuals' benefits from acts.

Prior literature on the burden of proof largely considers different issues. One set of articles focuses on adjudication itself, such as regarding the resources devoted to the presentation of evidence or bringing suit (usually with exogenous underlying behavior).<sup>2</sup> Some of this work has emphasized the placement of the burden to go forward – that is, the specification of which side loses if little or no evidence is presented – rather than the strength of the burden of proof.<sup>3</sup> Another set of literature does consider underlying behavior but has largely been confined to models in which there is only one type of decision: whether or not to commit a harmful act (or,

---

<sup>2</sup>See Bernardo, Talley, and Welch (2000), Demougin and Fluet (2008), Hay and Spier (1997), Miceli (1990), Posner (1999), and Rubinfeld and Sappington (1987).

<sup>3</sup>In legal usage, the single phrase “burden of proof” is sometimes used to refer to the persuasion burden (how convinced the factfinder must be, such as under the preponderance rule) and also the production burden (which party has the burden of going forward, i.e., who loses if little or no evidence is presented), although recently the former reference has become more common. See Garner (1999).

similarly, whether or not to take due care).<sup>4</sup> In such a model, the harm of mistaken imposition of sanctions is simply to reduce deterrence by making the socially preferable option less attractive. Often, deterrence is taken to be inadequate, so the problem reduces to choosing the burden of proof that maximizes deterrence. As already mentioned, these models (which are examined in section 3.E) are much simpler than the ones considered here, and their qualitatively different results provide little indication of optimal policy in the present settings.

## 2. Model I: Monitoring

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.<sup>5</sup> 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 might initially be confused with harmful acts, for other benign acts do not face the risk of sanctions and thus are unaffected by the enforcement instruments under consideration. (For many purposes, it would be possible to suppose 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.<sup>6</sup>) Individuals' benefits from committing an act are  $b$ , with differentiable density and cumulative distribution functions  $f^i(b)$  and  $F^i(b)$ , respectively, where  $i = H, B$ . Individuals know what type of act they are able to commit and its benefits to them, but the government initially has no knowledge of acts' type and never learns individuals' benefits from acts.

Throughout this article, the timing of the model is:

1. The government sets enforcement effort, the sanction, and the evidence threshold.
2. Individuals learn their type of act and their private benefit.
3. Individuals decide whether to act.
4. A portion of those who act are identified and brought before a tribunal.
5. In adjudication, individuals are sanctioned if and only if the evidence in their case exceeds the evidence threshold.

In this first version of the model, enforcement is by a government authority that scrutinizes the fraction  $\pi(e)$  of acts, where  $e$  indicates enforcement effort (expenditures) and

---

<sup>4</sup>See Demougin and Fluet (2006, 2008), Kaplow and Shavell (1994), Lando (2002), and Polinsky and Shavell (2007, pp. 427–29). Png (1986) has an additional behavioral dimension (activity level, in an accident setting) but does not analyze the burden of proof.

<sup>5</sup>It would be straightforward to adjust the analysis to allow it to involve a positive externality or a negative externality at a different level from that of the  $H$  type of act.

<sup>6</sup>The primary caveat pertains to setting the optimal sanction: specifically, the notion of setting the sanction at its maximum may be problematic if individuals might be sanctioned for multiple acts. Another variation would be to allow individuals to choose one of the two acts or inaction. This variation would significantly complicate the exposition but have only a modest effect on the qualitative results. Note that, in this setting, deterrence of harmful acts induces some individuals to switch to benign acts rather than inaction, and chilling of benign acts causes some individuals to switch to harmful acts rather than inaction.

$\pi'(e) > 0$ .<sup>7</sup> (Allowing type-specific  $\pi$ 's would not much affect the analysis, so is avoided for simplicity.<sup>8</sup>) One might think of enforcement as involving the posting of monitors (such as with much traffic enforcement or other police patrols) who will identify as suspicious some fraction of the acts that are committed.

If an individual's act is scrutinized, thereby leading to adjudication, the probability of being subject to the sanction  $s$  – taken for now to be a monetary penalty that is socially costless to impose – is  $p^i$ . The government, unfortunately, is not free to choose  $p^H$  and  $p^B$  independently; if it could, it would set the former to 1 (or sufficiently high to achieve first-best deterrence of harmful acts) and the latter to 0, thereby sanctioning all before the tribunal (or enough) who commit harmful acts and none who commit benign acts. Rather, it is supposed that there is some set of evidence produced when an individual is scrutinized. The government can choose a lower evidence threshold, which yields a higher  $p^H$ , but only at the cost of a higher  $p^B$ . Let us now specify this relationship.

The evidence in a given case will be represented by the variable  $x$ , a signal of the underlying act.<sup>9</sup> For each type of act, the densities and cumulative distribution functions for  $x$  are given by  $g^i(x)$ , assumed to be positive<sup>10</sup> and differentiable, and  $G^i(x)$ , respectively, where  $G^B(x) \geq G^H(x)$ , with strict inequality except when the  $G^i$  equal 0 and 1 (a strong form of first-order stochastic dominance).

Suppose the government sets the evidence threshold at some  $x^T$ , which is to say that sanctions are applied if and only if the value of the signal  $x$  exceeds  $x^T$ . Since  $G^i(x^T)$  denotes the probability that the signal falls below the threshold, it follows that  $p^i(x^T) = 1 - G^i(x^T)$ . Our assumption about the relationship between the two distribution functions implies that  $p^H(x^T) > p^B(x^T)$  – except when the  $G^i$  equal 0 and 1. Regarding the choice of  $x^T$ , note that it is equivalent (and sometimes useful for thought and to reduce notation somewhat) to suppose that the government instead chooses  $p^H$ . By inverting the function  $p^H(x^T)$ , this implies a value of  $x^T$ , and hence of  $p^B(x^T)$ . Accordingly, we can also define an implicit function for  $p^B$  that will be denoted

<sup>7</sup>We could further assume that  $\pi''(e) \leq 0$ , which would ordinarily be done to ensure that the second-order condition for welfare maximization with respect to  $e$  holds. As is apparent from the analysis in the appendix, however, it would still be necessary to confine attention to situations in which the second-order condition holds because there are other terms (involving the derivatives of the density functions) that may have the wrong sign.

<sup>8</sup>If the  $\pi$ 's differed by a constant factor (for example, the hit rate on benign acts may be proportionally lower than that on harmful acts), then the only effect would be that this constant would appear analogously in each first-order condition and in the condition for the preponderance rule, without affecting any qualitative statements or results. With general differences in the  $\pi(e)$  functions, allowed in section 4's extension for enforcement by investigation, the only further effects are precisely those presented there: the Becker argument for maximal sanctions need not hold, as discussed following expression (12), and equation (6) on optimizing both enforcement effort and the evidence threshold would become identical to equation (17).

<sup>9</sup>The optimal method of combining evidence (relevant in practice since multiple signals ordinarily are present) is given by likelihood ratios, as indicated by the Neyman-Pearson (1933) Lemma.

<sup>10</sup>They may equal zero at the lowest and highest possible values of  $x$ , if such exist (i.e., if the interval of possible values of  $x$  is closed at either end). The supports may be taken to be the same for signals associated with harmful and benign acts; for the problem to be interesting, all that is required is that the two distributions overlap.

$P^B(p^H)$ .<sup>11</sup> Because, from the definition of  $p^i(x^T)$ , we have  $dp^i(x^T)/dx^T = -g^i(x^T)$ , note that we can also write  $P^{B'} = dP^B/dp^H = [dp^B(x^T)/dx^T]/[dp^H(x^T)/dx^T] = g^B(x^T)/g^H(x^T)$ , which is positive. As will be seen, the first-order condition for the optimal evidence threshold  $x^T$ , (4) or (5), can be written using only  $P^{B'}$ , without explicit reference to the  $p^i(x^T)$ , for all that matters is the rate at which  $p^B$  changes per unit change in  $p^H$ . For some purposes, it is natural to assume further that  $P^{B''} > 0$ , i.e., that the implicit function  $P^B(p^H)$  is convex. This feature, which is unnecessary for the results in this paper, is implied if the density  $g^H(x)$  satisfies a strict version of the monotone likelihood ratio property with respect to  $g^B(x)$ .<sup>12</sup>

Finally, observe that choosing the evidence threshold  $x^T$  seems akin to choosing a burden of proof; a higher burden appears to correspond to a higher evidence threshold, and conversely. However, there is an important conceptual difference in that the threshold  $x^T$  is defined in terms of the evidence (the value of the signal) whereas conventional notions of a burden of proof are defined in terms of Bayesian posterior probabilities. Put another way, the  $p^i(x^T)$  refer to the probabilities of being sanctioned conditional on having committed an act of type  $i$  and then being scrutinized. These are ex ante probabilities that depend on one's type of act. By contrast, in discussions of the burden of proof, it is ordinarily understood that some act is before a tribunal, and we are inquiring into the probability that it is of the harmful type rather than the benign type. This ex post probability depends on the particular signal received (a specific value of  $x$  in the foregoing terminology, which is used to update one's prior) and on the base rates of the activity (the relative levels constituting one's prior), the latter of which are endogenous (notably, they depend on how the proof burden itself is set). We shall return to this matter in section 3.D.<sup>13</sup>

Individuals are assumed to be risk neutral, so those whose type of act is  $i$  commit their act if and only if  $b > \pi(e)p^i(x^T)s$ , that is, their benefit exceeds the expected sanction for their type of act.<sup>14</sup> Social welfare,  $W$ , is taken to be the aggregate of individuals' benefits from acting minus

<sup>11</sup>An equivalent expression is commonly used in signal detection theory, the graph of  $P^B(p^H)$  corresponding to what is called the receiver operating characteristic (ROC) curve. See Luce (1963).

<sup>12</sup>See Milgrom (1981). If  $P^{B''} < 0$  for some values of  $p^H$ , then a term in the second-order conditions (see (A1) and (A6) in the appendix) is positive, but since other terms may also be positive, first-order conditions are only necessary conditions and, for the social optimum, attention must in any case be confined to cases in which the corresponding second-order conditions are confirmed to hold. Part of why the assumption that  $P^{B''} > 0$  is natural to impose is that, otherwise, the optimal evidence rule does not consist of a single threshold  $x^T$  because one should apply sanctions for signal values with the highest likelihood ratios first (as stated by the Neyman-Pearson (1933) Lemma. Note that, when the assumption fails, one could reproduce the optimality of a single cutoff rule by reordering the values of the signal in accordance with the likelihood ratio, in which case the monotone property would hold for the transformed signal. On the relationship between the monotone likelihood ratio property and the optimality of simple cutoff decision rules, see Karlin and Rubin (1956).

<sup>13</sup>To foreshadow the analysis, any evidence threshold will in equilibrium generate a resulting, conventionally defined, burden of proof; that is, at the evidence threshold  $x^T$ , there is some corresponding ex post likelihood that the act is of the harmful type. The converse (moving from a stipulated ex post likelihood to an evidence threshold) is trickier – inversion would be possible if the relationship between the two was monotonic, continuous, and spanned the relevant range. The key complication is that the application of conventionally understood proof burdens depends on behavior, which itself depends on the proof burden.

<sup>14</sup>Introducing risk aversion would have two main effects. First, raising  $s$  would increase the deterrence and chilling punches of sanctions at an increasing rate. Second, sanctions would then be socially costly, as considered in the extension in section 5.B (although with risk aversion the social cost per unit of the sanction would be nonlinear

the harm from the commission of acts of type  $H$  and the cost of expenditures on enforcement.

$$(1) W = \int_{\pi(e)p^H(x^T)s}^{\infty} (b-h)f^H(b)db + \gamma \int_{\pi(e)p^B(x^T)s}^{\infty} bf^B(b)db - e.$$

The first term indicates the benefits and harm attributable to harmful acts, and the second the benefits from benign acts. In each case, the lower limit of integration is the benefit of the individual just at the margin, with all individuals having greater benefits committing the act.

### 3. Analysis of Model I

The government's problem is to choose  $e$ ,  $x^T$ , and  $s$  in a manner that maximizes social welfare (1). It is immediate that the optimal sanction is maximal. Following the idea in Becker (1968), if the sanction is not maximal, one could raise  $s$  and lower  $e$  by an amount that keeps the product  $\pi(e)s$  constant, in which case there would be no effect on either of the first two terms in expression (1), but welfare would rise on account of  $e$  being lower. For most of the analysis, the sanction  $s$  will simply be taken as given. It might be supposed to be at its maximal level, or, alternatively, it may be fixed for other reasons.

