ISSN 1045-6333

ACCURACY IN THE DETERMINATION OF LIABILITY

Louis Kaplow

and

Steven Shavell

Discussion Paper No. 117

10/92

Harvard Law School
Cambridge, MA 02138

The Program in Law and Economics is supported by a grant from the John M. Olin Foundation.

Accuracy in the Determination of Liability

Louis Kaplow and Steven Shavell*

Abstract

Many legal rules, notably rules of procedure and evidence, are concerned with achieving accuracy in the outcome of adjudication. In this article, we study accuracy in the conventional model of law enforcement. We consider why reducing error in determining liability is socially valuable and how error and its reduction affect the optimal probability and magnitude of sanctions.

*Both of Harvard University and the National Bureau of Economic Research. We are grateful to Marcel Kahan, A. Mitchell Polinsky, and participants in seminars at Harvard and New York University Law Schools for comments and to the Harvard Law School Program in Law and Economics, which is supported by a grant from the John M. Olin Foundation, for financial support.

1. Introduction

The degree of accuracy is a central concern of adjudication. Procedural rules in the civil, criminal, and administrative contexts, rules of evidence, and other features of the legal system are motivated to a substantial extent by concerns for achieving accurate outcomes, although it is recognized that greater accuracy usually comes at a higher cost. Similar considerations are reflected in features of alternative forms of dispute resolution, often selected or designed by contract.[1]

In this article, we examine accuracy and error in the standard model of law enforcement.[2] Specifically, we consider the possibility that individuals may be mistakenly found liable for acts they did not commit (false positives) and that they may be exonerated when they did in fact commit the acts in question (false negatives).[3] We study the social value of reducing these errors -- that is, of greater accuracy. In addition, we investigate how

---

[1] The issues we address can be seen as part of the principal-agent problem, as the accuracy with which information is observed may be an important aspect of an incentive scheme. See Grossman and Hart (1983), Holmström (1979), and Shavell (1979).

[2] In Kaplow and Shavell (1992b), we analyze mistakes in determining damages. The analysis there focuses on how greater accuracy ex post affects individuals behavior and incentives to become informed ex ante, and on whether private parties' incentives to present information in adjudication are socially optimal. There is a growing literature addressing aspects of legal error, although it does not emphasize the questions analyzed here: the optimal degree of accuracy and how accuracy affects the optimal probability and magnitude of sanctions. See Kaplow (1991) (optimal complexity of legal rules), Kaplow and Shavell (1992a) (predictable error and ex ante legal advice), Polinsky and Shavell (1989) (errors affect incentive to sue and thus incentive to obey the law), Posner (1973) (how injurers' behavior is affected when not all are held liable), and Rubinfeld and Sappington (1987) (how burden of proof affects defendants' litigation expenditures). Some of the present article's concerns are examined in Png (1986) (discussed in note 21) and Ehrlich (1982). Error in determining due care under a negligence rule is addressed in Craswell and Calfee (1986).

[3] Mistakes in determining liability may involve difficulties in determining causation or other issues, in addition to problems of identifying who committed an act. For example, errors may be made in deciding whether a firm's toxic substance or some other factor caused an individual's illness, or whether a seller's inability to meet contractual obligations was caused by factors deemed to excuse performance.

familiar results concerning the optimal probability and magnitude of sanctions are affected when accuracy is a problem.

In section 2, we present a model in which risk-neutral actors decide whether to commit harmful acts. Sanctions may be either costless (monetary) or costly (nonmonetary). We assume that greater enforcement effort increases the total number of individuals who are sanctioned, while greater expenditure on accuracy increases the number of truly guilty among them and decreases the number of innocent.

Our first result is that the optimal sanction is the maximum feasible sanction, even though error is possible and sanctions may be costly. The explanation is that of Becker (1968): if the sanction is not maximal, enforcement costs can be saved by raising the sanction and reducing enforcement effort.

Second, with regard to the appropriate level of investment in accuracy, we emphasize that accuracy and enforcement effort are alternative ways of increasing deterrence. A higher level of enforcement results in a higher probability of sanctions and thus increases deterrence. But so does a higher level of accuracy, as it raises the expected sanction for those who commit harmful acts (by reducing false negatives) and decreases the expected sanction for those who do not (by reducing false positives). Since accuracy and enforcement effort are substitute means of increasing deterrence, it is optimal to invest resources in them in a manner that reflects their relative effectiveness. If, for example, it is expensive to increase accuracy (suppose it is difficult to determine confidently whether accidents were due to poor maintenance), it will be efficient to raise enforcement effort instead (to investigate a higher fraction of accidents).

Third, we show that, for any given level of deterrence, the optimal level of accuracy is higher and the optimal level of enforcement is lower when sanctions are socially costly than when they are costless. The reason is that, when accuracy is raised and enforcement effort reduced, fewer people (both innocent and guilty) are sanctioned in achieving deterrence.[4] This reduction is advantageous to the extent sanctions are socially costly.

In section 3, we consider the case in which individuals are risk-averse and sanctions are monetary. The main difference in result from section 2 is that optimal sanctions may be less than maximal, a generalization of the conclusion of Polinsky and Shavell (1979). But this conclusion does not depend on the presence of inaccuracy, and we show that optimal sanctions may either rise or fall as inaccuracy increases.

We conclude in section 4 by offering some extensions and remarks on the analysis.

2. Analysis

2.1. The Model

Risk-neutral individuals decide whether to commit an act. If an individual commits the act, he causes an external harm h and also obtains a benefit b, where individuals' benefits are distributed according to $f(\cdot)$ on $[0,\infty)$, with a cumulative distribution function $F(\cdot)$.

