

NBER WORKING PAPER SERIES

OPTIMAL MULTISTAGE ADJUDICATION

Louis Kaplow

Working Paper 23364

<http://www.nber.org/papers/w23364>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

April 2017

I am grateful to Andrew Daughety (the editor), three referees, Steven Shavell, Holger Spamann, Kathryn Spier, Abraham Wickelgren, and workshop participants at Harvard, NBER, and Yale for comments and to the John M. Olin Center for Law, Economics, and Business at Harvard University for financial support. Disclaimer: I occasionally consult on antitrust cases, and my spouse is in the legal department of a financial services firm. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Louis Kaplow. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Optimal Multistage Adjudication
Louis Kaplow
NBER Working Paper No. 23364
April 2017
JEL No. D81,D82,K14,K41,K42

ABSTRACT

In many settings, there are preliminary or interim decision points at which legal cases may be terminated: e.g., motions to dismiss and for summary judgment in U.S. civil litigation, grand jury decisions in criminal cases, and agencies' screening and other exercises of discretion in pursuing investigations. This article analyzes how the decision whether to continue versus terminate should optimally be made when (A) proceeding to the next stage generates further information but at a cost to both the defendant and the government and (B) the prospect of going forward, and ultimately imposing sanctions, deters harmful acts and also chills desirable behavior. This subject involves a mechanism design analogue to the standard value of information problem, one that proves to be qualitatively different and notably more complex. Numerous factors enter into the optimal decision rule – some expected, some subtle, and some counterintuitive. The optimal rule for initial or intermediate stages is also qualitatively different from that for assigning liability at the final stage of adjudication.

Louis Kaplow
Harvard University
Hauser 322
Cambridge, MA 02138
and NBER
meskridge@law.harvard.edu

1. Introduction

Existing economics literature on decision-making in adjudication focuses on the final decision at the conclusion of a trial, i.e., on how stringent the burden of proof should be.¹ However, more cases are formally terminated earlier. In U.S. civil litigation, a court can grant a motion to dismiss after a case is filed or a motion for summary judgment after discovery but before trial. As an indication of the importance of these rules, the Supreme Court cases most cited by federal courts are a 1986 trilogy on summary judgment and one on motions to dismiss that was recently reversed in important respects.² In U.S. criminal cases, grand juries decide whether to issue indictments. In continental legal systems, adjudication has a sequential character, and some arbitration systems have the flexibility to operate in such a manner. More broadly, investigators, prosecutors, and agencies screen cases at various points and make ongoing decisions whether to pursue cases or to terminate them.³

This article analyzes how optimally to make continuation decisions at each point in a multistage process. At first glance, this problem seems like a familiar one in decision analysis concerning the value of information: at any stage, one can make an ultimate decision or else choose to expend resources to obtain further information upon which a later decision can be based. In contrast to these standard settings, however, the present context involves decisions that centrally influence *ex ante* behavior.⁴ Hence, the present investigation can be understood as developing a solution to the value of information problem in a mechanism design setting.

A central justification for the legal system is to deter harmful acts, and most prior work on the economics of law enforcement has focused on how to set policy instruments in light of deterrence effects and system costs. However, a recent strand of literature, which this paper follows, incorporates the important prospect of false positives, which chills desirable behavior. This consideration is featured in many analyses of substantive law. The most widely known illustration is medical malpractice, where the prospect of false positives may result in defensive medicine, denial of care to high-risk patients, and reduced physician supply in certain fields. In antitrust, this concern is often central. Prices below short-run marginal cost may involve predatory pricing, which the law forbids, but they may instead reflect promotional pricing to penetrate new markets or establish networks, or quantity expansion to move more rapidly down a learning curve. (Tech giants running large losses for years come to mind.) The challenge has

¹See, for example, Kaplow (2011), Lando (2002), and Rubinfeld and Sappington (1987). Some literature considers other aspects of adjudication not relevant for present purposes, such as regarding the resources devoted to the presentation of evidence or bringing suit (usually with exogenous underlying behavior). See, for example, Bernardo, Talley, and Welch (2000) and Hay and Spier (1997).

²On summary judgment, see *Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574 (1986), *Anderson v. Liberty lobby, Inc.*, 477 U.S. 242 (1986), and *Celotex Corp. v. Catrett*, 477 U.S. 317 (1986). On motions to dismiss, see *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544 (2007) (in the Court's language, "retir[ing]" the previously dominant version of the legal test from *Conley v. Gibson*, 355 U.S. 41 (1957)), and *Ashcroft v. Iqbal*, 556 U.S. 662 (2009). For further discussion, see Kaplow (2013).

³Some literature, notably in antitrust, has addressed information that may be used in screening but not how screening decisions should optimally be made in light of available information. See, for example, Abrantes-Metz et al. (2006) and Harrington (2007).

⁴The more typical and straightforward version of the value of information problem would, however, be applicable in legal contexts that, unlike those examined here, have as their main consequence the regulation of future conduct (licensing, zoning decisions, drug authorization, merger approval) or determination of eligibility for subsequent government transfer payments. See Kaplow (2013).

also been central in securities regulation where, for example, mistaken liability for IPOs may discourage them and raise the cost of capital. Indeed, in this realm, such concerns prompted the Private Securities Litigation Reform Act of 1995 and subsequent amendments, which impose a higher first-stage hurdle to continuation. The ubiquity of potential false positives and, as important, significant early-stage litigation costs borne by defendants who behaved properly, also motivated the aforementioned Supreme Court cases that make it easier to dismiss cases both at the outset and before trial.

When outcomes matter because of ex ante behavioral effects, the optimal decision problem – and, implicitly, determination of the value of information – is qualitatively different and considerably more complex than under standard decision analysis. For example, Bayesian prior probabilities, instead of playing a central role, are irrelevant to the core behavioral components, and they enter additively (rather than by a difference or ratio) in a system cost component. In addition, the cost of information itself is endogenous because the frequency with which the cost must be incurred depends on behavior, which in turn is influenced by the anticipated decisions, including those about when to obtain additional information. Accordingly, familiar lessons regarding when it is optimal to obtain additional information need to be confirmed, revised, extended, or supplanted in the present setting.

Section 2 examines the optimal first stage decision in a model with two stages: a preliminary decision and, conditional on continuation to trial, a final judgment on liability. This initial modeling choice is made for tractability and clarity; a number of insights that prove to be robust can be gleaned from this simpler formulation. The model has harmful acts that the legal system aims to deter as well as benign acts that, unfortunately, the legal system may mistake for harmful ones and therefore may cause to be chilled. Individuals' decisions for each type of act reflect its associated expected sanctions as well as the adjudication costs expected to be borne. A fraction of committed acts of each type enters the legal system (determined by a regime of enforcement, taken as given in the first part of the section). Based on a preliminary signal, the tribunal decides whether to let a case continue to final adjudication or to terminate it.⁵ If the case proceeds, costs are incurred both by the individual and the government, and an additional signal is observed. Final adjudication uses all the information to determine whether to find liability and thus impose a sanction. Taking the functioning of this final stage as given in this section, the focus is on how optimally to make the first-stage continuation/termination decision.

Continuation relative to termination in a given state – i.e., for a given realization of the preliminary signal – has two types of effects. Most directly, there is an inframarginal cost regarding both harmful and benign acts that are committed: individuals and the government bear additional adjudication costs. Moreover, there are marginal effects on behavior: greater deterrence and more chilling, as a consequence of keeping alive the possibility of imposing sanctions in an additional situation and also the certainty that, in the state under consideration, defendants in the system will bear additional adjudication costs. These inframarginal and marginal effects are qualitatively different. The former depend (among other factors) on the undeterred and unchilled masses for the two types of acts, which in turn depend on the cumulative distribution functions for individuals' benefits from the two types of acts. The latter

⁵It is obvious that an unconstrained optimal system would also allow adjudication to conclude at the first stage or any intermediate stage with a finding of liability. Because this either is not permitted or tends to be relatively unimportant for most of the legal system features being modeled, this possibility is omitted. It will be clear how one could extend the model to incorporate it.

depend instead on the densities of these distributions for the marginal actor of each type. Accordingly, some factors affect just one or the other and some factors affect both but in different ways.

Some of the results for the optimal rule are expected while others are more counterintuitive. When the degree of underdeterrence is greater and when the cost of chilling benign acts is lower, continuation is more favorable. An implication is that decisions in other states affect the optimal decision in the situation at hand. Continuation is also favored by a relatively higher rate of harmful rather than benign acts entering the legal system and by higher diagnosticities of the signals received when entering the first stage and when moving to final adjudication. Higher adjudication costs, however, have surprisingly complex effects on the optimality of continuation. The direct cost of continuation, of course, favors termination. However, the greater are adjudication costs, the greater is the social benefit of deterring a marginal harmful act and the less is the social cost of chilling a marginal benign act, so continuation's discouragement of both types of acts makes it more favorable when system costs are larger. Last, defendants' adjudication costs contribute to deterrence as well as to chilling, so higher defendants' costs can favor or oppose continuation on this behavioral dimension.

The analysis also compares optimal decisions across states (signals). It is explained that these are not characterized by a likelihood ratio test, in which all cases with a signal above some threshold are optimally continued and all below the threshold are terminated. The reason is that other parameters depend on the signal. For example, it may be optimal to continue in one state with a strong signal and low continuation costs but to terminate in another state with an even stronger signal but high continuation costs.

Section 2 also extends the analysis to allow the government to choose the level of enforcement effort and sanctions along with first-stage continuation/termination decisions. Many determinants of optimal effort differ qualitatively from those in prior literature. Optimal sanctions are maximal, a feature of many law enforcement models, but much of the basis here differs from the familiar argument, relating instead to the relative targeting precision of different enforcement instruments.

Section 3 presents a more general model for adjudication that allows any number of stages. This model is analyzed using backward induction and recursion. The optimality condition for decisions in any given stage, except for the final stage, is similar to that for the first stage in section 2's two-stage model. Nevertheless, the more general formulation allows one to see how optimal decisions can differ at different stages and how they are interdependent (in both directions) across stages. An important application of this sort of analysis relates to the fact that actual legal systems may be subject to institutional constraints on decisions at some stages, which importantly influence how decisions at other stages, which may be less constrained, should optimally be made. For example, one can determine how an agency would optimally take into account subsequent rules governing adjudication in courts, or how courts would optimally adjust their rules if the inflow of cases (reflecting, for example, prosecutors' overzealousness) may deviate systematically from what is socially optimal.

Section 3 also analyzes the optimal decision rule for the final stage of adjudication. This rule differs from that applicable to all earlier stages because there is no meaningful future continuation. As a consequence, this rule can be represented as a standard likelihood ratio test, unlike at prior stages. More broadly, optimal decision rules at nonfinal stages cannot readily be compared in terms of stringency with each other or with that at the final stage. Conventional legal wisdom favoring increased stringency at later stages is difficult to rationalize as a general

matter, among other reasons because more adjudication costs are sunk as cases proceed. Section 4 offers concluding remarks.

Before proceeding, it is worth noting that the present analysis abstracts from settlement and, relatedly, plea bargaining in the criminal context. As is familiar, negotiated resolutions reflect parties' expectations about how cases would otherwise be resolved. Hence, the model developed here could be employed to determine what settlements litigants would find attractive. The expectations of those settlements, in turn, would determine the deterrence and chilling effects of adjudication. Furthermore, the prospect of settlement reduces the magnitude of the expected adjudication costs that play a role in the analysis to follow.⁶

2. First Stage of Adjudication in a 2-Stage Model

A. Model

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.⁷ A mass of (risk neutral) individuals normalized to 1 may commit the harmful type of act.⁸ Those who may commit the benign type of act have a mass of γ . One interpretation is that γ indicates the relative quantity 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 entering the legal system and being subject to sanctions. The analysis is unchanged if, instead, the same individuals may commit both types of acts and γ indicates the relative frequency of opportunities to commit the benign type of act. (Note that γ does not indicate the portion of benign acts, which is endogenous, but merely allows for the possibility that the opportunities for the two types of acts differ.)

Before proceeding, a rather different version of this model should be mentioned. In some settings, it may be more natural to suppose that the same individual may be able to choose among three options: act H , act B , and inaction. (For example, act H may be a version of act B , but without taking a precaution.) This variation has been partly analyzed and it turns out that, although it greatly complicates the exposition, there is only a modest effect on the results. Specifically, deterrence of harmful acts becomes more beneficial because some deterred individuals would switch to benign acts rather than to inaction. Likewise, the chilling of benign acts becomes more detrimental because some chilled individuals would switch to harmful acts rather than to inaction. Because the results are largely the same, the model analyzed here is the aforementioned, simpler version in which a given individual chooses only between one type of act and inaction.

Returning, then, to this model, individuals' benefits from committing an act are b , with

⁶Specifically, the greater the fraction of cases that settle, the lower would be ex ante expected adjudication costs (the κ^i) and the inframarginal expected adjudication costs from continuation (the final terms in expressions 6 and 12). Of central interest in modeling settlement is the extent to which parties can anticipate the particular signals adjudicators will learn at subsequent stages and the nature of asymmetric information between the parties in this regard.