#### A. Optimal Enforcement Effort

To determine optimal enforcement effort, we can differentiate social welfare (1) with respect to  $e$ . Because  $\pi$  is a function of  $e$ , enforcement's effect on behavior is reflected in raising the lower limits of integration, and this is the source of deterrence benefits and chilling costs. The first-order condition (suppressing some arguments of functions) may be expressed as

$$(2) \pi'p^Hsf^H(\pi p^Hs)(h - \pi p^Hs) = \gamma\pi'p^Bsf^B(\pi p^Bs)\pi p^Bs + 1.$$

It is convenient to define  $b^i \equiv \pi p^is$ , the benefits of the individuals just indifferent about whether to act, in which case expression (2) can be restated as

$$(3) \pi'p^Hsf^H(b^H)(h - b^H) = \gamma\pi'p^Bsf^B(b^B)b^B + 1.$$

The left side represents the marginal gain from raising enforcement effort, which is the product of the deterrence punch  $\pi'p^Hs f^H(b^H)$  (the rise in expected sanction times the density of individuals just deterred per unit increase in expected sanction) and the net gain from deterring the marginal harmful act: avoiding the social harm  $h$  but losing the private benefit  $b^H$  from the act. The right side indicates the marginal cost from raising enforcement effort. The first term is the cost of chilling desirable behavior, the product of the relative mass  $\gamma$  of the benign act in the

---

in the sanction). For a discussion of prior literature on how risk aversion influences optimal sanctions, see Polinsky and Shavell (2007, pp. 413-16).

population times the chilling punch  $\pi' p^B s f^B(b^B)$  (the rise in expected sanction on the benign act times the density) times the benefit  $b^B$  from the marginal benign act that is forgone due to the higher (mistaken) expected sanction. The final term is just the cost of the added unit of enforcement effort.

*Proposition 1: Given  $x^T$ , at an interior optimum for the level of enforcement,  $e^*$ :*

- a. Raising  $h$  raises  $e^*$ ;*
- b. Raising  $\gamma$  reduces  $e^*$ ;*
- c. Raising  $s$  has an ambiguous effect on  $e^*$ ; and*
- d. At  $e^*$ ,  $h > b^H$  (i.e., there is underdeterrence relative to the first-best), and this would be so even if enforcement were costless.*

Propositions 1.a and 1.b are intuitive but the demonstrations are not entirely straightforward. Raising  $h$  naturally favors higher effort because deterrence of harm is more important. Raising  $\gamma$  favors lower effort because chilling effects are relatively more significant. Some complexity arises, however, because the pertinent derivatives of the first-order condition include a number of terms of differing and in some cases indeterminate signs; relatedly, multiple interior optima for  $e$  are possible. Much of the complication is due to the possibility that the distributions of benefits are not uniform, which also implies that the second-order condition (expression (A1) in the appendix) need not hold globally. Suppose, for example, that both densities are single-peaked, with the benefits of benign acts concentrated at a low level and those of harmful acts at a higher level. Between the peaks,  $f^B$  would be falling and  $f^H$  rising, the former implying that the chilling punch of greater enforcement is falling and the latter that the deterrence punch of greater enforcement is rising. If these effects from the density functions are sufficiently large relative to the others, one could have a local minimum in such a region. Nevertheless, as demonstrated in the appendix, the intuitive results obtain because, given that effort is at an optimum, the second-order condition implicitly limits the magnitude of effects due to the nonconstancy of the density functions.<sup>15</sup>

Regarding proposition 1.c, raising  $s$  has multiple influences. First, it raises the  $b^i$ , both of which favor lower enforcement effort. For harmful acts, the marginal deterrence gain is falling with the expected sanction (and thus with  $e$  or  $s$  considered separately, holding everything else constant) because the harm  $h$  is the same across harmful acts but the forgone benefit from deterring the marginal individual who otherwise commits the harmful type of act,  $b^H$ , is rising. For benign acts, the marginal chilling cost is rising with the expected sanction because the forgone benefit from chilling the marginal individual who otherwise commits the benign type of act,  $b^B$ , is rising. Second, raising  $s$  boosts the productivity of greater enforcement effort, which favors higher enforcement. This second effect, by itself, is not ambiguous at the optimum: although both marginal deterrence and marginal chilling rise, the welfare effect of the former necessarily dominates that of the latter because the former must equal the sum of the latter and the “plus one” that consists of the effort cost itself. Third, as was the case with propositions 1.a and 1.b, nonconstant density functions create additional effects, each of which in general can

---

<sup>15</sup>Note also that it is possible for there to be a corner solution involving  $e^* = 0$  if the marginal return from enforcement,  $\pi'(0)$ , is sufficiently low relative to the magnitude of  $h$ .

have any sign. In the appendix, examples are constructed in which raising  $s$  raises  $e^*$  and in which it reduces  $e^*$ .

Proposition 1.d follows from inspection of expression (3). Because the right side is necessarily positive, so must be the left side at the optimum, which indicates that  $h > b^H$ , which is to say that the harmful type of act is underdeterred relative to first-best behavior. This result is common in models of optimal enforcement due to the positive cost of raising enforcement effort. Here, there is the additional reason, the chilling cost (the first term on the right side), which implies that the optimal degree of underdeterrence is greater. Even if additional enforcement was free (in which case there would be no “plus one” on the right side of 3), one would stop short of first-best deterrence.

### B. Optimal Evidence Threshold

Our main policy variable of interest is the evidence threshold  $x^T$ . If we take the derivative of social welfare (1) with respect to  $x^T$  – keeping in mind that the  $p^i$  are functions of  $x^T$  so that the evidence threshold’s effect is through reducing the lower limits of integration – we obtain a first-order condition that can be expressed as

$$(4) \quad \pi s f^H(\pi p^H s)(h - \pi p^H s) = \gamma \pi s P^B f^B(\pi p^B s) \pi p^B s, \text{ or}$$

$$(5) \quad f^H(b^H)(h - b^H) = \gamma P^B f^B(b^B) b^B.$$

Lowering  $x^T$  is costless in terms of direct resource costs; its effects on social welfare are exclusively through behavior. On the left side of (4) or (5) is the social gain: the deterrence punch, indicated by the height of the density for those just deterred, times the net gain from deterring the marginal harmful act. (Note that  $\pi s$  was divided out of both sides in moving from (4) to (5), for convenience of later comparisons.) The right side indicates the social cost: the chilling punch – the relative mass of population that contemplates committing the benign act times the rate that  $p^B$  rises per unit of increase in  $p^H$  (as a consequence of the reduction in  $x^T$ ), times the density – all multiplied by the net benefit forgone from chilling the marginal benign act.

*Proposition 2: Given  $e$ , at an interior optimum for the evidence threshold,  $x^{T*}$ :*

- a. Raising  $h$  reduces  $x^{T*}$ ;
- b. Raising  $\gamma$  raises  $x^{T*}$ ; and
- c. If the distributions  $F^i$  are uniform, raising  $s$  raises  $x^{T*}$ .

These proofs are also in the appendix, but the core basic reasoning is articulated here. Raising  $h$  naturally favors a lower level of  $x^T$  because deterrence is more important. Raising  $\gamma$  favors a higher  $x^T$  because chilling is more significant. And, as with proposition 1, these intuitive results hold despite complications regarding nonconstancy of the density functions.

For proposition 2.c, note first that if  $\pi s$  is higher – whether due to higher enforcement effort or sanctions – then the optimal evidence threshold  $x^T$  will tend to be higher because, as explained with regard to proposition 1.c, the marginal net benefit of deterring harmful acts is reduced when deterrence is already higher and the marginal cost of chilling benign acts is raised when the marginal chilled individual’s benefit is higher. This claim is proved in the appendix with regard to changes in  $s$  for constant density functions. Nonuniform distributions of the benefits from either of the two types of acts complicate this phenomenon (but not so much 2.a and 2.b because, again, satisfaction of the second-order condition, expression (A6) in the appendix, guarantees the claimed properties). A higher  $f^H$  (locally) will favor a lower  $x^T$ , and a higher  $f^B$  (locally) will favor a higher  $x^T$ . The analysis of how changing the sanction influences the problem therefore depends on how the densities are changing locally. Specifically, if  $f^H$  is rising rapidly, then a slight increase in  $s$  notably increases the marginal deterrence punch, favoring a lower  $x^T$ ; likewise, if  $f^B$  is falling rapidly, then a slight increase in  $s$  notably reduces the marginal chilling punch, favoring a lower  $x^T$ . Accordingly, statement of the unambiguous result in proposition 2.c is limited to the case of a uniform density function.

### *C. Comparison of Raising Enforcement Effort and Lowering the Evidence Threshold*

It is illuminating to compare directly the choice between raising enforcement effort  $e$  and lowering the evidence threshold  $x^T$  as ways to deter harmful acts. Laxer proof requirements and greater enforcement effort are substitutes, but imperfect ones that differ along two dimensions. Compare the first-order conditions (5) and (3). First, as previously noted, reducing  $x^T$ , which raises  $p^H$ , is free of resource costs in the present setting, unlike enforcement effort. Second, abstracting from the densities, which are analogous in the two conditions, the behavioral punches are in the proportion  $p^H$  to  $p^B$  in (3) for enforcement effort but they are in the proportion 1 to  $P^B$  in (5) for the evidence threshold.

To represent these ideas explicitly, use the first-order condition (3) for optimal enforcement effort  $e$  to substitute into condition (5) for the optimal evidence threshold  $x^T$ . One way of doing so yields

$$(6) \quad \gamma \left( P^{B'} - \frac{p^B}{p^H} \right) f^B(b^B) b^B = \frac{1}{\pi' p^H s}.$$

The term in parentheses on the left side depicts the difference in the relative “effective” rise in  $p^B$  as one “effectively” increases  $p^H$ : that for raising  $p^H$  directly as a consequence of reducing  $x^T$  minus that for raising enforcement effort, which does not raise  $p^H$  or  $p^B$  per se but does increase their common multiplier  $\pi$ . This term is then multiplied by the marginal importance of chilling effects (the leading  $\gamma$  indicating relative importance of the benign type of act, also multiplied by the density and the forgone benefit from the marginal benign act that is chilled). In other words, the left side is the relative difference in chilling effects weighted by their importance. (Note that, in this formulation, the deterrence of harmful acts is implicitly held constant.) On the right side is the cost disadvantage of raising enforcement effort, normalized in units of deterrence punch (rise in the expected sanction on harmful acts). When both enforcement effort and the evidence

threshold are optimized, these magnitudes are equated.

*Proposition 3: Higher enforcement effort and a lower evidence threshold are substitutes in achieving a given level of deterrence, and the optimal mix of enforcement effort and evidence threshold is given by expression (6).*

Note further that the right side of (6) is necessarily positive, so it must be that, at the optimum,  $P^{B'} > p^B/p^H$ .<sup>16</sup> This feature indicates that, for a given increase in  $\pi p^H$ , there is a relatively greater rise in  $\pi p^B$  when one raises  $p^H$  directly, by lowering the evidence threshold  $x^T$ , than when one raises its ultimate impact by raising  $\pi$ , that is, by increasing enforcement effort. In this respect, lowering the evidence threshold  $x^T$  is a less advantageous means of increasing the deterrence of harmful acts than is raising enforcement effort  $e$ . The intuition is that lowering the evidence threshold assigns liability in cases of lower quality (a lower ratio of ones with harmful acts to ones with benign acts) than those previously giving rise to sanctions (which had higher values of  $x$ ). By contrast, raising enforcement effort brings more cases into the system across the evidence-quality spectrum, producing liability in cases of the same average quality as before.

This result can also be related to the comment at the outset of section 3 that optimal sanctions are maximal in this model. Raising sanctions dominates lowering the evidence threshold because it has the same superior discriminating power as does raising enforcement effort (but it is costless). Specifically, for a given increase in  $s$ , if  $x^T$  is increased to keep  $p^H s$  constant, the requirement that  $P^{B'} > p^B/p^H$  implies that  $p^B s$  falls. This point can be seen as a new version of the argument associated with Becker (1968) that favors maximal sanctions – traditionally, because it is generally optimal to raise sanctions and lower enforcement effort in achieving a given level of deterrence.<sup>17</sup> Here, the savings involve a reduction in chilling effects rather than in enforcement expenditures. It is interesting that the possibility of mistaken imposition of sanctions provides an additional reason favoring high sanctions even for acts causing low harm, which adds to the already serious tension between this prescription of many models of optimal law enforcement and observed practice.