The sanction for individuals identified as having committed the act is s, where $s \leq \bar{s}$.[5] Sanctions that are imposed involve a social cost of σs, where $\sigma \geq 0$.[6] Enforcement effort is p, which may be interpreted as an audit rate, a level of monitoring, or an intensity of investigation of particular harmful acts reported to authorities. Individuals who are detected may or may not bear a sanction. Specifically, individuals who have committed the act --

[4] That expected sanctions fall for the innocent is obvious when accuracy is raised and enforcement effort is reduced. And, since expected sanctions for the innocent fall, it must be that expected sanctions for the guilty fall as well, for deterrence will be the same if and only if the difference between the expected sanction for the guilty and the expected sanction for the innocent is unchanged.

[5] The upper limit $\bar{s}$ may be interpreted as the wealth of individuals if sanctions are monetary or as life imprisonment if sanctions are nonmonetary.

[6] When the sanction is a fine, it is customary to assume that $\sigma = 0$, since transfers have no social cost (when individuals are risk-neutral). For nonmonetary sanctions (or monetary sanctions with collection costs), $\sigma > 0$.

referred to as the "guilty" for convenience -- and who are detected erroneously escape sanctions with probability $q_0(k)$; individuals who have not committed the act -- the "innocent" -- and who are detected erroneously bear sanctions with probability $q_1(k)$. The variable k is the effort devoted to enhancing accuracy, where $q_i'(k) < 0$ and $q_i''(k) > 0$, for $i = 0,1$. Therefore, guilty individuals bear sanctions with probability $p(1-q_0(k))$ and innocent individuals bear sanctions with probability $pq_1(k)$.[7]

Enforcement costs take the form $c(p,k)$, where $c_p > 0$ and $c_k > 0$.[8] We also assume that $c_{pk} > 0$, which means that it is more costly to increase accuracy when the enforcement level is higher. For example, increasing the accuracy of audits raises total costs by a greater amount when more audits are being conducted.

2.2. Individual Behavior and Social Welfare

Individuals who commit the harmful act obtain an expected net benefit of $b - p(1-q_0)s$ and those who do not commit the act bear an expected sanction of pq_1s. Thus, an individual will commit the act if and only if

$$(1) \quad b \geq (1 - q_0)ps - q_1ps = (1 - q)ps = b^*,$$

where $q = q_0 + q_1$. Note that the threshold b^* is determined by the gap between the expected sanction for committing the act, $(1-q_0)ps$, and the expected sanction for not doing so, q_1ps.

---

[7] The interpretation for enforcement by monitoring or investigation is less straightforward than for enforcement by random audit, as detection and identification as guilty may be a single act, or an act based on the same information. Related, some enforcement actions inevitably affect p and k simultaneously: for example, better detectives may catch more individuals in total, with a higher fraction of those apprehended being the truly guilty.

[8] In the case of random audits, $c(p,k)$ may simply equal pk: p is the number of audits and k is the cost per audit (more expensive audits are more accurate audits). For monitoring, the same expression might be appropriate, or, alternatively, $\gamma(p)k$, where $\gamma' > 0$. For enforcement by investigation of particular harmful acts, one might modify the cost function to reflect the fact that the number of investigations, and thus total costs, will depend on the number of acts committed. We omit this from our formulation for simplicity; most of our analysis holds deterrence constant, so this consideration would have no effect.

The enforcement authority chooses p, s, and k to maximize social welfare, defined as the benefits individuals obtain from committing the act, less the harm done, sanction costs, and enforcement costs:

$$(2)\quad W = -\int_0^{b*} pq_1(k)\sigma sf(b)db + \int_{b*}^{\infty} [b - h - p(1-q_0(k))\sigma s]f(b)db - c(p,k).$$

The first term is the expected social cost from sanctions imposed on the innocent. The second term is the effect on welfare associated with the guilty: each obtains a benefit, causes harm, and generates an expected social cost from imposition of sanctions.

2.3. The Optimal Sanction

It is straightforward to demonstrate the following result.

Proposition 1. The optimal sanction is $\bar{s}$, the maximum feasible sanction. The reasoning from Becker (1968) applies in this model. For any s less than $\bar{s}$, one can raise s and reduce p so as to keep ps and thus b* unchanged. This modification in p and s does not affect the first two terms in (2), since $b* = (1-q)ps$, but, by reducing p, reduces the enforcement cost c(p,k) and thus increases welfare.[9]

Remark: The fear of imposing sanctions on the innocent not only fails to alter the Becker argument, it reinforces his argument when one takes into account that the level of accuracy is endogenous. If s is reduced and p increased (in a manner that keeps ps constant), the optimal k falls. To demonstrate this, recall that the first two terms of (2) are unaffected when ps is constant. The only effect would be on enforcement costs, c. Because $c_{pk} > 0$, the marginal cost of accuracy is higher when p is higher, so it would

---

[9] In some contexts, p is effectively fixed, as when an enforcement technique generally applies to many acts. In a standard model where perfect accuracy (given detection) is costless, the optimal sanction is less than $\bar{s}$ unless the harm is such that $h \geq p\bar{s}$. See Mookherjee and Png (1992), Shavell (1991). In this model, however, an extreme sanction might be optimal even when p is fixed and the harm is low. For any $s \leq \bar{s}$, one can raise s and reduce k so that $(1-q(k))ps$ and thus behavior (b*) remains unchanged. This modification reduces the enforcement cost and will be optimal unless sanction costs rise by an amount sufficiently large to offset this savings.

be optimal to reduce k. (For example, when there are many audits, being careful in each one is, in total, more expensive.) Thus, when greater accuracy is costly, an independent desire to avoid mistakes represents an additional reason for an enforcement authority to employ a high sanction, low probability enforcement strategy.