⁷One could readily allow a smaller negative externality or a positive externality.

⁸Introducing risk aversion would have two main effects. First, deterrence and chilling effects would rise nonlinearly with s (which is of little interest here since s is taken as given). Second, sanctions would then be socially costly, the effect of which is discussed in note 11.

density functions $f^i(b)$ (which are positive for positive values of b) and cumulative distribution functions $F^i(b)$, where $i = H, B$. Individuals know what type of act they are able to commit and its benefits to them, but the government initially has no knowledge of an act's type and never learns individuals' benefits from acts.

Individuals who commit these acts enter the legal system with probabilities π^i .⁹ The government's policy choice is whether to allow cases to proceed ($\delta = 1$) or instead to terminate them ($\delta = 0$). Terminating a case may be thought of as a court granting a defendant's motion to dismiss or for summary judgment, a grand jury's decision not to issue an indictment, or an agency's or prosecutor's decision not to continue an investigation (screening out a case). For cases that continue to adjudication, two costs are incurred: c is borne by the defendant and k by the state.¹⁰ These cases are ultimately tried and result in findings of liability with probabilities p^i , conditional on reaching the final stage of adjudication. Individuals found liable are subject to the sanction s , taken here to be a socially costless fine.¹¹ At this point, all legal system attributes except δ are taken as given; one may suppose that they are optimized or simply are fixed for other reasons. (Of particular interest, the p^i can be understood to depend, in a manner suppressed in this section, on the use of an optimal decision rule at the final stage, as analyzed explicitly in section 3.C. It turns out that using the features of the optimal final-stage rule to further restrict the p^i does generate additional, sharper characterizations of the first-stage rule.)

Cases entering the system generate a signal σ , with density $z^i(\sigma)$, positive on the real line, that is informative about the type of case.¹² The p^i and the continuation costs, c and k , each depend on σ .¹³ (It is natural to think of σ as a vector, some elements of which may bear on the type of act and others on costs; a scalar representation is employed to simplify notation.) To motivate this important feature, cases at early stages often vary tremendously not only in how strong they appear to be but also in how costly it is likely to be to gather further information. In

⁹The model takes as given some enforcement technology and (until section 2.C) the level of enforcement effort that generates these probabilities. The present formulation is most akin to monitoring (the posting of agents on the lookout for what appear to be harmful acts) or auditing (including inspections and the like). In contrast, an investigation can be triggered by the observation of a harmful act, which involves qualitative differences (which are described briefly in note 29). For further discussion of different enforcement technologies, see Shavell (1991), Mookherjee and Png (1992), and Kaplow (2011).

¹⁰One could readily introduce defendant and government costs of entering the initial stage of adjudication, a possibility allowed in the more general formulation in section 3.

¹¹One could introduce costly sanctions, as in Kaplow (2011), which would have two competing effects on the optimal rules. On one hand, continuation (and, in the final stage, liability) would be more favorable because deterring and chilling marginal acts would have the added benefit of reducing expected sanction costs. On the other hand, for inframarginal acts (those that remain undeterred or unchilled, as the case may be), continuation (or ultimate liability) would be more costly.

¹²No particular requirement, such as satisfaction of the monotone likelihood ratio property, is imposed on the $z^i(\sigma)$; as will emerge with Proposition 4, little would be gained in the present setting because the optimal rule does not take the form of a likelihood ratio test. All that is assumed is that the functions are such that the integrals below are well defined.

¹³Regarding the p^i , the intended interpretation is that, for different σ , there are different expected distributions of evidence that will be used in adjudication and, for the given decision rule, this difference will imply differing probabilities of liability. See section 3.C. That the π^i do not depend on σ is without loss of generality. (If one did allow this dependence, one could simply redefine variables so that what is here called π^i would be the mean of that variable, with the variation absorbed in the corresponding density functions; below, π^i and $z^i(\sigma)$ always appear as a product.) One could also extend the model to allow the f^i and h (as well as the policy instrument s) to depend on σ , the implications of which would largely be straightforward.

U.S. civil litigation, continuation at the first stage is followed by discovery, which is the most costly phase, and the anticipated discovery expense depends significantly on the particulars revealed thus far. Even within a given field of law and type of case – for example, predatory pricing in antitrust – anticipated discovery costs vary greatly with the particulars of the case, and even for a particular case, the cost of a trial, after discovery, will depend significantly on what is learned at that stage. Sometimes discovery effectively eliminates a significant portion of the case; other times, it opens new channels.¹⁴

The government’s problem is to choose the function $\delta(\sigma)$: that is, for each σ , to indicate whether the case is allowed to proceed or is to be terminated. The model can be summarized by reference to its timing:

1. The government sets all policy instruments, notably, the function $\delta(\sigma)$.
2. Individuals learn their type of act (H or B) and their private benefit b .
3. Individuals decide whether to act.
4. A portion of those who commit each type of act, π^H and π^B , are identified and brought before a tribunal.
5. The signal σ is realized.
6. The tribunal either allows a type of case to continue (if $\delta(\sigma) = 1$) or instead terminates it (if $\delta(\sigma) = 0$).
7. If the case continues to final adjudication, the costs $c(\sigma)$ and $k(\sigma)$ are incurred by the defendant and the government, respectively.
8. If the case continues to final adjudication, individuals pay the sanction s with probabilities $p^H(\sigma)$ and $p^B(\sigma)$, as appropriate for their type of act.

It is helpful to introduce some auxiliary notation for expositional convenience. Let

$$(1) \rho^i(\sigma) = \pi^i z^i(\sigma),$$

$$(2) \lambda^i(\sigma) = c(\sigma) + p^i(\sigma)s.$$

The expression $\rho^i(\sigma)$ is the unconditional (ex ante) probability that a case involving an act of type i enters the legal system and falls into state σ . The expression $\lambda^i(\sigma)$ is the legal system’s expected burden on an act of type i , conditional on continuation, when in state σ . Accordingly, those whose type of act is i commit their act if and only if:

$$(3) b > \int_{-\infty}^{\infty} \rho^i(\sigma) \delta(\sigma) \lambda^i(\sigma) d\sigma \equiv b^i.$$

That is, individuals act when their benefit exceeds their expected legal system burden for their type of act, where their expected burden is (the integral, for each realization of σ , of) the product of (A) the likelihood that they will be in the legal system in state σ , (B) the likelihood that their

¹⁴This point can also be seen in the more familiar medical diagnosis context. The first-ordered test depends on the particular symptoms a patient initially presents. The costs of the next follow-up may be huge or small, depending on what the first test reveals, and so forth. It is also of interest that, in this context (and with legal cases as well), there is no general, monotonic relationship (whether positive or negative) between the strength of the initial signal (or a subsequent signal) and the likely cost of the next test. Sometimes an expensive biopsy is appropriate only with a sufficiently strong signal of a given risk, but other times a very strong signal may make the biopsy unnecessary.

case will continue rather than be terminated (which likelihood is one or zero, depending on the decision rule), and (C) the legal system's burden in that condition. As indicated on the right side of expression (3), it is convenient to define b^i as the value of this expected cost for each type of act, that is, the benefit level of an individual who is just indifferent about whether to act for each type of act, H and B .

Social welfare, W , is taken to be the aggregate of individuals' benefits from acting minus the harm from the commission of acts of type H and the defendants' and the government's expenditures on adjudication.

$$(4) W = \int_{b^H}^{\infty} \int_{-\infty}^{\infty} (b - h - \kappa^H(\sigma)) z^H(\sigma) d\sigma f^H(b) db \\ + \gamma \int_{b^B}^{\infty} \int_{-\infty}^{\infty} (b - \kappa^B(\sigma)) z^B(\sigma) d\sigma f^B(b) db.$$

where

$$(5) \kappa^i(\sigma) = \pi^i \delta(\sigma) (c(\sigma) + k(\sigma)).$$

The first term indicates the benefits and harm attributable to harmful acts, where, in addition to the direct external harm, there is also the expected adjudication cost of defendants and the government, captured by $\kappa^H(\sigma)$ for each scenario. Similarly, the second term indicates the benefits from benign acts and the expected adjudication cost for them. In each instance, the lower limit of integration on the outer integral is the benefit of the individual just at the margin, with all individuals having greater benefits committing the act, as per expression (3).

B. Analysis

The government's problem is to select (in advance; see the timeline, step 1) the decision rule for each possible realization of σ so as to maximize W in expression (4), where behavior is given by expression (3). This can be determined by pointwise optimization, wherein the optimal decision rule provides for continuation in any particular state σ° if and only if the value of W in expression (4) is greater when $\delta(\sigma^\circ) = 1$ than when $\delta(\sigma^\circ) = 0$.¹⁵ It is understood that, because each point has no mass, the resulting analogue to the first-order condition must hold almost everywhere (which qualification will be omitted in the exposition that follows).¹⁶ Also note that the resulting necessary condition for an optimum also characterizes a constrained optimum,

¹⁵This yields a result equivalent to what would be produced if one takes the derivative of W with respect to $\delta(\sigma^\circ)$. Because each point is massless, this derivative is a constant: if it is positive, the optimum is at the boundary $\delta(\sigma^\circ) = 1$, and if it is negative, at the boundary $\delta(\sigma^\circ) = 0$.

¹⁶Consider as well a discrete version of the problem, where the variable σ designates partitions. For that problem, one might be concerned about whether other system variables would be optimized (in general, differently) for different choices of a particular $\delta(\sigma)$. For the continuous version examined here, such would not matter (and, in any case, for present purposes, other policy variables are taken to be fixed, although in this formulation we can imagine them to be at their optimal values, which are invariant when characterizing the optimal δ for a given σ°). Also, in the discrete version, the optimum may involve randomization, there being an optimal fraction of cases in some partition that would be continued rather than terminated.

wherein the decision rule for some mass of states is fixed (perhaps due to an institutional constraint outside the model) in a nonoptimal fashion, because the pointwise optimization for any state σ° yields an expression that holds for any given settings of the decision rule for other states and not just when those other settings are optimal.

Examination of expression (4) indicates that the difference in the pertinent values of W will involve two types of effects. First, the values of the b^i will differ: if the case is allowed to continue ($\delta(\sigma^\circ) = 1$), these magnitudes will be higher, which is to say, the deterrence of harmful acts and the chilling of benign acts will be greater, in a manner determined by expression (3). Second, the values of the adjudication cost terms in both integrands will be larger since, in state σ° , defendant and government adjudication costs are incurred if $\delta(\sigma^\circ) = 1$ but not if $\delta(\sigma^\circ) = 0$, as indicated by expression (5). Setting $\delta(\sigma^\circ) = 1$ is strictly optimal – taking as given the decision rule for other states – if and only if:

$$(6) \quad \rho^H(\sigma^\circ) \lambda^H(\sigma^\circ) f^H(b^H) (h + \kappa^H - b^H) > \gamma \rho^B(\sigma^\circ) \lambda^B(\sigma^\circ) f^B(b^B) (b^B - \kappa^B) \\ + \left((1 - F^H(b^H)) \rho^H(\sigma^\circ) + \gamma (1 - F^B(b^B)) \rho^B(\sigma^\circ) \right) (c(\sigma^\circ) + k(\sigma^\circ)),$$

where

$$(7) \quad \kappa^i = \int_{-\infty}^{\infty} k^i(\sigma) z^i(\sigma) d\sigma.$$

The variable κ^i defined in expression (7) is the expected adjudication cost associated with each of the two types of acts (that is, averaged over all scenarios).

On the left side of the inequality in expression (6) is the deterrence gain from allowing cases with the signal σ° to continue. First, we have the deterrence punch, which is the product of the three factors before the larger parentheses. The first two are familiar from expression (3) that characterizes individuals' behavior: the likelihood that they will be in the legal system in state σ° , $\rho^H(\sigma^\circ)$; and the legal system's burden in that condition, $\lambda^H(\sigma^\circ)$. The final factor is the density evaluated at the marginal act, $f^H(b^H)$; this indicates, for a unit rise in individuals' expected cost for the harmful act, how many harmful acts are deterred. This deterrence punch is multiplied by the term in larger parentheses indicating the effect on social welfare per marginal act that is deterred: the sum of the external harm that is avoided and the expected adjudication cost savings, minus the benefit of the marginal act that is forgone. To elaborate on the middle term (which, from expression (7), is an integral over σ rather than being evaluated at σ°): for acts that are deterred, they will not enter the system for any realization of σ , and thus the adjudication costs will be avoided for all such σ (not just for σ°).

The first term on the right side of (6) is the corresponding chilling cost, reflecting that there is now a greater likelihood that benign acts will be subject to defendants' costs of adjudication and the possibility of being formally sanctioned. (Obviously, this term does not contain h because there is no external harm associated with benign acts.) This term is weighted by γ , indicating the relative mass of potential benign acts that may be committed.

