Michigan Law Review

Volume 97 | Issue 7

1999

Deterrence and Damages: The Multiplier Principle and Its Alternatives

Richard Craswell
Stanford University

Follow this and additional works at: https://repository.law.umich.edu/mlr

Part of the Law and Economics Commons, Law and Psychology Commons, and the Legal Remedies Commons

DETERRENCE AND DAMAGES: THE MULTIPLIER PRINCIPLE AND ITS ALTERNATIVES

*Richard Craswell**

TABLE OF CONTENTS

One purpose of fines and damage awards is to deter harmful behavior. When enforcement is imperfect, however, so the probability that any given violation will be punished is less than 100%, the law's deterrent effect is usually thought to be reduced. Thus, it is often said that the ideal penalty (insofar as deterrence is concerned) equals the harm caused by the violation multiplied by one over the probability of punishment. For example, if a violation faces only a 25% (or one-in-four) chance of being punished, on this view the optimal penalty would be four times the harm caused by the violation.

This prescription, which I will call the "multiplier principle," has a long pedigree.[1] It figures prominently in texts on law and economics,[2] and has been discussed in many scholarly works.[3] Indeed, in the law review literature the multiplier principle is now routinely cited as part of standard deterrence theory.[4] The

---

1. In modern economics, the multiplier principle owes much of its prominence to Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968). An early version of the multiplier principle can be found in CESARE BECCARIA, ON CRIMES AND PUNISHMENTS 46 (David Young ed., 1986) (1766), and more explicitly in JEREMY BENTHAM, A THEORY OF LEGISLATION 325-26 (R. Hildreth trans., London, Trübner & Co. 1864) (1802). For a further discussion of these historical roots, see Keith N. Hylton, *Punitive Damages and the Economic Theory of Penalties*, 87 GEO. L.J. 421, 425-27 (1998).

2. *See, e.g.*, ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 390-95 (1988); A. MITCHELL POLINSKY, AN INTRODUCTION TO LAW AND ECONOMICS 77-78 (2d ed. 1989); RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW § 7.1, at 241 (5th ed. 1998).

3. The most recent and most comprehensive of these is A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869 (1998). Earlier analyses, in addition to the textbooks cited *supra* note 2, include WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 160 (1987); Bruce Chapman & Michael Trebilcock, *Punitive Damages: Divergence in Search of a Rationale*, 40 ALA. L. REV. 741, 808-19 (1989); Robert D. Cooter, *Punitive Damages for Deterrence: When and How Much?*, 40 ALA. L. REV. 1143, 1149-61 (1989); and Dorsey D. Ellis, Jr., *Fairness and Efficiency in the Law of Punitive Damages*, 56 S. CAL. L. REV. 1, 25-26 (1982).

4. Examples published in just the last two years include Tom Baker, *Reconsidering Insurance for Punitive Damages*, 1998 WIS. L. REV. 101, 106-07; Roger D. Blair & Thomas F. Cotter, *An Economic Analysis of Damage Rules in Intellectual Property Law*, 39 WM. & MARY L. REV. 1585, 1619-20 (1998); David Crump, *Evidence, Economics, and Ethics: What Information Should Jurors Be Given to Determine the Amount of a Punitive-Damage Award?*, 57 MD. L. REV. 174, 187 nn.82-83 (1998); Antony W. Dnes & Jonathan S. Seaton, *An Exploration of the Tort-Criminal Boundary Using Manslaughter and Negligence Cases*, 17 INTL. REV. L. & ECON. 537, 539 (1997); C. Douglas Floyd, *Antitrust Victims Without Antitrust Remedies: The Narrowing of Standing in Private Antitrust Actions*, 82 MINN. L. REV. 1, 5 (1997); Thomas Koenig & Michael Rustad, *"Crimtorts" as Corporate Just Deserts*, 31 U. MICH. J.L. REFORM 289, 315 & nn.115-16 (1998); Dale A. Nance, *Guidance Rules and Enforcement Rules: A Better View of the Cathedral*, 83 VA. L. REV. 837 app. at 934 (1997); Paul H. Rubin et al., *BMW v. Gore: Mitigating the Punitive Economics of Punitive Damages*, 5 SUP. CT. ECON. REV. 179, 186 (1997); John K. Setear, *Responses to Breach of a Treaty and Rationalist International Relations Theory: The Rules of Release and Remediation in the Law of Treaties and the Law of State Responsibility*, 83 VA. L. REV. 1, 81-86 (1997); Cass R. Sunstein et al., *Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)*,

multiplier principle has also begun to be recognized by courts — especially by economically sophisticated judges — as a possible rationale for punitive damages.[5]

What is less widely appreciated, however, is that the multiplier principle is almost never *necessary* to achieving optimal deterrence. Even when the probability of punishment is less than 100%, more recent work in law and economics has identified several other remedies that could also achieve optimal deterrence.[6] These alternative remedies are often significantly less than those called for by the multiplier principle. In some cases, the alternative remedies could even be less than the harm caused by the violation, implying that optimal deterrence could be achieved if damages were reduced.

My principal aims in this article are to explain why the multiplier principle is not necessary for optimal deterrence and to begin a discussion of the alternatives. While the mathematical analysis behind the recent work is often quite technical, the basic principles are not hard to grasp, and they can be illustrated with simple numerical examples. Thus, a secondary aim is to familiarize a larger audience with the conclusions of this technical body of work. Since this work identifies alternatives to the multiplier principle, its significance is potentially as broad as that of the multiplier principle itself.

Part I begins by reviewing how optimal deterrence is achieved by what I will call the "traditional multiplier principle," in which the harm caused by a defendant's offense is multiplied by one over the probability of punishment. When the probability of punishment is the same no matter how badly a defendant has behaved, such a multiplier is relatively easy to administer. In most contexts in which enforcement is imperfect, however, the probability of punishment any particular defendant faces depends in part on the nature of his or her violation. That is, in most legal regimes, defendants who commit only marginal offenses are less likely to be punished than those who commit more serious or egregious ones. Whenever this is the case, Part I shows that the traditional multiplier principle can

---

107 YALE L.J. 2071, 2082-84 (1998); and W. Kip Viscusi, *The Social Costs of Punitive Damages Against Corporations in Environmental and Safety Torts,* 87 GEO. L.J. 285, 312-13 (1998).

5. *See, e.g.,* BMW of N. Am., Inc. v. Gore, 517 U.S. 559, 592-94 (1996) (Breyer, J., concurring); Kemezy v. Peters, 79 F.3d 33, 35 (7th Cir. 1996) (Posner, C.J.); Zazú Designs v. L'Oréal, S.A., 979 F.2d 499, 508 (7th Cir. 1992) (Easterbrook, J.); FDIC v. W.R. Grace & Co., 877 F.2d 614, 623 (7th Cir. 1989) (Posner, J.).

6. I discuss this work *infra* in Part II.

achieve optimal levels of deterrence *only if the multiplier is calculated on a case-by-case basis,* so that the multiplier varies with each defendant's probability of punishment.

Significantly, few legal regimes follow the traditional multiplier in this respect, for few (if any) use multipliers that are calculated case-by-case. Often no multiplier is used and only compensatory damages are awarded, as in most civil suits under the common law. When the law does use a multiplier, it is often set at a single value that is the same for all defendants, as in the treble damage rule of antitrust law.[7] And when criminal or administrative penalties are used, it is common to set a single fine for all violations of a certain type (e.g., $100 for failing to stop at a stop sign), regardless of either the harm caused or the probability of punishment. Obviously, none of these systems of punishment satisfies the traditional, case-by-case multiplier principle.

The fact that real legal systems usually use some alternative to a case-by-case multiplier raises two questions for those interested in deterrence. First, when (if ever) might these alternatives be superior to the case-by-case multiplier? Second, whether or not they are superior, at what level should the fines or penalties be set (under each of these alternatives) to get the best deterrence possible? A naïve view might hold that when the law uses one of the alternatives, the best policy would still set the fine or other penalty *as close as possible* to the level of the case-by-case multiplier. But this naïve view is incorrect, for — as recent work in law and economics has also shown — each of these alternatives requires a different level of penalties to achieve optimal deterrence. Unfortunately, these differences are not widely understood in mainstream legal analysis, which continues to be fixed on the more traditional theory in which the multiplier is adjusted case by case.

Part II explores these differences by identifying the optimal fines or damage awards[8] under each of the possible alternatives. If the law uses the same multiplier for all defendants, optimal deterrence will usually require a penalty below the traditional multiplier principle. If the law instead uses a single *fine* for all defendants, as opposed to a single *multiplier,* the optimal fine could be either above or below the level of the multiplier principle,

---

7. 15 U.S.C. § 15(a) (1994).

8. In most of the article, I will speak of fines and damage awards interchangeably, without regard to whether they are paid to the victims (the usual rule for damages) or to the state (the usual rule for fines). While there are of course important differences between each of these penalties, most of those differences do not alter the resulting deterrent effect.

depending on (among other things) how rapidly the probability of punishment changes if a defendant commits a more or a less egregious offense. Indeed, this last factor — the responsiveness of the probability of punishment to *changes* in the egregiousness of an offense — is an important determinant of the optimal penalty under each of the alternatives discussed here. For this reason, Part III of the article discusses the factors that could make the probability of punishment more or less responsive to the degree of a defendant's offense.