2.4. Optimal Enforcement Effort and Accuracy

To characterize the solution to the problem of choosing p and k optimally, we find it useful begin by determining the condition for the optimal p and k for a given level of deterrence associated with a given b*. This, of course, is a necessary condition for maximizing social welfare. Then, we will discuss the optimal level of deterrence, b*.

To derive the condition for the optimal p and k given b*, we differentiate social welfare (2) with respect to p, where k is implicitly determined as a function of p by the constraint $(1-q(k))ps = b^*$. This constraint, which keeps deterrence unchanged, implies that $k'(p) = (1-q)/pq'(k)$. Because the limits of integration in (2) do not change, the derivative is simply[10]

$$(3)\quad \frac{dW}{dp}\Big|_{k=k(p)} = -\int_0^{b^*} (pq_1'k' + q_1)\sigma sf(b)db - \int_{b^*}^{\infty} (-pq_0'k' + (1-q_o))\sigma sf(b)db$$

$$-(c_p + c_k k').$$

The first two terms are the inframarginal changes in sanction costs; changes in costs are positive, decreasing welfare. To explain, an increase in p must be accompanied by a reduction in k, since b* is held constant. Both the increase in p and the reduction in k cause expected sanctions borne by the innocent to increase. (Their expected sanction is pq_1s; both p and q_1 increase.) Moreover, this increase in the expected sanction for the innocent

[10] The assumption that the q_i'' are positive -- i.e., that effort to increasing accuracy is subject to diminishing returns -- implies that the terms of the second derivative of W corresponding to the first two terms of (3) are negative (or zero if $\sigma = 0$). Thus, a sufficient condition for the second-order condition to hold is that the technology be convex in p and k. While this holds for an audit or monitoring technology in which $c(p,k) = pk$, it need not hold generally.

implies that expected sanctions for the guilty must rise as well if deterrence is to remain constant. (Deterrence, b^*, is determined by the gap between the expected sanctions for the guilty and innocent, which is constrained to remain constant.) Hence, expected sanctions and thus sanction costs rise for both the innocent and guilty when p rises. The third term is the change in enforcement costs: greater enforcement effort increases costs, but the reduction in accuracy decreases costs.

We now state two results.

Proposition 2. Assume that sanctions are costless ($\sigma = 0$). Then, for any given level of deterrence, the optimal level of enforcement (p) and accuracy (k) are those that minimize enforcement costs.
To demonstrate this, observe that when $\sigma = 0$, (3) reduces to the third term, which is the derivative of enforcement costs, $c(p,k)$. Thus, the first-order condition for maximizing W (3) is the condition for minimizing c.[11]

Remark: This result captures the point that enforcement effort (p) and accuracy (k) are to be regarded as substitutes in achieving a given level of deterrence. For example, if it is expensive to increase accuracy further, the most efficient way to increase deterrence would involve increasing p.

Proposition 3. Assume that sanctions are costly ($\sigma > 0$). Then, for any given level of deterrence, the optimal level of accuracy (k) is higher and the optimal level of enforcement (p) is lower than if sanctions are costless.
This follows because, when $\sigma > 0$, the first two terms in (3) are negative, so that dW/dp is less than it is when $\sigma = 0$, and this in turn implies that the optimal p must be lower (and thus k higher) when $\sigma > 0$ than when $\sigma = 0$.[12] The explanation for this result is that the substitution of k for p in achieving a

---

[11] The result can be seen directly from the objective function. When $\sigma = 0$, (2) simplifies to

$$\int_{b^*}^{\infty} (b - h)f(b)db - c(p,k).$$

When b^* is held constant, only the enforcement cost is affected by changing p and k.

given b* reduces expected sanctions borne by the innocent and guilty. When sanctions are socially costly, this is an advantage. Thus, with costly sanctions, accuracy and enforcement effort are substitutes with regard to achieving deterrence, but not with regard to minimizing sanction costs.

We do not present derivations for the optimal level of deterrence, b*, in the presence of inaccuracy. Once the social problem has been reduced to selecting b*, it differs little from the problem of determining the optimal b* in the enforcement model without inaccuracy. For instance, when $\sigma = 0$, in the usual model without inaccuracy, raising b* increases deterrence but also raises enforcement costs, so the optimal b* reflects this trade-off (with the result that the optimal b* is less than h). When inaccuracy is present, the trade-off is qualitatively similar.[13]

3. Extension: Risk-Aversion

In this section, we consider briefly the case in which individuals are risk-averse and sanctions are monetary. (The model is presented in the appendix, as the analysis is tedious.) It is apparent that proposition 1, which states that the optimal sanction is maximal, need not hold when individuals are risk-averse.[14] This follows essentially by the logic in

[12] A function that has a derivative everywhere lower than that of another function must reach its maximum at a lower value of its argument than the other function.

[13] With nonmonetary sanctions in a model without inaccuracy, increased deterrence has two marginal effects (improved behavior, if underdeterrence is involved, and reducing the number of individuals subject to sanctions applicable to the guilty), an inframarginal sanction cost effect with regard to the rest of the population, and an enforcement cost. See Kaplow (1990), Polinsky and Shavell (1984). The only qualitative difference with inaccuracy is that, to the extent that deterrence is increased by raising k, the inframarginal effect is that greater expected sanctions are borne by the guilty and lesser by the innocent, so aggregate sanction costs may be higher or lower on this account.

