

## Optimal Proof Burdens, Deterrence, and the Chilling of Desirable Behavior

*Louis Kaplow*\*

Determination of the stringency of the burden of proof is a central design question in any system of adjudication, including in many private settings (employee promotion and firing, product launches). Prior work oriented towards ex ante behavior focuses on deterrence, using models in which individuals' decide whether or not to commit a harmful act – versus abstention or instead of a less harmful act, such as when a higher level of care is taken.<sup>1</sup> The optimal proof burden maximizes deterrence, which involves maximizing the difference between the ex ante likelihoods of correct and incorrect imposition of sanctions. Section I briefly derives this result and shows how, contrary to some suggestions, it differs from the familiar preponderance of the evidence rule under which sanctions are applied if and only if it is more likely than not that the individual before the tribunal committed the harmful act.

Such analysis, however, omits a central concern with the mistaken imposition of sanctions: that the prospect of such errors will chill desirable behavior. In competition law, efficient, pro-competitive behavior (e.g., promotional pricing) may be discouraged; with securities regulation, the cost of capital may be raised if honest IPOs often lead to successful class action suits; under tort law, health care providers might avoid treatment of high-risk patients for fear of erroneous malpractice liability. This dimension of the problem seems widespread and is likely to be

---

\* Harvard Law School, Cambridge, MA 02138 (meskridge@law.harvard.edu). Financial support received from Harvard's John M. Olin Center for Law, Economics, and Business.

<sup>1</sup> See, for example, Dominique Demougin and Claude Fluet (2008) and Henrik Lando (2002).

Ivan P.L. Png (1986) has an additional behavioral dimension but does not analyze how to set the burden of proof.

empirically important.<sup>2</sup> Section II models this case and shows that the determinants of the optimal burden of proof are radically different from those in the simpler model, indicating that prior work fails to provide a good guide for system design. They also diverge greatly from those for the preponderance of the evidence rule and other commonly employed formulations of the burden of proof.

### I. Deterrence

Individuals choose between a harmful act,  $H$ , and a benign one,  $B$ . The former yields private benefits  $b$ , distributed in the population according to the density function  $f^H(b)$  and distribution function  $F^H(b)$ , and it also imposes external social harm of  $h$ . In this model, the benign act is associated with no private benefits or external harm (which may be viewed as normalizations).

The fraction  $\pi$  of each type of act is scrutinized by an enforcement authority; this probability may be interpreted as an audit rate or a detection frequency due to the posting of monitors (such as police). Scrutinized acts proceed to adjudication, which applies the sanction  $s$  with probabilities  $p^i$ ,  $i=H, B$ , where  $p^H > p^B$ . Therefore, an individual of type  $b$  commits the harmful act if and only if  $b - \pi p^H s > -\pi p^B s$ , or  $b > \pi \Delta p s$ , where  $\Delta p = p^H - p^B$ .

For present purposes,  $\pi$  and  $s$  are taken to be fixed at levels such that first-best deterrence cannot be achieved (i.e.,  $\pi \Delta p s < h$ ), and enforcement and sanction costs are taken to be zero. To model the choice of the burden of proof, a reduced form is employed for simplicity. Let  $\chi$  denote the minimum strength of evidence required to apply the sanction. Both  $p^H(\chi)$  and  $p^B(\chi)$  are taken to be strictly decreasing in  $\chi$ . Furthermore, we can think of the choice variable as

---

<sup>2</sup> See, for example, Thomas J. Walker et al. (2009) on IPO litigation risk and David M. Studdert et al. (2005) on the avoidance of medical procedures and patients posing high litigation risks.

being  $p^H$  since any chosen value of it implies a value of  $\chi$ , which in turn implies a value of  $p^B$ , allowing us to write  $p^B(p^H)$ , with  $p^{B'} > 0$ ; we shall also assume convexity:  $p^{B''} > 0$ .<sup>3</sup>

Social welfare can be written

$$(1) \quad W = \int_{\pi\Delta ps}^{\infty} (b-h)f^H(b)db.$$

Undeterred individuals commit the harmful act, realizing private benefits  $b$  and causing harm  $h$ .

Since more deterrence is better (recall the assumption that  $\pi\Delta ps < h$ ), it is obvious that the optimum maximizes  $\Delta p$ , which is accomplished when

$$(2) \quad p^{B'} = 1.$$

Despite the impression given by some of the literature (Demougin and Fluet 2008 and Lando 2002), this result differs from the preponderance rule because the latter is understood as a Bayesian posterior – it asks whether it is more likely than not that the individual before the tribunal committed the harmful type of act – and thus depends on the priors. Specifically, here it would involve the condition:

$$(3) \quad 1 - F^H(b^H) = p^{B'} F^H(b^H),$$

where  $b^H \equiv \pi\Delta ps$  denotes the type just indifferent whether to commit the harmful act.

## II. Deterrence and Chilling

To introduce the chilling of desirable behavior, suppose that some individuals decide whether or not to commit the harmful act,  $H$  – inaction involves no private benefit or external harm – and some decide whether or not to commit the benign act,  $B$ , which is now assumed to

---

<sup>3</sup> For details of the underlying derivation from evidence (signals), see Kaplow (2011a).