Finally, Part IV addresses the first question identified above, by assessing the advantages and disadvantages of these alternatives relative to a case-by-case multiplier. The alternatives clearly differ in their computational complexity, and in the informational and other demands they place on judges and juries (or on legislatures). They may also differ in other key respects, such as their effect on litigation costs or on defendants' overall levels of activity, or on the law's symbolic or expressive effects. In the end, I conclude that remedies based on the traditional multiplier principle may well be useful, and may even dominate the alternatives, in a fairly small set of cases. In other cases, however, the balance of advantages and disadvantages is harder to assess.

I. The Traditional Multiplier Principle

At the outset, it will help to distinguish between two ways of achieving optimal deterrence: by adjusting the penalty, and by adjusting the substantive legal standard. Suppose, for example, that we wish to reduce the risk of leaks of toxic waste, and that the optimal level of precautions would reduce that risk to exactly 1%. One way to achieve this goal is to set a substantive legal standard under which defendants are liable if but only if their risk exceeds that level. If the substantive standard is set correctly, and if it can be applied with no risk of error — two qualifications that will become important below — it may not matter if the penalty for violating the standard is set according to the multiplier principle. In such a regime, all that matters is that the penalty be large enough to deter defendants from violating the substantive legal standard, so any penalty set at this level *or higher* will achieve optimal deterrence. Indeed, as long as defendants who comply with the substantive standard can be assured of not paying any penalties at all, the penalty could (in theory) be increased to infinity without inducing overdeterrence, because defendants could always avoid any penalty by complying with the substantive standard. In such a regime,

therefore, any increases in the penalty above the minimum level needed to induce compliance with the standard would not affect the behavior chosen by defendants.[9]

In most legal regimes, however, the size of the penalty will have a continuous effect on defendants' behavior. For example, if the law holds defendants strictly liable for *every* leak, it will then be impossible for defendants to insulate themselves from liability simply by complying with any substantive legal standard. The same is true if there is no official rule of strict liability but if the substantive standard is applied with some risk of error, so that even a defendant whose leaks are within the officially permitted level still faces some chance of being held liable. As long as defendants face some chance of being held liable, every increase in the penalty will strengthen the incentive to reduce their number of leaks. At some point, then, if the penalty is set too high, the incentive will grow too strong and there will be too much deterrence. In short, optimal deterrence in such a regime requires that the penalty be set at exactly the right level, not merely that it be at or above some minimum.

My focus in this article is on regimes of the second type, for that is where the multiplier principle is most relevant.[10] Whenever defendants cannot avoid any chance of liability, the *average* or *expected* liability is what governs the deterrence incentives. For example, if there is only one chance in four that they will actually be found liable, defendants will discount the penalty by 25% in calculating how much they could save by reducing their leaks. This discounting supplies the rationale for the multiplier principle, for it seems to imply that deterrence will be reduced unless the penalty is multiplied by four, to offset the 25% probability of punishment.

However, this conclusion (that the law's deterrent effect will be reduced without a multiplier) requires one additional assumption: that the probability of punishment is unaltered by any changes in a defendant's behavior. Subsection I.A below shows the role this assumption plays in the traditional analysis of multipliers. Subsection I.B then relaxes this assumption, to show why the multiplier principle is not necessary for optimal deterrence if the assumption is invalid.

---

9. This point is developed at more length in Robert Cooter, *Prices and Sanctions,* 84 COLUM. L. REV. 1523, 1524-27 (1984). I will return to it *infra* in section II.C.

10. The application of the multiplier principle (and of punitive damages generally) to regimes of the first sort is discussed in Hylton, *supra* note 1.

A. *Assuming the Probability of Punishment Stays Constant*

Suppose that a firm's activities pose a 1% risk of leaking toxic waste in a way that would cause $6 million worth of damages.[11] The average or expected costs of this activity would thus be $60,000 (.01 × $6,000,000). Suppose, though, that the firm could alter its operations to eliminate that risk and replace it with a 1% risk of a less serious leak that would cause only $5 million in damages, thus reducing the expected costs to $50,000 (.01 × $5,000,000). This means that the alteration would reduce the expected social costs of this activity by $10,000 ($60,000 − $50,000). The alteration is therefore socially efficient if, but only if, its cost is less than $10,000.

If enforcement is perfect — that is, if the firm knows that it will have to pay for all damages its leaks cause — then the firm's average liability will be $60,000 if it does not alter its operations and $50,000 if it does. By altering its operations, the firm can thus save $10,000 in expected liability ($60,000 − $50,000). Since $10,000 is also the social benefit from the alteration, the firm's incentives will be socially optimal. That is, the firm will have an incentive to make the alteration if, but only if, the cost of doing so is less than $10,000.

Now suppose that enforcement is instead imperfect, and that even if there is a leak the firm faces only a 25% chance of being held liable. The firm might then reason as follows: "If we do not alter our operations, then whenever there is a leak (i.e., in 1% of the cases) we will face a 25% probability of having to pay $6 million in damages. This is equivalent to an expected liability of $15,000 (.01 × .25 × $6,000,000). If we instead alter our operations, we will face a 25% chance of having to pay only $5 million, so our expected liability will decline to $12,500 (.01 × .25 × $5,000,000). But this shows that altering our operations would reduce our expected liability by only $2,500 ($15,000 − $12,5000)." This provides too little incentive for the firm to alter its operations, since the social gain from the alteration would still be $10,000 (not $2,500). This is the reasoning behind the multiplier principle: if the chance of having to pay damages is only 25%, the firm will discount the expected penalties undesirably, and the incentive to improve its behavior will be weakened.[12]

---

11. This example is similar to one used by Polinsky & Shavell, *supra* note 3, at 879-80. Like them, I assume that all the relevant costs and benefits can be measured (or at least approximated) numerically.

12. More precisely, the deterrent effect will be reduced if the firm *believes* that the probability of punishment is only 25%: it is the firm's beliefs about the probability that matter. In the discussion that follows, this qualification will be assumed. For a further dis-

This example also shows that if the measure of damages were multiplied by four, the firm's incentives would be optimal again. If damages were multiplied by four, the firm would then have to pay $24 million (4 × $6,000,000) if it caused a leak and had not altered its operations, and would have to pay $20 million (4 × $5,000,000) if it had altered its operations. This gives the firm an average or expected liability of $60,000 without the alteration (.01 × .25 × $24,000,000), compared to $50,000 with the alteration (.01 × .25 × $20,000,000), so the alteration would reduce the firm's expected liability by $10,000 ($60,000 – $50,000). A multiplier of four thus restores the equality between the private gain to the firm and the gain to society at large. This is because a multiplier of four exactly compensates for the 25% chance of being held liable.

Examples such as this are what give the multiplier principle its air of inevitability. Notice, though, that this example assumed that the probability of punishment stayed fixed at 25% *whether or not* the firm altered its operations. The firm's expected liability was calculated on the assumption that if it failed to make any alterations, it faced a 25% probability of paying $24 million in damages; while if it did make the alterations, it faced the same 25% probability of paying $20 million in damages. In other words, the example assumed that the probability of having to pay damages stayed the same regardless of the firm's actual behavior.

As noted earlier, more recent work in law and economics has relaxed this assumption by considering legal regimes in which the probability of punishment depends in part on how well the defendant has behaved.[13] The following subsection explains the implica-

---

cussion of this aspect of deterrence theory, see Steven Garber, *Product Liability, Punitive Damages, Business Decisions and Economic Outcomes,* 1998 Wis. L. Rev. 237, 247-51.

13. The earliest analyses include CHARLES J. GOETZ, CASES AND MATERIALS ON LAW AND ECONOMICS 299-303 (1984); STEVEN SHAVELL, ECONOMIC ANALYSIS OF ACCIDENT LAW 93-99 (1987); John E. Calfee & Richard Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards,* 70 VA. L. REV. 965 (1984); and Richard Craswell & John E. Calfee, *Deterrence and Uncertain Legal Standards,* 2 J.L. ECON. & ORG. 279 (1986). Other work will be cited below where it is relevant.

Another (and much larger) line of literature addresses the fact that the probability of punishment can also be altered by increased expenditures on detection and enforcement, or by other improvements to the enforcement system as a whole. For examples of this literature — a literature that also traces its origins to Becker, *supra* note 1 — see A. Mitchell Polinsky & Steven Shavell, *The Optimal Tradeoff between the Probability and Magnitude of Fines,* 69 AM. ECON. REV. 880 (1979); or Steven Shavell, *Specific Versus General Enforcement of Law,* 99 J. POL. ECON. 1088 (1991). My concern in this article, however, is with the extent to which the probability of punishment varies with changes *in an individual defendant's behavior.* I therefore will not address any policies that might alter the probability of punishment across the board, while still leaving that probability unaffected by changes in an individual defendant's behavior.

tions of this work, and shows why the multiplier principle may no longer be necessary for optimal deterrence.

B. *Relaxing the Constant Probability Assumption*

In many contexts, the probability of punishment declines if defendants reduce the riskiness of their behavior or if they behave better along any other dimension. Some reasons for the lower probability of punishment will be discussed in Part III — for example, the public authorities may be less likely to prosecute a firm that appears to be making a good-faith effort to reduce the severity of its leaks, or such a firm might be better able to convince a court that it has not violated the applicable legal standard. For now, the exact reason for the reduced probability of punishment does not matter, as long as the probability does in fact decline.