[14] The remark to proposition 1 -- that if one reduces s and increases p such that deterrence is unchanged, a lower level of accuracy will be optimal -- does, however, still apply, as it depends on the form of the cost function (the assumption that $c_{pk} > 0$) rather than the form of individuals' utility functions. When individuals are risk-averse, another factor further reduces the optimal level of accuracy when p is raised and s reduced: lowering s and raising p reduces risk-bearing costs, while one of the benefits of using a higher k rather than a higher p to achieve a given level of deterrence is that

Polinsky and Shavell (1979),[15] that imposing maximal sanctions involves risk-bearing costs that may be worth reducing by lowering sanctions and raising enforcement effort.

The reason that risk aversion may result in less than maximal monetary sanctions while optimal nonmonetary sanctions are maximal is as follows. The costs of nonmonetary sanctions are assumed to be linear. Thus, a reduction in s and increase in p that kept behavior unchanged did not change total sanction costs. But risk-bearing costs are not linear; rather, they are increasing in s.

It is also apparent that, when sanction costs arise from risk aversion, p and k are substitutes in achieving deterrence, but not in minimizing enforcement costs, as indicated by propositions 2 and 3. Beginning at the p and k that minimize enforcement costs for a given level of deterrence, a slight increase in k and decrease in p that held b* constant would have no first-order effect on enforcement costs, but would reduce the number of innocent and guilty individuals who bear the costly sanction, which would tend to be advantageous. There is, however, the qualification that the resulting distribution of wealth would change, which could affect the desirability of improving accuracy.[16]

Finally, we ask whether inaccuracy is an independent reason to reduce sanctions and increase enforcement effort when individuals are risk-averse.

---

sanctions are imposed less often. Lowering s and raising p to keep deterrence constant also affects the marginal utility of wealth, which in turn may have subtle effects on the optimal level of accuracy.

[15] For further exploration of this case in a model closer to the one here, see Kaplow (1992).

[16] When fewer individuals are sanctioned, fine revenues are lower, so the lump-sum tax (see (A3) in the appendix) must be higher. If the tax increases could be confined to the groups (those who act or those who do not act) that previously paid the fines, welfare would unambiguously increase, because the resulting redistribution would simply be one that involved a reduction in risk-bearing costs. But this need not be the case. Assume, for example, that most fines are paid by a small fraction of the population that acts and receives large benefits from acting -- and thus has a low marginal utility of wealth. Because the lump-sum tax is the same for all individuals, much of the fine revenue is distributed to individuals who do not act and thus have a high marginal utility of wealth. In this instance, reducing the incidence of risky fines on whose who act could be undesirable because the fines serve as redistributive taxes.

(We have already noted that risk aversion even in the absence of inaccuracy is sufficient for the possible optimality of less than extreme sanctions.) We examine how the optimal investment in accuracy may change (which accordingly affects the optimal p and s) for a given level of deterrence, how different error rates affect the optimal p and s for a given investment in accuracy and level of deterrence, and how a change in the actual and optimal levels of deterrence affect the optimal p and s for a given investment in accuracy.

First, if accuracy were to change exogenously (i.e., the $q_i(k)$ functions were to shift up), the optimal investment in accuracy would likely change, which in turn would involve adjustments in p and s. Observe, for example, that when accuracy is exogenously lower, increases in p and s produce less deterrence than otherwise. (More of the effect will fall on the innocent and less on the guilty.) Thus, it may be that in achieving a given level of deterrence, it would be optimal to use a higher k (which would offset some of the exogenous reduction in accuracy).[17] A higher k, in turn, would make enforcement effort more costly at the margin (because $c_{pk} > 0$), which would favor reducing p and raising s relative to what would have been optimal.

Second, consider the possibility that accuracy changes exogenously but that the investment in accuracy (k) remains fixed. What would be the direct effect of, say, greater error on the optimal p and s? In the appendix, we examine this problem under some simplifying assumptions, including that the level of deterrence is held constant. We find that whether the presence of error calls for lower or higher sanctions (and a higher or lower level of enforcement effort) than otherwise depends on whether errors of mistakenly convicting the innocent are more numerous than errors of mistakenly acquitting the guilty. If higher error involves sanctions more often being imposed on the innocent, risk-bearing costs are higher so there is more reason to reduce s and increase p. But if higher error involves sanctions less often being imposed on the guilty, risk-bearing costs are lower so a higher s and lower p would be optimal.

---

[17] Whether it is indeed optimal to raise k would depend on whether and how the slopes of the $q_i(k)$ functions changed.

Third, if accuracy changes exogenously, the actual and optimal levels of deterrence may change. In particular, if error were greater, deterrence would fall. To restore the level of deterrence may involve raising both p and s. But, when error is greater, the optimal level of deterrence may fall; this decline may be more or less than the decrease in actual deterrence. Finally, when the optimal levels of p and s change, the optimal relative use of p and s may differ as well. Consider the case when the optimal p and s fall. On one hand, because risk-bearing costs are nonlinear in s, there would be less of a benefit to marginal reductions in s the more s is reduced. On the other hand, because enforcement costs may be nonlinear, there may be less of a benefit to marginal reductions in p the more p is reduced. Either effect could dominate.[18]

4. Discussion

(a) *Summary of results.* In the model of law enforcement supplemented by errors in determining liability, greater accuracy can be valuable in many ways. First, increasing accuracy is a method, other than increasing enforcement effort, of increasing deterrence. Thus, expenditures on accuracy and on the level of enforcement are substitutes. When high accuracy can be achieved at very low cost (as with parking and many traffic violations), therefore, a low probability of enforcement may be employed. Second, increasing accuracy (and reducing enforcement effort) allows a given level of deterrence to be achieved while imposing sanctions less often on both the innocent and the guilty.[19] When sanctions are socially costly (nonmonetary sanctions, or fines when individuals are risk-averse) rather than mere transfers, this is a further benefit of increasing expenditures on accuracy

---

[18] One case of interest is a simple audit or monitoring technology, $c(p,k) = pk$. In this case, $c_{pp} = 0$, so enforcement costs are linear in p (for a given investment in accuracy). Then, if greater error made it optimal to reduce deterrence by more than it fell on account of the error itself, this would be achieved optimally by reducing p, so the optimal mix would involve relatively heavier use of the sanction when error was greater.