The second term on the right side (row two) is the rise in expected adjudication costs

associated with state σ° : for all acts (of both types) that are committed – the $1 - F^i(b^i)$ terms indicate the undeterred and unchilled fractions of those who have opportunities to commit harmful and benign acts, respectively – the portion $\rho^i(\sigma^\circ)$ enter the legal system as being in state σ° , and the decision to continue rather than terminate in state σ° means that both the defendants' and the government's adjudication costs associated with that state are incurred.

If the deterrence gain exceeds the sum of the chilling cost and supplemental adjudication costs, it is optimal to allow the case to continue. Observe that the chilling cost can readily be a net benefit: specifically, if the marginal benign act that is chilled has a benefit, b^B , lower than the expected adjudication cost associated with the act being committed, κ^B , social welfare would be higher as a consequence of greater chilling. (Note importantly that the benefit of the pertinent act chilled depends on the marginal benign act, that for which individuals are just indifferent, which in turn depends on the overall legal system, notably, how continuation/termination decisions are made in all scenarios.)

Because of this point, it is not necessarily true, as in standard enforcement models (see Polinsky and Shavell 2007), that at the optimum there will be underdeterrence (relative to the first best, wherein acts are committed if and only if the private benefit exceeds h). Suppose, however, that the right side of (6) is positive – a possibility that may well hold even if chilling did happen to be desirable, as a consequence of the other term, reflecting adjudication costs in state σ° if the decision is to continue. Because the deterrence benefit on the left side includes not only avoiding h but also avoiding κ^H , the left side could be positive even if there was overdeterrence (that is, if $h < b^H$).¹⁷

Reflection on expression (6) suggests a number of features regarding how the optimal first-stage decision rule should be set for a given state (a given signal σ°), some of which may readily have been anticipated and others that are not as obvious and may initially seem counterintuitive. Regarding the Propositions to follow, two important observations should be noted. First, as just explained, it is possible that the chilling effect, the first term on the right side of (6), involves a net benefit rather than a cost. Some of the statements are qualified accordingly (by the further stipulation that $b^B > \kappa^B$).

Second, each of the propositions stipulates that “all else is equal.” The meaning is straightforward when the claim involves an exogenous parameter that, moreover, has no influence on any of the endogenous variables, notably, the b^i and the κ^i (the former of which, in turn, affects the f^i and the F^i). When an exogenous parameter does influence an endogenous variable, the meaning is that the values of other exogenous variables in other states, which do not directly appear in expression (6), are taken to be adjusted so as to keep those endogenous variables constant. Accordingly, the claim will involve a comparison of two different worlds (sets of exogenous variables) in which, in the state in question, only the posited variable differs. Similarly, when a claim concerns an endogenous variable, it is assumed that exogenous variables in other states differ so that, in the state under consideration, the only resulting difference regards the endogenous variable that is the subject of the claim. In these latter instances, it will be explained (in notes) how such situations in which “all else is equal” can be constructed. Formal proofs of the actual propositions are omitted because, given this proviso, the derivation of each result is reasonably straightforward.

¹⁷It is also possible that, considering all three elements, the gain per act deterred can be negative, which case will not be considered explicitly (and would not be optimal in typical settings).

The first set of results pertains to the benefits and costs of deterrence and chilling.

Proposition 1: When all else is equal,

a: a higher level of harm, h , favors $\delta(\sigma^\circ) = 1$,

b: a higher portion of potential benign acts, γ , favors $\delta(\sigma^\circ) = 0$ (assuming that $b^B > \kappa^B$)

c: a higher achieved degree of deterrence (b^H being larger) and a higher achieved degree of chilling (b^B being larger) each favor $\delta(\sigma^\circ) = 0$, and

d: a higher level of expected adjudication costs associated with either type of act, κ^H and κ^B , favors $\delta(\sigma^\circ) = 1$.

Proposition 1.a is obvious: raising h increases the deterrence benefit on the left side of expression (6) and has no other effect. Likewise for Proposition 1.b: raising γ (the mass of opportunities for benign acts) increases two of the costs (the chilling cost and a component of adjudication costs) on the right side of expression (6). (The reason for the additional proviso, a sufficient condition for the result, is given above.)

Proposition 1.c reflects that both deterrence and chilling involve the cost of forgoing individuals' benefits from their acts, and the greater the forgone benefit per act deterred or chilled, the larger is that cost. (For elaboration on the "all else equal" proviso regarding this result and the next, see the note.¹⁸) An important implication is that the optimal decision in a given state depends on the characteristics of other states. For example, anticipating Propositions 2.a and 2.b, if the signal in many other states strongly indicates a harmful act, so as to lead to continuation, the achieved degree of deterrence may be great, making continuation in the present state less advantageous. Similarly, if continuation in many other states involves substantial chilling, termination in the present state is favored. This result casts in a different light one of the central understandings of courts and legal scholars: that continuation (to allow discovery and the like) is necessarily favored in cases in which information is primarily in the possession of the defendant. On one hand, if this is generally true and it leads to termination in most states, then achieved deterrence and chilling may tend to be low, favoring continuation; on the other hand, if it is generally false but idiosyncratically true in the case (state) at hand, then the opposite situation may prevail, favoring termination. (And, as will be discussed in a moment, the result in Proposition 1.d also cuts against the standard view.)

Note further that the levels of achieved deterrence and chilling described in Proposition 1.c may in practice depend on considerations outside this model. If adjudication involves private suits but there is also substantial public enforcement that operates independently or there are strong market forces that contribute to deterrence (perhaps through reputation), then achieved deterrence may be high, favoring termination. The overall lesson is that, because ex ante behavioral effects are central, the aggregation of all forces – these external factors and, as

¹⁸For 1.c, the variables b^i , from expression (3), are endogenous, determined by other parameters. Note that the values of the $p^i(\sigma)$ only appear directly in (6) for the specific signal σ° , so it is possible to have different settings in which only the b^i differ by supposing different values of $p^i(\sigma)$ solely in other states. There is another complication with the "all else equal" proviso mentioned earlier in the text: the b^i also enter (6) through the values of the corresponding density and distribution functions. Here, one could postulate different benefit distributions such that, comparing two situations with different values of the b^i , these two values would be the same. For Proposition 1.d, where the κ^i differ, one could allow the value of k to differ in states other than σ° . (In this instance, note that differing k are not borne by actors and hence do not feed back on the b^i .)

emphasized just above, decisions in other states – importantly influences the deterrence benefit and the chilling cost of continuation in a given state. This conclusion contrasts with that in conventional, purely forward-looking decision problems involving the value of information, where hypothetical decisions in other states tend to be irrelevant to the decision at hand.

The result in Proposition 1.d, that greater expected costs of adjudication favor continuation over termination, may at first seem surprising. The explanation is that, the larger are these costs, the more is saved as a consequence of both greater deterrence and greater chilling. Recall from expression (7) that these κ^i are the aggregate expected costs associated with each type of act, of which the costs in a given state σ° are infinitesimal. The effects of the state-specific costs are considered below. Of course, higher expected (average) costs are associated with higher costs in at least some states. The point of Proposition 1.d, juxtaposed with the others, is to distinguish average adjudication costs from those in the particular situation at hand – which costs may be atypically high or low. Propositions 1.d, 3.a, and 3.b together indicate that the relationship among components of adjudication costs – average versus state-specific and, for the latter, defendants’ versus the government’s – matters a good deal and that some of the effects are contrary to what one might intuitively have supposed.

Observe that many of the factors that influence the b^i also influence the κ^i . In (6), these factors appear as a difference: for example, $b^B - \kappa^B$ on the right side. Subtracting the terms in the two integrands in (3) and (7) – which is more apparent when one reverses the notational simplifications to state explicitly the underlying components – one can determine that the net effect depends on the difference between the government’s adjudication cost in each state, $k(\sigma)$, and the pertinent expected sanction in each state, $p^i(\sigma)s$. (The actor’s adjudication costs, $c(\sigma)$, cancel, reflecting that, for each act discouraged, this cost is saved to society, but it also equals a subcomponent of the forgone benefit from the marginal act that is discouraged – each weighted by the same requisite probability.) Relatedly, it is this difference (along with the avoided external harm, in the case of acts of type H) that determines the net welfare effect of discouraging acts. Note that, in the preceding discussion of cross-state dependence of optimal decision rules, it was noted that lower preexisting deterrence – say, due to terminations in most other states – made deterrence more valuable at the margin. We can now see that, when we do not hold all else equal, it also means that expected adjudication costs per undeterred act will be less, which makes deterrence less valuable at the margin. Whether the net effect of, say, more terminations in other states, makes deterrence more or less valuable will depend on which of these forces is greater. (In this regard, note that h is a constant, independent of the level of achieved deterrence.) And, of course, the same logic applies to chilling.

The next set of results pertains to the relative magnitude of the deterrence punch and the chilling punch that are generated when cases are allowed to continue. Note that the proviso that chilling is net desirable is included for all of these claims; in each instance, it is a sufficient condition.

Proposition 2: When all else is equal and $b^B > \kappa^B$,

- a: a higher likelihood ratio, $\rho^H(\sigma^\circ)/\rho^B(\sigma^\circ)$, favors $\delta(\sigma^\circ) = 1$,*
- b: a higher probability that the harmful act will result in liability in final adjudication, $p^H(\sigma^\circ)$, and a lower probability of liability for the benign act, $p^B(\sigma^\circ)$, each favor $\delta(\sigma^\circ) = 1$, and*
- c: a higher density of harmful acts that are at the margin of being deterred, $f^H(b^H)$, and a lower density of benign acts that are at the margin of being chilled, $f^B(b^B)$, each favor $\delta(\sigma^\circ) = 1$.*

Proposition 2.a asserts that, the greater the relative likelihood that the signal in question, σ° , is associated with harmful rather than benign acts, the more valuable it is to allow the case to proceed, because the relative magnitude of deterrence versus chilling is greater. Note that this fraction is the analogue to the likelihood ratio for the present problem (although, as Proposition 4 will indicate, the optimal rule is not a standard likelihood ratio test). Moreover, if we divide both sides of our optimality condition (6) by $\rho^B(\sigma^\circ)$, we can plainly see that the $\rho^i(\sigma^\circ)$ only enter as the likelihood ratio. This ratio multiplies the deterrence benefit but only one component of the cost, making continuation more attractive. (The foregoing oversimplifies in a manner that may not be obvious in that the $\rho^i(\sigma^\circ)$ each depend on the respective π^i , while both the b^i 's and the κ^i 's depend on the π^i as well, which complicates the “all else is equal proviso,” as elaborated in the margin.¹⁹)

Proposition 2.b indicates that higher diagnosticity of stage-two adjudication – the more (less) likely it is that sanctions will be applied to harmful (benign) acts – favorably influences the magnitudes of the deterrence punch and the chilling punch and thus favors continuation. (Here, recall from expression (2) that $\lambda^i(\sigma^\circ) = c(\sigma^\circ) + p^i(\sigma^\circ)s$.)

Regarding Proposition 2.c, in determining the deterrence and chilling punches, what matters are not only the changes in individuals' expected costs for the two types of acts but also how many harmful and benign acts will be deterred and chilled, respectively, for given increases in their expected costs. This feature, in turn, is indicated by the magnitudes of the densities for individuals' benefits from acts, each evaluated for the pertinent marginal act. When near the peaks of the distributions, these effects are large, whereas at other points the effects can be small. In general, either of these densities might be higher. Which it is – and how large any difference might be – will depend on the nature of the two distributions as well as on all the features of the legal system. Regarding the latter, even if the two distributions were the same, since individuals' expected costs for the two types of acts will differ, perhaps substantially, so may the values of these density functions.

The next set of results concerns adjudication costs from continuation, the last term in expression (6).

Proposition 3: When all else is equal,

- a. a higher state-specific continuation cost borne by the defendant, $c(\sigma^\circ)$, has an ambiguous effect on the optimal $\delta(\sigma^\circ)$,*
- b. a higher state-specific continuation cost borne by the state, $k(\sigma^\circ)$, favors $\delta(\sigma^\circ) = 0$,*
and
- c. a higher mass of undeterred harmful acts, $1 - F^H(b^H)$, and a higher mass of unchilled benign acts, $1 - F^B(b^B)$, each favors $\delta(\sigma^\circ) = 0$.*

Proposition 3.a may seem surprising, especially when juxtaposed with Proposition 3.b, which is obvious. Just as with $k(\sigma^\circ)$, a higher $c(\sigma^\circ)$ raises the last term in expression (6),

¹⁹Even if one considers separately the ratio of the $z^i(\sigma^\circ)$ and of the π^i , the statement holds. For the former, this is obvious because the complication mentioned in the text does not arise. For the latter, it is possible to construct comparisons in which all else is indeed equal by adjusting other exogenous parameters in other states in a manner that keeps them constant. Starting with the b^i 's, one could adjust the c 's or p^i 's in other states. Then, for the κ^i 's, one could adjust the k 's. (If the former is done by adjusting the c 's, that also influences the κ^i 's, so the needed adjustment to the k 's would need to reflect that as well.)

indicating that it is more expensive to go forward. However, a higher $c(\sigma^\circ)$ also raises the deterrence and chilling punches (reflected on the left side of expression (6) and in the first term on the right, respectively) because individuals thereby anticipate paying greater costs when their cases will proceed to final adjudication. (Recall again from expression (2) that $\lambda^i(\sigma^\circ) = c(\sigma^\circ) + p^i(\sigma^\circ)s$.) In an optimal system, this deterrence gain will tend to exceed the chilling cost (because if it did not, it would probably be true that too few cases were being terminated, in light of adjudication costs). Accordingly, we can imagine situations under which a higher defendant's adjudication cost would tip the balance in favor of continuation. Specifically, suppose that the factors in Propositions 1 and 2 would favor continuation in a wide range of states were it not for the fact that, in most of those states, $k(\sigma)$ was so high that termination was best. Now, in the state σ° under consideration, suppose that $k(\sigma^\circ)$ is barely high enough that termination is optimal. In that setting, a higher $c(\sigma^\circ)$ might produce a deterrence gain that is relatively large compared to both the chilling cost and the added adjudication cost, swinging the balance in favor of continuation.