If the probability of punishment does decline when a defendant improves its behavior, this produces two offsetting effects. The fact that the probability of punishment is still less than 100% will weaken the law's deterrent effect, just as in the earlier example. But if the probability of punishment falls even lower when the defendant improves its behavior, this will *strengthen* the law's deterrent effect, because improved behavior will then bring an extra "reward" in the form of a lower probability of punishment. And since there is no logical connection between the size of these two effects, it is possible for either one to outweigh the other, leading either to a net weakening effect (and net underdeterrence) or a net strengthening effect (net overdeterrence). Indeed, it is even possible for both effects to exactly offset, leaving deterrence at just the optimal level.[14]

These possibilities can all be illustrated using the toxic waste example. Let us now suppose that the probability of punishment is still 25% if the firm does not alter its operations, but only 10% if it does. If no multiplier is used, the firm will still face an expected liability of $15,000 if it does not alter its operations (.01 × .25 × $6,000,000), but if it does alter its operations its expected liability will fall to $5,000 (.01 × .10 × $5,000,000). This means that altering its operations will reduce the firm's expected liability by $10,000 ($15,000 − $5,000). But $10,000 is also the social gain from the alteration, so the firm's private savings will once again equal the social savings. In other words, even with imperfect enforcement

---

14. For a mathematical demonstration of all three possibilities, see Craswell & Calfee, *supra* note 13, at 284 tbl.1.

and no damage multiplier at all, it is still possible for the firm's in-
centives to be optimal.

Of course, it is only by coincidence (or by careful selection of
the hypothetical numbers) that the firm's incentives would be ex-
actly optimal. If extra precautions did not reduce the probability of
punishment quite as much, so that probability fell from 25% to
some figure above 10%, the firm's savings in expected liability
would then be less than $10,000, which would give the firm too
weak an incentive (net underdeterrence). Conversely, if the extra
precautions reduced the probability of punishment even more, to
some figure even less than 10%, the firm's savings would then be
more than $10,000, and the firm would have too strong an incentive
(net overdeterrence). The net effect on the firm's incentives thus
depends critically on the *rate* at which the probability of punish-
ment falls as the firm improves its behavior. In economic terms, it
is the marginal *change* in the expected penalty, and not its absolute
level, that governs the firm's incentives.

C. *Adjusting the Multiplier Case-by-Case*

Even when the multiplier principle is not strictly *necessary* for
optimal deterrence, it could still provide one way of achieving that
goal. To be sure, we just saw that if extra precautions reduced the
probability of punishment from 25% to 10%, then defendants
would have optimal incentives without any multiplier at all. In such
a case, it might seem that any multiplier would be a bad idea,
because it could only strengthen the deterrent effect and lead to
overdeterrence. However, this intuition must be discarded when
the probability of punishment is not a constant. In this section, I
show that optimal deterrence can still be achieved using the tradi-
tional multiplier principle, but only *if the multiplier is calculated sep-
arately for each defendant,* based on each defendant's actual
probability of punishment.

Consider again the example from the preceding subsection. In
that example, defendants who did not take extra precautions faced
a 25% probability of punishment. If the multiplier is recalculated
case by case, these defendants will still be given a multiplier of four,
so they will still have to pay $24 million ($4 \times \$6,000,000$) if they are
found liable. Their average or expected liability will thus be
$60,000 ($.01 \times .25 \times \$24,000,000$). However, if defendants who do
take the extra precautions face only a 10% probability of punish-
ment, they would be assigned a multiplier of ten (if the multiplier is
recalculated case by case). These defendants will therefore have to

pay $50 million (10 × $5,000,000) whenever they are found liable, so their expected liability will be $50,000 (.01 × .10 × $50,000,000).[15] This means that defendants' incentives will be optimal again, because taking the extra precautions will reduce their expected liability by $10,000 ($60,000 − $50,000), which just equals the social gain from those precautions.

In short, even when the multiplier principle is not absolutely *necessary* for optimal deterrence, it may still be *sufficient* to achieve that goal. As long as the probability of punishment (whether 25% or 10%) is offset by just the right multiplier, the effect will be the same as if enforcement were perfect, and defendants will be made to feel the full costs of their behavior.

1. *Case-by-Case Multipliers as "Taxes" on Improvements*

It is instructive to compare this conclusion with the one reached in the preceding subsection, when optimal deterrence was achieved even without any damage multiplier. In the preceding subsection, we saw that the decline in the probability of punishment (from 25% to 10%) could itself be enough to create optimal incentives, thus achieving optimal deterrence even though defendants paid only the actual costs of their behavior ($6 million and $5 million, respectively).[16] Yet the example just discussed showed that optimal deterrence could also be achieved if the penalties were increased through a multiplier — and increased quite dramatically, to $24 million and $50 million respectively. But this might seem paradoxical: How can a dramatic increase in penalties leave the law with the same deterrent effect?

The explanation lies in the fact that, while a case-by-case multiplier does increase penalties overall, it also introduces a new factor that checks any tendency to overdeterrence. If the multiplier is recalculated in every case, defendants who improve their behavior will (in effect) be "taxed" on that improvement, by being subjected to a larger multiplier if and when they are found liable. In the example just discussed, defendants who took extra precautions reduced their probability of punishment (from 25% to 10%) and thereby *increased* their multiplier (from four to ten). This tax, in the form of a higher multiplier, counteracted what would otherwise

---

15. Notice that the penalty assessed against these defendants ($50 million) will be larger than the penalty assessed against defendants who did *not* take the extra precautions ($24 million). The possibility that this relationship might offend notions of corrective justice will be discussed *infra* in section IV.G.

16. *See supra* text following note 14.

be an incentive for overdeterrence. As long as the multiplier is re-calculated whenever the probability of punishment declines, the net effects should exactly offset and defendants should have just the optimal incentives.[17]

Indeed, the same analysis would apply even if the original incentives (in the absence of any multiplier) had favored *over*deterrence. For example, if taking extra precautions reduced the probability of punishment from 25% to less than 10%, the reward for taking those precautions would then be even greater (in the absence of a multiplier), leading to too much deterrence. But if a multiplier is used, and if the multiplier is calculated separately for each defendant, defendants who take extra precautions would then face an even higher tax, for their new multiplier would be even greater than ten (because their probability of punishment would be less than 10%). Paradoxical as it may seem, *increasing* penalties can sometimes *reduce* the law's deterrent effect, as long as the extent of the increase itself responds (through changes in the multiplier) to each change in a defendant's behavior.

2. *The Effect of a Constant Multiplier*

The preceding analysis also shows why the multiplier, to have this effect, *must* be recalculated in every case, so that it changes with each defendant's actual behavior. If the multiplier is instead fixed at some constant level, it no longer leads to optimal deterrence.

The toxic waste example can illustrate this point as well. In that example, the average probability of punishment was somewhere between 10% and 25%. Suppose, for the sake of concreteness, that this average was exactly 20% (as would be the case if one-third of all defendants took the extra precautions but two-thirds did not). If the damage multiplier were set at a constant based on this average probability of punishment, all defendants would be given a multiplier of five. This means that defendants who did not take extra precautions would have to pay $30 million in damages (5 × $6,000,000) each time they were caught, while defendants who did take extra precautions would have to pay only $25 million in damages (5 × $5,000,000).

Unfortunately, these penalties will create too much deterrence. Defendants who do not take extra precautions will now face expected liabilities of $75,000 (.01 × .25 × $30,000,000), while defend-

---

17. For a mathematical proof, see Craswell & Calfee, *supra* note 13, at 292 n.18.

ants who do take extra precautions will face expected liabilities of $25,000 (.01 × .10 × $25,000,000). Thus, defendants who take the extra precautions will be rewarded with a $50,000 reduction in their expected liability ($75,000 − $25,000). But the true social savings from these precautions are still only $10,000 ($60,000 − $50,000), so this $50,000 reward is too large and will lead to overdeterrence.

Notice, too, that the problem is not just that a constant multiplier will not be sufficiently fine-tuned, in the sense that any constant will be too high for some defendants and too low for others.[18] To the contrary, overdeterrence can still be a problem even if the constant multiplier is set at a level that is not "too high" (according to the traditional principle) for any individual defendant. In the example just considered, the highest probability of punishment was 25%, for defendants who did not take the extra precautions. If we follow the traditional multiplier analysis, a multiplier of four would be just right for those defendants and too low for all others (those for whom the probability of punishment was only 10%). And if a multiplier of four is just right for some and too low for others, it might seem as though that multiplier would have to yield too little deterrence on balance.

In fact, though, a multiplier of four will still produce too much deterrence *if it is applied as a constant.* With a constant multiplier of four, defendants who do not take the precautions will have to pay $24 million in damages (4 × $6,000,000) each time they are caught, while defendants who do take the precautions will have to pay $20 million (4 × $5,000,000). This means that defendants who do not take the precautions now face an expected liability of $60,000 (.01 × .25 × $24,000,000), while defendants who do take the precautions face an expected liability of $20,000 (.01 × .10 × $20,000,000). As a result, defendants can now save $40,000 in expected liability ($60,000 − $20,000) by taking the precautions, and this is still far more than the $10,000 social value of the precautions. Thus, a constant multiplier can still lead to overdeterrence even when it seems "just right" for some defendants (according to the traditional multiplier principle) and "too low" for all others.

While this conclusion, too, may seem paradoxical, it follows directly from the earlier analysis. The reason is that constant multipliers — even ones set at relatively low levels — eliminate the "tax" referred to earlier. As we have seen, a case-by-case multiplier taxes

---

18. Polinsky and Shavell emphasize this objection to constant multipliers. *See* Polinsky & Shavell, *supra* note 3, at 893.

defendants whose behavior improves by giving them a higher multiplier (to correspond to their lower probability of punishment). A constant multiplier eliminates this "tax," but it still increases the overall deterrent effect by increasing all penalties across the board. By thus increasing the size of all penalties without any offsetting "tax," constant multipliers can easily produce too much deterrence, just as in the example above. In fact, to achieve the optimal level of deterrence a constant multiplier will usually have to be set *below* the traditional multiplier principle (as the following Part will discuss).