[19] For an explanation of why sanctions are less for the guilty, see the discussion of expression (3).

rather than on the level of enforcement. Thus, greater accuracy is appropriate in criminal proceedings involving sanctions of imprisonment or fines likely to be a large fraction of individuals wealth than in civil disputes between large corporations.[20] A third benefit, involving improvements in actors' choices among acts in ways that cannot be achieved by simply increasing the level of enforcement effort, did not arise in our model, but would in others one could construct.[21]

In simple models of law enforcement without inaccuracy, the optimal sanction is maximal both when sanctions are costless (fines with risk-neutral actors) or when sanctions are costly but social costs are linear in the amount of the sanction (as might be the case with nonmonetary sanctions), while the optimal sanction may be less than maximal when individuals are risk-averse. Introducing inaccuracy does not fundamentally alter these conclusions. The only affect of inaccuracy on the optimal sanction arises in the model with risk aversion, in which case the effect on the optimal use of sanctions versus enforcement effort is ambiguous. One reason for the ambiguity is that error is of two types: false convictions of the innocent, which increases sanction costs, and false acquittals of the guilty, which decreases sanction costs. The former effect favors a lower sanction and the latter a higher sanction. Another reason is that error reduces deterrence, which may make it optimal to raise the sanction

In contrast with our results concerning optimal sanctions, it is often believed that the possibility that sanctions will be imposed on the innocent is a reason to reduce their level. In addition to the preceding remarks, it should be emphasized that if s is lowered, p must be raised if deterrence is

---

[20] In fact, large institutions often opt in advance for inexpensive forms of adjudication for disputes that may arise between them.

[21] For example, if our model were modified so that there were two forms of innocent (i.e., harmless) activity, only one of which (call it the "first") results in a risk of sanctions, enforcement effort and accuracy would no longer be substitutes with regard to behavior. An increase in p and decrease in k that kept an individual's incentive to commit the harmful act unchanged would increase the expected sanction for the first harmless act, which would inefficiently induce individuals who do not commit the harmful act to choose the second harmless act rather than the first. If, as in Png (1986), one could subsidize the harmless activity that is subject to sanctions, increased accuracy would have no behavioral benefit in this regard.

to be maintained, and an increase in p will result in the innocent being punished more often.[22] We also noted that if, for whatever reason, an enforcement authority is required to employ a lower sanction and higher level of enforcement than would be optimal, it will be desirable to reduce the level of accuracy, which increases the rate of mistakes in the imposition of sanctions. Thus, an independent concern for reducing mistakes is not generally a reason to rely less on sanctions and more on enforcement effort.

(b) *Different types of inaccuracy.* The model analyzed here concerns error in determining liability. The analysis does not depend on whether the error involves misidentification of who committed an act or mistake with regard to whether a given act is indeed the cause of the harm. The analysis, however, assumes that the error involves whether an individual is being properly sanctioned for a prior act, rather than whether the proper magnitude of sanction is applied to the act or when adjudications involve determining future obligations, such as an entitlement to receive public benefits.[23]

(c) *Policies to which our analysis is applicable.* Our analysis does not depend on whether the context involved criminal sanctions or even whether it involved the formal legal system rather than contractually created dispute resolution. The trade-off between cost and accuracy is relevant, for example, to piecemeal reforms (how much to limit pretrial discovery), judge's decisions concerning the conduct of cases (whether to deem inadmissible evidence that is largely redundant), and major restructuring of the legal system (mandatory small claims courts, substitution of administrative proceedings for court trials). Similarly, parties drafting contracts to govern their future

---

22 While the optimal level of deterrence might fall on account of inaccuracy, the argument in the text, of course, remains applicable because it applies to *any* level of deterrence. For example, when actors are risk neutral, if reducing s increased welfare, welfare could be further increased by raising s to $\bar{s}$ and reducing p: enforcement costs would be reduced, while keeping unchanged behavior, total sanction costs, and total error in imposing sanctions on the innocent.

23 The former problem, which is studied in Kaplow and Shavell (1992a, 1992b), involves different behavioral effects that depend primarily on the extent of individuals' information at the time they act concerning the true character of their activities and the errors an adjudicator is likely to make. In the latter problem, one would expect the importance of accuracy to depend upon social evaluations of different distributions of wealth rather than on ex ante incentive considerations.

relationships (for example, two firms, an employer and employees, members of a trade association) need to decide upon how accurate they wish resolution of their disputes to be.

The analysis suggests that the optimal set of procedures undoubtedly will vary greatly by context and will depend upon the how other instruments of enforcement are being used. Our legal system does have different rules in the criminal context and in small claims courts, and it uses other specialized tribunals; in addition, adjudicators no doubt make ad hoc adjustments in particular cases. Yet, across wide ranges of legal disputes, most rules concerned with accuracy are largely invariant.