Convexity is implied by a strict version of the monotone likelihood ratio property with respect to the densities for the underlying signals.

generate a private benefit  $b$ , with density function  $f^B(b)$  and distribution function  $F^B(b)$ .

Furthermore, let  $\gamma$  indicate the relative frequency of opportunities to commit the benign act. It does not matter whether distinct parts of the population have the different types of opportunity or the same individuals sometimes have each. (It is ruled out that the two opportunities can arise simultaneously, with individuals choosing between  $H$ ,  $B$ , and inaction – a possibility that would complicate the analysis but not change the results qualitatively.)

Social welfare can now be written

$$(4) \quad W^* = \int_{\pi p^H s}^{\infty} (b - h) f^H(b) db + \gamma \int_{\pi p^B s}^{\infty} b f^B(b) db.$$

Undeterred individuals, those with  $b > \pi p^H s = b^H$ , commit the harmful act, and unchilled individuals, those with  $b > \pi p^B s = b^B$ , commit the benign act. Raising  $p^H$  (through lowering  $\chi$ , the minimum strength of evidence) desirably enhances deterrence, but it also implies a higher  $p^B$ , which increases the chilling of desirable behavior. The first-order condition characterizing an interior optimum can be expressed as

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

On the left side, raising  $p^H$  raises the expected sanction on harmful acts ( $\pi p^H s$ ), which raises deterrence; the magnitude of the density for those just on the margin of committing harmful acts,  $f^H(b^H)$ , indicates the size of this deterrent effect; and  $h - b^H$  is the net social benefit from deterring a marginal act since harm of  $h$  is avoided but the private benefit  $b^H$  is forgone. For chilling costs, on the right side, we have corresponding terms, except that the loss of the private benefit  $b^B$  from the marginal benign act is not offset by any social benefit since no externality is avoided.

Moreover, these chilling terms are preceded by two additional factors:  $\gamma$ , the relative frequency of opportunities to commit benign acts, and  $p^{B'}$ , the rate at which  $p^B$  increases as we raise  $p^H$  by

requiring stronger evidence.

Comparison of expressions (5) and (2) reveals that the optimums for the two problems have almost nothing in common. Not surprisingly, both share the term  $p^{B'}$  because, for both, the greater the rate at which  $p^B$  rises with  $p^H$ , the lower the optimal level of  $p^H$ , which is to say the stronger the minimum required level of evidence  $\chi$ . But that is it for similarities. For the single-act, deterrence-only model from section I (which roughly corresponds to the models in prior papers), this factor alone determines the optimal evidence threshold.

For the model in this section, with deterrence and the chilling of desirable behavior, we have a much richer set of considerations.<sup>4</sup> First, we care about the rates of deterrence and chilling, denoted by the two density functions. Clearly, a higher (lower) density for the harmful act favors a lower (higher) evidence threshold, and a higher (lower) density for the benign act favors a higher (lower) threshold. Next, the greater the marginal benefits of the two types of act, the higher a threshold is optimal. Note that, for both types of act, society loses private benefits on account of discouragement. By contrast, as we would expect, the greater the external harm  $h$ , the lower the optimal threshold. Finally, and also in accord with intuition, the greater the relative frequency of opportunities to commit benign acts,  $\gamma$ , the higher the optimal threshold.

It may seem surprising that such a modest change in the model has such large, fundamental effects on the character of the optimum. One might be even more surprised by the fact that the level of social harm,  $h$ , the magnitude of the forgone benefit from the harmful act,  $b^H$ , and the deterrence impact,  $f^H(b^H)$ , have no effect in section I's model. The reason is that, when the only

---

<sup>4</sup> As is familiar, these comments on the first-order condition are not, strictly speaking, comparative statics since many values are endogenous. For the exogenous variables  $h$  and  $\gamma$ , the suggested comparative statics results are demonstrated in Kaplow (2011a).

effect of the false imposition of sanctions is to reduce the desirability of abstention from the harmful act, we have essentially offsetting effects from these three factors. Both types of error affect the same deterrence margin. Fewer false negatives – a higher rate of correctly imposing sanctions on those who do in fact commit harmful acts – raises deterrence, whereas more false positives – a higher rate of incorrectly imposing sanctions on those who do not commit harmful acts – reduces deterrence. The density of individuals at the deterrence margin as well as the social benefits and costs of deterrence are the same for both types of error. Therefore, all that matters is the relative rates that the two probabilities change, as expression (2) indicates.