Indeed, this consideration nullifies two of the main results in important papers by Mookherjee and Png (1992) and Shavell (1991). In their models, the same monitors observe multiple kinds of harmful acts whose degrees of harm differ (perhaps littering and mugging). When the single level of enforcement effort is optimized, the less harmful types of acts might be fully deterred by a less-than-maximal sanction, in which event the optimal sanction for such acts is set to achieve first-best deterrence. However, if one introduces benign acts that might be chilled and takes the evidence threshold to be an enforcement instrument, it instead would be optimal to raise sanctions to the maximum even for low-harm acts, while simultaneously raising the evidence threshold for them, in order to reduce chilling costs, as just explained. Chilling

---

<sup>16</sup>If one assumes that  $P^{B''} > 0$ , this condition is necessarily satisfied (even if not at the optimum, as long as the probabilities are not 0).

<sup>17</sup>And, just as the traditional Becker argument is known to imply that, as the maximum feasible sanction goes to infinity, optimal enforcement effort approaches zero, we have here that one can indefinitely raise the evidence threshold, driving down chilling effects.

costs also make underdeterrence optimal even though first-best deterrence can be achieved costlessly (recall proposition 1.d).

#### *D. Comparison to Conventional Burden of Proof Notions*

The foregoing results should be compared to those under, say, a preponderance of the evidence standard that is used in civil adjudication in the United States (and which has been the subject of some previous literature using quite different models, examined in section E).<sup>18</sup> Under a typical burden of proof notion, one asks whether, given the evidence, the likelihood that the individual before the tribunal committed the harmful act (the Bayesian posterior probability) exceeds some level, 50% under the preponderance standard. Suppose, then, that the tribunal observes signal  $x$ . Conditional on the individual's act being of type  $i$  (and the individual being scrutinized, where individuals committing both types of acts are scrutinized with the same probability  $\pi$ ), the likelihood of the signal is given by the density  $g^i(x)$ . The probability that a given individual before the tribunal committed the harmful rather than the benign type of act also depends on the base rates of the activities,  $1 - F^i(b^i)$ , further multiplied by  $\gamma$  for the benign act. The  $b^i$ 's, of course, are not parameters but are determined endogenously by individuals' decisions about acts, which in turn depend on the enforcement parameters, including the evidence threshold (burden of proof) itself.

To identify the proper threshold for the signal  $x$  if one wishes to implement a preponderance of the evidence rule, we can ask what value of  $x^T$  implicitly solves the equation:

$$(7) \quad g^H(x^T) \left(1 - F^H(b^H)\right) = \gamma g^B(x^T) \left(1 - F^B(b^B)\right),$$

where the  $b^i$ 's are determined assuming the use of this  $x^T$  (and some given  $\pi$  and  $s$ ), with the implied values of the  $p^i(x^T)$ . In expression (7), the left side indicates the frequency with which individuals who committed harmful acts (conditional upon being scrutinized and thus appearing before the tribunal) will be associated with the signal  $x^T$ , and likewise for the right side for individuals who committed benign acts. If the frequencies represented by the two sides of equation (7) are equal, then indeed  $x^T$  is a value of the signal such that it is equally likely that the individual before the tribunal committed each type of act. Accordingly, insisting that  $x$  exceed  $x^T$  satisfies the more-likely-than-not (preponderance) standard (with additional assumptions on the  $g^i$ ).<sup>19</sup> Note that if we desired a different likelihood at the threshold  $x^T$ , we could multiply by a further constant to the right side. For example, for a requisite likelihood of 75%, implying that the threshold signal was three times more likely to have been emitted from an individual who committed the harmful type of act than one who committed the benign type of act, one would multiply the right side by 3. Or if proof beyond a reasonable doubt were understood to require, say, a likelihood of at least 90% or 95%, one would multiply the right side by 9 or 19,

---

<sup>18</sup>For extensive discussion of how the present analysis applies to legal discussions of the burden of proof, see Kaplow (2012a).

<sup>19</sup>A sufficient condition is that  $P^{B''} > 0$ , which, recall from section 2, is implied by the density  $g^H(x)$  satisfying a strict version of the monotone likelihood ratio property with respect to  $g^B(x)$ . See note 12.

respectively.

To facilitate comparison with the equality determining the optimal evidence threshold, it is helpful to restate expression (7). Recall from section 2 that  $P^{B'} = dP^B/dp^H = [dp^B(x^T)/dx^T]/[dp^H(x^T)/dx^T] = g^B(x^T)/g^H(x^T)$ . Using this fact and rearranging terms yields

$$(8) \quad 1 - F^H(b^H) = \gamma P^{B'} \left(1 - F^B(b^B)\right).$$

Now compare expression (5), reproduced here for ease of comparison:

$$(5) \quad f^H(b^H)(h - b^H) = \gamma P^{B'} f^B(b^B)b^B.$$

Both have the constant  $\gamma$  on the right side, and both have the term  $P^{B'}$ . In all other respects, however, these expressions differ markedly. The densities on each side of (5) are for the distribution of benefits evaluated for the marginal individual, but no corresponding terms appear in (8). By contrast, each side of (8) is weighted by the base rate of the activity, indicated by one minus the value of the cumulative distribution function; no such terms appear in (5). This particular difference reflects the Bayesian, information-inference character of the conventional burden of proof notion, in contrast to the optimizing, behavioral orientation of the first-order condition, where it is the density functions for benefits that matter because they determine the marginal behavioral effects. Additionally, the level of harm and the benefits of the two types of marginal individuals appear directly in (5) but not in (8). In short, setting an optimal evidence threshold  $x^T$  and finding the cutoff that corresponds to any given conventional burden of proof notion are very different enterprises.

*Proposition 4:*

*a. The optimal evidence threshold, given implicitly by (5), generally differs from the preponderance of the evidence rule, given implicitly by (8) (or any similar rule denominated in terms of an ex post likelihood that an individual before the tribunal must have committed the harmful type of act).*

*b. At the optimal evidence threshold, the minimum ex post likelihood that an individual must have committed the harmful type of act in order to be subject to sanctions can take on any (admissible) value.*

*c. Raising the evidence threshold  $x^T$  may increase the frequency with which individuals who commit benign acts are mistakenly subject to sanctions.*

*d. Raising the evidence threshold  $x^T$  may result in a lower ex post likelihood that an individual must have committed the harmful type of act in order to be subject to sanctions.*

Proposition 4.a having been established, turn to 4.b, stating that when the evidence threshold is set optimally, the minimum ex post likelihood that an individual must have committed the harmful type of act in order to be subject to sanctions can take on any admissible value (i.e., between zero and one). To confirm this conclusion, note that condition (5) for the optimal burden of proof does not depend on the cumulative distribution functions  $F^i$  whereas

condition (8) depends on these functions and on little else. Hence, for a given solution to condition (5), one could make inframarginal adjustments to the distributions of benefits for the two types of acts to produce different values for the minimum ex post likelihood.<sup>20</sup> Perhaps more intuitively, observe that condition (5) implies that, as  $h$  goes to zero, the optimal evidence threshold will approach infinity (when all acts produce positive net social benefits, or at worst negligible net social harm, it should be essentially impossible for evidence to be strong enough to apply sanctions). Likewise, as  $h$  approaches infinity, the optimal evidence threshold will approach negative infinity (or the lower bound of the support).<sup>21</sup>

Proposition 4.c states that raising the evidence threshold  $x^T$  can increase on the frequency of mistaken applications of sanctions. Following traditional thinking about proof burdens, one might suppose that they would fall, but this need not be so. On one hand, raising  $x^T$  does make it less likely that unchilled individuals will be sanctioned for the benign acts that they commit. On the other hand, raising  $x^T$  reduces the chilling effect, so more individuals commit benign acts and thus may be subject to sanctions. Specifically, it can be shown that, if and only if  $b^B f^B(b^B) > 1 - F^B(b^B)$ , the latter effect is greater and thus a higher evidence threshold raises the number of false positives.

Before demonstrating proposition 4.d, consider first whether, for any stipulated requisite ex post likelihood, there exists an  $x^T$  that yields that likelihood. As one varies the value of  $x^T$ , a range of ex post likelihoods will be traced; since the model's assumptions guarantee that all behavior changes continuously, any ex post likelihood will be possible, subject to the minimum and maximum values that result from this exercise. The full range of ex post likelihoods, however, cannot necessarily be produced for a given set of parameters. Most obviously, for a requisite ex post likelihood of 100%, there would have to be no benign acts committed. Note further that the relationship between  $x^T$  and the requisite ex post likelihood need not be monotonic (given our lack of tight restrictions on the shape of the  $f^i$ ), which implies that, for a given likelihood, there may exist more than one choice of  $x^T$  that would generate it. To illustrate this point, suppose that the distribution of benign acts was concentrated at a very low level, but that for harmful acts was not. As  $x^T$  is reduced, which increases  $p^B$ , eventually most benign acts will be chilled, so the ex post likelihood will approach 100%. From there, demanding stronger evidence, that is, raising the threshold  $x^T$ , would in such a case (for at least some range) reduce the ex post likelihood. This establishes proposition 4.d.

#### *E. Modification: Model with No Benign Acts*

As mentioned in the introduction, prior literature (cited in note 4) offers a prescription for the optimal evidence threshold that differs markedly from that derived here. Moreover, some of this work purports to find that the preponderance rule is optimal. To explain these differences, consider a simpler model in which there are no benign acts – only harmful acts and a failure to

---

<sup>20</sup>One way to do this is to consider a case in which the two distributions are uniform; the endpoints of each of the intervals can then be shifted (while keeping the pertinent  $b^i$ 's in the intervals) to produce any desired result.

<sup>21</sup>As explained just below with regard to proposition 4.d, spanning the full range of  $x^T$  need not be sufficient in a given case to guarantee that all admissible ex post likelihoods are spanned.

act. Suppose further, as in these other papers, that scrutiny covers inaction. Accordingly, the probability of mistaken sanctions,  $p^B$ , is now taken to apply to inaction (but for convenience the same notation is retained).<sup>22</sup>

Specifically, all individuals are scrutinized with probability  $\pi(e)$ . Those who commit the harmful act, as before, are (conditional on being scrutinized) subject to the sanction  $s$  with probability  $p^H$  whereas those who do not commit the act (but are scrutinized) now face the sanction with probability  $p^B$ . Accordingly, an individual will commit the act if and only if  $b - \pi p^H s > -\pi p^B s$ , which is equivalent to the condition  $b > \pi(p^H - p^B)s$ . Because individuals who do not commit the harmful act might be sanctioned in any event, deterrence is reduced. (There is no other cost to mistaken sanctions; the only benign behavior that is chilled is inaction.) It is convenient to represent  $p^H - p^B$  by  $\Delta p$ , since only the difference in the two probabilities matters.

Social welfare in this modified problem, denoted by a “ $\sim$ ”, is

$$(9) \quad \tilde{W} = \int_{\pi(e)\Delta p(x^T)s}^{\infty} (b - h) f^H(b) db - e.$$

The optimal  $x^T$  is determined implicitly by

$$(10) \quad \pi \left( 1 - P^B \right) f^H(b^H) (h - b^H) = 0.$$

Obviously, the optimum is characterized by  $P^B = 1$ , which is to say that  $\Delta p$ , the gap between the  $p^i$ , is maximized (assuming that  $\pi$  and  $s$  are such that there is underdeterrence, which is optimal since raising  $\pi$  is costly; recall proposition 1.d). As mentioned in the introduction, this result is essentially equivalent to the core finding on optimal proof burdens in prior models that in a relevant sense have only one act.<sup>23</sup>

---

<sup>22</sup>See also Kaplow (2011).