Proposition 3.c is straightforward from examination of the final term in expression (6). Note that these factors involve inframarginal effects; specifically, they refer to the undeterred and unchilled portions of the population. By contrast, both the deterrence and chilling effects depend on the densities (see Proposition 2.c), reflecting that they involve *changes* in behavior due to switching the decision from terminate to continue, rather than the existing *stock* of behavior, in aggregate.²⁰ (Note that it is only here – really, in the last line of expression (6), showing the direct costs of continuation – that inframarginal effects matter, whereas they are central in the more familiar, forward-looking, value of information problem. Moreover, here the two inframarginal measures enter additively, which is qualitatively different from how they affect the standard problem.)

The results so far focus on the determination of the optimal rule in a given state σ° . Using these results, we can also compare optimal rules across states. In this respect, the following result is of interest:

Proposition 4: The optimal $\delta(\sigma)$ do not take the form of a likelihood ratio test. That is, when $\rho^H(\sigma^\circ)/\rho^B(\sigma^\circ) > \rho^H(\sigma^1)/\rho^B(\sigma^1)$ for some σ° and σ^1 , it is possible that it is optimal for $\delta(\sigma^\circ) = 0$ and $\delta(\sigma^1) = 1$.

In many settings, optimal decision rules do take the form of a likelihood ratio test, which is to say that there exists some cutoff for the likelihood ratio such that one decision is always optimal whenever the likelihood ratio exceeds the critical value and the other decision is always optimal whenever the likelihood ratio is below the critical value.²¹ Here, such is not true because other factors than the signal strength are a function of the signal: specifically, both continuation costs, $c(\sigma^\circ)$ and $k(\sigma^\circ)$, and also the conditional probabilities of liability, $p^i(\sigma^\circ)$ – again recalling that $\lambda^i(\sigma^\circ) = c(\sigma^\circ) + p^i(\sigma^\circ)s$. In this regard, Proposition 4 can be regarded as a corollary of Propositions 2.b, 3.a, and 3.b.

²⁰Observe further that it is easy to construct changes in the F^i that do not affect the f^i by moving some of the mass from one end of the support to the other (that is, not at the b^i in question).

²¹On likelihood ratio tests, see, for example, Neyman and Pearson (1933), Karlin and Rubin (1956), and Milgrom (1981).

Concretely, we can illustrate this case by supposing that, in the two compared states, the one with the higher likelihood ratio has an extremely high $k(\sigma^\circ)$, so termination is optimal, but the one with the lower likelihood ratio has a tiny $k(\sigma^\circ)$, and continuation is optimal. (Recall from Proposition 3.b that a higher $k(\sigma^\circ)$ unambiguously favors termination, which is obvious from expression (6).) This possibility is of practical importance. When all continuation costs are near zero, continuation can readily be optimal even if the likelihood ratio based on the initial signal is quite low because one might learn (essentially for free) that it would be optimal to assign liability. And continuation costs are notoriously variable, depending on the particular facts of an individual case. Accordingly, attempting to state non-final-stage decision rules purely in probabilistic terms ignores important features embodied in an optimal continuation/termination rule.

In spite of Proposition 4's important lesson, likelihood ratios are indeed important in the present model, as we would expect. First, as Proposition 2.a already states, a higher likelihood ratio, *ceteris paribus*, does favor continuation (under the stated conditions). Second, as will be demonstrated in section 3.C, below, the optimal final-stage decision rule does take the form of a likelihood ratio test, the key difference being that all of the other factors that depend on the signal pertain to features of continuation that are moot in the final stage.

Finally, it is natural to inquire whether there are some simple sufficient conditions under which the first-stage decision rule would be a likelihood ratio test. It turns out that there are not. Even if one assumes that c and k are constant (and thus independent of the signal σ), our first-order condition (6) also depends on the $p^i(\sigma^\circ)$ (through the $\lambda^i(\sigma^\circ)$, as indicated by (2)). Of course, it is not natural to consider the case in which those probabilities are independent of the signal, for then the signal tells us nothing about case strength. We could take these first-stage probabilities of liability to be optimally determined (by their own likelihood ratio test, as mentioned, which appears in section 3.C) and impose further restrictions on the $p^i(\sigma^\circ)$, but even then (because $p^H(\sigma^\circ)$ and $p^B(\sigma^\circ)$ multiply different expressions), the resulting condition need not constitute a pure likelihood ratio test.²²

C. Additional Enforcement Instruments

The present analysis can be extended and thereby more closely compared with prior literature on the economics of law enforcement by allowing the government also to choose the level of sanctions and enforcement effort. The discussion in this section will be brief, with derivations in the appendix.

Begin with the sanction s . It does not appear directly in W (see expression 4), its influence being through deterrence and chilling, as indicated by expression (3), via expression

²²To pursue this more formally, one can divide both sides of condition (6) by $\rho^B(\sigma^\circ)$ (so that the likelihood ratio then appears in two places, on the left side and in a component of the cost term on the right side) and then isolate the likelihood ratio by placing it on the left side and everything else on the right. This representation indicates that there should be continuation if and only if the likelihood ratio exceeds the value of a cumbersome expression. Previously (when discussing Proposition 4), each of the terms in that expression depended on σ° . The present assumptions, as mentioned, still leave the probabilities, the $p^i(\sigma^\circ)$ in this expression. In addition to the right side being non-constant, further analysis suggests that it is not possible to show that appropriate restrictions on the $p^i(\sigma^\circ)$ – including that they were determined optimally at the final stage – imply that the expression was (nonstrictly) monotone decreasing, which would have been sufficient to prove the existence of a critical likelihood ratio.

(2): $\lambda^i(\sigma) = c(\sigma) + p^i(\sigma)s$. The first-order condition for an optimal interior sanction equates the deterrence gain to the chilling cost. (See expression (A2).) Obviously, if both involve a net benefit (recall the above discussion), the optimal sanction is maximal. More broadly, if the government can adjust all policy instruments – for the moment, continuation/termination decisions and the sanction – the optimal sanction will be maximal regardless, a result reminiscent of that suggested by Becker (1968), although the analysis differs in the present setting.

Proposition 5: If $p^H(\sigma) \geq p^B(\sigma)$ for all σ , $b^B > \kappa^B$, and the government may choose the $\delta(\sigma)$, then the optimal sanction, s , is maximal.

To see why this result arises, suppose that s is not maximal. Consider the experiment of marginally raising s and also switching from continuation to termination in states with the worst effective diagnosticity, that is, the highest $(c(\sigma)+p^B(\sigma)s)z^B(\sigma) / (c(\sigma)+p^H(\sigma)s)z^H(\sigma)$.²³ (See the discussion of expression (A5), and note that the substitution using notation $\lambda^i(\sigma)$, as given by expression (2), is omitted here because the components have differential significance for some of the argument to follow.) Furthermore, make these adjustments to an extent that keeps deterrence, that is, b^H , constant. It can be shown that there are three effects, all of which are favorable.

First, terminations save adjudication costs. Second, chilling falls due to the better targeting of the sanction s across states. The heightened s applies (probabilistically) in all states with continuation – an average effect – whereas the now-terminated states had the worst (and hence necessarily a below-average) targeting of expected system costs. Third, chilling also falls because (assuming that $p^B(\sigma)/p^H(\sigma) < 1$)²⁴ sanctions are better targeted on harmful versus benign acts than are defendants' continuation costs, $c(\sigma)$. (See the discussion of expression (A4).) Keep in mind that a key component of both deterrence and chilling punches, for each state, is $c(\sigma)+p^i(\sigma)s$: the sum of the defendant's continuation costs and the probability of the sanction times its magnitude. Continuation costs in a given state are the same for harmful and benign acts, whereas the likelihood of being sanctioned at the final stage is taken to be higher for the former. Hence, switching from continuations to a higher s as the source of a given level of deterrence relies more on the sanction and less on continuation costs, which is relatively favorable for benign acts. Interestingly, the latter two effects indicate that a concern for chilling, which arises from the legal system's incidental inclusion of innocent behavior, bolsters the case

²³Interestingly, this concept of effective diagnosticity used in the proof of Proposition 5 is materially different from the more familiar likelihood ratio, which further illustrates why the likelihood ratio is not a sufficient statistic, as indicated by Proposition 4.

²⁴One might have suspected that, had we assumed that final-stage liability was determined optimally (as indicated by condition (14) or (15) in section 3.C), this assumption would follow, in which case it would be appealing to substitute an optimality assumption rather than this raw assumption. However, this is not the case. Suppose, for example, that when some σ is generated by a harmful act, the subsequent signal will be dispositive, one way or the other (so the likelihood ratio with both signals realized will be infinite or zero), with liability being optimal when the subsequent signal is positive – which will have some associated probability, $p^H(\sigma)$. Suppose further that, when the same initial signal is generated by the benign act, the subsequent signal is very noisy. Then, the associated probability, $p^B(\sigma)$, will depend in significant part on how high is the critical likelihood ratio, which optimally depends on many other parameters of the problem (such as the relative importance of deterrence and chilling); if the critical likelihood ratio is sufficiently low, we will have $p^B(\sigma) > p^H(\sigma)$.

for a maximal sanction in an optimized enforcement regime.

Now consider enforcement effort. This may be introduced by supposing that the π^i are each a function of expenditures, e . Assume that $\pi^{i'}(e) > 0$, $\pi^{i''}(e) \leq 0$, for $i = H, B$. (It may be natural for some interpretations that there be a maximum level of e , such as when an audit rate reaches one hundred percent.) In expression (4) for social welfare, one would insert an additional term, “ $-e$,” at the end. The first-order condition for an interior optimum includes deterrence and chilling effects (through the influence of e on the π^i (a component of the $\rho^i(\sigma)$) in expression (3) for the b^i), inframarginal costs due to the fact that all undeterred and unchilled acts are now more likely to enter the legal system (see the integrands in expression (4)), and a “ -1 ” reflecting the cost of enforcement effort itself. (See expression (A6).)

In achieving, say, a given level of deterrence, it is interesting to compare the level of e with the choice of the set of states in which cases are continued rather than terminated, in a manner analogous to the foregoing discussion. As with raising s and terminating in more states, raising e and terminating in more states again saves continuation costs and can also be beneficial with regard to chilling effects by terminating in those states with the weakest effective diagnosticity (because greater enforcement effort increases the flow in all states, a sort of average effect). However, the benefit with regard to chilling that s operates only through the p^i is inapplicable because, like continuation, raising e , which increases the π^i , increases both deterrence and chilling also through defendants’ adjudication costs c .

By comparison to raising s , raising e has additional disadvantages. Most obvious is the cost of the expenditure itself. In addition, by raising the π^i , more undeterred and unchilled acts enter the legal system in all states, as mentioned, which thereby raises total system costs in all states where continuation is still employed. A final point concerns the targeting of increases in e . If the enforcement technology involves purely random audits, raising e would raise the π^i in proportion. However, for other means of enforcement – or for audits that are prioritized, say, by some preliminary signal – it will be optimal to target the most promising cases first, so there will be diminishing returns in terms of targeting precision. Put another way, raising e would tend to raise π^B relatively more than π^H , and this will be disadvantageous. Taken together, some factors favor raising e and terminating in more states, and others are opposed. Accordingly, the optimal system will ordinarily involve an intermediate level of enforcement effort combined with continuation in some but not all states.

Anticipating section 3, note that the enforcement effort decision can in many respects be interpreted as an additional (earlier) stage of adjudication. When enforcement actions are guided by some signal, the analogy is fairly precise. In that event, the choice between enforcement effort and continuation/termination decisions at stage 1 can instead be understood as one about optimal continuation/termination decisions at different stages, a subject addressed in section 3.B.