It should also be noted that the level of accuracy is often not chosen directly in our legal system. Rather, parties introduce into evidence whatever information they find it in their interest to develop and present. There is no reason to suppose that the private incentive to produce information systematically equals its social value. Private parties are motivated by the desire to improve the ex post result; society is concerned with the resulting incentives for ex ante behavior, enforcement costs, and sanction costs. Thus, the problem of designing an efficient dispute resolution system involves the added complication of creating appropriate incentives for litigants.[24]

(d) *Burden of proof*.[25] One instrument of the legal system of particular importance with regard to accuracy is the burden of proof for conviction. When the burden of proof rises, the probability of false convictions falls and that of false acquittals rises. In terms of our notation, then, if α is the burden of proof, we have q_1 decreasing in α and q_0 increasing in α. Our results, therefore, apply for a given burden of proof, and the enforcement

---

[24] See Kaplow and Shavell (1992b). Of course, the choice to rely largely on privately motivated litigants to produce the relevant information rather than having the adjudicator collect information could be made differently, and this is done in some forms of dispute resolution -- as when an arbitrator of a construction dispute is an expert and resolves most factual controversy by personally inspecting the construction site.

[25] See Posner (1973). Rubinfeld and Sappington (1987) examine how setting the burden of proof effects defendants' litigation efforts.

problem with inaccuracy would also involve determining the optimal proof burden.[26]

When sanctions are more costly, it seems plausible that a higher burden of proof is appropriate. The reason is that raising the burden of proof means that both the innocent and guilty are sanctioned less often.[27] One would also expect that, at the optimal burden of proof, the marginal effect of increasing the proof burden is more favorable to the guilty who are detected than to the innocent. The reason is that, if raising the proof burden has the benefit of reducing sanctions, at the optimum it must have a cost, reduced deterrence.

[26] Our analysis assumed that greater expenditures on accuracy reduced both types of errors, whereas some strategies -- i.e., increasing resources available to indigent criminal defendants -- may reduce false convictions but increase false acquittals. But if one were simultaneously to increase the burden of proof to keep the fraction of the types of errors constant (and if one assumes that more resources for indigent defendants increase information rather than noise), then the strategy would reduce both types of errors. Thus, concerns for minimizing false convictions would sometimes be addressed most efficiently by increasing resources for the most informative strategies (even if they help the prosecution), combined with adjusting the burden of proof, rather than funneling more resources to defendants.

[27] Of course, raising the proof burden may affect behavior; if deterrence fell, more individuals would choose to be guilty and the guilty have a higher expected sanction, so it would be possible for total sanctions imposed to increase.

References

Becker, Gary S., "Crime and Punishment: An Economic Approach," Journal of Political Economy 76, 169-217 (1968).

Craswell, Richard and John E. Calfee, "Deterrence and Uncertain Legal Standards," Journal of Law, Economics, & Organization 2, 279-303 (1986).

Ehrlich, Isaac, "The Optimum Enforcement of Laws and the Concept of Justice: A Positive Analysis," International Review of Law and Economics 2, 3-27 (1982).

Grossman, Sanford J. and Oliver D. Hart, "An Analysis of the Principal-Agent Problem," Econometrica 51, 7-45 (1983).

Holmström, Bengt, "Moral Hazard and Observability," Bell Journal of Economics 10, 74-91 (1979).

Kaplow, Louis, "A Note on the Optimal Use of Nonmonetary Sanctions," Journal of Public Economics 42, 245-247 (1990).

__________, "The Optimal Probability and Magnitude of Fines for Acts That Definitely Are Undesirable," International Review of Law and Economics 12, 3-11 (1992).

__________ and Steven Shavell, "Private versus Socially Optimal Provision of Ex Ante Legal Advice," Journal of Law, Economics, & Organization 8, 306-320 (1992a).

__________ and __________, "Accuracy in the Assessment of Damages," Harvard Law School Program in Law and Economics Discussion Paper No. 116 (1992b).

Mookherjee, Dilip and I.P.L. Png, "Monitoring vis-à-vis Investigation in Enforcement of law," American Economic Review 82, 556-565 (1992).

Png, I.P.L., "Optimal Subsidies and Damages in the Presence of Judicial Error," International Review of Law and Economics 6, 101-105 (1986).

Polinsky, A. Mitchell and Steven Shavell, "The Optimal Tradeoff Between the Probability and Magnitude of Fines," American Economic Review 69, 880-891 (1979).

__________ and __________, "The Optimal Use of Fines and Imprisonment," Journal of Public Economics 24, 89-99 (1984).

__________ and __________, "Legal Error, Litigation, and the Incentive to Obey the Law," Journal of Law, Economics, & Organization 5, 99-108 (1989).

Posner, Richard A., "An Economic Approach to Legal Procedure and Judicial Administration," Journal of Legal Studies 2, 399-458.

Rubinfeld, Daniel L. and David E.M. Sappington, "Efficient Awards and Standards of Proof in Judicial Proceedings," Rand Journal of Economics 18, 308-315 (1987).

Shavell, Steven, "Risk Bearing and Incentives in the Principal and Agent Relationship," Bell Journal of Economics 10, 55-73 (1979).

__________, "Specific versus General Enforcement of Law," Journal of Political Economy 99, 1088-1108 (1991).

Appendix: Risk Aversion

Assume that individuals are risk-averse, with utility functions $u(\cdot)$ that are strictly concave in wealth. Wealth is taken to be the benefit obtained from an act (if an individual acts) minus the sanction if it is imposed and a lump-sum tax. (For simplicity, it is assumed that the effect of the external harm h is additively separable, rather than entering wealth.[28]) The tax is the enforcement cost minus expected fine revenue. In this case, an individual acts if and only if

$$\text{(A1)}\quad (1-pq_1)u(-t) + pq_1u(-s-t) \geq (1-p(1-q_0))u(b-t) + p(1-q_0)u(b-s-t).$$

Again, it is useful to denote the type of individual just indifferent as to whether to act as the type with benefit b*. (Because the right side of (A1) is increasing in b and the left side is unaffected, individuals who act are those for whom $b \geq b^*$.)