<sup>23</sup>In some, the choice variable is a dichotomous level of care – engagement in the pertinent activity being taken as given – where it is always desirable to take greater care and underdeterrence is present, making it optimal to maximize the difference between the likelihood of the sanction when taking low care versus high care. A related feature of these sorts of models is that the costs of false positives and of false negatives are the same: reducing  $p^H$  hurts deterrence by the same degree as does raising  $p^B$  by the same amount. Thus, the practice in some work of arbitrarily assigning the same costs to each type of error is unproblematic in such a model, although in the model explored in the rest of this article, the costs of the errors are not symmetric. Moreover, it is sometimes remarked that it is problematic to assign a given cost to a type of error since the social cost of the error is nonconstant. That is true even in the model in this subsection: the cost of either type of error is reduced deterrence, which has a net social cost of  $h - b^H$  for the marginal act. As one changes the evidence threshold (or other control variables), this net marginal cost is changing. But that change is also irrelevant because, throughout, it remains symmetric (so stipulating equal losses for each type of error yields the correct answer for the optimum even if the assigned levels of the losses are incorrect – for setting the burden of proof, but not for setting, say, the optimal level of enforcement effort). By contrast, in the main model used in the rest of section 3, the error costs are not merely nonconstant but most typically change in opposite directions: as one raises enforcement effort or reduces the evidence threshold, the benefit of

We can now compare this result with the first-order condition (5). Although (5) has  $P^{B'}$  on the right side, it also has six other components. Therefore, the optimal evidence threshold in prior literature indeed differs radically from that derived in this article's model, as a consequence of introducing the possibility of chilling benign acts.

Finally, consider implications regarding the preponderance of the evidence rule. The analogue to expression (8) for the present case is

$$(11) \quad 1 - F^H(b^H) = P^{B'} F^H(b^H).$$

Some literature has stated that the preponderance rule implies the maximization of  $\Delta p$ , which we just saw entails choosing  $x^T$  such that  $P^{B'} = 1$ . However, this result is not implied by expression (11), which instead equates  $P^{B'}$  with the ratio of the base rates. For example, for acts that few people commit, this ratio would be near zero, implying an extremely high evidence threshold (assuming that  $P^{B''} > 0$ ), that is,  $P^{B'}$  much less than 1. What is really contemplated, if deterrence is to be maximized, is that base rates be ignored. Indeed, sometimes reference is made to application of a modified preponderance rule under which equal base rates (or flat or unbiased priors) are stipulated.<sup>24</sup> In that case, expression (11) would be equivalent to  $P^{B'} = 1$ . Nevertheless, the preponderance rule is an ex post Bayesian concept, one that fundamentally conflicts with ignoring ex post likelihoods and considering instead deterrence maximization, which here depends only on ex ante probabilities, the  $p^i$  (to be precise, the ratio of their derivatives). This semantic point is reinforced by considering the main model and the analysis of the preponderance rule in section D. If one examines expression (8) and assumes equal base rates, one again produces the conclusion that  $P^{B'} = 1$ , which, as mentioned, bears little resemblance to the optimal rule given by expression (5).

#### 4. Model II: Investigation

##### A. Model

The model presented in section 2 supposes that the method of enforcement is akin to the posting of monitors that enable the authority to scrutinize the fraction  $\pi(e)$  of all acts (of both types, since initially it cannot tell the difference). Now suppose instead that enforcement involves undertaking investigation when a harmful act is committed, the assumption being that whether a harmful act occurred can be readily observed (for example, murder, automobile theft, or ground water contamination) but not who committed the harmful act. Let  $\pi^H(e)$  denote the fraction of harmful acts in which the correct individual is scrutinized, i.e., the person who actually committed the harmful act. Further suppose that, for each harmful act, the likelihood that an individual who committed the benign type of act is subject to scrutiny is

---

detering an additional harmful act is falling whereas the cost of chilling an additional benign act is rising (with the total marginal social consequence also depending on the magnitude of the pertinent density function, each of which may be rising or falling).

<sup>24</sup>See, for example, Demougin and Fluet (2006, pp. 970–71).

$\pi^B(e)\gamma(1-F^B(b^B))$ ). The rationale for the additional factors, elaborated further just below, is that an investigation is more likely to lead to the apprehension of an individual who committed a benign act when there are more such individuals. It is further assumed that  $\pi^{i'} > 0$ ,  $\pi^{i''} < 0$  for  $i = H, B$ .<sup>25</sup> As before, conditional on scrutiny, the probability of being subject to the sanction  $s$  is  $p^H$  if the individual before the tribunal actually committed the harmful act and  $p^B$  if the individual before the tribunal committed a benign act, with the magnitude of each determined by the evidence threshold  $x^T$  (and so forth). Finally, the timing is also as in model I: the government sets  $e$ ,  $s$ , and  $x^T$ ; next, individuals learn their type of act and their private benefits, after which they decide whether to act; then, a portion are scrutinized and, in adjudication, are sanctioned if and only if  $x > x^T$ .

Once again, the expected sanction for those who commit the harmful type of act is  $\pi^H p^H s$ , which likewise indicates the benefit level  $b^H$  of the marginal individual. For the benign type of act, the analysis is more involved. Investigations are only potentially triggered with probability  $1-F^H(b^H)$ . (This is also true for those committing harmful acts, but when an individual commits a harmful act, that person knows that an investigation may be triggered, leading to scrutiny of the actor with probability  $\pi^H(e)$ .) Conditional upon an investigation, the probability that *some* individual who committed a benign act is scrutinized is, as mentioned,  $\pi^B(e)\gamma(1-F^B(b^B))$ . Accordingly, the ex ante frequency with which individuals who commit benign acts are subject to sanctions is  $(1-F^H(b^H))\pi^B(e)\gamma(1-F^B(b^B))p^B$ . Now, since  $\gamma(1-F^B(b^B))$  individuals commit benign acts, the ex ante probability that *any given* individual who commits a benign act is sanctioned is  $(1-F^H(b^H))\pi^B(e)\gamma(1-F^B(b^B))p^B / \gamma(1-F^B(b^B)) = (1-F^H(b^H))\pi^B(e)p^B$ , which will sometimes be referred to as  $\rho^B$ .<sup>26</sup> Thus, the expected sanction for those who commit the benign act is  $(1-F^H(b^H))\pi^B(e)p^B s$ , or  $\rho^B s$ , either of which likewise indicates the benefit level  $b^B$  of the marginal individual.

At this point, we can state the social welfare function for this modified problem.

---

<sup>25</sup>It may also be natural to suppose that  $\pi^H + \gamma\pi^B < 1$ .

<sup>26</sup>The assumption that the likelihood an individual committing the benign act is scrutinized is  $\pi^B(e)\gamma(1-F^B(b^B))$  is chosen in part for convenience, which is now apparent. It does seem natural to allow this probability to be rising (even if not linearly) in  $\gamma(1-F^B(b^B))$ . A more general approach, which also incorporates the later need to divide the expression by  $\gamma(1-F^B(b^B))$ , would employ the function  $\pi^B(e, \gamma(1-F^B(b^B)))$ . If this was done instead, the only effect on the analysis would be that expressions (14) and (16) for the derivatives of  $\rho^B$  with respect to  $e$  and  $p^H$ , respectively, would each have an additional term (the same term) dividing them, indicating that the magnitude of the derivative would be dampened (enlarged) if the derivative of this modified  $\pi^B$  with respect to its second argument was positive (negative), basically allowing for a feedback effect of marginal changes in the chilling effect on the ex ante likelihood that individuals committing the benign act are scrutinized. Accordingly, little of analytical consequence is sacrificed by employing the particular functional form adopted in the text.

$$(12) \hat{W} = \int_{\pi^H(e)p^H(x^T)s}^{\infty} (b-h)f^H(b)db + \gamma \int_{(1-F^H(\pi^H(e)p^H(x^T)s))\pi^B(e)p^B(x^T)s}^{\infty} bf^B(b)db - (1-F^H(\pi^H(e)p^H(x^T)s))e.$$

This expression differs in two ways from expression (1) for social welfare in the original (monitoring) version of the model. First, the lower limit of integration for the second term reflects the immediately preceding discussion concerning the likelihood that individuals who commit the benign act will be subject to sanctions; in particular, this probability includes the factor  $1-F^H(b^H)$  because investigations are only triggered when harmful acts are committed. Second, the enforcement effort cost  $e$  has a similar weighting.

Before proceeding, note that the standard argument associated with Becker (1968) that demonstrates the necessary optimality of maximal sanctions need not hold in this extension. As one raises  $s$  and lowers  $e$  so as to keep the deterrence of harmful acts constant, it is not necessarily true that the chilling of benign acts is also kept constant (and it may rise) because it is not assumed that  $\pi^{H'}/\pi^H = \pi^{B'}/\pi^B$ . The new variant of the argument for maximal sanctions introduced in section 3.C likewise does not suffice to guarantee the optimality of maximal sanctions for similar reasons, unless further assumptions are introduced.<sup>27</sup> In any case, for the remainder of this section, the sanction  $s$  will be taken to be fixed. The government's problem will be to choose  $e$  and  $x^T$  to maximize social welfare (12).

### B. Optimal Enforcement Effort

The first-order condition for enforcement effort  $e$  may be expressed (by analogy to expression (3) for the original model) as

$$(13) \pi^{H'} p^H s f^H(b^H) [(h-b^H) + e] = \gamma \frac{d\rho^B}{de} s f^B(b^B) b^B + (1-F^H(b^H)),$$

where

---

<sup>27</sup>If one raises a nonmaximal sanction  $s$  and raises the evidence threshold  $x^T$  so as to keep deterrence constant (specifically,  $p^H s$  is kept constant, and  $\pi^H$  is unaffected), the value of the first integral and of the third term do not change; the resulting reduction in  $p^B$ , however, may not be great enough to offset the effect of the increase in  $s$  on  $p^B s$ , so chilling can rise. In model I, condition (6) sufficed to guarantee that  $P^{B'} > p^B/p^H$  at the optimum and thus that  $p^B s$  would fall. However, the analogue to expression (6) for the investigation model, expression (17) below, differs. As stated in the text, the traditional Becker argument can fail when  $\pi^{H'}/\pi^H > \pi^{B'}/\pi^B$ . In that situation,  $\pi^B/\pi^H > \pi^{B'}/\pi^{B'}$ , so the term in large parentheses in (17) being positive does not guarantee that  $P^{B'} > p^B/p^H$ .

$$(14) \frac{d\rho^B}{de} = \pi^{B'} p^B (1 - F^H(b^H)) - \pi^B p^B \pi^{H'} p^H sf^H(b^H).$$

These expressions differ from expression (3) in a number of ways. On the left side of expression (13), there is the additional “plus  $e$ ” term in brackets, which indicates that there is a greater marginal benefit from deterring harmful acts, namely, that fewer investigations are undertaken, which saves enforcement expenditures. Second, on the right side, instead of a “plus 1” term, we have  $1 - F^H(b^H)$ , indicating that the additional enforcement effort is weighted by the fraction of undeterred harmful acts, since only those trigger investigations.

Inspection of expression (14) reveals two additional differences. First, the effect of increased effort on chilling, reflected in the term leading with  $\pi^{B'}$ , is likewise weighted by  $1 - F^H(b^H)$ , reflecting that chilling is muted since only acts causing harm trigger investigations (that might in turn go awry by ultimately resulting in sanctions on benign acts). Moreover, the chilling effect that does exist is reduced, as reflected in the second term, on account of greater deterrence of harmful acts. That is, increased effort, by deterring harmful acts, reduces investigations at the margin, which further reduces chilling costs. (Because of its genesis, it is natural to think of the second term in expression (14) as associated with the left side of (13) rather than the right side, a point reinforced by comparing the terms explicitly.)

In light of these differences, it seems natural to view enforcement effort as more valuable at the margin when enforcement is by investigation. However, the two models are, at heart, not comparable, because they reflect different technologies applicable in different circumstances. For example, nothing has been said about the relationship between  $\pi(e)$  in our original model and the  $\pi^i(e)$ 's in the present model. Nevertheless, these qualitative differences do suggest that the different enforcement methods have different relative strengths and weaknesses, many of which pertain to chilling effects in particular.

Finally, observe that, unlike in our original setup and in many standard law enforcement models, it need not be true that there is underdeterrence (relative to first-best behavior) at the optimum. Even if  $h < b^H$ , there remain the other two marginal benefits of greater deterrence: reduced enforcement expenditures and reduced chilling, both on account of there being fewer harmful acts to investigate.