Finally, compare the choice between enforcement effort and sanctions, a central topic in the literature on the economics of law enforcement (Polinsky and Shavell 2007) and the locus of Becker’s (1968) original argument favoring raising s and lowering e , keeping deterrence constant but saving enforcement resources. (For the analogue here, see expression (A7).) The present model adds benign activity that may be chilled and multiple stages of adjudication. Drawing on the preceding discussion, we can see that Becker’s experiment plausibly remains favorable, and for a number of additional reasons. The direct cost point (reducing e saves resources but raising s does not) continues to hold. There is the additional inframarginal savings because fewer undeterred and unchilled acts enter the legal system and thus (probabilistically,

depending on the state) result in continuation costs. In addition, because of chilling effects, there may well be additional advantages: e influences behavior through defendants' continuation costs c , which are less diagnostic than s , and for some enforcement technologies, e is subject to diminishing returns in targeting efficiency. As shown in the appendix, however, the point that continuation costs more poorly target harmful versus benign acts is more complex and, without further assumptions, is ambiguous, in contrast to the earlier construction involving an increase in s accompanied by termination in additional states. (See the discussion of expression (A8).)

3. Multistage Adjudication

Legal systems often employ more than two stages. Most familiar in the United States, civil litigation can be terminated at the outset, by granting a motion to dismiss; after costly discovery, by granting a motion for summary judgment; or at the end of a trial (or midway through, at the end of the plaintiff's presentation). But the application is much broader, often including a larger number of steps, where their nature may be informal. Regulatory agencies employ internal procedures that periodically assess whether an investigation should be continued or terminated. And other legal systems, including many continental courts and some modes of arbitration, proceed in smaller steps, deciding along the way what evidence to consider next and when there is enough clarity to reach a decision rather than to continue. Accordingly, it is of interest to extend section 2's model to this setting, particularly for purposes of seeing what light can be shed on how optimal decision rules differ across stages. This section also considers how intermediate-stage rules differ qualitatively from the optimal final-stage rule.

A. Model

Before plunging into the details – and because some of the analysis is relegated to the appendix – it is helpful to discuss intuitively how to think about this problem. Suppose that a case is at some intermediate stage. Two sets of considerations will be relevant. The first concerns how the case got there. Although prior costs are sunk, it is necessary to know the probability of arriving at the current stage along the particular signal path that has been realized so far because the current decision will depend on those prior signals as well as the one just realized. The second concerns continuation. Because there may (or may not, depending on later signal realizations and consequent decisions) be several subsequent steps, all variables concerned with the future will now take the form of expected values. Notably, continuation costs – those borne by the defendant and by the government – will include not only the certain cost of moving to the next stage (if indeed the decision at the current stage is to continue) but also, with varying probabilities (depending on the future signal realizations and subsequent-stage decision rules), costs of moving to later stages as well. Likewise, instead of there simply being some probability of liability if the case continues, we need to take the expected value of the final-stage probability (which itself is, of course, a probability), which incorporates as well the probabilities of making it through all of the intervening stages.

With this understanding in mind, let us now describe the multistage version of the model explicitly. Let t denote the stage of adjudication, with the final stage designated by T , where $T > 2$. The analysis in section 2 can be understood as applicable to $t = 1$, where the stage was suppressed and all variables involving continuation can now be understood as implicitly

referring to pertinent expected values, as will now be made explicit.

Consider first the decision whether to allow a case that has entered stage $t < T$ to continue to stage $t+1$ ($\delta_t = 1$) or terminate at that point ($\delta_t = 0$). For a case that continues, the two adjudication costs in moving to the next stage are now designated as c_t borne by the defendant and k_t by the government. It is also useful to introduce the notation $E_t^i(c)$ and $E_t^i(k)$, where $E_t^i(\cdot)$, the expected value operator as of stage t , refers to the expected sum of the pertinent costs borne in all future stages, conditioning on the fact that one has reached stage t and decides to continue. Note that the costs c_t and k_t , in moving from stage t to $t+1$, are certain when $\delta_t = 1$, whereas for all subsequent stages (if any), they are conditional on the subsequent signals and on the decision rules contingent on those signals being such that the case is continued yet again rather than terminated. (All costs incurred to reach stage t are sunk and thus are excluded from this expression for expected continuation costs.) Also, the superscript i appears on the expectation operator because, although the costs incurred in moving to the next stage are taken to be the same for the two types of acts, expected costs for subsequent stages depend on subsequent decisions, which in turn depend on the subsequent signals, the likelihoods of which generally differ for the two types of acts.

A case that reaches stage T is tried and results in a finding of liability ($\delta_T = 1$) or instead is terminated ($\delta_T = 0$). As before, individuals found liable are subject to the socially costless sanction s . At nonfinal stages t , the pertinent probability of ultimately being sanctioned is designated $E_t(p^i)$, which here refers to the probability that the final stage T is ultimately reached and that liability is also found at that stage, all conditional on a case already having arrived at stage t .

As before, individuals who commit their acts enter the legal system (at the first stage) with probabilities π^i . For this multistage version of the problem, it is supposed that, at the outset of each stage t , there is a signal θ_t . In any stage t , the realizations of signals up to and including that stage are known. It is convenient to adjust the previous notation to use, for each stage t , the symbol σ_t to denote the vector of these signals: i.e., $\sigma_t = (\theta_1, \dots, \theta_t)$. At each stage, the signal θ_t has a conditional density – i.e., it is conditioned on $\sigma_{t-1} = (\theta_1, \dots, \theta_{t-1})$ – given by the function $z_t^i(\sigma_t)$, positive on the real line, where the superscript i again indicates that these densities depend on the type of act. (For elaboration, see the footnote.²⁵)

The $E_t^i(c)$, $E_t^i(k)$, and $E_t(p^i)$ each depend on σ_t . The government's problem is to choose, at the outset, for each stage t , the function $\delta_t(\sigma_t)$: that is, to indicate, for each σ_t , whether the case is to continue to stage $t+1$ (if $t = T$, to assign liability) or is instead to terminate.

The model for adjudication with T stages can be summarized by reference to its timing:

1. The government sets all policy instruments, notably, the functions $\delta_t(\sigma_t)$, for all $1 \leq t \leq T$.

²⁵Specifically, for any stage t , let $Z_t^i(\theta_1, \dots, \theta_t)$ be the joint density function, positive on the domain \mathbb{R}^t . Then, the conditional density function specified in the text is defined by this joint density function divided by the pertinent marginal density (which is just the joint density on the conditioned variables from the prior stage):

$$z_t^i(\sigma_t) = z_t^i(\theta_1, \dots, \theta_t) \equiv Z_t^i(\theta_t | \theta_1, \dots, \theta_{t-1}) = \frac{Z_t^i(\theta_1, \dots, \theta_t)}{Z_{t-1}^i(\theta_1, \dots, \theta_{t-1})}$$

(For $t = 1$, the denominator on the right side is taken to equal 1.) For convenience, this further notation is suppressed throughout.

2. Individuals learn their type of act (H or B) and their private benefit b .
3. Individuals decide whether to act.
4. A portion of those who commit each type of act, π^H and π^B , are identified and brought before a tribunal, in which event costs c_0 and k_0 are incurred by the defendant and the government, respectively.
5. The stage is $t = 1$, and the signal θ_1 is realized.
6. The tribunal either allows the case to continue (if $\delta_t(\sigma_t) = 1$) or instead terminates it (if $\delta_t(\sigma_t) = 0$); in the latter case, the game ends.
7. If the case continues:
 - a. Costs $c_t(\sigma_t)$ and $k_t(\sigma_t)$ are incurred by the defendant and the government, respectively.
 - b. The case enters the next stage (t is incremented by 1).
 - c. The new signal θ_t is realized.
 - d. If $t < T$, the case reenters step 6; else, it goes to step 8.
8. In final adjudication, stage T , the tribunal finds liability (if $\delta_T(\sigma_T) = 1$) and thus applies the sanction s or instead finds no liability (if $\delta_T(\sigma_T) = 0$).

Paralleling section 2, let us first introduce the some auxiliary notation:

$$(8) E_0(\lambda^i) = E_0^i(c) + E_0(p^i)s,$$

the legal system's expected burden on each type of act, at the outset. The pertinent expectation operators in expression (8) are assessed at what is here referred to as stage 0. Recall that stage 1 is the stage at which a case enters the legal system. Stage 0 is to be interpreted as the preceding point in time, at which individuals decide whether to act (step 3 in the above time line). The appendix derives the expected values in expression (8) and others that will be needed below, using backward induction and recursion.

Those whose type of act is i commit their act if and only if:

$$(9) b > E_0(\lambda^i) \equiv b^i.$$

Expression (9) is the multistage analogue to expression (3), when the process was taken to involve only a single preliminary stage ($t = 1$) and a final stage ($T = 2$).

As in section 2, social welfare is taken to be the aggregate of individuals' benefits from acting minus the harm from the commission of acts of type H and the costs of defendants' and the government's expenditures on adjudication.

$$(10) W = \int_{b^H}^{\infty} (b - h - \kappa^H) f^H(b) db + \gamma \int_{b^B}^{\infty} (b - \kappa^B) f^B(b) db,$$

where

$$(11) \kappa^i = E_0^i(c) + E_0^i(k).$$

As mentioned, the government's problem is to define a rule, for each possible realization σ_t , at each stage t , that indicates whether the case is to continue or terminate. Just as when analyzing the first stage ($t = 1$) in the two-stage setting, examination of expression (10) indicates that the difference in the pertinent values of W will involve two types of effects. First, the values of the b^i will differ: if the case is allowed to continue, these magnitudes will be higher, which is to say, the deterrence of harmful acts and the chilling of benign acts will be greater. Second, for any

nonfinal stage ($1 \leq t < T$), the values of the adjudication cost terms in both integrands will be larger since, if the case continues rather than terminates, defendants' and the government's adjudication costs are incurred both in moving to the next period, $t+1$, and, with a probability, to subsequent periods as well.

B. Intermediate Stages

The condition at stage t in state σ_t° for continuation ($\delta_t(\sigma_t^\circ) = 1$) to be strictly optimal (which must hold almost everywhere), expression (A16) in the appendix, is quite similar to expression (6) for the first-stage decision ($t = 1$) in the two-stage model:

$$(12) \quad \rho_t^H(\sigma_t^\circ) E_t(\lambda^H)(\sigma_t^\circ) f^H(b^H) (h + \kappa^H - b^H) > \gamma \rho_t^B(\sigma_t^\circ) E_t(\lambda^B)(\sigma_t^\circ) f^B(b^B) (b^B - \kappa^B) \\ + (1 - F^H(b^H)) \rho_t^H(\sigma_t^\circ) E_t(\kappa^H)(\sigma_t^\circ) + \gamma (1 - F^B(b^B)) \rho_t^B(\sigma_t^\circ) E_t(\kappa^B)(\sigma_t^\circ),$$

where

$$(13) \quad E_t(\kappa^i(\sigma_t^\circ)) = E_t^i(c)(\sigma_t^\circ) + E_t^i(k)(\sigma_t^\circ).$$

All legal system attributes except the $\delta_t(\sigma_t^\circ)$ under consideration are taken as given. The other δ_t can be taken to be optimal or simply as stipulated. To determine the optimal sequence of the δ_t for each possible sequence of signals, one would use backward induction.²⁶

To reinforce the similarity and to solidify understanding of the decision at intermediate stages, it is useful to examine briefly the terms in expression (12). On the left side of the inequality is the deterrence benefit of allowing a case that has reached stage t in state σ_t° to proceed rather than to terminate at that point. The deterrence punch is the product of the likelihood that one committing the act will be in the legal system in state σ_t° , $\rho_t^H(\sigma_t^\circ)$ (which takes into account all of the preceding decisions in a manner captured by expression (A15) in the appendix); the legal system's burden in that condition, $E_t(\lambda^H(\sigma_t^\circ))$; and the magnitude of the density function evaluated for the marginal type, $f^H(b^H)$ (indicating the quantity of acts deterred per unit increase in the expected legal burden). This deterrence effect is, as before, multiplied by the deterrence benefit per act deterred for the marginal act. The avoided expected adjudication cost now reflects, as expressed in (11), the expected costs through all stages for a harmful act.

The first term on the right side of the inequality for the chilling cost is analogous. The final two terms (the second row) are, as in section 2's model, the increment to the expected adjudication costs incurred with respect to inframarginal acts of both types. For each type of act, we have the product of: the mass of acts of that type that are committed, $1 - F^i(b^i)$; the probability that the type of act enters stage t and also is associated with the pertinent signal path, $\rho_t^i(\sigma_t^\circ)$; and the sum of expected defendants' and government costs, going forward, $E_t(\kappa^i(\sigma_t^\circ))$. Note that this latter component excludes the costs incurred in reaching stage t : for inframarginal

²⁶Note that in general there can exist multiple local optima. To illustrate, suppose that adjudication costs in continuing from stage 1 are a relatively large factor. There might exist a local optimum in which most cases are terminated at stage 1 (which, note, implies that expected system costs per undiscouraged act, the κ 's in expression (11), will be low, which makes deterrence and chilling less valuable, suggesting the desirability of termination). And there may exist a local optimum in which most cases are continued at stage 1 (which reverses the foregoing logic).

acts, these prior-stage costs are sunk; specifically, they would not be avoided if a case was terminated at stage t rather than continued, which is the decision under consideration.