The enforcement authority chooses p, s, and k to maximize welfare, which is now

$$\text{(A2)}\quad W = \int_0^{b^*} [(1-pq_1)u(-t) + pq_1u(-s-t)]f(b)db$$

$$+ \int_{b^*}^{\infty} [(1-p(1-q_0))u(b-t) + p(1-q_0)u(b-s-t) - h]f(b)db,$$

subject to the constraint that

---

[28] Because the analysis will keep behavior constant, the level of harm will be unaffected. Moreover, because the analysis will use a constant-absolute-risk-aversion utility function, the effect of the harm if it entered wealth would be multiplicatively separable, and thus would not affect any of the derivations.

(A3) $t = c(p,k) - ps[q_1F(b^*) + (1-q_0)(1-F(b^*))]$.

Our analysis here is confined to considering whether inaccuracy is an independent reason to reduce s and increase p. The inquiry is complicated by three considerations. First, the presence of inaccuracy makes it more costly to achieve a given level of deterrence, so the optimal level of deterrence may be lower. The optimal mix of p and s may depend on the level of deterrence without regard to error.[29] But this would be true regardless of why deterrence should be lower. (For example, it would be true if deterrence should be lower because the harm caused by the act is less.) Thus, our analysis will hold behavior constant, and ask how the degree of inaccuracy affects the optimal mix of p and s.

Second, the problem need not have an interior solution -- that is, a maximal sanction may be optimal both with and without inaccuracy.[30] We will assume that an interior solution exists and ask how the optimal intermediate sanction and probability of enforcement change when the level of accuracy is changed exogenously.

Third, wealth effects complicate comparative statics in problems like the present one. To illuminate the problem without providing an exhaustive analysis, we will consider the case in which individuals have a constant-absolute-risk-aversion utility function. In particular, we will assume that $u(w) = -e^{-\alpha w}$, where w denotes total wealth and $\alpha > 0$.

We begin by restating welfare (A2) for this utility function.

---

29 There are competing effects. When deterrence must be higher, the use of a higher sanction implies greater risk-bearing costs, which favors a greater emphasis on enforcement effort. Yet greater enforcement effort is increasingly costly at the margin for most technologies, which favors a greater emphasis on high sanctions.

30 It is also possible that $p = 0$ is optimal (in which case the level of the sanction is irrelevant).

$$\text{(A4)}\quad W = -e^{\alpha t}[1 - pq_1 + pq_1 e^{\alpha s}]F(b*) - (1 - F(b*))h$$

$$- e^{\alpha t}[1 - p(1-q_0) + p(1-q_0)e^{\alpha s}] \int_{b*}^{\infty} e^{-\alpha b}f(b)db.$$

The first term is the expected utility of an individual who does not act weighted by the fraction of the population that does not act. (The first component reflects the utility cost of the tax and the second the utility cost of the sanction.) The second term is the expected harm (the fraction of the population who are not deterred times the harm). The final term is the expected utility of those who act. (The components reflect the tax, sanction, and benefit of acting.)

The method of analysis begins by setting the derivative of welfare (A2) with respect to s equal to zero, where p is given by p(s) so that b* is unchanged.

$$\text{(A5)}\quad \frac{dW}{ds}\Big|_{p=p(s)} = e^{\alpha t}[p'q_1(1 - e^{\alpha s}) - pq_1\alpha e^{\alpha s}]F(b*) - \alpha t'e^{\alpha t}[1 - pq_1 + pq_1e^{\alpha s}]F(b*)$$

$$+ e^{\alpha t}[p'(1-q_0)(1 - e^{\alpha s}) + p(1-q_0)\alpha e^{\alpha s}] \int_{b*}^{\infty} e^{-\alpha b}f(b)db$$

$$- \alpha t'e^{\alpha t}[1 - p(1-q_0) + p(1-q_0)e^{\alpha s}] \int_{b*}^{\infty} e^{-\alpha b}f(b)db = 0,$$

where $p' = dp(s)/ds$. It can be shown that the first and third terms equal zero. The reason is that the bracketed components reflect the marginal utility effect of raising s and lowering p so as to keep b* constant (in the first term for those who do not act, and in the third term for those who act); with the stipulated utility function, this marginal utility effect is zero.[31]

The second and fourth terms both weight the component $-\alpha t'e^{\alpha t}$ by positive components. We rewrite (A5) as

---

[31] One can show that $p' = \alpha pe^{\alpha s}/(1-e^{\alpha s})$ by equating the utility of those who act and are of type b* with that of those who do not act and taking the derivative with respect to s, varying p according to p(s), so that the equality continues to hold. Substituting this expression for p' into the first and third terms yields the result in the text.

$$\text{(A6)}\ \frac{dW}{ds}\Big|_{p=p(s)} = -\alpha t' e^{\alpha t}\Omega = 0,$$

where Ω is the sum of the positive components. Thus, if there is an interior optimum, it must be that t' equals zero. The tax term captures the effects of the increase in s and reduction in p on both enforcement costs and risk-bearing costs.[32] Raising s and lowering p reduces enforcement costs and increases risk-bearing costs; these effects must just offset each other at the margin when at the optimum.