### *C. Optimal Evidence Threshold*

By analogy to expressions (4) and (5), the first-order condition for  $x^T$  is

$$(15) \pi^H sf^H(b^H) [(h - b^H) + e] = \gamma \frac{d\rho^B}{dp^H} sf^B(b^B) b^B,$$

where

$$(16) \frac{d\rho^B}{dp^H} = \pi^B \left[ P^{B'} (1 - F^H(b^H)) - p^B \pi^H s f^H(b^H) \right].$$

The differences between this first-order condition and expression (5) largely parallel those given for the optimal level of enforcement effort. In expression (15), we have the additional “plus  $e$ ” on the left side, indicating the additional advantage from deterrence of reducing enforcement expenditures. Through expression (16), we have two additional effects that pertain to chilling. First, the direct chilling effect, reflected by  $P^{B'}$  as in the original model (where the ratio of effects on harmful versus benign acts again depends on how the rate of mistaken sanctioning rises with the rate of correct sanctions as the evidence threshold is relaxed), is now weighted by  $1 - F^H(b^H)$  because the mistaken imposition of sanctions can only happen when an investigation is triggered in the first instance. Relatedly, the second term in (16) indicates the marginal reduction in chilling effects on account of enhanced deterrence from the increase in  $p^H$  that results from a reduction in  $x^T$ . As with the interpretation of the first-order condition for optimal enforcement effort, we cannot really say that, because we have three more favorable factors, a lower  $x^T$  is optimal when enforcement is by investigation. However, we again see that the nature of the optimization problem differs qualitatively and in important respects that relate directly to the concern with the chilling of benign acts.

One can also construct the analogue to expression (6), which compares raising enforcement effort and lowering the evidence threshold in achieving a given level of deterrence, by combining the first-order condition for  $e$ , expressions (13) and (14), and that for  $x^T$ , expressions (15) and (16), which yields

$$(17) \gamma \left( P^{B'} \left( \frac{\pi^B}{\pi^H} \right) - \frac{p^B}{p^H} \left( \frac{\pi^{B'}}{\pi^{H'}} \right) \right) f^B(b^B) b^B = \frac{1}{\pi^{H'} p^H s}.$$

Note that the fractions  $\pi^B/\pi^H$  and  $\pi^{B'}/\pi^{H'}$  were both equal to one in model I.

Additionally, the first-order condition for  $x^T$  may usefully be compared to the analogue to expression (8) for the preponderance of evidence rule:

$$(18) 1 = \gamma \frac{\pi^B}{\pi^H} P^{B'} (1 - F^B(b^B)).$$

This expression differs from expression (8) for the monitoring technology in two ways. Most apparent, the term  $1 - F^H(b^H)$  no longer appears on the left side because, with the investigation technology, it is known that a harmful act was committed. (Put another way, although the aggregate flow of individuals who commit harmful acts and end up before the tribunal has the factor  $1 - F^H(b^H)$ , this factor also appears in the aggregate flow of individuals who commit benign acts and end up before the tribunal, so the effects cancel.) Second, the fraction  $\pi^B/\pi^H$  on the right

side here simply equalled one under the assumptions for the monitoring technology. (The origin of this fraction is that the base rate for those committing the harmful act – on the left side to the analogue to expression (7), which is not produced here – has the factor  $\pi^H$ , and the right side has  $\pi^B$ .) The differences between the preponderance of the evidence rule (18) and the first-order condition (15 and 16) are even more striking with the investigation technology. For example, the preponderance rule, as before, depends on the distribution function for benign acts whereas the first-order condition again does not; however, the preponderance rule here does not depend on the distribution function for harmful acts whereas the first-order condition here does, both reversed from the prior case. Let us now state:

*Proposition 5: When enforcement is by investigation:*

a. *The optimal evidence threshold, given implicitly by (15) and (16), generally differs from the preponderance of the evidence rule, given by (18) (or any similar rule denominated in terms of an ex post likelihood that an individual before the tribunal must have committed the harmful type of act).*

b. *Raising the evidence threshold  $x^T$  may increase the frequency with which individuals who commit benign acts are mistakenly subject to sanctions.*

For proposition 5.b, consider for this model the effect of raising the evidence threshold  $x^T$  on the frequency of mistaken applications of sanctions. For the monitoring model, the result was indeterminate (recall proposition 4.c) because benign acts that are committed face a lower frequency of sanctions but the number of benign acts that are committed increases. In the present model, this net effect is multiplied by another factor, itself of indeterminate sign, which just equals  $d\rho^B/dp^H$ .<sup>28</sup> The intuition is that, in the monitoring model, raising  $x^T$  unambiguously reduces  $b^B$ , the benefit level of the marginal chilled individual; here, whether that happens or the opposite depends on how  $\rho^B$  changes, which depends not only on the sign of  $P^{B'}$  as before (which is unambiguously positive), but also on additional terms. Specifically, examining expression (16), we can see two competing effects: on one hand, raising  $x^T$ , which reduces  $p^H$  and also  $p^B$ , will reduce the frequency with which individuals who commit the benign act are sanctioned; on the other hand, raising  $x^T$ , which reduces  $p^H$ , will reduce the deterrence of harmful acts, thereby resulting in more investigations and thus more opportunities for innocent individuals to be mistakenly subject to sanctions. Combining the above, we have the product of two terms of indeterminate sign – the composite term from the monitoring case (summing the competing inframarginal and marginal effects from a change in the expected sanction on those committing benign acts) and  $d\rho^B/dp^H$  – which is itself indeterminate in light of the fact that each depends on largely separate factors. (The former, recall, has its sign indicated by whether  $b^B f^B(b^B) > 1 - F^B(b^B)$ , which can be contrasted with expression (16).) One might suppose that a common case will be one in which the component common with the monitoring model will be

---

<sup>28</sup>The change in the expected number of mistaken applications of sanctions as one raises  $p^H$ , which has a sign opposite to that of raising  $x^T$ , is given by

$$\gamma \frac{d\rho^B}{dp^H} \left[ \left( 1 - F^B(b^B) \right) - f(b^B) b^B \right].$$

In the monitoring model, in place of  $d\rho^B/dp^H$ , one has  $\pi^{P^{B'}}$ .

such that the dominant effect involves the existing pool of unchilled individuals (perhaps most individuals commit their benign acts), with the change in the size of the pool being relatively unimportant. For that case, in model I with monitoring, raising  $x^T$  would indeed reduce the frequency of false positives. However, in model II with investigation, it seems plausible that, especially when the evidence threshold is fairly high, deterrence effects regarding harmful acts would dominate so that tighter proof requirements would result in more benign acts being mistakenly subject to sanctions.

## 5. Extensions

### *A. Auditing (Model III)*

This subsection briefly describes a third model: enforcement by auditing. Regarding behavior, the analysis is like that in model I. Regarding enforcement costs, there are similarities to model II: there, the pool of acts triggering costly scrutiny consists of harmful acts, which likewise trigger costly audits here. The difference here is that benign acts also lead to enforcement costs being incurred because they too are subject to audits.

To be explicit, modify model I as follows. The probability of scrutiny (here, an audit) will continue to be denoted by  $\pi$ , so as before we have  $b^i \equiv \pi p^i s$  for  $i = H, B$ . The enforcement variable  $e$  will now be interpreted as the expenditure required per audit. (It is natural to consider the case in which  $\pi'$  is constant, corresponding to costs per audit being invariant to the level of audits, but we will not restrict attention to that case for the sake of generality and to facilitate comparability across cases.) Social welfare for the auditing model is as in expression (1), except that  $e$  in the last term is weighted by the number of audits (the audit probability  $\pi$  times the size of the pool subject to audit, which is the total number of individuals who commit either type of act). See expression (A10) in the appendix.

Note that, as with the monitoring model (but unlike with the investigation model), the standard Becker (1968) argument applies. Note further that the new version of the Becker argument introduced in section 3.C does not hold in this case: even though raising a nonmaximal sanction  $s$  while raising the evidence threshold  $x^T$  so as to keep deterrence constant will, as before, reduce the chilling effect, which in and of itself is desirable, reduced chilling also means that more individuals commit benign acts, which increases the size of the audit pool, thereby raising enforcement costs (even though  $e$  and thus  $\pi$  are held constant).

The first-order conditions corresponding to expressions (3), (5), and (6) for this model also appear in the appendix (A11–A13). There are two sets of modifications. Regarding marginal effects of the enforcement instruments on behavior, deterrence is now more valuable (making it possible that the optimum involves overdeterrence relative to first-best behavior) and chilling less costly, in both cases because fewer acts lead to fewer audits, which reduces enforcement expenditures. (Indeed, chilling is net desirable when  $b^B < \pi e$ , that is, when the benefit of the marginal benign act is less than the expected audit cost.) This pair of effects makes both greater enforcement effort and a lower evidence threshold more attractive. Inframarginally, greater enforcement expenditures (and thus a higher audit rate) raise costs for

those who commit either type of act, which makes raising enforcement less attractive. One should, however, keep in mind the caveat presented in section 4's analysis of the investigation model: that one cannot formally compare results across models because different technologies and settings are envisioned. Finally, observe that because, as noted at the outset, behavioral effects themselves are the same as in model I with monitoring, raising the evidence threshold  $x^T$  in this auditing model can increase the frequency of mistakenly sanctioning benign acts. In light of the foregoing analysis – supplemented by that in section 3 on monitoring, where behavior is determined identically – we can state (by analogy to proposition 4) that:

*Proposition 6: When enforcement is by auditing:*

*a. The optimal evidence threshold, given implicitly by (A12), generally differs from the preponderance of the evidence rule, given implicitly by (8) (or any similar rule denominated in terms of an ex post likelihood that an individual before the tribunal must have committed the harmful type of act).*

*b. At the optimal evidence threshold, the minimum ex post likelihood that an individual must have committed the harmful type of act in order to be subject to sanctions can take on any (admissible) value.*

*c. Raising the evidence threshold  $x^T$  may increase the frequency with which individuals who commit benign acts are mistakenly subject to sanctions.*

*d. Raising the evidence threshold  $x^T$  may result in a lower ex post likelihood that an individual must have committed the harmful type of act in order to be subject to sanctions.*

Before concluding this subsection, it is worth observing that the additional features of this auditing model are also pertinent in realistic settings with monitoring. For example, when police patrols identify possible perpetrators, there may need to be follow-up investigation or at least some sort of adjudication in order to generate and assess evidence, which obviously entails costs, such as are included here. Specifically, the costs are triggered in all cases where the monitor ultimately picks out an act, which is likewise true with regard to the decision to conduct an audit of some act. Thus, monitoring is also likely to be more costly when more acts (of both types) are committed, giving rise to the additional effects here attributed to a model with auditing.

### *B. Socially Costly Sanctions*

Suppose that sanctions, instead of being costless transfers, entail direct social costs.<sup>29</sup> Notably, imprisonment is socially costly both because individuals' utility loss is not offset by a corresponding gain and because the operation of prisons consumes resources. In particular, assume that each unit of the sanction  $s$  that is imposed entails a social cost of  $\sigma$ . The resulting expression for social welfare (in the monitoring model) is:

---

<sup>29</sup>For prior analyses of socially costly sanctions (in models without errors), see Polinsky and Shavell (1984) and Kaplow (1990); and for discussion with error but in a single-decision framework like that in section 3.E, see Kaplow (1994). On risk aversion, see note 12.

$$(19) \tilde{W} = \int_{\pi(e)p^H(x^T)s}^{\infty} (b - h - \pi(e)p^H(x^T)s\sigma) f^H(b) db$$

$$+ \gamma \int_{\pi(e)p^B(x^T)s}^{\infty} (b - \pi(e)p^B(x^T)s\sigma) f^B(b) db - e.$$

Compared to our original expression (1), each integrand now subtracts a term  $\pi(e)p^i(x^T)s\sigma$ , which is the expected sanction for the type of act multiplied by the social cost per unit of sanction imposed. This modification obviously makes the commission of both the harmful and the benign type of act less socially desirable.<sup>30</sup>

The first-order conditions corresponding to expressions (3), (5), and (6) for, respectively, the optimal choices of  $e$  and  $x^T$  and the combination indicating the relative desirability of enforcement effort versus evidence threshold adjustment are as follows:

$$(20) \pi' p^H s f^H(b^H) (h - (1 - \sigma)b^H) = \gamma \pi' p^B s f^B(b^B) (1 - \sigma)b^B + 1$$

$$+ \pi' s \sigma \left[ p^H (1 - F^H(b^H)) + \gamma p^B (1 - F^B(b^B)) \right].$$

$$(21) f^H(b^H) (h - (1 - \sigma)b^H) = \gamma P^{B'} f^B(b^B) (1 - \sigma)b^B$$

$$+ \sigma \left[ (1 - F^H(b^H)) + \gamma P^{B'} (1 - F^B(b^B)) \right].$$

$$(22) \gamma \left( P^{B'} - \frac{P^B}{p^H} \right) \left[ f^B(b^B) (1 - \sigma)b^B + \sigma (1 - F^B(b^B)) \right] = \frac{1}{\pi' p^H s}.$$

*Proposition 7: In the monitoring model, introducing or raising a social cost of sanctions has ambiguous effects on optimal enforcement effort, the optimal evidence threshold, and the relative desirability of increasing enforcement effort versus reducing the evidence threshold in achieving a given level of deterrence.*

---

<sup>30</sup>Also of interest is that the traditional Becker (1968) argument for maximal sanctions holds because, since the product  $\pi s$  is held constant, sanction costs are also held constant. By contrast, the new variation presented in section 3.C need not hold because raising a nonmaximal sanction and increasing the evidence threshold in a way that keeps deterrence constant continues to have the benefit of reducing chilling costs but this reduction means that more benign acts are committed and thus greater sanction costs will be incurred. Put more formally, expression (22) does not (as with expression (6)) guarantee that  $P^{B'} > p^B/p^H$ .