II. Alternative Routes to Optimal Deterrence

As noted in the introduction,[19] most legal systems do not use multipliers that are calculated case by case. Instead, they use multipliers set at the same level for all defendants, or fines set at the same level for all defendants, or compensatory damages with no multipliers at all. Accordingly, it is important to understand how penalties should be set to achieve optimal deterrence under each of these alternative approaches — and how those penalties would differ from the more familiar, case-by-case multiplier principle.

In this Part, I show that constant multipliers can achieve optimal deterrence if they are set at some suitably lower level (lower, that is, than the traditional multiplier principle). Constant fines can also achieve optimal deterrence, usually by being set below the traditional multiplier principle, but sometimes by being set above that level. Optimal deterrence might also be achieved by reforms that do not involve any multiplier at all — for example, by caps on the highest possible damage awards, or by changing the substantive legal standard. Thus, while the traditional multiplier principle may perhaps be *sufficient* for optimal deterrence, it is not at all *necessary*.

A. *The Optimal Constant Multiplier*

The toxic waste example has already shown how a constant multiplier set at 1.0 could still achieve optimal deterrence. In that example, extra precautions reduced the probability of punishment from 25% to 10%, and that effect by itself was enough to reduce defendants' expected liability by $10,000.[20] Since $10,000 was also the social benefit produced by the extra precautions, this meant

---

19. *See supra* text accompanying note 7.
20. *See supra* text following note 14.

that defendants had the optimal incentives even with no multiplier at all. But using no multiplier at all, and merely awarding compensatory damages, is equivalent to multiplying the damages by one in every case. And if enforcement is imperfect, a multiplier of one will be below the level recommended by the traditional multiplier principle, which in this example would have been between four and ten (because the probability of punishment ranged from 25% to 10%). In short, we have already seen one example where optimal deterrence was achieved using a constant multiplier below the optimal case-by-case multiplier.

Of course, there is no reason to think that the optimal constant multiplier will always equal one. Indeed, if we change the example slightly, so extra precautions reduce the probability of punishment from 25% to only 20%, the constant multiplier will then have to be larger. In this revised example, defendants who take no extra precautions will face expected liability of $15,000 (.01 × .25 × $6,000,000) if no multiplier is used, but defendants who do take the precautions will face expected liability of $10,000 (.01 × .20 × $5,000,000). Defendants who take the precautions will thus be rewarded with only a $5,000 reduction in expected liability ($15,000 − $10,000), and this reward is less than the social value of the precautions, which is still $10,000 ($60,000 − $50,000). In this case, then, defendants will have too little incentive to take the extra precautions. To correct this problem, the multiplier must be greater than one in order to magnify the reward for taking extra precautions. In this example, the optimal constant multiplier happens to be exactly two.[21]

To be sure, a constant multiplier of two is still less than the optimal *case-by-case* multiplier under the traditional multiplier principle, for that multiplier would now be somewhere between four and five (based on a probability of punishment ranging from 25% to 20%). In this new example, though, the optimal constant multiplier is not as far below the lower end of this range as it was in the original example, where the optimal constant multiplier was only one. This is because the new example features a probability of punishment that is less responsive to improvements in defendants' behav-

---

21. The full calculations are as follows. With a constant multiplier of two, defendants who do not take extra precautions will have to pay $12 million when they are caught (2 × $6,000,000), while defendants who do take extra precautions will have to pay $10 million (2 × $5,000,000). This gives defendants an expected liability of $30,000 if they do not take precautions (.01 × .25 × $12,000,000), as compared to an expected liability of $20,000 (.01 × .20 × $10,000,000) if they do take precautions. The savings in expected liability is therefore $10,000 ($30,000 − $20,000), which again equals the social benefit from the precautions.

ior: taking extra precautions in the new example brings just a slight reduction in the probability of punishment, from 25% to 20%. The lesson here is that the rate at which the probability of punishment declines is a key factor in determining how low a constant multiplier should be.

Indeed, if the probability of punishment were to decline even faster with each improvement in a defendant's behavior, the optimal constant multiplier would be even smaller — and could even be less than one, implying that even compensatory damages would be too high. To see this, let us alter the example so that extra precautions reduce the probability of punishment from 25% all the way down to 5%. If no multiplier is used, defendants who do not take the precautions would still face expected liabilities of $15,000 (just as before), but defendants who do take the precautions would face expected liabilities of $2,500 (.01 × .05 × $5,000,000). This gives each defendant who takes the precautions a $12,500 reduction in its expected liability ($15,000 − $2,500). But this $12,500 reward is greater than the social value of those precautions (which is still $10,000), so these penalties will lead to overdeterrence. To dampen defendants' incentives and achieve optimal deterrence, any constant multiplier would have to be set at some value less than one. In this example, the optimal constant multiplier happens to be 0.80, which is equivalent to holding defendants liable for only 80% of the harm caused by their behavior. Even though defendants are clearly paying less than the full social costs of their behavior, their incentives are still socially optimal.[22]

Of course, a multiplier of 0.80 is *much* less than the traditional case-by-case multiplier, which in this example would be somewhere between four and twenty (based on probabilities of punishment between 25% and 5%). In general, the more the probability of punishment declines with any improvement in a defendant's behavior, the greater will be the divergence between the optimal constant multiplier and the optimal case-by-case multiplier. As long as the probability of punishment declines at all, though, the optimal constant multiplier will almost always be less than the optimal case-by-case multiplier. More precisely, the optimal constant multiplier will

---

22. With a constant multiplier of 0.80, defendants who do not take extra precautions will have to pay only $4.8 million each time they are caught (.80 × $6,000,000), while defendants who do take extra precautions will have to pay $4 million (.80 × $5,000,000). This gives defendants an expected liability of $12,000 if they do not take the extra precautions (.01 × .25 × $4,800,000), compared to an expected liability of only $2,000 (.01 × .05 × $4,000,000) if they do take extra precautions. This makes the savings in expected liability again equal to $10,000 ($12,000 − $2,000), which is still the social value of the extra precautions.

never be greater than the optimal case-by-case multiplier, and will usually be less.[23]

The reason for this follows from the analysis given earlier. We have already seen that the traditional multiplier will create optimal incentives *if* that multiplier is adjusted on a case-by-case basis to reflect each improvement in a defendant's behavior. We have also seen that this case-by-case adjustment operates as a kind of a tax, which penalizes defendants whose behavior improves by making them face a larger multiplier (to correspond to their reduced probability of punishment). Making the multiplier a constant takes away this tax on improvements and thus strengthens the incentives to make such improvements. But if the case-by-case multiplier (including the tax on improvements) had been creating exactly the optimal incentives, any change that strengthens those incentives will make them too strong, leading to net overdeterrence. The only way to correct this problem (while still employing a constant multiplier) is to reduce the size of the multiplier, reducing the deterrent effect. This is why the optimal constant multiplier will usually be less than, and can never be greater than, the optimal case-by-case multiplier.

The size of the optimal constant multiplier will also depend on the expected damage award facing a defendant whose behavior is at or very near the socially optimal level. In the original toxic waste example, defendants who took the socially efficient level of precautions expected that a leak would occur in 1% of all cases, and that they would then face a 10% chance of having to pay damages of $5 million (before any multiplier was applied). This prospect left these defendants facing an expected liability of $50,000 (.01 × .10 × $5,000,000). But if that figure were either higher or lower, the law's deterrent effect would be altered, requiring a corresponding adjustment in the size of the constant multiplier.

To see this, let us change the original example to make the toxic waste slightly more hazardous, so that a leak now causes $8 million in damages (rather than $6 million) if no extra precautions are taken, and $7 million in damages (rather than $5 million) if the defendant does take extra precautions. Since both figures have increased by the same amount, the social efficiency of the precautions will be just what it was before: the precautions will still reduce

---

23. For a mathematical proof, see Craswell & Calfee, *supra* note 13, at 297 & n.25.

expected social costs by $10,000.[24] However, this change in the ab-
solute levels of harm reduces the optimal constant multiplier. In
the earlier example where the damage caused by a leak was either
$6 million or $5 million, and extra precautions reduced the
probability of punishment from 25% to only 20%, we saw that the
optimal constant multiplier was exactly two.[25] If we use these same
probabilities in this new example, though, the optimal constant
multiplier falls to 1.67.[26]

The reason why the absolute size of the damages matter is that
this affects the *significance* (to the defendant) of the rate at which
the probability of punishment declines. In both of these examples,
the extra precautions reduced the probability of punishment by the
same amount, from 25% to 20%. In the first example, though, this
represented a reduction of five percentage points in the probability
of paying $6 or $7 million (before any multiplier was applied), while
in the second case it represented a reduction of five percentage
points in the probability of having to pay $7 or $8 million. Obvi-
ously, the second reduction is more valuable to defendants than the
first, so the second offers stronger incentives for defendants to im-
prove their behavior. As a result, any constant multiplier that was
optimal in the first case will be too strong in the second case, and
therefore will have to be reduced. This is just what happened in the
example itself: the optimal constant multiplier fell from 2.0 to 1.67.
This shows that the optimal level for a constant multiplier depends
both on the rate at which the probability of punishment declines
and on the absolute level of damages that are expected when de-
fendants behave optimally.

Of course, the absolute level of these damages depends partly
on the technology of the defendant's activity — for example, what
kind of waste is involved, and just how toxic is it? But the level of
damages may also be affected by other legal rules, especially the
rules governing causation. For example, if a defendant who failed

---

24. In this new example, the expected social costs of the defendants' behavior are $80,000
(.01 × $8,000,000) if they do not take the extra precautions, and $70,000 (.01 × $7,000,000) if
they do. The social value of the extra precautions is therefore $10,000 ($80,000 − $70,000).