In spite of the greater complexity embedded in the recursive expressions for various of the terms, the decision rule represented in expression (12) is qualitatively the same at nonfinal stages in the T -stage model as it is at the first stage in the two-stage framework (expression (6)). Hence, the elaboration on the analytics appearing in section 2.B does not need to be repeated. Specifically, analogues to all of the first five Propositions hold.

Additional features: This framework with multiple stages prior to the final stage of adjudication offers a number of lessons that go beyond those in the simpler version of the problem. Specifically, these relate to how the optimal decision rule differs at different stages (e.g., does optimal stringency rise or fall as one moves to later stages?) and how the stringency at a given stages depends on that at other stages (e.g., if stringency is loosened at one stage – perhaps due to an external constraint or the presence of agency problems – how is the optimal stringency at other stages affected?). Let us consider these two questions in turn.

We can answer the first question, of how optimal stringency may change as we move to later stages, by examining how each of the individual components of expression (12) differs across stages. Begin with the deterrence and chilling effects. Consider first the determinants of $\rho_t^i(\sigma_t^\circ)$, which (recall from expression (1)) includes the probability $\pi_t^i(\sigma_t^\circ)$ that an act of the pertinent type enters stage t along the posited path of signal realizations. When entering the first stage, we simply have π^i . From expression (A15), we can see that in subsequent stages, the magnitude of $\pi_t^H(\sigma_t^\circ)$ relative to $\pi_t^B(\sigma_t^\circ)$ will depend on the discriminating power of the signals in prior stages. Accordingly, the more it is true on a signal path that prior decisions to continue were not close calls, the more likely it is that it would be optimal to continue in stage t because the deterrence effect will be larger relative to the chilling effect, *ceteris paribus*.

The next factor in the deterrence and chilling effect terms, $E_t(\lambda^i)$, represents the increment to individuals' expected costs from a decision to proceed at stage t . On one hand, as a case proceeds to later stages, subsequent adjudication costs are lower since prior stages' costs are sunk. (See expression (A9).) On the other hand, the expected probability of liability with continuation tends to be higher as the case gets closer to the final stage and thus a possible finding of liability. (See expression (A13).) Taken together, the aggregate expected legal burden on an actor imposed by continuation decisions could be rising or falling as cases progress to later stages. For example, for deterrence it may be rising, because the expected sanction is relatively more important than expected litigation costs, but for chilling it may be falling, because the opposite is true.

By contrast, our next factor – the magnitude of the density function for the distribution of individuals' benefits from each of the two types of acts, evaluated for the marginal act, $f^i(b^i)$ – is independent of what stage a case is in. Similarly, neither the gain per harmful act that is deterred (avoided harm and adjudication costs minus the benefit from the act forgone) nor the cost per benign act that is chilled (forgone benefit minus avoided adjudication costs) depends on the stage a case is in. Also, as expression (11) depicts, the avoided adjudication costs in question here are from discouraged acts, so they are the expected adjudication costs of the entire process, not just that for a particular stage.

Continuation also increases inframarginal adjudication costs for undeterred harmful acts and for unchilled benign acts, as indicated in the second row of expression (12). The masses of

activity for the two types of acts, $1-F^i(b^i)$, do not depend on the stage, whereas the $\rho_i^i(\sigma_i^\circ)$ (discussed above) do. Regarding the final terms, the expected defendants' and government adjudication costs of going forward: in later stages, more of both components of total adjudication costs are sunk, so it takes less of a deterrence gain relative to the chilling cost to warrant continuation. That is, this factor – the direct cost of continuation – favors an increasingly lenient approach toward continuation as cases progress through the legal system.

Having now explained each of the reasons that an optimal decision at one stage may be more or less stringent than the decision at, say, the preceding stage, we are in a position to address the conventional wisdom that stringency should optimally rise as we move to later stages. It seems clear that, as a general proposition, this view is highly problematic. Most obviously, it was just explained that, because more adjudication costs are sunk as one proceeds to later stages, there is a reason that optimal rules become more generous toward continuation, not more stringent. Moreover, it was also noted that it was plausible (although not necessarily true) that deterrence effects may be rising and chilling effects falling as we proceed further, which also would favor more generous continuation. Yet other factors are less determinate. Although the common conjecture cannot be definitively rejected, it seems quite difficult to endorse it as even approximately correct in most settings. (This conclusion would be bolstered if we compared the final-stage decision rule, considered in the next section, to the optimal rule at stage $T-1$. At the very end, continuation costs are zero, entirely eliminating two of the three costs on the right side of expression (12).)

An important caveat must be offered with regard to the foregoing discussion because there is an important ambiguity in characterizing decision rules at one stage as more or less stringent than those at another stage. Proposition 4 (which, like the others in section 2, is applicable here) indicates that the optimal rule does not take the form of a standard likelihood ratio test.²⁷ Because a given stage does not have a common threshold likelihood ratio, one cannot describe (in any simple manner) the threshold at one stage as higher or lower than that at another. However, all else equal – holding the magnitudes of other stage-specific variables constant – one can address differential stringency. This point casts further light on such phenomena as the fact that, in U.S. civil litigation, it is considered to be relatively easy for a case to survive a motion to dismiss (at the outset), harder to survive a defendant's motion for summary judgment (after discovery), and harder still to win at trial – and conventional wisdom regards this state of affairs to be appropriate. We can now see that this understanding is not entirely coherent and, if refined to make it so, may well not be optimal because so many factors change at subsequent stages of adjudication. Most obviously, more adjudication costs are sunk as cases proceed through the system, providing an important reason to suppose that optimal stringency (properly interpreted) may be falling, not rising, as cases move to later stages.

Let us now turn to our second question: How does greater generosity or stringency in making continuation decisions at some stages influences how decisions are optimally made at others? To answer this question, we now consider decisions as a whole rather than just on a particular path of signal realizations because a number of key variables depend on the overall operation of the system. Propositions *1.c* and *1.d* draw our attention to two variables. First, the values of the b^i – i.e., the benefits of the acts of each type that are just at the margin – depend on

²⁷As shown in section 3.C, the optimal rule is a likelihood ratio test for the final stage, but that is insufficient for unambiguous comparisons because none of the rules at other stages take this form.

the extent of deterrence and chilling: when they are larger, the benefit of the marginal forgone private benefit is greater, so deterrence is less valuable and chilling is more costly. Second, the κ^i likewise depend on the aggregate behavior at the system. Hence, these factors are influenced not only by decisions in other states at the same stage (as discussed in section 2) but also by decisions at other stages. As before, these two sets of considerations (marginal forgone benefits and expected adjudication costs) often cut in the opposite direction: On one hand, greater overall leniency raises the value of deterring a marginal harmful act and reduces the cost of chilling a marginal benign act because the forgone private benefit is lower. On the other hand, greater overall leniency results in less deterrence and chilling, so more cases flow into the legal system, which raises expected system costs and hence makes deterrence and chilling more beneficial, all else equal. Finally, because greater leniency means that each of the two density functions for actors' private benefits in general have different values, there is an additional source of indeterminacy. To fix thinking, the following discussion will focus on the case in which the first consideration is dominant, so that greater overall leniency makes deterrence more desirable and chilling less damaging. (Implications for the opposite case will be the reverse.)

Suppose now that, in the first stage, a large portion of cases are terminated, say because the government's cost of continuation to stage two is particularly large. This stage one leniency, under the assumption just stated, will tend to favor continuation at subsequent stages. In addition, the stringent test at stage one implies that the mix of cases remaining at later stages tends to involve a heavier concentration of harmful acts, which also favors subsequent continuation. Conversely, if in the first stage a large portion of cases are continued, perhaps because the costs of continuation to stage two are particularly low, then termination will be relatively more favorable at subsequent stages.

In considering these statements, however, an important distinction must be drawn concerning the basis for the stringency or generosity of treatment at various stages. The foregoing was motivated by stage-one-specific considerations, in particular, the continuation costs of moving to stage two. But other factors that may favor, say, termination in many states at stage one – e.g., a low h or high γ – will tend to favor similar, not opposing, decisions at later stages.

Even so, three important points remain true. First, as mentioned, decisions at prior stages will affect the mix of cases that remain and hence what is optimal going forward. Second, some factors, as noted, are stage specific. Third, it is of practical relevance to entertain the possibility that decisions at some stages are not made optimally because legal systems may impose institutional constraints on how cases are handled at particular points in adjudication. For example, it may be impermissible to terminate cases at some stages (perhaps because some agents are not trusted), or decisions may be made by other agents with different objectives (perhaps prosecutors, who may be the pertinent stage-one gatekeepers, are overzealous). In such circumstances, if there is flexibility at other stages, counterbalancing action may tend to be optimal. In this regard, keep in mind that expression (12) indicates when continuation at a given stage, for a given signal path, is optimal taking as given how decisions are made at other stages and on other paths. That is, the optimality condition did not assume that those given decisions were made optimally.

Interdependence of decisions across stages also applies in reverse: optimal decisions at stage t (including, e.g., stage one) depend on the conditional expectations regarding future decisions, a point reflected by the expectation operators that appear throughout expression (12).

However, the implications need not be the same as with dependence on prior-stage decisions. Notably, if at a later (including the final) stage, it were known that institutional constraints required termination in some subset of cases, it may be unwise to be more generous in allowing continuation at early stages: if the continued cases will probably be terminated later in any event, with adjudication costs being incurred along the way, the attempted counterbalance may primarily waste resources. (Proposition 3.a offers a caveat, reflecting that defendants' adjudication costs serve a deterrence function that is independent of whether liability is ultimately imposed.) On the other hand, if later stages will be too lenient due to institutionally constrained decisionmaking, then earlier decisions may best tilt toward termination.

C. Final Stage

Consider now the final stage, T . To do so, we will interpret expression (12) for $t = T$. This expression indicates when the optimal rule is to continue ($\delta_T(\sigma_T) = 1$) rather than to terminate ($\delta_T(\sigma_T) = 0$), but in this instance continuation means assigning liability, i.e., imposing the sanction s . Therefore, $E_T(p^i) = 1$, for $i = H, B$. Also, because there is no subsequent stage, $E_T^i(c) = 0$ and $E_T^i(k) = 0$, so $E_T(\kappa^i) = 0$. Together these also imply that $E_T(\lambda^i) = s$. Accordingly, for stage T , expression (12) reduces to

$$(14) \quad \rho_T^H(\sigma_T^\circ) s f^H(b^H)(h + \kappa^H - b^H) > \gamma \rho_T^B(\sigma_T^\circ) s f^B(b^B)(b^B - \kappa^B).$$

This expression simply compares the deterrence benefit to the chilling cost (there being no additional adjudication costs of continuation associated with inframarginal acts, as there were in expression (12), for $t < T$). The deterrence benefit is the rise in individuals' expected cost for harmful acts (the probability of harmful acts reaching this stage, with the pertinent signal, times the sanction) times the density of harmful acts (together giving the incremental quantity of harmful acts deterred) times the benefit per harmful act that is deterred (which is the same as before). Likewise for the chilling cost.

By contrast to the optimal rules for intermediate stages, we can state here:

Proposition 6: The optimal $\delta(\sigma_T)$ in the final stage of adjudication takes the form of a likelihood ratio test. That is, there exists a ψ^ (determined by (15)) such that liability is assigned ($\delta(\sigma_T^\circ) = 1$) if and only if $\rho_T^H(\sigma_T^\circ)/\rho_T^B(\sigma_T^\circ) > \psi^*$.*

This result can most readily be seen if we rewrite expression (14) as follows:

$$(15) \quad \frac{\rho_T^H(\sigma_T^\circ)}{\rho_T^B(\sigma_T^\circ)} > \frac{\gamma f^B(b^B)(b^B - \kappa^B)}{f^H(b^H)(h + \kappa^H - b^H)} \equiv \psi^*.$$

The left side of expression (15) gives the likelihood ratio for state σ_T° . The right side of the inequality is the critical likelihood ratio, ψ^* , in the Proposition. The key point is that ψ^* does not depend on σ_T° . The critical ratio is the chilling cost per unit increase in individuals' expected cost for benign acts over the deterrence benefit per unit increase in individuals'

expected cost for harmful acts.²⁸

It is useful to compare expressions (14) and (15) to expression (12), the latter for cases in which $t < T$. Expression (12) also could have been put in the form of a likelihood ratio test. However, the resulting critical likelihood ratio, analogous to the right side of the inequality in expression (15), would depend on σ_t° through a large number of terms, so we would not have a conventional likelihood ratio test. (Recall the discussion of Proposition 4.) In particular, note that all of the expectation operators depend on the signal distributions in subsequent periods, which in turn depend on the prior realizations, just as do the $\rho_t^i(\sigma_t^\circ)$. (See expressions (A9), (A11), (A13), and (A15).) We can also see that the determinants of the optimal decision rule for intermediate stages (including the first stage) are a good deal more complex than those for final adjudication. Reviewing the results in section 2.B, it is apparent that Propositions 2*b*, 3*a*, 3*b*, and 3*c* have no analogue for the final stage.