By contrast, in the model with chilling effects, the two types of error operate on different margins, the deterrence margin and the chilling margin. Not surprisingly, this makes all factors pertinent to the chilling of desirable behavior (those on the right side of expression (5)) relevant. But, as just explained, all the factors pertaining to the net marginal social gain from the deterrence of harmful acts (those on the left side of expression (5)) also become relevant in a way that they were not previously. The implication is that the setting of optimal proof burdens is done in an almost entirely different way in the two models. Moreover, since it seems that in a wide range of contexts (most?), the second model is the relevant one, the deterrence maximization lesson from prior work is a quite misleading basis for system design.<sup>5</sup>

One might also wish to compare the first-order condition (5) with the analogue to the preponderance of the evidence rule for the present model, which is

$$(6) \quad 1 - F^H(b^H) = \gamma p^B (1 - F^B(b^B)).$$

---

<sup>5</sup> Another implication is that the practice in some work (e.g., Daniel L. Rubinfeld and David E.M. Sappington 1987) of using constants to stand in for the marginal costs of false positives and of false negatives is an unsatisfactory reduced form because both magnitudes are endogenous.

This expression is similar to expression (3) for the first model except that the prior for the benign act now consists of unchilled individuals who take advantage of opportunities to commit benign acts rather than deterred individuals who refrain from the harmful act. It is immediate that there is almost no correspondence between the factors determining whether the preponderance rule (6) is satisfied and those determining the optimal rule (5). Moreover, this conclusion applies not only to a rule requiring a 50% likelihood, but equally to rules requiring other likelihoods, such as 75% (sometimes associated with the “clear and convincing evidence” standard) or 95% (sometimes used to approximate “proof beyond a reasonable doubt”).

Note further that the informational requirements for the optimal rule (5) and the preponderance rule (6) are essentially the same. For both, one must determine behavior, which depends on the evidence threshold  $\chi$  that is selected (and also on  $\pi$  and  $s$ ); this determines the marginal types, and thus the values of the densities and cumulative distribution functions. To implement the optimal rule (5), one must also know  $h$ , the external harm caused by the activity, whereas the preponderance rule (6), and others like it, are independent of how harmful is the behavior being regulated.

### **III. Discussion**

The burden of proof is an important feature of adjudication. The optimum reflects a tradeoff of deterrence and chilling effects, the latter of which are important in many settings yet have not been modeled in prior work. The introduction of chilling effects greatly changes the nature of the optimum; indeed, only one of the many relevant factors appears in the first-order condition for the simpler model without chilling. In addition, there is almost no overlap between determinants of the optimum and of the condition for the preponderance of the evidence rule or other commonly used rules, suggesting that current practice is optimal only by chance – and only occasionally, since the optimum depends on context-specific considerations ignored by the

preponderance rule.

This work can be extended in many ways. Kaplow (2011a, 2011b) considers how determination of the optimal proof burden interacts with optimal choices of the degree of enforcement effort (which determines  $\pi$ ) and the sanction  $s$ , how the optimal proof burden differs with the mode of enforcement (notably, when enforcement is by investigation that is triggered by the commission of a harmful act), the relevance of socially costly sanctions (such as imprisonment), and other matters. Among the findings are: Laxer proof requirements and greater enforcement effort are substitutes in achieving deterrence, with different comparative disadvantages (the former is worse regarding chilling effects whereas the latter entails direct resource costs). The optimality of a maximal sanction varies across settings and can arise for different reasons from those in models with an exogenous proof burden (the prospect of chilling innocent behavior can provide an additional reason for maximal sanctions because they enable tougher proof burdens). Demanding stronger evidence can raise the likelihood that individuals who commit benign acts are sanctioned (the reduction in chilling increases the flow of benign acts into adjudication). When enforcement is by investigation, greater deterrence helps reduce chilling effects (fewer investigations are triggered). And sanction costs can favor a more relaxed proof requirement (greater deterrence and chilling both reduce the frequency of imposition of sanctions).

#### REFERENCES

- Demougin, Dominique, and Claude Fluet.** 2008. "Rules of Proof, Courts, and Incentives." *Rand Journal of Economics*, 39(1): 20–40.
- Kaplow, Louis.** 2011a. "On the Optimal Burden of Proof." Unpublished.
- Kaplow, Louis.** 2011b. "Burden of Proof." Unpublished.
- Lando, Henrik.** 2002. "When is the Preponderance of Evidence Standard Optimal?" *Geneva*



*Papers on Risk and Insurance – Issues and Practice*, 27(4): 602–08.

**Png, Ivan P.L.** 1986. “Optimal Subsidies and Damages in the Presence of Judicial Error.”

*International Review of Law and Economics*, 6(1): 101–05.

**Rubinfeld, Daniel L., and David E.M. Sappington.** 1987. “Efficient Rewards and Standards of Proof in Judicial Proceedings.” *Rand Journal of Economics*, 18(2): 308–15.

**Studdert, David M., Michelle M. Mello, William M. Sage, Catherine M. DesRoches, Jordon Peugh, Kinga Zapert, and Troyen A. Brennan.** 2005. “Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment.” *Journal of the American Medical Association*, 293(21): 2609–17.

**Walker, Thomas J., Harry J. Turtle, Kuntara Pukthuanthong-Le, and Dolruedee**

**Thiengtham.** 2009. “Legal Opportunism, Litigation Risk, and IPO Underpricing.”

<http://69.175.2.130/~finman/Reno/Papers/LegalOpportunismLitigationRiskAndIPOUnderpricingFMA0109n.pdf>.