25. *See supra* text accompanying note 21.

26. With a constant multiplier of 1.67, defendants who do not take the extra precautions
will have to pay $13.33 million each time they are caught (1.67 × $8,000,000), while defend-
ants who do take the precautions will have to pay $11.67 million (1.6 × $7,000,000). This
gives defendants an expected liability of $33,333 if they do not take the precautions (.01 × .25
× $13,333,333), compared to an expected liability of $23,333 (.01 × .20 × $11,666,667) if they
do, so defendants can save $10,000 in expected liability by taking the precautions ($33,333 −
$23,333). This $10,000 savings exactly equals the precautions' social value, so defendants'
incentives will be socially optimal.

to take the appropriate precautions can prove that the same leak would have occurred even if all appropriate precautions had been taken, and that the same victims would still have suffered $7 million in damages (rather than the $8 million in damages they actually suffered), the rules of causation will sometimes reduce such a defendant's liability to only $1 million, reflecting the *incremental* harm attributable to its failure to take proper care.[27] If damages are consistently reduced in this way, the optimal constant multiplier will be larger, because a larger multiplier will be needed to achieve the same deterrent effect. In this example, the optimal constant multiplier would equal four.[28] Of course, a multiplier of four would also satisfy the traditional multiplier principle, based on a probability of punishment of 25%.

This conclusion holds more generally: whenever a causation rule of this sort is applied without error, the optimal constant multiplier will always be just as great as (though no greater than) the optimal case-by-case multiplier.[29] If such a causation rule is perfectly applied, a defendant who takes the socially optimal level of care should never have to pay any damages at all. Such a defendant will therefore be unconcerned with the probability that it might be held liable, and will thus be unaffected by the rate at which that probability might change if the defendant were to deviate slightly from the socially efficient level of care. But if the deterrent effect will be the same regardless of the rate at which that probability changes, this means that the deterrent effect will be the same as it would have been in a world where the probability of punishment did not change at all. And in a world where the probability of punishment did not change at all, there would be no difference between the optimal constant multiplier and the optimal case-by-case multiplier, because in such a world no case-by-case adjustments would ever be required. This is why a perfectly applied causation rule sup-

---

27. *See* sources cited *infra* note 30.

28. The full calculations are as follows. Defendants who do not take extra precautions will be charged with incremental damages of $1 million, for the reasons discussed in the text. A multiplier of four means that these defendants will have to pay $4 million each time they are caught (4 × $1,000,000), leaving them with an expected liability of $10,000 (.01 × .25 × $4,000,000). However, defendants who do take extra precautions will not be charged with any losses at all (under this incremental damage rule), so their expected liability will be zero. This means that defendants can save $10,000 in expected liability ($10,000 minus 0) by taking the precautions — which is just what the socially optimal incentives require.

29. For mathematical analyses of this relationship, see Craswell & Calfee, *supra* note 13, at 295-97; Marcel Kahan, *Causation and Incentives to Take Care Under the Negligence Rule,* 18 J. LEGAL STUD. 427, 437-39 (1989). For a more qualitative discussion, see Mark F. Grady, *Punitive Damages and State of Mind: A Positive Economic Theory,* 40 ALA. L. REV. 1197, 1201-09 (1989).

plies the limiting case in which there is no effective difference between the optimal constant multiplier and the optimal case-by-case multiplier.

It should be stressed, however, that this is indeed a limiting case. If the causation rule is not applied perfectly, but instead involves any risk of error, then even defendants who take socially efficient precautions will always face at least some expected liability, so they will care about the rate at which the probability of punishment changes. More important, the rules of causation often are not applied in precisely the way that the earlier example assumed.[30] For instance, if the efficient level of care would have led to a different *kind* of accident, or to an accident whose losses would have fallen on different victims, defendants' liability generally is not limited to the incremental losses above and beyond those that would have been imposed if they had taken the efficient level of care. The same is true when the expected losses (at any level of care) are entirely probabilistic, so the most that defendants can show is that a similar accident *might* have happened even if they had taken the efficient level of care. In such a case, defendants' liability usually is not reduced by the statistical or expected value of the losses that might have occurred even if they had behaved efficiently. Obviously, there will also be no reduction of their liability in any regime of strict liability, where defendants are held liable for the full costs regardless of whether they behaved efficiently. And if (for any of these reasons) defendants' liability is not limited to the purely incremental losses, the optimal constant multiplier will then be *less* than the multiplier principle. Often it will be less by a considerable margin, as in most of the examples considered above.

To summarize, the optimal constant multiplier can never exceed the optimal case-by-case multiplier, and only rarely will it equal that level. Instead, in most cases the optimal constant multiplier will be less than the optimal case-by-case multiplier, meaning that a constant multiplier can achieve optimal levels of deterrence without satisfying the multiplier principle. Moreover, the extent to which the optimal constant multiplier falls below the multiplier principle depends on two factors: (a) the rate at which the probability of punishment falls with improvements in a defendant's behavior, and (b) the absolute level of expected damages facing a defendant who

---

30. For discussions of the actual legal rules, see Cooter, *supra* note 3; Grady, *supra* note 29; David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L. REV. 849, 862-66 (1984). I discuss the deterrent effect of other causation rules at more length *infra* in section III.D.

behaves in the socially optimal way. Only when one of these factors is eliminated — that is, when the probability of punishment never changes at all, or when defendants who behave optimally can be assured of paying no damages — only then will the optimal constant multiplier be the same as the optimal case-by-case multiplier.

B. *The Optimal Constant Fine*

Another route to optimal deterrence involves the use of a constant fine, in which every defendant who is found liable pays the same total amount. For example, traffic laws often impose the same fine for every violation of a certain type, regardless of the actual damage caused by the violation. A constant fine is slightly different from the constant multiplier discussed in the preceding section, for a constant multiplier allows the total penalty (the multiplier times the harm caused) to vary from defendant to defendant. By contrast, a constant fine keeps the total penalty the same for all defendants.

The significance of this distinction is that a constant multiplier gives defendants two reasons to improve their behavior, while a constant fine gives them only one. Under a constant multiplier, defendants who improve their behavior will be rewarded with a lower probability of being punished, as emphasized in the preceding sections of this article. But defendants who improve their behavior will also be rewarded with a reduction in the damages to which any multiplier will be applied, because improved behavior should (if it is truly an "improvement") inflict lower social costs. Indeed, both of these effects can be seen in the toxic waste example discussed earlier, because in every example there were two reasons why defendants who took the extra precautions faced a lower expected liability. Defendants who took extra precautions benefited from a reduction in the probability of punishment, from 25% to some lower value; but they also benefited from a reduction in the social losses to which any multiplier would be applied, from $6 million down to $5 million (or from $8 million down to $7 million, depending on the example).

The key point for present purposes is that a constant fine eliminates the second of these effects. If the law uses a constant fine, defendants who improve their behavior may still be rewarded with a reduction in the probability of being penalized at all, but this will be their *only* reward: the amount they will have to pay if they are penalized will stay constant. It follows that the deterrent effects of a constant fine will always be less than the deterrent effects of an

equivalent constant multiplier. To offset this reduction in deterrent effect, the optimal constant fine will have to be larger than the optimal constant multiplier.[31]

This conclusion, too, is easy to see using a concrete example. In the toxic waste scenario, we calculated earlier that if extra precautions reduced the probability of punishment from 25% to only 20%, the optimal constant multiplier would be exactly two.[32] Using such a multiplier, defendants who do not take the extra precautions will have to pay $12 million each time they are caught ($2 \times \$6,000,000$), while defendants who do take the precautions will have to pay $10 million ($2 \times \$5,000,000$). If the legal system instead uses a constant fine, however, optimal deterrence will be achieved only if the fine is set at $20 million.[33] Obviously, this $20 million fine is larger than the $12 million and $10 million penalties that would have been optimal under a constant multiplier regime.

At the same time, a $20 million fine is below the penalty called for by the traditional multiplier principle. In this example, defendants who do not take extra precautions still face a 25% probability of punishment, so the traditional multiplier principle would require a multiplier of four, resulting in a total penalty of $24 million ($4 \times \$6,000,000$). Similarly, defendants who do take the extra precautions face only a 20% probability of punishment in this example, so for these defendants the multiplier principle requires a multiplier of five and a total penalty of $25 million ($5 \times \$5,000,000$). As we have seen, the optimal constant fine is only $20 million, which is less than either $24 million or $25 million.

However, the optimal constant fine may not always be less than the optimal case-by-case multiplier. The optimal constant fine will always be greater than the optimal *constant* multiplier, and the optimal *case-by-case* multiplier will also be greater than (or equal to) the optimal constant multiplier. However, the relationship between the optimal constant fine and the optimal case-by-case multiplier is more difficult to characterize. The most that can be said is that the optimal constant fine will be less than the optimal case-by-case mul-

---

31. For a mathematical proof, compare Craswell & Calfee, *supra* note 13, at 294 (deriving the value of the optimal constant multiplier) with *id.* at 297 (deriving the value of the optimal constant fine).

32. *See supra* text accompanying note 21.