To see why these ambiguous results follow, observe that, in each of these three optimality conditions, we have two new sets of effects. All conditions have an added term that includes the factors  $\sigma$  and  $1-F^i(b^i)$ . These indicate that both greater enforcement effort  $e$  and a lower evidence threshold  $x^T$  raise the expected sanction on individuals who continue to commit both the harmful and the benign types of acts, which is now socially costly in itself. From the final terms on the right sides of expressions (20) and (21), this effect favors a lower  $e$  and a higher  $x^T$ , as one might have expected.

However, there is a second sort of effect regarding the marginal consequences of deterrence and chilling, indicated by the  $1-\sigma$  before  $b^H$  on the left sides of (20) and (21) and before  $b^B$  on the right sides. As expected sanctions rise, more individuals with both types of acts are discouraged from acting, producing now a marginal social gain per individual who no longer acts equal to the expected sanction (which just equals the benefit of the respective marginal individual) times the social cost per unit of sanction  $\sigma$ . This effect favors a higher  $e$  and a lower  $x^T$ , which may be contrary to what one would have anticipated (compare, for example, Lando 2002, p. 606). Even the chilling of desirable acts now has a benefit: that fewer individuals are subject to the socially costly sanction.

Because the first effect, changing the level of sanctions imposed on individuals who nevertheless act, is inframarginal, depending on the cumulative distribution functions,  $F^i$ , and the second effect just described is marginal, depending on the densities,  $f^i$ , among other factors, costly sanctions have an ambiguous effect on optimal enforcement effort and the evidence threshold. To confirm that either set of competing effects could be larger, observe that, at  $\sigma = 0$ , the terms with the  $F^i$  vanish. Then, as we raise  $\sigma$  slightly from 0, the magnitude of the inframarginal effects can be small or large in a way that depends on parts of the distribution that do not affect the original optimum.<sup>31</sup>

Expression (22) indicates that, for the same reasons, the effect on the relative desirability of using enforcement effort versus a lower evidence threshold to achieve a given level of deterrence is ambiguous. On one hand, lower evidence thresholds (which are free) disproportionately target benign acts relative to greater enforcement effort (which is costly), and this now has a further disadvantage of increasing sanction costs on individuals who are not chilled. On the other hand, there is a new advantage in that the net social marginal cost of chilling benign acts is reduced on account of the resulting saving in sanction costs.

The foregoing analysis, particularly of expression (21) for the optimal evidence threshold, thus upsets conventional wisdom that costly sanctions in the criminal context provide a clear justification for the higher burden of proof employed there. One can also slightly modify the foregoing extension to capture additional considerations. For example, another ex post consequence of imprisonment and certain other sanctions (e.g., suspensions of permission to engage in activities, such as through license revocation) is to reduce the incidence of harmful acts by making individuals who might otherwise commit them unable to do so for a period of

---

<sup>31</sup>Furthermore, in each expression, (20) and (21), each marginal term,  $b^i f^i(b^i)$ , has precisely the same weight as its corresponding inframarginal term,  $1-F^i(b^i)$ .

time. Consider the case in which the reduction in social harm is proportional to the magnitude of the sanction, in which event the  $\sigma$  in the first integrand of expression (19) for social welfare would be replaced with a lower value,  $\sigma^-$ . (If the incapacitation gain were sufficiently great,  $\sigma^-$  would be negative.) The value of  $\sigma$  remains unaltered in the second term because incapacitation of individuals who commit benign acts is assumed not to reduce the commission of harmful acts. The result of this change would be to accordingly reduce the magnitudes of  $\sigma$  to  $\sigma^-$  on the left sides of expressions (20) and (21), and this lower  $\sigma^-$  would multiply the first term in square brackets on the right sides (with  $\sigma$  continuing to weight the second term). We can immediately see that, as the consequence of these two sets of changes, the effects of an incapacitation benefit on optimal enforcement effort and on the optimal evidence threshold are ambiguous. Just as raising the social cost of sanctions has an ambiguous effect on how these instruments should be set, so does reducing this social cost, even to a negative value (a net social benefit) for just those committing harmful acts.

Finally, although the analysis throughout this article has taken a standard welfarist approach under which social welfare is determined by the benefits and harms of acts, enforcement costs, and now the costs of imposing sanctions, the present extension allows one as well to characterize optimal policy if society cares per se about the correct and incorrect imposition of sanctions, as is commonly suggested (especially by noneconomists). Specifically, suppose that social welfare is higher by some amount per unit of correctly imposed sanction and lower by some other amount per unit of mistakenly imposed sanction (without regard to behavioral effects, resource costs, and so forth). By simply altering the magnitudes of  $\sigma^-$  and  $\sigma$  to reflect these considerations, the reinterpreted versions of expressions (20) through (22) would apply; in particular, the modified expression (21) would indicate the optimal evidence threshold. Yet again, the effects on optimal policy are ambiguous, contrary to the conventional view that, for example, associating a greater cost with mistakenly sanctioning the innocent favors a higher evidence threshold. Finally, if one cared *only* about such ex post effects, one could set  $\sigma^-$  to a negative value equaling the social gain per unit of correctly imposed sanction,  $\sigma$  to the positive value for the social cost per unit of mistakenly imposed sanctions (regarding the signs, recall that the  $\sigma$ 's were originally presented as costs, not benefits), and remove the other terms relating to behavior and enforcement costs. It is not suggested that this is a plausible social objective, and the results have some peculiar implications,<sup>32</sup> but this discussion indicates how the present framework readily allows one to analyze a variety of objective functions.

---

<sup>32</sup>For example, under the posited ex post objective function, deterrence of harmful acts turns out to be detrimental because society obtains its benefit from the correct imposition of sanctions less often, and chilling desirable acts is beneficial because society less often suffers the welfare loss from incorrectly sanctioning benign acts. For criticism of nonwelfarist approaches, including that they can violate the Pareto principle, see Kaplow and Shavell (2001, 2002). Some have suggested to me an objective function that only cares about the relative effects in cases that actually come before the tribunal. Again, deterrence could be undesirable and chilling desirable because both “worsen” the priors in that a lower fraction of cases before the tribunal involve true positives. Moreover, the optimal overall enforcement scheme would involve virtually no prosecutions (and hence essentially no deterrence), bringing only the occasional case in which one was truly certain that the individual committed a harmful act. Also, welfare would improve by rounding up many innocent individuals and subsequently acquitting them, because the outcomes ratio would improve. Accordingly, even by reference to objective functions that focus on ex post treatment, it is not that easy to attempt to rationalize the existing legal system, including an ex post Bayesian burden of proof concept.

### C. Accuracy<sup>33</sup>

The accuracy of the decision-making process depends on the quality of the signal  $x$ , which in turn is indicated by the densities  $g^i(x)$  or, equivalently, the distribution functions  $G^i(x)$ . There are two preliminary difficulties in analyzing the effects of the degree of accuracy on the optimal level of enforcement effort and the optimal evidence threshold. First, in comparing two signals, it may not be that one is unambiguously more or less accurate than another because the relative error rates may depend on the choice of the evidence threshold. Second, it is somewhat unclear what it means to hold the evidence threshold constant when comparing two tests. For example, one could hold  $x^T$  fixed, but such would be incoherent if the two signals are distributed on different, non-overlapping intervals; more generally, the meaning of a given  $x^T$  will differ across tests.

Accordingly, for concreteness let us consider a simple, special case where the comparison is likely to be as clear as one might hope, in order to see what can be learned. Specifically, let the distribution functions be such that  $p^i(x^T) = (1 - x^T)^{\alpha^i}$ , for  $x \in [0,1]$ , where  $\alpha^B > \alpha^H > 1$ . It is convenient to rewrite this as  $p^B = (p^H)^\alpha$ , where  $\alpha = \alpha^B/\alpha^H$ . As discussed previously, the choice of  $x^T$  can be understood equivalently as a choice of  $p^H$ . This particular functional form relating  $p^B$  to  $p^H$  is convex (see section 2 for further discussion). With this formulation, an increase in  $\alpha$  represents an unambiguous increase in accuracy (except at 0 and 1): for any given  $p^H$ ,  $p^B$  falls. Analysis will be confined to choices where, at least initially, the evidence threshold  $x^T$  under the higher  $\alpha$  is set such that  $p^H$  is higher and  $p^B$  is lower.

Further restricting attention to the case of uniform distributions of benefits (constant densities), reconsider the first-order conditions (3) and (5). When  $p^H$  is higher and  $p^B$  is lower, the marginal gain from deterrence falls (since  $b^H$  is higher) and the marginal cost from chilling also falls (since  $b^B$  is lower), which have opposing effects on both the optimal level of enforcement and the optimal evidence threshold. Regarding enforcement effort, there is an additional set of effects: when  $p^H$  is higher, the deterrence punch is greater (a given increase in  $e$  has a larger effect in raising the expected sanction on harmful acts), and when  $p^B$  is lower, the chilling punch is less (a greater increase in  $e$  has a smaller effect in raising the expected sanction on benign acts), both of which favor greater enforcement effort. Regarding the optimal evidence threshold, one might also wonder about the effect of increased accuracy on  $P^{B'}$ . This will depend on how  $p^H$  and  $p^B$  are chosen under the more accurate test, although the initial values and other parameters have implications for its magnitude (when it is chosen optimally). It might seem natural to focus on the special case in which  $P^{B'}$  remains the same (although this may be inconsistent with supposing that  $p^H$  is higher and  $p^B$  is lower).<sup>34</sup>

---

<sup>33</sup>Prior treatments of accuracy and law enforcement (in a model like that in section 3.E) include Kaplow and Shavell (1994) and Kaplow (1994).

<sup>34</sup>Regarding condition (5) for the optimal evidence threshold, consider first the case where the initially optimal threshold is high, implying a low  $p^H$ . With greater accuracy,  $p^B$  will rise slowly, so chilling costs will be small, and one will not reach a given level of  $P^{B'}$  until  $p^H$  is higher, indicating the optimality of a lower evidence threshold (subject to the usual qualifications depending on the  $f^i$ 's). However, if the optimum is initially far enough to the right (a low evidence threshold), a more accurate signal could have the opposite effect, favoring a lower  $p^H$ ,

It is also instructive to revisit the comparative first-order condition (6) that indicates the relative virtues of lowering the evidence threshold compared to raising enforcement effort in achieving a given level of deterrence. Again, even assuming a constant density function, we have competing effects. The reduction in  $p^B$  implies, as mentioned, a lower  $b^B$ , a reduced marginal social loss from chilling the marginal individual, which reduces the relative disadvantage of lowering the evidence threshold (which, recall, involves relatively more increase in the marginal chilling punch in achieving a given level of deterrence than does raising enforcement effort.) Other factors, by contrast, make the use of enforcement effort relatively more favorable. The reduction in  $p^B$  and increase in  $p^H$  both reduce the fraction  $p^B/p^H$ : on account of higher accuracy, increased enforcement effort now has less chilling punch and more deterrence punch than before, as noted; i.e., enforcement effort's relative benefit in terms of the chilling–deterrence tradeoff is enhanced. The  $p^H$  in the denominator on the right side indicates that greater accuracy makes it less expensive to achieve a given level of deterrence augmentation through increased enforcement effort; i.e., enforcement effort's resource cost disadvantage is reduced. And, again, we have  $P^{B'}$ , about which little in general may be said. Considering as well that we are limiting analysis to the case of a constant density function and a particular formulation of the signal and how it changes with increased accuracy, what can be concluded about the effect of accuracy on the optimal setting of other enforcement instruments is limited.