4. Conclusion

Most systems of formal adjudication have multiple stages. More broadly, investigators, prosecutors, and regulatory agencies have internal processes that screen cases out at various points and make interim decisions whether to expend resources to gather additional information. This article formally models multistage adjudication in order to determine how decisions are optimally made at initial and interim stages as well as at the final stage. As mentioned in the introduction, the present analysis can be understood as addressing the value of information problem in a mechanism design context. Most of the complexity is due to two factors: first, decisions at all stages influence *ex ante* behavior – both the deterrence of harmful acts and the chilling of benign acts – making the flow of cases endogenous; and second, decisions in other states and at other stages influence what decision is optimal in a given state at a given stage.

Optimal decision-making at nonfinal stages reflects a large number of factors, some pertaining to marginal effects on behavior (deterrence and chilling), some to inframarginal effects on adjudication costs (continuation raises these costs for undeterred and unchilled acts), and some to both. And some factors can have subtle and counterintuitive effects. For example, individual actors' prospective adjudication costs are a social cost that is ideally minimized but also are part of the total cost of committing acts, including harmful acts, and thus contribute to deterrence (as well as to chilling); in addition, more generous continuation, while raising total costs per case, also reduces the aggregate number of cases through deterrence and chilling. And in the extension that allows the level of the sanction to be adjusted, the possibility of mistakes that burden innocent behavior actually favors higher sanctions because, when combined with more stringent termination decisions, deterrence can be maintained at a lower chilling cost due to improved targeting on two dimensions.

Another implication for system design concerns how decisions at any given stage influence optimal decisions at other stages – both at subsequent stages, by influencing the mix of

²⁸The discussion in the text supposes that both the numerator and denominator on the right side are positive. If (only) the numerator is negative – which indicates that it is net desirable to chill benign acts because the marginal benign act has a benefit less than the expected adjudication costs it generates – then expression (14) indicates that liability would optimally be assigned in every state. If (only) the denominator is negative (which indicates overdeterrence), then, in moving from (14) to (15), the inequality would reverse, and the interpretation is that liability would not optimally be assigned in any state.

cases that remain in the system and also the levels of deterrence and chilling, and at prior stages, since optimal decisions there depend on the consequences of continuation. Furthermore, because many actual legal systems impose institutional constraints that may render decisions at some stages suboptimal, the present analysis allows one to assess how decisions at other stages should be adjusted to compensate. It is also explained how optimal final stage decisions involve a conventional likelihood ratio test whereas those at nonfinal stages do not. In addition, conventional legal wisdom and practice that seems to favor increasing stringency as one proceeds from the initial stage to the final stage is difficult to rationalize. This latter point reflects that many factors change (and in different directions) as cases proceed through the system, including that more costs are sunk, which reduces the direct cost of continuation at later stages.

The models analyzed here, although general on many dimensions, are oversimplified on others, suggesting potential extensions. The method of enforcement embedded here is akin to the posting of monitors or auditing (including inspections), whereas in some settings, notably crime, investigation (information gathering triggered by the observation of particular harmful acts) is more commonly employed.²⁹ Another important variant concerns adjudication aimed primarily at regulating future conduct rather than at influencing *ex ante* behavior (deterrence), a simpler problem mentioned briefly in note 4. In addition, for some legal systems, the structuring of the stages is itself a decision variable, so one could analyze, for example, whether consecutive stages should be combined (balancing cost reduction due to economies of scope in information collection against forgone option value) and how stages should optimally be ordered (information with a high ratio of diagnosticity to cost would optimally be gathered first). Finally, although atypical in most legal settings, it is natural to consider as well the possibility of assigning liability (as a third alternative to termination and continuation) at nonfinal stages, the analysis of which would closely mirror that presented here.³⁰

A different sort of extension would be to model the behavior of those who initiate and pursue cases. For the government, this would include police, prosecutors, and agency officials. For private litigation, the focus would be on the incentives of plaintiffs and their lawyers.³¹ These actors influence which cases enter the legal system and also incentives to make expenditures to gather information at each stage. The present analysis is complementary in this regard because it analyzes how optimally to make decisions at every stage – from the initial one to final determinations of liability – taking all other decisions and information as given, that is, without supposing that other parts of the system operate optimally.

²⁹In preliminary notes, I have analyzed enforcement by investigation. The results are more complicated in ways that parallel the corresponding extension in Kaplow (2011) on the optimal burden of proof in final adjudication, although most of the qualitative effects derived here are preserved. Many of the differences concern the fact that greater deterrence has the added effect that fewer investigations are triggered (for a given enforcement probability), which in turn reduces the number of benign acts that enter the legal system.

³⁰These extensions and others are examined informally in Kaplow (2013).

³¹For example, a literature (surveyed in Spier 2007) considers the credibility of suits – usually in models with exogenous behavior and with only one, final stage of adjudication and in which the decision rule there is taken as given. Allowing multiple stages, with termination possible at each, affects litigants' filing decisions and their willingness to continue. Relatedly, as mentioned in the introduction, the present analysis also abstracts from settlement (including plea bargaining).

Appendix A. Extensions

A.1 Additional Enforcement Instruments

Following the text in section 2.C, begin with the problem in which the government chooses not only the $\delta(\sigma)$ but also the sanction s . The effect of the sanction on deterrence and chilling can be assessed by taking the derivative of expression (3) with respect to s :

$$(A1) \quad \frac{db^i}{ds} = \int_{-\infty}^{\infty} \rho^i \delta(\sigma) p^i(\sigma) d\sigma.$$

Using expression (4), we can determine that:

$$(A2) \quad \frac{dW}{ds} = \left(h + \kappa^H - b^H \right) f^H(b^H) \frac{db^H}{ds} - \gamma \left(b^B - \kappa^B \right) f^B(b^B) \frac{db^B}{ds}.$$

Throughout, the net benefit per deterred act (the leading parenthetical expression in the first term on the right side of (A2)) is taken to be positive. As mentioned, it is obvious if the net welfare impact per chilled act (the leading parenthetical expression in the second term) is also favorable (i.e., $b^B < \kappa^B$), this derivative must be positive, implying that the optimal sanction is maximal if it holds throughout. Hence, the main case of interest in attempting to establish Proposition 5 is that involving an imagined optimum with a nonmaximal sanction, with the social impact of chilling the marginal act being negative.

The construction suggested in the text can be presented more precisely as follows. To begin, let the set Σ denote all σ such that, at the hypothesized optimal s , we have $\delta(\sigma) = 1$. (Note that it is not necessary that the $\delta(\sigma)$ have been set optimally.) In the first step of the experiment, raise s slightly and, for each $\sigma \in \Sigma$, reduce $\delta(\sigma)$ slightly such that the contribution to deterrence from the state is constant. That is, as one raises s , the $\delta(\sigma)$ change as follows:

$$(A3) \quad \frac{d\delta(\sigma)}{ds} = - \frac{p^H(\sigma)}{c(\sigma) + p^H(\sigma)s}.$$

The derivation of (A3) is straightforward from expression (3): the numerator on the right side indicates how much, conditional on a case entering the legal system and being in state σ , the expected cost of a harmful act rises per unit increase in s , and the denominator indicates the change in the expected cost of committing a harmful act per unit change in $\delta(\sigma)$. (Note that, unlike in section 2 of the text, here the notational substitution using $\lambda^H(\sigma)$ for this denominator – and similar substitutions below – are omitted because the decomposition is explicitly relevant to the argument.) Because this relationship holds for all $\sigma \in \Sigma$, it is obvious that, when we integrate over these states, the value of b^H given by expression (3) remains constant.

Next, to determine how chilling, the level of b^B , changes, take db^B/ds , using expression (3), where the $\delta(\sigma)$ adjust according to expression (A3). For each state σ (i.e., conditional on a case entering the legal system and being in that state), the effect on b^B is given by:

$$(A4) \quad p^B(\sigma) - \left(\frac{p^H(\sigma)}{c(\sigma) + p^H(\sigma)s} \right) (c(\sigma) + p^B(\sigma)s) =$$

$$p^H(\sigma) \left(\frac{p^B(\sigma)s}{p^H(\sigma)s} - \frac{c(\sigma) + p^B(\sigma)s}{c(\sigma) + p^H(\sigma)s} \right) \leq 0,$$

with strict inequality for any state in which $p^H(\sigma) > p^B(\sigma)$. Intuitively, when we increase s and terminate rather than continue just often enough to keep the contribution to deterrence in each state constant, chilling is reduced to the extent that there is any diagnosticity in the application of the explicit sanction s . The reason is that defendants' continuation cost, $c(\sigma)$, influences deterrence and chilling but lacks any diagnosticity (conditional on being in the state σ); hence, relying less on it and more on s improves targeting. Finally, what is true in each state is true in aggregate, so if deterrence is held constant by this experiment, chilling must fall. (If $p^H(\sigma) = p^B(\sigma)$ for all $\sigma \in \Sigma$, chilling remains constant, which is sufficient in establishing the Proposition.)

In the second step of the construction, we will now raise these $\delta(\sigma)$ back to 1.0, and keep deterrence constant by terminating (with certainty) in those states that have the worst overall targeting. Specifically, as suggested in the main text, we can order all the $\sigma \in \Sigma$ from highest (worst) to lowest (best) in terms of the ratio:

$$(A5) \quad r(\sigma) = \frac{(c(\sigma) + p^B(\sigma)s)z^B(\sigma)}{(c(\sigma) + p^H(\sigma)s)z^H(\sigma)}.$$

Now, having first restored all the $\delta(\sigma)$ back to 1.0, we can remove states from Σ , starting from those with the highest $r(\sigma)$, until we have reached the point at which deterrence is back to its initial level. It should be clear that this step reduces chilling (or keeps it constant if the $r(\sigma)$ are uniform across all σ). The reason is simply that $r(\sigma)$ indicates the contribution to chilling – to the level of b^B – per unit contribution to deterrence – the level of b^H – for changes in the level of $\delta(\sigma)$. See expression (3). Because (combining the components of this second step) we raise $\delta(\sigma)$ (to 1) in all $\sigma \in \Sigma$ with $r(\sigma) \leq r^*(\sigma)$, for some critical value $r^*(\sigma)$, and reduce $\delta(\sigma)$ (to 0) in all $\sigma \in \Sigma$ with $r(\sigma) > r^*(\sigma)$, it follows that chilling falls further. (If all the $r(\sigma)$ are equal, this step has no effect on chilling, which is sufficient in establishing the Proposition.)

Finally, observe that, after this construction is fully implemented, we have terminations in some states in Σ (and no continuations in any states not in Σ). Hence, on this account, continuation costs fall. Taken together, we have deterrence constant, a (weak) decline in chilling (which was assumed to be costly in this part of the proof), and a strict decline in continuation costs. Hence, welfare must be higher. Therefore, a nonmaximal sanction s cannot be optimal. ■

Again paralleling section 2.C, we now consider enforcement effort, where we have $\pi^i(e)$ such that $\pi^{i'}(e) > 0$, $\pi^{i''}(e) \leq 0$, for $i = H, B$. Using expressions (3) and (4), modified accordingly, the first-order condition for (an intermediate) e is given by:

$$(A6) \frac{dW}{de} = (h + \kappa^H - b^H) f^H(b^H) b^H \frac{\pi^{H'}(e)}{\pi^H(e)} - \gamma (b^B - \kappa^B) f^B(b^B) b^B \frac{\pi^{B'}(e)}{\pi^B(e)}$$

$$- (1 - F^H(b^H)) \frac{\pi^{H'}(e)}{\pi^H(e)} \kappa^H - \gamma (1 - F^B(b^B)) \frac{\pi^{B'}(e)}{\pi^B(e)} \kappa^B - 1 = 0.$$

The discussion in section 2.C of this first-order condition and what immediately follows (which considers raising e and switching to termination in some states) can be related to expression (A6) in a straightforward manner.

Finally, consider the classic Becker (1968) experiment of raising s and lowering e so as to keep deterrence (b^H) constant. Using expression (3), the requisite adjustment in e is as follows:

$$(A7) \frac{de}{ds} = - \frac{\int_{-\infty}^{\infty} \pi^H(e) \delta(\sigma) p^H(\sigma) z^H(\sigma) d\sigma}{\int_{-\infty}^{\infty} \pi^{H'}(e) \delta(\sigma) (c(\sigma) + p^H(\sigma) s) z^H(\sigma) d\sigma}.$$

The numerator indicates the rise in deterrence on account of a unit increase in s , and the denominator on account of a unit increase in e . As we can see, raising e , because it increases the flow of cases into the legal system, raises the expected costs of those who commit harmful acts due to both the increase in the expected (explicit) sanction and the increase in defendants' expected adjudication costs. We have already seen that the latter is, in an important sense, less well targeted, so it might seem that we could prove that chilling must fall.