33. With a constant fine set at $20 million, defendants will face an expected liability of $50,000 ($.01 \times .25 \times \$20,000,000$) if they do not take the extra precautions, but an expected liability of only $40,000 ($.01 \times .20 \times \$20,000,000$) if they do take the extra precautions. The savings in expected liability therefore equals $10,000 ($50,000 − $40,000), which is also the social value of the extra precautions.

tiplier whenever the probability of punishment is relatively responsive to improvements in a defendant's behavior. In particular, if the probability of punishment declines quite rapidly with each improvement in a defendant's behavior — so much that the defendant's incentives would be optimal, or not too suboptimal, with no multiplier at all — then the optimal constant fine will always be less than the optimal case-by-case multiplier.[34] But if the probability of punishment is somewhat less responsive to improvements in a defendant's behavior, the deterrent effect of a constant fine will then be reduced, which means that the absolute level of the fine will have to be raised in order to keep deterrence optimal. At some point, if the probability of punishment is extremely *un*responsive to improvements in a defendant's behavior, the optimal constant fine could even surpass the optimal case-by-case multiplier. In short, the *rate* at which the probability of punishment responds is again a key variable — in this case, the variable that determines the size of the optimal fine.

C. *Adjustments to the Substantive Standard*

Another method of achieving optimal deterrence, at least under some legal regimes, is to raise or lower the substantive threshold for liability.[35] For example, suppose that the optimal risk of toxic leaks is exactly 1% — but suppose that the net incentives with no multiplier at all favor underdeterrence, so firms choose instead to permit a 1.5% risk. One way of correcting this underdeterrence is to make the substantive standard more strict, to permit no more than (say) a 0.75% risk of toxic leaks. At this lower substantive standard, any firm that continued to permit a 1.5% risk would now face a higher probability of detection and/or conviction, as its behavior would more obviously violate the new substantive standard. This increase in the probability of detection or conviction would give firms a stronger reason to reduce their levels of risk, thus increasing the law's deterrent effect and moving firms back toward the socially optimal risk of 1%. If firms still permitted a risk of more than 1%,

---

34. A mathematical proof is available from the author upon request.

35. The possibility of changing the substantive standard to adjust for imperfect enforcement is discussed in Calfee & Craswell, *supra* note 13, at 997-99; and in Jason S. Johnston, *Punitive Liability: A New Paradigm of Efficiency in Efficiency in Tort Law,* 87 COLUM. L. REV. 1385 (1987) [hereinafter Johnston, *Punitive Liability*]. Note, too, that the substantive standard could also be raised or lowered indirectly by (for example) changing the evidentiary rules or the burdens of proof. For discussions of this approach, see Craswell & Calfee, *supra* note 13, at 290-92; Jason S. Johnston, *Bayesian Fact-Finding and Efficiency: Toward an Economic Theory of Liability Under Uncertainty,* 61 S. CAL. L. REV. 137 (1987).

then the substantive standard could be lowered even further, until the resulting probability of punishment gave defendants an incentive to choose exactly the 1% level.

Of course, if the net incentives without any multiplier instead favored overdeterrence, the appropriate response would require relaxing rather than tightening the substantive standard. Either way, though, we could still achieve optimal deterrence without satisfying the traditional multiplier principle. That is, there should be some adjustment to the substantive legal standard that can achieve optimal deterrence even if the probability of punishment remains less than one, and even if the fines or damage awards remain at purely compensatory levels (and thus fail to satisfy the traditional multiplier principle). This shows, again, that satisfaction of the multiplier principle is not always necessary for optimal deterrence.

In addition, if the substantive standard were relaxed sufficiently, optimal deterrence might also be achieved with penalties *above* the multiplier principle. This can be achieved if the substantive standard is defined in such a way that it could never be applied, even through judicial error, to a defendant who had behaved efficiently. Penalties of this sort — referred to by Robert Cooter as "sanctions"[36] — thus require a triggering test such as "egregious behavior," "gross negligence" or "reckless disregard for human safety," to ensure that they are never applied to efficient behavior.[37] If the probability of the higher penalty is literally zero for any defendant who behaves efficiently, there will be no danger of inducing too many precautions, because a defendant who is already behaving efficiently will (by hypothesis) face no risk of having to pay the higher penalty. As further increases in the penalty should have no effect on these defendants, the penalty could (in theory) be increased indefinitely without deterring beyond the optimal level.[38]

Since this approach permits the imposition of penalties that are *larger* than the multiplier principle (rather than penalties that are smaller, as under most of the other alternatives discussed here), it

---

36. *See* Cooter, *supra* note 9, *passim.* By contrast, penalties calculated according to the multiplier principle are referred to by Cooter as "prices," because they achieve deterrence by making defendants pay the full social costs of their behavior (no more, and no less) as the "price" of that behavior. *See id.* at 1528.

37. *Compare* Polinsky & Shavell, *supra* note 3, at 905-08 (arguing that the reprehensibility of the defendant's conduct should be irrelevant when damages are calculated by the multiplier principle, because under the multiplier principle the only relevant factors are the amount of the harm and the probability of punishment).

38. *See* Cooter, *supra* note 9, at 1524-27.

raises distinct issues that I will not address further.[39] Still, the possibility of optimal deterrence with even larger penalties does reinforce the general lesson of this article — i.e., that the multiplier principle is sufficient but not necessary for optimal deterrence.

D. *Caps on Damages, and Other Possible Adjustments*

Finally, optimal deterrence can also be achieved by leaving the substantive standard unchanged, but starting with compensatory damages and adding or subtracting (rather than multiplying) a constant amount to each award. The size of the optimal addition or subtraction is harder to express mathematically,[40] but its general character depends on the same factors that characterize the optimal constant multiplier. That is, whenever the optimal constant multiplier would be greater than one (implying that compensatory damages need to be increased), optimal deterrence can be achieved by *adding* some amount to each compensatory award. Similarly, whenever the optimal constant multiplier would be less than one, optimal deterrence can be achieved by *subtracting* some amount from each compensatory award. Either way, the size of the amount to be added or subtracted will be greater or less depending on just how much the optimal constant multiplier would have been greater or lesser than one.

For a similar reason, optimal deterrence could also be achieved by placing a cap on the maximum size of the award — a move that several states have enacted or considered in recent years.[41] Placing a cap on the largest possible award is similar in many respects to subtracting something from the *expected value* of the damage award. That is, if defendants face uncertainty about how damages will be measured (in addition to uncertainty about whether they will be liable at all), an upper limit on the maximum possible award will cut off the upper end of that distribution. This will reduce the average award, thereby reducing the deterrent effect.

To be sure, these reductions would make little sense if optimal deterrence were being pursued by means of the traditional multiplier principle. Under the traditional multiplier principle, we must be prepared to increase damage awards to almost any level, espe-

---

39. For further discussions of this approach as it might apply to punitive damages, see Hylton, *supra* note 1; and Johnston, *Punitive Liability, supra* note 35.

40. For a mathematical model, see A. Mitchell Polinsky & Daniel Rubinfeld, *The Welfare Implications of Costly Litigation for the Level of Liability,* 17 J. LEGAL STUD. 151, 159-60, 163 app. (1988).

41. For citations to the various legislative proposals, see *BMW of North America, Inc. v. Gore,* 517 U.S. 559 app. (1996) (Ginsburg, J., dissenting).

cially when the probability of punishment is relatively small (so that the optimal case-by-case multiplier is large). As Polinsky and Shavell have recently argued, caps on the largest possible award could interfere with this goal, leading to less than optimal deterrence.[42]

If optimal deterrence is instead being pursued by some other route, however, it is harder to condemn such caps a priori. For example, if the law instead aims to achieve optimal deterrence by approximating the optimal constant multiplier, we have already seen that this typically requires awards that are lower than those called for by the traditional multiplier principle.[43] Viewed with this goal in mind, a cap on the highest possible awards could be seen as moving the average or expected award in just the right direction. In other words, a reform that seems obviously unsound from the standpoint of one route to optimal deterrence could be perfectly sound from the standpoint of another.

Of course, to say that caps on damage awards *could* be justified is not to say that any particular proposal is a good idea. Calculating the exact size of the optimal adjustment will often be difficult, in which case it may be hard to say whether any particular cap goes too far or not far enough. Indeed, calculation problems arise under each of the alternatives — as, of course, they arise to some extent under the multiplier principle itself. These calculation problems, and other issues surrounding the administration of each set of remedies, will be discussed below in Part IV.

For now, my point is simply that the existence of alternative routes must first be recognized before we can even begin to consider which is easiest to calculate, or which is best on any other grounds. If we instead assume that the traditional multiplier represents the *only* means of achieving optimal deterrence — as much of the legal literature implicitly assumes today — we will never reach the question of which route is on balance superior, because only one route will even be considered. This accounts for my goal in the first two parts of this article: to show that the multiplier principle is sufficient, but not necessary, for optimal deterrence.

III. THE RESPONSIVENESS OF THE EXPECTED PUNISHMENT

Part IV will discuss the actual advantages and disadvantages of the alternative routes to optimal deterrence. Before beginning that

---

42. *See* Polinsky & Shavell, *supra* note 3, at 900.

43. *See supra* section II.A.

discussion, though, this Part will say more about just when (and why) the expected punishment might respond to changes in a defendant's behavior. As we have already seen, the rate at which the probability of punishment declines is a key factor in determining just how much lower a constant multiplier or a constant fine should be to achieve optimal deterrence. Thus, the responsiveness of the probability of punishment will play a key role in the comparisons in Part IV of this article.

To see how the expected punishment responds to changes in a defendant's behavior, it will be helpful to isolate several components. Before a defendant can be punished, several events must occur: (a) the defendant's offense must be detected by someone; (b) a plaintiff, or a government prosecutor, must decide to file suit; (c) the court, or some other adjudicative body, must find the defendant liable; and (d) the appropriate fine or measure of damages must be assessed. Since all four events must occur before the defendant will be punished, the expected penalty depends on the combined probability of all four — that is, on the probability of detection *times* the probability of prosecution *times* the probability of a finding of liability *times* the expected fine or damage award. Each of these component probabilities will be discussed below.