Finally, it is interesting to inquire into the value of accuracy in the present context and how optimal accuracy is determined. In our special case in which  $p^B = (p^H)^\alpha$ , and supposing that all other instruments are set optimally, the marginal value of accuracy can be determined by taking the derivative of social welfare in expression (1) with respect to  $\alpha$ , while holding  $p^H$  fixed. (By the envelope theorem, we can ignore as second-order further influences on welfare due to the optimal adjustment of  $p^H$ , which was a focus of the preceding discussion.) This marginal value, which involves reduced chilling costs, can be expressed as  $-[p^B \ln p^H] \pi s \gamma f^B(b^B) b^B$ , which is fairly similar to the right side of expression (4). The terms after the brackets are the product of an expected sanction effect,  $\pi s$ , the (population-mass-weighted) density for the benign act,  $\gamma f^B(b^B)$ , and the benefit of the marginal act no longer chilled,  $b^B$ . The term in brackets is the derivative of  $p^B$  with respect to  $\alpha$  for the posited functional form, which is negative because  $[\ln p^H]$  is negative (in light of the fact that  $p^H < 1$ ). Not surprisingly, the value of enhanced accuracy is greater the larger is the drop in  $p^B$  and the greater is the marginal chilling cost. Regarding the cost of accuracy, suppose that the enforcement expenditure term  $e$  in expression (1) is replaced by  $ec(\alpha)$ , where  $c'(\alpha) > 0$ . At the optimal level of accuracy, then, the foregoing expression for the marginal value of accuracy would be equated to  $ec'$ .

#### *D. Relative Concentration of Chilling Effects*

Model I assumes that harmful acts and benign acts are scrutinized at a common rate  $\pi$ , with  $\gamma$  indicating the mass of opportunities to commit benign acts (with that for harmful acts normalized to one). Consider now the possibility that, relative to this benchmark, the chilling potential of law enforcement is more concentrated (or dispersed) by introducing a factor  $\lambda$  that multiplies  $\pi$  for benign acts and also divides  $\gamma$ :  $\lambda$  greater (less) than one indicates more highly

---

that is, a higher evidence threshold.

concentrated (dispersed) chilling effects. (If  $\lambda = 2$ , only half as many individuals have opportunities to commit benign acts that may become subject to scrutiny, but such acts are scrutinized with twice the frequency.) The expression for social welfare becomes

$$(23) \widehat{W} = \int_{\pi p^{H_s}}^{\infty} (b - h) f^H(b) db + \frac{\gamma}{\lambda} \int_{\lambda \pi p^{B_s}}^{\infty} b f^B(b) db - e.$$

In the analogues to the first-order conditions (3) and (5) for the optimal level of enforcement effort  $e$  and the optimal evidence threshold  $x^T$ , as well as condition (6) indicating the optimal relative use of the two instruments to achieve a given level of deterrence, the  $\lambda$ 's from the second term of expression (28) cancel (we have  $1/\lambda$  times as many potential benign acts and  $\lambda$  times as many are subject to sanctions, *ceteris paribus*), so all the expressions are the same as before. Nevertheless, the results differ because we now have  $b^B = \lambda \pi p^{B_s}$ . A higher  $\lambda$  (more concentrated chilling effects) means that the social cost of chilling the marginal act is higher: those individuals whose benign acts are potentially subject to mistaken sanctions face a higher expected sanction, so the marginal type has higher benefits, raising the marginal social cost. (By analogy, it is familiar that applying a tax to half as many people but at twice the rate causes more marginal and total distortion; the prospect of being fined in this model is tantamount to being subject to a probabilistic tax on the activity.) Just as with propositions 1b and 2b giving comparative statics for raising  $\gamma$  (indicating a relatively greater importance of chilling effects), so a higher  $\lambda$  on this account tends to make optimal a lower level of enforcement effort and a higher evidence threshold, and also a greater relative reliance on enforcement effort than on a lower evidence threshold to achieve a given level of deterrence.

This characterization is incomplete, however, because in each of our three optimality conditions, the term  $b^B$  is multiplied by  $f^{B'}(b^B)$ : there is a reinforcing or countervailing effect depending on whether  $f^{B'}(b^B)$  is positive or negative. The originally stated sign of the net effect continues to govern (for a marginal change in  $\lambda$ ) if and only if  $f^B/b^B > -f^{B'}(b^B)$  (which obviously holds when  $f^{B'}(b^B) \geq 0$ ).

## 6. Conclusion

Setting the evidence threshold or burden of proof is an important and often overlooked policy instrument, along with the traditionally studied tools of enforcement effort and sanctions. Incorporating it into a framework that allows for errors not only to reduce deterrence but also to chill desirable behavior casts additional and sometimes quite different light on previously examined questions and presents new and challenging ones. Further interest is added by the fact that there are important, qualitative differences across the enforcement technologies of monitoring, investigation, and auditing. Among the distinctions are that, when enforcement is by investigation, optimal sanctions need not be maximal, there need not be underdeterrence at the optimal level of enforcement, and raising the evidence threshold may raise the frequency with which individuals who commit benign acts are mistakenly sanctioned but possibly under opposite circumstances compared to when enforcement is by monitoring or auditing.

Characterizations are offered for the optimal level of enforcement effort and for the optimal evidence threshold in all three models. It is explained how these two instruments are substitutes in achieving the deterrence of harmful acts. The key differences are that raising effort involves resource costs, unlike lowering the evidence threshold, but the latter tends to be relatively less discriminating regarding benign acts and thus entails greater chilling costs (at least in the neighborhood of the optimum). Briefer attention is devoted to the often simpler question of optimal sanctions, including formulation of a new sort of Becker (1968) argument – favoring raising sanctions while raising the evidence threshold, which can keep deterrence and enforcement costs fixed while reducing chilling effects. This new argument for maximal sanctions holds in some cases in which the traditional one does not, thereby upending important results in prior literature.

Attention is also devoted to the relationship between the setting of an evidence threshold – a minimal level of evidence (value of a signal) required to impose sanctions, the direct policy lever analyzed throughout – and conventional understandings about the burden of proof. The latter, embodied for example in the preponderance of the evidence rule that requires it to be more likely than not that the individual before the tribunal committed the harmful act, is an ex post Bayesian notion that has very little connection to the first-order condition that characterizes the optimal evidence threshold in any of the three models. One implication is that, at the optimal evidence threshold, the requisite ex post likelihood that an individual must have committed the harmful type of act in order to be subject to sanctions can take on any value. Also, raising the evidence threshold may increase the frequency with which individuals who commit benign acts are mistakenly sanctioned and may reduce the ex post likelihood that an individual before a tribunal must have committed the harmful type of act in order to be sanctioned. Prior work, by focusing on a much simpler case (with only one type of act and thus no possibility of chilling effects), has tended to obscure the magnitude of the gulf between the different approaches to the burden of proof.

An extension considers costly sanctions and finds that the effects are ambiguous; for example, raising the evidence threshold, which may seem attractive because, conditional on behavior, fewer individuals who have committed both harmful and benign acts are sanctioned, has the offsetting effect that more of both types of acts are committed ex ante, thereby expanding the pool who might receive the costly sanctions. Incapacitation is introduced; in the case considered, it is tantamount to reducing (possibly to a negative level) the social cost of sanctions as applied to actually harmful acts. Social benefits or costs associated with the correct or incorrect imposition of sanctions, viewed as ends in themselves, are analyzed similarly. Enhanced accuracy has ambiguous effects on the optimal setting of enforcement instruments. And, when chilling effects are more concentrated, their social consequences tend to be worse (but not necessarily so), in which case less enforcement effort and a higher evidence threshold are optimal and relatively greater reliance on the former instrument is desirable.

Even though three different enforcement technologies are analyzed, much remains to be

done to provide a comprehensive view of evidence thresholds (or proof burdens).<sup>35</sup> One of the most significant unexamined dimensions concerns enforcers' behavior in gathering evidence and bringing cases, regarding both public officials, such as government prosecutors or those in administrative agencies, and private litigants. Incorporating this feature – aspects of which have been examined in basic settings, often with exogenous behavior and stipulated, constant social costs to errors (see the references in note 2; Prendergast (2007) on optimal biases of government officials; and Kamenica and Gentzkow (2011), who more generally investigate persuasion) – renders endogenous both enforcement levels and costs, making the burden of proof an even more momentous and more difficult to analyze enforcement instrument. For informal discussion, including greater attention to the application to civil litigation, see Kaplow (2012a). Another possibility is that the legal system might impose one sort of proof burden at an initial stage but a different one after further examination.<sup>36</sup> Such a distinction is involved in prosecutors' or agencies' screening processes and in many systems of adjudication that eliminate cases at early stages if what can be established at that point fails to meet some minimum standard.

The present inquiry, although preliminary in significant respects, helps illuminate these and other questions. It provides characterizations of the optimal use of different enforcement instruments, including the burden of proof, and shows how they interact – how best to achieve a target level of deterrence through adjusting multiple instruments. It demonstrates the large difference in the analysis when errors may chill desirable behavior, a phenomenon absent in past work. It exposes how results change qualitatively when enforcement is by investigation or by auditing rather than by monitoring. And it presents the stark differences between determinants of an optimal evidence threshold and of conventional notions of the burden of proof, most of which do not arise in simpler models.

---

<sup>35</sup>For further extensions and for relation of the analysis to legal perspectives on the burden of proof, see Kaplow (2012a, 2012b).

<sup>36</sup>This extension is the subject of work in progress, tentatively entitled “Optimal Multistage Adjudication.”

## Appendix

*Proof of Proposition 1.* To begin, it is helpful to state the second-order condition for an interior optimum for enforcement effort.

$$(A1) \frac{d^2W}{de^2} = -\pi'^2 p^{H^2} s^2 f^H + \left( \pi'' p^H s f^H + \pi'^2 p^{H^2} s^2 f^{H'} \right) (h - b^H) \\ - \gamma \left( \pi'' p^B s f^B b^B + \pi'^2 p^{B^2} s^2 f^{B'} b^B + \pi'^2 p^{B^2} s^2 f^B \right) < 0,$$

where the arguments of the  $f^i$  (each evaluated for the pertinent marginal benefit level  $b^i$ ) are suppressed for convenience.

To prove part *a*, we take the derivative of the first-order condition (2) or (3) with respect to  $h$  and solve for  $de/dh$ , yielding the following:<sup>37</sup>

$$(A2) \frac{de}{dh} = \frac{\pi' p^H s f^H}{-d^2W/de^2} > 0.$$

Moreover, the numerator, reflecting the rise in deterrence of harmful acts, accords with the intuition given for why raising  $h$  raises  $e^*$ .

To prove part *b*, we examine  $de/d\gamma$ :

$$(A3) \frac{de}{d\gamma} = \frac{\pi' p^B s f^B b^B}{d^2W/de^2} < 0.$$

Again, the result accords with the stated intuition, as the numerator indicates the marginal loss from deterring the marginal benign act.

---

<sup>37</sup>Although it is not in general sufficient to determine comparative statics by differentiating the first-order condition in problems in which the second-order condition does not hold globally (which is true here), examination of the first-order condition (2) or (3), moving all the terms to the left side (which gives the expression for  $dW/de$ ) makes apparent that there is no concern. Specifically, for part *a*, it is obvious that raising  $h$  raises the value of  $dW/de$  throughout, so if there were two local optima that were each a global optimum, it is clear that raising  $h$  would make that associated with a higher  $e$  the unique global optimum. Likewise, for part *b*, raising  $\gamma$  has the opposite effect. Analogous reasoning is applicable with regard to all parts of proposition 2.

Establishing part  $c$  is more complicated. Solving for  $de/ds$  yields:

$$(A4) \frac{de}{ds} = \left[ \begin{array}{l} -\left(\pi' p^H f^H + \pi \pi' p^{H^2} s f^{H'}\right)(h - b^H) + \pi \pi' p^{H^2} s f^H \\ + \gamma \left(\pi' p^B f^B b^B + \pi \pi' p^{B^2} s f^{B'} b^B + \pi \pi' p^{B^2} s f^B\right) \end{array} \right] / \frac{d^2 W}{de^2}.$$

It is easy to assert that expression (A4) cannot unambiguously be signed, but it requires further effort to confirm that this is so. It suffices to construct examples consistent with the model's assumptions that yield opposite signs for the expression.