It is now possible to explore the effect of accuracy on the choice of p and s. Let θ denote an exogenous parameter affecting accuracy, such that $dq_i/d\theta < 0$. That is, an exogenous increase in θ corresponds to an exogenous increase in accuracy (holding k constant).[33] Moreover, assume that an increase in accuracy reflected by a change in θ is accompanied by a shift in the distribution of benefits, with all individuals benefits rising by $\beta(\theta)$, so as to keep behavior constant. (As noted previously, if accuracy changes, the level of deterrence will change, which itself would call for adjustments in p and s. This construction allows us to consider changes in p and s caused by different levels of accuracy when deterrence and enforcement costs are otherwise unaffected.)

Taking the derivative of (A6) with respect to θ, where $\beta(\theta)$ is given as described, we have

---

[32] If behavior is held constant, it must be that the utility cost of the expected sanction is the same. Thus, if the lump-sum tax were the same, welfare (A2) would be unaffected. But the lump-sum tax (A3) is higher on account of increased enforcement costs and lower because more fine revenue is collected. The reason for the latter is that the term ps must increase if the same utility cost is imposed with a higher p and a lower s, because $u(\cdot)$ is strictly concave. The intuition connecting the fine revenue to risk-bearing costs is that, the higher the fine revenue for a given utility cost of the sanction, the less of the utility cost is a welfare loss as a result of the imposition of risk and the more is simply a transfer.

[33] This approach is used rather than exploring a change in k because the latter would involve an effect on costs which itself may call for some adjustment in p and s. The construction in the text allows us to focus solely on the effects of an exogenous change in accuracy.

(A7) $\frac{d}{d\theta}\left[\frac{dW}{ds}|_{p=p(s)}\right]|_{\beta=\beta(\theta)} = -\alpha t' e^{\alpha t}\Omega' - \Omega(\alpha^2 t_\theta t' e^{\alpha t} + \alpha e^{\alpha t} t'_\theta)$.

Because $t' = 0$ at the optimum, the first term and first component of the second term equal zero. Thus the sign of (A7) is the opposite of the sign of t'_θ. First, we note that

(A8) $t' = \frac{dt}{ds}|_{p=p(s)} = c_p p' - ps(0 + 0) - (p's + p)[q_1 F(b*) + (1-q_0)(1-F(b*))]$.

Next,

(A9) $t'_\theta = \frac{dt'}{d\theta}|_{\beta=\beta(\theta)} = -(p's + p)[q_{1\theta}F(b*) - q_{0\theta}(1-F(b*))]$,

where $q_{i\theta}$ denotes $dq_i/d\theta$. (Expression (A9) follows from (A8) because p' is independent of the level of accuracy.[34]) From (A8), it follows that $p's + p$ is negative at the optimum. (Because c_p is positive and p' is negative, the first term in (A8) is negative, so at the optimum, where $t' = 0$, it must be that $p's + p$ is negative.) Thus, the sign of expression (A9) is given by the last component on the right side. And, as noted above, the sign of (A9) is the opposite of the sign of (A7), so we have

(A10) sign $\frac{d}{d\theta}\left[\frac{dW}{ds}|_{p=p(s)}\right]|_{\beta=\beta(\theta)} = -$ sign $[q_{1\theta}F(b*) - q_{0\theta}(1-F(b*))]$.

That is, if the bracketed expression in the right side of (A10) is positive, a lower sanction is optimal if accuracy increases.[35]

To interpret this condition, consider the two components on the right side of (A10). An increase in θ increases accuracy by reducing q_1, the rate at which sanctions are mistakenly imposed on those who did not commit the act. Because $q_{1\theta}$ is negative, this indicates that a higher sanction is optimal. That is, when there is less error with regard to mistakenly convicting the innocent, the optimal enforcement strategy shifts in the direction of a higher s and a lower p. Error of this type is indeed a possible reason to moderate the level of the sanction and raise enforcement effort, all other things equal. The intuition is that when, due to error, sanctions are more often

---

[34] See note 31.

[35] This conclusion holds only locally. If one is at a unique interior optimum, a small increase in θ would reduce the optimal s if the component is positive.

imposed on the innocent, risk-bearing costs are greater for a given level of the sanction, so the trade-off between p and s changes in a manner that favors reducing s.

An increase in θ also increases accuracy by reducing q_0, the rate at which those who commit the act (and are detected) mistakenly escape sanctions. This effect is in the opposite direction of that just discussed. The intuition is that when sanctions are imposed less often on the guilty, the risk-bearing costs are less for a given level of the sanction, which favors reducing p and raising s to achieve a given level of deterrence.

Whether the presence of error calls for lower or higher sanctions (and a higher or lower level of enforcement effort) than otherwise depends on whether errors of mistakenly convicting the innocent are more numerous than errors of mistakenly acquitting the guilty. The magnitudes of the two effects in (A10) are simply the fraction of the relevant group in the population times the amount by which the error rate changes.[36]

---

[36] For example, assume initially that $F(b^*) = .99$ -- i.e., only 1% of the population optimally commits the act. Also assume that $q_0 = .5$ -- i.e., those who commit the act are sanctioned half of time once detected -- and that $q_1 = .001$ -- i.e., one in a thousand of those who do not commit the act (weighted by the detection rate) are sanctioned. To illustrate, if $p = .5$ and there are 200,000 people, 99 innocent people are convicted and 500 guilty people are convicted. Now assume that both error rates increase by 1%. One has (substituting population totals for fractions)

$$q_{1\theta}F - q_{0\theta}(1-F) \approx .00001 \times 198{,}000 - .005 \times 2{,}000 = 1.98 - 10 = -8.02.$$

Thus, the increase in the rate of both types of error would make it optimal to raise s and lower p. If, instead, the fraction not committing the act were .999, the evaluation would be $.00001 \times 199{,}800 - .005 \times 200 = 1.998 - 1 = .998$, in which case an increase in the rate of both types of error would make it optimal to lower s and raise p.