A. *The Probability of Detection*

In some contexts, the probability of detecting any given offense will be completely *un*responsive to the defendant's behavior. For example, whether a speeding violation is detected typically depends entirely on whether a police officer is present. Moreover, the probability of a police officer being present is normally unaffected by how fast the driver drives, and thus is unresponsive to any improvement in the driver's behavior. In such a case, where the probability of punishment depends *only* on the probability the offense is detected, that probability will not respond to any improvements in a defendant's behavior.[44]

In some cases, though, even the probability of detection may respond to the seriousness of the defendant's behavior. For example, if it is easy to detect that toxic waste has been leaked but hard to detect who was responsible for the leak, the investigating authorities may spend more time tracking down large leaks than tracking

---

44. Interestingly, many analyses of the traditional multiplier principle assumed the only reason for imperfect enforcement was a low probability of *detection*. This is clearly true of the earliest writers such as Beccaria and Bentham, cited *supra* in note 1. It is also true of most of the examples in the recent article by Polinsky & Shavell, *supra* note 3.

down small ones. If so, then manufacturers who permit the risk of large leaks will effectively face a higher probability of detection than will manufacturers who take more precautions and whose leaks are likely to be small.[45] In that event, even the probability of detection will respond to changes in a defendant's behavior.

More important, even when the probability of *detection* is unresponsive to the defendant's behavior, the final probability of punishment will also depend on the probability of prosecution and conviction. As the following subsections demonstrate, these probabilities are almost always responsive to changes in a defendant's behavior.

B. *The Probability of Litigation*

1. *The Probability of Prosecution in Systems of Public Enforcement*

Public prosecutors have limited resources, so the probability that a detected offense will be prosecuted is typically less than 100%. While this will reduce the law's deterrent effect, the important point (for present purposes) is that the lower probability will not normally be the same for all defendants, and thus will not be a *constant.* Prosecutors usually have discretion to decide which cases they pursue, and they often try to concentrate their limited prosecutorial resources on the most serious offenses. If so, defendants who improve their behavior will be rewarded with a lower probability of prosecution.

Indeed, it is hard to imagine systems of public prosecution in which this is not true. Police officers, for example, are more likely to stop a speeder who is thirty miles per hour over the speed limit than one who is only five miles per hour over; and pollution authorities are more likely to seek penalties for a huge leak of toxic waste than they are for a small one. Some of this responsiveness may be because it will be easier to win a case against the larger offender than against the marginal one, thus implicating the probability of success at trial (to be discussed in the following subsection). But even if conviction were certain, one would still expect prosecutors to devote the most resources to prosecuting those defendants

---

45. On the other hand, manufacturers who permit large leaks (or commit other serious offenses) may also take greater pains to try to conceal their offense, thereby reducing the probability of punishment. For a formal model of this effect — but one assuming that the probability of punishment does not also vary with the egregiousness of the defendant's behavior, and considering deterrence only by means of a constant fine — see Arun S. Malik, *Avoidance, Screening and Optimum Enforcement,* 21 RAND J. ECON. 341 (1990).

whose violations were the most serious. In any system of public enforcement, then, we should expect the probability of prosecution to decline with improvements in a defendant's behavior. The exact rate of the decline will depend on the extent of the prosecutorial discretion, and on the political and other factors that influence how that discretion is exercised.[46]

2. *The Probability of Litigation Under Private Enforcement*

If enforcement instead depends on private lawsuits, there may be no public-spirited reasons for plaintiffs to concentrate their efforts on the most serious offenses. There may, however, be private motives for them to do so. In particular, if the most serious offenses are also the ones that cause the most damage to their victims, then the victims of the most serious offenses will expect the largest damage awards. All else equal, then, victims of the most serious offenses will also be the ones most likely to find it worthwhile to hire a lawyer and sue. If so, then the most serious offenders will again face the highest probability of suit.[47]

Of course, some kinds of improvements may not have this effect, if they reduce the probability of an accident but do not alter the injury that results if an accident occurs.[48] In that event, since the improvements will not affect the amount that is likely to be recoverable at trial, they also will not affect the probability of litigation. In many cases, though, extra precautions *will* reduce the injury likely to be caused by an accident, and thus will also reduce the amount likely to be recoverable at trial. For example, driving at a lower speed typically reduces the magnitude of any injuries likely to be suffered, as well as reducing the likelihood of any accident at all. Similarly, improvements in a manufacturer's system of quality control often will reduce the average severity of any defects that happen to slip through the system, as well as reducing the likelihood of any defect at all. Whenever improvements in a defendant's

---

46. This suggests that a fruitful line of research might investigate the extent to which, under different systems of punishments, it would be optimal for prosecutors to *try* to make the probability of litigation more or less responsive to changes in a defendant's behavior. I am grateful to Howard Chang for this suggestion.

47. For a mathematical model of just such a system, see Polinsky & Rubinfeld, *supra* note 40. The same argument is also made in A. Mitchell Polinsky, *Are Punitive Damages Really Insignificant, Predictable, and Rational? A Comment on Eisenberg et al.*, 26 J. LEGAL STUD. 663, 675-76 (1997).

48. For a mathematical model employing this alternative assumption, see Keith N. Hylton, *The Influence of Litigation Costs on Deterrence Under Strict Liability and Under Negligence*, 10 INTL. REV. L. & ECON. 161 (1991).

care do reduce the damages victims will suffer, they should also reduce the probability that any victim will bother to sue.

I should note that, when extra precautions do reduce the probability that a victim will sue, there is an additional reason why optimal penalties *could* depart from the traditional multiplier principle. My analysis so far has interpreted the socially efficient level of precautions as the level that minimizes the total accident costs plus the total cost of precautions. However, if litigation is costly then it may be more plausible to define the efficient level of precautions as the level that minimizes the sum of accident costs, precaution costs, *and* litigation costs.[49] If so, reducing the size of the penalty may produce additional benefits by reducing the number of suits that are filed, thus reducing total litigation costs. On the other hand, if the reduced penalty led to fewer precautions and hence a larger number of accidents, then the total number of suits might actually rise (even if the probability that *any given accident* would lead to a suit had declined). As a result, it is hard to say which way penalties should be adjusted to optimize the effect on the total amount of litigation.[50]

Nevertheless, even if the effect on the absolute level of litigation is indeterminate, it remains true that the probability of litigation facing any individual defendant will usually be somewhat responsive to that defendant's behavior. This is enough to give rise to all of the effects analyzed in earlier sections of this article. That is, if the probability of litigation is itself responsive to defendants' behavior, then the overall probability of punishment will also be responsive, because the probability of litigation is simply one component of the overall probability of punishment. And if the overall probability of punishment is responsive to defendants' behavior, then the optimal constant multiplier will again diverge from the optimal case-by-case multiplier, and each of these will diverge from the optimal fine (or the optimal adjustment in any substantive legal standard). In other words, if the probability of litigation responds to improvements in a defendant's behavior, satisfaction of the mul-

---

49. This is the goal assumed in *id.* at 164; and by Polinsky & Rubinfeld, *supra* note 40. For a discussion of these (and other) goals more generally, see GUIDO CALABRESI, THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS 26-33 (1970).

50. As Polinsky and Rubinfeld conclude, "[t]he optimal adjustment to compensatory damages takes both of these considerations into account, and may be positive or negative." Polinsky & Rubinfeld, *supra* note 40, at 153. For a qualitatively similar analysis, see David D. Friedman, *An Economic Analysis of Punitive Damages*, 40 ALA. L. REV. 1125, 1133-34 (1989).

Litigation costs could also rise if the amount spent on each suit increased, even if the total number of suits did not. This possibility will be discussed *infra* in section IV.F.

tiplier principle will again be sufficient but not necessary for optimal deterrence.

C. *The Probability of Conviction*

The effects just discussed will be even more pronounced if improvements in a defendant's behavior also affect the probability of an adverse decision in any case in which a lawsuit has been filed. For simplicity, I will refer to this as "the probability of conviction," though I intend it to include unfavorable decisions in both public prosecutions and private damage actions (where "probability of a finding of liability" would be a more appropriate but more cumbersome label). As the probability of conviction depends largely on the underlying legal standard, I discuss legal regimes based on strict liability separately from those based on negligence.

1. *The Probability of Conviction Under Strict Liability*

If the legal regime is truly one of strict liability, the probability of conviction may be unresponsive to a defendant's behavior. The probability of conviction may not be 100%, for even regimes of "strict liability" usually require proof of certain elements (was the defendant engaged in the activity? did that activity cause this plaintiff's harm?), and the chance of judicial error on one of these elements may leave defendants facing a probability of punishment below 100%. Still, as long as the chance of error is not correlated with the defendant's level of care, or with the social desirability of any other dimension of the defendant's behavior, the probability of conviction will still be unresponsive to any *improvements* in the defendant's behavior.[51] If so, then any responsiveness in the overall probability of punishment will have to come from other components of that probability, such as the probability of detection or the probability of litigation.

2. *The Probability of Conviction Under Negligence*

If the legal regime is based on negligence, though, a different picture emerges. A "negligence" regime, as I use that term, is one that conditions liability on whether the defendant conformed to some legally determined standard of behavior. Under such a regime, the probability of conviction should be extremely responsive to improvements in a defendant's behavior.