To determine the effect on chilling (b^B), we can also use expression (3), taking the derivative with respect to s , wherein e changes according to expression (A7). The sign of the net effect on chilling is given by the sign of:

$$(A8) \frac{\int_{-\infty}^{\infty} \delta(\sigma) p^B(\sigma) z^B(\sigma) d\sigma}{\int_{-\infty}^{\infty} \delta(\sigma) p^H(\sigma) z^H(\sigma) d\sigma} - \frac{\pi^{B'}(e) / \pi^B(e)}{\pi^{H'}(e) / \pi^H(e)} \frac{\int_{-\infty}^{\infty} \delta(\sigma) (c(\sigma) + p^B(\sigma) s) z^B(\sigma) d\sigma}{\int_{-\infty}^{\infty} \delta(\sigma) (c(\sigma) + p^H(\sigma) s) z^H(\sigma) d\sigma}.$$

(For convenience, the $\pi^i(e)$ elements in the numerator and denominator of the first term were divided out, so they could be included with the $\pi^{i'}(e)$ elements in the second term.) The first term (subject to the stated adjustment) gives the ratio of chilling to deterrence per unit increase in s and the second term (subject to the stated adjustment) gives the ratio for the calibrated change in e . If the latter is larger, chilling falls.

The point in section 2.C about diminishing returns in terms of targeting precision as one increases e , for some enforcement technologies, is apparent from the leading component of the second term. It indicates the relative rise in the probability of benign acts entering the legal

system, compared to that in the probability of harmful acts entering the legal system. If this ratio is constant, let us say, we are left, in terms of chilling, with the fact that defendants' continuation costs play a greater role when using e rather than s to achieve a given level of deterrence.

In the earlier analysis of raising s and terminating in the weakest states so as to keep deterrence constant, this differential was favorable. While there may well be such a tendency as a practical matter here, however, this is not necessarily true. The reason is that, when we raise s and reduce e , the latter reduces the extent to which defendants bear adjudication costs and (conditional) expected (explicit) sanctions in all $\sigma \in \Sigma$, not just in the worst states in terms of targeting. Recall, moreover, how "worst" is defined in the present context, as indicated by the ratio $r(\sigma)$ given by expression (A5). We have not only $[c(\sigma)+p^B(\sigma)s] / [c(\sigma)+p^H(\sigma)s]$ – which was at the core of the argument based on the difference in expression (A4) – but also $z^B(\sigma)/z^H(\sigma)$, which can confound the preference for relying more on s and less on the $c(\sigma)$.

To see this point, assume that the $p^i(\sigma)$ are barely diagnostic, so that the targeting advantage of greater reliance on s rather than the $c(\sigma)$ is slight. Furthermore, suppose that in some states with substantial mass, we have very low $z^B(\sigma)/z^H(\sigma)$ and very large $c(\sigma)$. In that event, the $c(\sigma)$ in those states are very well targeted, more so than is s *on average*. Now, when s is raised, e is reduced, and perhaps substantially because the $p^i(\sigma)$ that are not very diagnostic may nevertheless be large in states with high $z^B(\sigma)/z^H(\sigma)$. This reduction in e , in turn, renders less important the powerful targeting from the states with very low $z^B(\sigma)/z^H(\sigma)$ but very large $c(\sigma)$. Taken together, chilling could rise. And if it rose enough to outweigh the other advantages of the classic Becker experiment, the net welfare impact could be negative.

One could advance an inelegant, although perhaps practically plausible, sufficient conditions to rule this possibility out. (Note, for example, that if $z^B(\sigma)/z^H(\sigma)$ is very low, it seems unlikely that, in a well-operating system, the $p^i(\sigma)$ would fail to be highly diagnostic in such states.) In any event, since Proposition 5's claim that the optimal sanction is maximal can be established through the earlier construction, further exploration is of limited interest.

A2 Multistage Model

Begin with prospective defendants' expected adjudication costs. For $t \geq 1$, recall that $E_t^i(c)(\sigma_t)$ is the expected sum of the defendant's adjudication costs borne in all future stages, conditioning on the fact that the case has reached stage t with the signal path σ_t , and the decision is to continue. In stage T , such expected costs are 0 because continuation in stage T means assignment of liability and application of the sanction, with no further adjudication. In stage $T-1$, continuation means that the cost $c_{T-1}(\sigma_{T-1}^\circ)$ is incurred, where the history of the signals to (and including) that stage is σ_{T-1}° . For all prior stages, generically t , the expected cost if there is continuation is the sum of (A) the (certain) cost of moving from stage t to stage $t+1$, which is $c_t(\sigma_t^\circ)$, and (B) the expected value of all subsequent costs. The latter, for each possible realization of the signal θ_{t+1} , either is zero if the case is terminated in stage $t+1$ or, otherwise, it is the pertinent probability of being in that situation times the expected continuation costs apropos for stage $t+1$ (in that case, in moving to stage $t+2$). Accordingly, for $1 \leq t < T$, we can write

$$(A9) \quad E_t^i(c)(\sigma_t^\circ) = c_t(\sigma_t^\circ) + \int_{-\infty}^{\infty} \delta_{t+1}(\sigma_{t+1}) E_{t+1}^i(c)(\sigma_{t+1}) z_{t+1}^i(\sigma_{t+1}) d\theta_{t+1}.$$

(When $t = T-1$, the right side of expression (A9) just equals $c_{T-1}(\sigma_{T-1}^\circ)$ since $E_T^i(c)(\sigma_T) = 0$, consistent with the prior explanation.) Note that the variable of integration is θ_{t+1} : although this variable may not be apparent in the integrand, recall that σ_{t+1} is a vector, the last element of which is θ_{t+1} . (Regarding the conditional density $z_{t+1}^i(\sigma_{t+1})$, see note 25.)

For use in expression (8), we need a parallel formulation for what is being called stage 0, when individuals contemplate whether to act, which is:

$$(A10) \quad E_0^i(c) = \pi^i \left(c_0 + \int_{-\infty}^{\infty} \delta_1(\sigma_1) E_1^i(c)(\sigma_1) z_1^i(\sigma_1) d\theta_1 \right).$$

Here, we add a preceding factor π^i , the probability that an individual who commits an act of a given type will enter the legal system in the first place. Also note that it is allowed that there be a cost, denoted c_0 , of initially entering the system. (For convenience, this latter possibility was omitted in section 2.)

The corresponding expressions for the government's cost k are analytically the same:

$$(A11) \quad E_t^i(k)(\sigma_t^\circ) = k_t(\sigma_t^\circ) + \int_{-\infty}^{\infty} \delta_{t+1}(\sigma_{t+1}) E_{t+1}^i(k)(\sigma_{t+1}) z_{t+1}^i(\sigma_{t+1}) d\theta_{t+1}.$$

$$(A12) \quad E_0^i(k) = \pi^i \left(k_0 + \int_{-\infty}^{\infty} \delta_1(\sigma_1) E_1^i(k)(\sigma_1) z_1^i(\sigma_1) d\theta_1 \right).$$

We can now use the same technique to derive the expected probability of liability. For $t \geq 1$, $E_t(p^i)(\sigma_t)$ is the conditional expected probability of liability. In stage T , this probability is 1 if there is continuation and 0 otherwise. In earlier stages, continuation carries no immediate consequence. The only effect is that, once in the next stage, there is the possibility of further continuation, ultimately reaching stage T . For $1 \leq t < T$, we can write

$$(A13) \quad E_t(p^i)(\sigma_t^\circ) = \int_{-\infty}^{\infty} \delta_{t+1}(\sigma_{t+1}) E_{t+1}(p^i)(\sigma_{t+1}) z_{t+1}^i(\sigma_{t+1}) d\theta_{t+1}.$$

(As just indicated, when $t = T$, we have $E_T(p^i)(\sigma_T) = 1$, with the interpretation that, when the decision in period T is to continue, this means that liability is assigned with certainty.)

Next, as with adjudication costs, we can state the appropriate expression for what is referred to as stage 0, when individuals contemplate whether to act:

$$(A14) \quad E_0(p^i) = \pi^i \int_{-\infty}^{\infty} \delta_1(\sigma_1) E_1(p^i)(\sigma_1) z_1^i(\sigma_1) d\theta_1.$$

Once again, we have the preceding factor π^i , the probability that an individual who commits an act of a given type will initially enter the legal system.

It is also useful to introduce one last bit of notation. Because we will contemplate decision choices at various stages after stage 1, it is convenient to have an expression for the probability that a given stage t (including final stage T) is entered along the particular path of signals embodied in σ_t° . (Note that we are interested in this probability for the particular path and not aggregated along all paths because the stage t continuation/termination decision in a given state, $\delta(\sigma_t^\circ)$, is made with full knowledge of the prior signals and, in general, will differ

depending on the realizations of those signals.) For entering the first stage, these probabilities, the π^i , were already postulated. More generally, define

$$(A15) \quad \pi_t^i(\sigma_t^\circ) = \pi^i \prod_{\tau=1}^{t-1} \delta_\tau(\sigma_\tau^\circ) z_\tau^i(\sigma_\tau^\circ).$$

(And, consistent with the previous statement, when $t = 1$, the expression involving the product operator on the right side of (A15) should be taken to equal 1.) Expression (A15) states that the probability of entering stage t along a particular signal path is the probability of entering the first stage, multiplied by the product of the indicator variable and of the density of the signal at each subsequent stage up to (but not including) stage t .³² Note that if any decision before stage t involves termination, the probability of entering stage t along that path is zero (making the decision at that stage or any subsequent stage moot).

Finally, analogous to the two-stage model, the condition at stage t in state σ_t° for continuation ($\delta_t(\sigma_t^\circ) = 1$) to be strictly optimal is:

$$(A16) \quad \rho_t^H(\sigma_t^\circ) E_t(\lambda^H)(\sigma_t^\circ) f^H(b^H)(h + \kappa^H - b^H) > \gamma \rho_t^B(\sigma_t^\circ) E_t(\lambda^B)(\sigma_t^\circ) f^B(b^B)(b^B - \kappa^B) \\ + (1 - F^H(b^H)) \rho_t^H(\sigma_t^\circ) E_t(\kappa^H)(\sigma_t^\circ) + \gamma (1 - F^B(b^B)) \rho_t^B(\sigma_t^\circ) E_t(\kappa^B)(\sigma_t^\circ).$$

³²Because what happens in stage t is excluded, the probability in expression (A15) does not actually depend on the signal in stage t ; however, it is appealing to use the notation for the signal vector through stage t since this is the pertinent vector for all the other variables in the optimality condition.

References

- Abrantes-Metz, Rosa M., Luke M. Froeb, John F. Geweke, and Christopher T. Taylor. 2006. "A Variance Screen for Collusion," *International Journal of Industrial Organization* 24: 467–486.
- 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.
- Harrington, Joseph E., Jr. 2007. "Behavioral Screening and the Detection of Cartels," in Claus-Dieter Ehlermann & Isabela Atanasiu, *European Competition Law Annual: 2006 – Enforcement of Prohibition of Cartels*: 51–67, Oxford: Hart Publishing.
- Hay, Bruce L., and Kathryn E. Spier. 1997. "Burdens of Proof in Civil Litigation: An Economic Perspective," *Journal of Legal Studies* 26: 413–431.
- Kaplow, Louis. 2011. "On the Optimal Burden of Proof," *Journal of Political Economy* 119: 1104–1140.
- _____. 2013. "Multistage Adjudication," *Harvard Law Review* 126:1179–1298.
- Karlin, Samuel, and Herman Rubin. 1956. "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *Annals of Mathematical Statistics* 27: 272–299.
- Lando, Henrik. 2002. "When is the Preponderance of Evidence Standard Optimal?" *Geneva Papers on Risk and Insurance – Issues and Practice* 27: 602–608.
- Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics* 12: 380–391.
- Mookherjee, Dilip, and Ivan P.L. Png. 1992. "Monitoring Vis-à-Vis Investigation in Enforcement of Law," *American Economic Review* 82: 556–565.
- 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.
- Polinsky, A. Mitchell, and Steven Shavell. 2007. "The Theory of Public Enforcement of Law," in A. Mitchell Polinsky and Steven Shavell, *Handbook of Law and Economics*, vol. 1: 403–454, Amsterdam: North-Holland.
- Rubinfeld, Daniel L., and David E.M. Sappington. 1987. "Efficient Rewards and Standards of Proof in Judicial Proceedings," *Rand Journal of Economics* 18: 308–315.
- Shavell, Steven. 1991. "Specific Versus General Enforcement of Law," *Journal of Political Economy* 99:1088–1108.
- Spier, Kathryn E. 2007. "Litigation," in A. Mitchell Polinsky and Steven Shavell, *Handbook of Law and Economics*, vol. 1: 259–342, Amsterdam: North-Holland.