We will confine attention to the numerator and consider the case where both of the densities  $f^i$  are constant on the unit interval and  $\gamma = 1$ . Furthermore, we can use the first-order condition (3) to substitute for the leading terms in each row of the numerator. When all this is done, the numerator simplifies to

$$(A5) -\frac{1}{s} + \pi \pi' s \left( p^{H^2} + p^{B^2} \right).$$

It appears that a low value of  $s$  renders expression (A5) negative and a high value makes it positive. However, since the values of  $\pi$  and  $\pi'$  are determined by the optimal choice of  $e$ , which in turn depends on  $s$ , some further analysis is required.

Suppose that  $s$  initially equals 0.5. We also know that  $\pi$  cannot exceed one and that the  $p^i$  cannot exceed one (nor can their squares). Hence, to obtain a negative value for expression (A5) and thus a positive value for (A4), it is sufficient that  $\pi' < 2$ . But we can choose a  $\pi(e)$  function and an  $h$  that is high enough to assure this at the optimal level of  $e$ .

For the other possibility, first observe that, for any burden of proof, the value of the term in parentheses will equal some constant. Thus, we can guarantee that expression (A5) is positive and hence (A4) negative if  $\pi \pi'$  is sufficiently large (at some given initial value of  $s$ ). Suppose that the function  $\pi$  is such that it obtains a value of 0.5 essentially for free (we can choose the function such that the level of  $e$  such that  $\pi(e) = 0.5$  is arbitrarily small), which places a floor on the value of  $\pi$ . We are left with demonstrating that it is possible that  $\pi'$  is sufficiently large given each of these other factors that can be treated as constants (or floors). Again, this can be accomplished with an appropriate function  $\pi(e)$  and value of  $h$ . Specifically, choose  $h$  high enough that it is beneficial to raise  $\pi$  to at least 0.5, but not so high to drive  $\pi'$  below the desired value. (Note that we have essentially chosen a  $\pi$  function such that  $\pi'$  is arbitrarily large until  $\pi = 0.5$  and then drops sufficiently rapidly since we know  $\pi$  cannot exceed 1. If  $h$  was chosen so that, at  $\pi = 0.5$ , one was indifferent to a free increase in  $\pi$ , one would increase  $\pi$  no further and thus the optimal  $\pi'$  can be made as high as we like.)

*Proof of Proposition 2.* As discussed in section 2, it is equivalent to view the problem of

choosing the optimal  $x^T$  as instead one of choosing the optimal  $p^H$ . The latter course will be adopted here because it is somewhat less cumbersome notationally. Obviously, all comparative statics in terms of  $p^H$  will, because  $dp^i(x^T)/dx^T = -g^i(x^T)$ , have the opposite sign of those for  $x^T$ , the subject of the proposition .

Accordingly, begin with the second-order condition for an interior optimum for  $p^H$ . (Throughout, the version of the first-order condition in expression (5), which cancels  $\pi s$  from each side, is employed.)

$$(A6) \quad \frac{d^2W}{dp^{H^2}} = -\pi s f^H + \pi s f^{H'} (h - b^H)$$

$$- \gamma \left( P^{B''} f^B b^B + \pi P^{B'} s f^{B'} b^B + \pi P^{B'} s f^{B'} \right) < 0.$$

To prove part *a*, we take the derivative of the first-order condition (4) or (5) with respect to  $h$  and solve for  $dp^H/dh$ .<sup>38</sup>

$$(A7) \quad \frac{dp^H}{dh} = \frac{\pi s}{-d^2W/dp^{H^2}} > 0.$$

As explained, raising  $h$  makes the marginal deterrence benefit greater to that extent, producing a higher  $p^{H*}$  and thus a lower  $x^{T*}$ .

For part *b*, we examine  $dp^H/d\gamma$ :

$$(A8) \quad \frac{dp^H}{d\gamma} = \frac{P^{B'} f^B b^B}{d^2W/dp^{H^2}} < 0.$$

Because there are more benign acts, the marginal chilling cost of a higher  $p^H$  is greater, so  $p^{H*}$  is lower and thus  $x^{T*}$  is higher.

For part *c*, we can solve for  $dp^H/ds$ :

---

<sup>38</sup>Regarding possible multiple optima, see note 36.

$$(A9) \frac{dp^H}{ds} = \frac{\pi p^H f^H - \pi p^H f^{H'} (h - b^H) + \gamma \left( \pi p^B P^{B'} f^{B'} b^B + \pi p^B P^{B'} f^B \right)}{d^2 W / ds^2}.$$

If the densities  $f^i$  are constant (uniform distributions), the middle terms in the numerator, which otherwise could have either sign, equal zero. The first and last terms are positive and the denominator is negative, so under the stated assumption,  $dp^H/ds < 0$ . The first term indicates the increase in the marginal forgone benefit from harmful acts times the deterrence punch (recall that  $\pi s$  has been cancelled from each side of the first-order condition before performing this derivation). Likewise, the final term is the increase in the forgone benefit from benign acts times the chilling punch. Since both forgone benefits are larger when  $s$  is higher,  $p^{H*}$  is lower and thus  $x^{T*}$  is higher.<sup>39</sup>

*Auditing (Model III).* The social welfare function is:

$$(A10) \bar{W} = \int_{\pi(e)p^H(x^T)s}^{\infty} (b-h)f^H(b)db + \gamma \int_{\pi(e)p^B(x^T)s}^{\infty} bf^B(b)db \\ - \pi(e) \left[ \left( 1 - F^H(\pi(e)p^H(x^T)s) \right) + \gamma \left( 1 - F^B(\pi(e)p^B(x^T)s) \right) \right] e.$$

As suggested in the text, the first two terms are the same as those in expression (1) for the monitoring model, whereas the last weights  $e$  by the number of audits, which in turn equals the audit probability  $\pi$  times the size of the pool subject to audit, that is, the total number of individuals who commit either type of act.

The first-order conditions corresponding to expressions (3), (5), and (6) for this model are as follows:

---

<sup>39</sup>Further exploration of (A9) can take advantage of the fact that the numerator is similar to the second-order condition (A6), particularly when the latter is multiplied by  $-p^H/s$ . The differences are that (A6) has an additional term (the middle one, with  $P^{B''}$ ), and the last two terms of (A6) (after the modification) have  $p^H P^{B'}$  whereas (A9) has  $p^B$  – the former exceeding the latter at an optimum, as discussed in section 3.C. Because modified expression (A6)'s additional term is positive and the total of the modified expression is positive (we did multiply through by negative one), it is natural to consider a case where, at the optimum, the second-order condition is barely satisfied, suggesting that the numerator of (A9) might therefore be negative, with the result that we would have  $dp^H/ds > 0$ . However, this possibility is not so easy to establish because the relative magnitude of that term in (A6) is affected (among other things) by the magnitude of  $P^{B''}$ , the value of which, at the optimum, depends on the level of  $p^H$ , which is also related to the difference between  $p^H P^{B'}$  and  $p^B$ . Accordingly, for the present it is left as a conjecture that this other case can arise.

$$(A11) \quad \pi'p^H sf^H(b^H)(h - b^H + \pi e) = \gamma\pi'p^B sf^B(b^B - \pi e) \\ + (\pi + \pi'e)\left[(1 - F^H(b^H)) + \gamma(1 - F^B(b^B))\right].$$

$$(A12) \quad f^H(b^H)(h - b^H + \pi e) = \gamma P^{B'} f^B(b^B)(b^B - \pi e).$$

$$(A13) \quad \gamma\left(P^{B'} - \frac{p^B}{p^H}\right) f^B(b^B)(b^B - \pi e) = \frac{(\pi + \pi'e)\left[(1 - F^H(b^H)) + \gamma(1 - F^B(b^B))\right]}{\pi'p^H s}.$$

## References

- Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76:169-217.
- Bernardo, Antonio E., Eric Talley, and Ivo Welch. 2000. "A Theory of Legal Presumptions," *Journal of Law, Economics, & Organization* 16: 1-49.
- Demougin, Dominique, and Claude Fluet. 2006. "Preponderance of Evidence," *European Economic Review* 50: 963-76.
- \_\_\_\_\_ and \_\_\_\_\_. 2008. "Rules of Proof, Courts, and Incentives," *Rand Journal of Economics* 39: 20-40.
- Garner, Bryan A., ed. 1999. *Black's Law Dictionary*, 7th ed., St. Paul: West Group.
- Hay, Bruce L., and Kathryn E. Spier. 1997. "Burdens of Proof in Civil Litigation: An Economic Perspective," *Journal of Legal Studies* 26: 413-31
- Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion," *American Economic Review* 101: 2590-615.
- Kaplow, Louis. 1990. "A Note on the Optimal Use of Nonmonetary Sanctions," *Journal of Public Economics* 42: 245-47.
- \_\_\_\_\_. 1994. "The Value of Accuracy in Adjudication," *Journal of Legal Studies* 23: 307-401.
- \_\_\_\_\_. 2011. "Optimal Proof Burdens, Deterrence, and the Chilling of Desirable Behavior," *American Economic Review: Papers & Proceedings* 101(3): 277-80.
- \_\_\_\_\_. 2012a. "Burden of Proof," *Yale Law Journal* 121: 738-859.
- \_\_\_\_\_. 2012b. "On the Optimal Burden of Proof," NBER Working Paper No. \_\_\_\_.
- \_\_\_\_\_, and Steven Shavell. 1994. "Accuracy in the Determination of Liability," *Journal of Law and Economics* 37: 1-15.
- \_\_\_\_\_ and \_\_\_\_\_. 2001. "Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle," *Journal of Political Economy* 109: 281-286.
- \_\_\_\_\_ and \_\_\_\_\_. 2002. *Fairness versus Welfare*, Cambridge: Harvard University Press.
- Karlin, Samuel, and Herman Rubin. 1956. "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *Annals of Mathematical Statistics* 27: 272-99.
- Kessler, Daniel P., William M. Sage, and David J. Becker. 2005. "Impact of Malpractice Reforms on the Supply of Physician Services," *Journal of the American Medical Association* 293: 2618-25.
- Lando, Henrik. 2002. "When is the Preponderance of Evidence Standard Optimal," *Geneva Papers on Risk and Insurance - Issues and Practice* 27: 602-08.
- Luce, R. Duncan. 1963. "Detection and Recognition," in R. Duncan Luce, Robert R. Bush, and Eugene Galanter, *Handbook of Mathematical Psychology*, vol. 1: 103-89, New York: John Wiley and Sons.
- Miceli, Thomas J. 1990. "Optimal Prosecution of Defendants Whose Guilt Is Uncertain," *Journal of Law, Economics, & Organization* 6: 189-201.
- Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics* 12: 380-91.

- Mookherjee, Dilip, and Ivan P.L. Png. 1992. "Monitoring Vis-à-Vis Investigation in Enforcement of Law," *American Economic Review* 82: 556–65.
- Neyman, Jerzy, and Egon Pearson. 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character* 231: 289–337.
- Png, Ivan P.L. 1986. "Optimal Subsidies and Damages in the Presence of Judicial Error," *International Review of Law and Economics* 6: 101–05.
- Polinsky, A. Mitchell, and Steven Shavell. 1984. "The Optimal Use of Fines and Imprisonment," *Journal of Public Economics* 24: 89–99.
- \_\_\_\_\_ and \_\_\_\_\_. 2007. "The Theory of Public Enforcement of Law," in A. Mitchell Polinsky and Steven Shavell, *Handbook of Law and Economics*, vol. 1: 403–54, Amsterdam: North-Holland.
- Posner, Richard A. 1999. "An Economic Approach to the Laws of Evidence," *Stanford Law Review* 51: 1477–546.
- Prendergast, Canice. 2007. "The Motivation and Bias of Bureaucrats," *American Economic Review* 97: 180–96.
- Rubinfeld, Daniel L., and David E.M. Sappington. 1987. "Efficient Rewards and Standards of Proof in Judicial Proceedings," *Rand Journal of Economics* 18: 308–15.
- Sanchirico, Chris William. 2008. "A Primary-Activity Approach to Proof Burdens," *Journal of Legal Studies* 37: 273–313.
- Shavell, Steven. 1991. "Specific Versus General Enforcement of Law," *Journal of Political Economy* 99:1088–108.
- Studdert, David M., et al. 2005. "Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment," *Journal of the American Medical Association* 293: 2609–17.
- Walker, Thomas J., Harry J. Turtle, Kuntara Pukthuanthong-Le, and Dolruedee Thiengham. 2010. "Legal Opportunism, Litigation Risk, and IPO Underpricing" (unpublished manuscript).