---

51. Some reasons why this probability might indeed be correlated with the social desirability of the defendant's behavior will be discussed *infra* in subsection III.D.

Indeed, the standard economic analysis of negligence rules has already recognized many of the points made in this article.[52] The earliest analyses focused on perfectly applied negligence rules, in which defendants' liability depended solely on whether they had complied with the legal standard of care. This meant that the probability of punishment declined instantly and dramatically, from 100% all the way to 0%, as soon as defendants came into compliance with the legal standard. A perfectly functioning negligence regime is thus the most extreme case of the phenomenon of interest here, for it yields a probability of punishment that is *extremely* responsive to improvements in a defendant's behavior.

Significantly, analyses of perfect negligence standards have long recognized that the expected measure of damages (if and when liability is found) need not satisfy the traditional multiplier principle.[53] To the contrary, if the probability of punishment falls instantaneously to zero, there will usually be a broad range of fines or damage awards that will suffice to induce compliance with the legal standard. Because defendants who comply with such a standard are rewarded by having their liability eliminated entirely, they will have an incentive to do so as long as the penalty (if they do not comply) equals or exceeds the cost of complying, so any penalty at *or above* that level should lead to optimal deterrence.[54] In other words, under a perfect negligence standard, satisfaction of the multiplier principle clearly is sufficient but not necessary for optimal deterrence.

Of course, most real-world negligence standards are not perfect, and compliance with the standard will not change the probability of conviction from 100% to zero. More realistically, even defendants who comply with the standard might still be held liable through legal error, though that risk will normally decline as they take more care (i.e., as they take precautions well in excess of the legal standard).[55] On the other side of the line, even defendants who do not comply will sometimes be exonerated through judicial error, though

---

52. For a nontechnical discussion, see Cooter, *supra* note 9, at 1526-27, 1538-39.

53. *See, e.g., id.*

54. *See id.* Of course, compliance with the legal standard will produce socially optimal deterrence only if the legal standard is set at the socially efficient level of care.

55. Note that if there were no error or uncertainty of this sort, then defendants would always comply with the negligence standard and would never be found liable, so it would never be worthwhile for plaintiffs to bring suit. In other words, a negligence system with perfect compliance and no uncertainty or error is not a sustainable equilibrium. For mathematical analyses of this aspect of a negligence system, see Keith N. Hylton, *Costly Litigation and Legal Error Under Negligence,* 6 J.L. Econ. & Org. 433 (1990); Janusz A. Ordover, *Costly Litigation in the Model of Single Activity Accidents,* 7 J. Legal Stud. 243 (1978).

this possibility will typically decline as they take less care (i.e., as they fall obviously short of the legal standard). In other words, under most real-world negligence standards, defendants face a continuously declining probability of conviction that varies with the amount of care they actually take.

This continuously declining probability of punishment is precisely what I have assumed throughout the body of this article. That is, in this sort of regime defendants who improve their level of precautions will see their probability of conviction fall from 25% to 20%, or 10%, or some other lower figure (depending on the rate of legal error). Under such a regime, there will no longer be an entire range of punishments that induce compliance with the optimal standard of care, as there would have been under a perfect negligence system. Instead, as we have already seen, there will only be *one* constant multiplier that will create the optimal incentives, just as there will only be one constant fine that is optimal (and only one optimal case-by-case multiplier). And since the optimal case-by-case multiplier will then diverge from the optimal constant multiplier (as well as from the optimal constant fine), there will again be more than one route to optimal deterrence.

It is important to remember, too, that "negligence" as it is used here is a term of art that covers more than its usual legal meaning. That is, in economic analyses of law, "negligence" is used to refer to any regime in which defendants are legally liable for the harm they cause if, but only if, some aspect of their behavior is judged by a court to fall short of some socially desirable level.[56] This term is usually used to mark a contrast with regimes of "strict liability" (another term of art) in which defendants are legally liable without regard to the social desirability of any dimension of their behavior, so the probability of conviction is truly unresponsive to improvements in a defendant's behavior.

Under these definitions, many bodies of law — probably most — are "negligence" regimes. For example, many pollution laws hold defendants liable if (but only if) their pollution exceeds a legally permitted level.[57] Similarly, the law of predatory pricing prohibits monopolists from cutting prices under certain circumstances, if (but only if) the new prices fall below a legally permitted level.[58]

---

56. For a rigorous mathematical definition, see SHAVELL, *supra* note 13, at 8.

57. *See, e.g.,* 40 C.F.R. pts. 425-71 (1998) (EPA effluent guidelines and standards under the Federal Water Pollution Control Act); 33 U.S.C. § 1319 (1994) (penalties for violation of same).

58. For an overview of the relevant legal standards, see Michael L. Denger & John A. Herfort, *Predatory Pricing Claims After* Brooke Group, 62 ANTITRUST L.J. 541 (1994).

And the law of fraud holds defendants liable if (but only if) the allegedly fraudulent statements would have been interpreted by a "reasonable" listener as asserting a false claim, where the judge or jury must decide whether any given statement violates that standard (or whether the statement should instead have been discounted as mere "puffing").[59] This makes each of these doctrines a "negligence" regime, according to the economic definition of that term.

Indeed, even legal doctrines described by courts as "strict liability" are sometimes negligence regimes for purposes of this definition. For example, it is often said that manufacturers are "strictly" liable for all defects that leave their products in an unreasonably dangerous state. When the alleged defect consists of a dangerous design, however, the court must decide whether the design was so bad as to produce an *unreasonably* dangerous product, and the manufacturer will be legally liable if (but only if) the product's design falls short of the standard adopted by the court.[60] This makes the regime one of "negligence" rather than "strict liability" (according to the economic definition), for the defendant's liability depends on the social desirability of its product design choices. By contrast, in strict liability regimes the defendant's liability may depend on how it behaved, but not on the social desirability of its behavior.[61]

My purpose here is not to quibble with the economic definitions, which do serve a useful analytic purpose. Instead, my point is simply that under the vast majority of legal standards — whatever

---

59. When a fraud claim rests on the defendant's failure to disclose information adequately, the court must also decide whether the undisclosed information would have been "material" to a reasonable listener — in other words, whether the defendant's behavior went far enough in disclosing all the information that might have been disclosed. Fleming James, Jr. & Oscar S. Gray, *Misrepresentation* (pt. 2), 37 MD. L. REV. 488, 497-502 (1978). The analogous issues raised in public prosecutions for false advertising are discussed in Richard Craswell, *Interpreting Deceptive Advertising*, 65 B.U. L. REV. 657, 679-81, 696-714 (1985).

60. The same is true if the manufacturer's liability rests on an alleged failure to provide users with an *adequate* warning of the product's risks (thus raising disclosure issues similar to those raised by some claims of "fraud," discussed *supra* in note 59). For more complete discussions of the relationship between "strict liability" and "negligence" in the products liability context see, e.g., Sheila L. Birnbaum, *Unmasking the Test for Design Defect: From Negligence [to Warranty] to Strict Liability to Negligence*, 33 VAND. L. REV. 593 (1980); David G. Owen, *Defectiveness Restated: Exploding the "Strict" Products Liability Myth*, 1996 U. ILL. L. REV. 743.

61. For example, strict liability for ultrahazardous activities (such as dynamiting) attaches only to defendants who engage in that activity, so even this liability is conditional on how the defendant behaved. However, it is not conditional on any judgment about the *social desirability* of the relevant behavior — for example, there is no scrutiny of the desirability of a defendant's decision to engage in dynamiting. This is what makes liability for ultrahazardous activities a regime of "strict liability" even under the economic definition.

label we use to characterize them — the probability of conviction will be highly responsive to changes in a defendant's behavior. Companies whose products impose only slight dangers, or dangers only slightly in excess of their acknowledged benefits, are surely more likely to escape liability than are manufacturers whose products impose greater dangers. Similarly, sellers who disclose all of the most important facts, and omit only a few arguably relevant ones, are more likely to escape liability than are those who disclose nothing whatsoever, omitting even those facts whose importance is obvious. In short, these are all regimes in which the probability of a conviction changes with improvements in a defendant's behavior, and this is enough to generate the effects discussed earlier in this article. (Of course, if the probability of detection and/or the probability of filing suit are *also* responsive to changes in a defendants' behavior, these same effects will be present even in regimes of true strict liability.)

D. *The Probability of Different Damage Awards*

Finally, the legal doctrines governing the measurement of damages can produce similar effects. Moreover, these effects may be present whether the regime is based on negligence or on strict liability. True, the measurement of damages will be irrelevant in any system employing a constant fine, where (by definition) the penalty does not depend on the damages in any particular case. Whenever the penalty does depend on the damages in each case, however, the rules for measuring damages can alter the deterrent effect.

The most relevant doctrines here are those that exclude certain elements from the legally recoverable damages. Sometimes losses are excluded if they were not reasonably foreseeable to the defendant,[62] or if their amount could not be proven with an acceptable degree of precision.[63] Sometimes whole categories of losses may be excluded, as with economic losses in some tort cases,[64] or damages for mental suffering and emotional distress in most breaches of contract.[65] In still other cases, if the defendant's behavior contributed to the victim's harm in only a probabilistic way (by increasing the

---

62. *See, e.g.,* Palsgraf v. Long Island R.R., 162 N.E. 99 (N.Y. 1928); Hadley v. Baxendale, 9 Exch. 341, 156 Eng. Rep. 145 (1854).

63. *See* 4 Fowler V. Harper et al., The Law of Torts § 25.3 (2d ed. 1986).

64. *See* 4 *id.* § 25.18a.

65. *See* Restatement (Second) of Contracts § 353 (1979).

risk of an injury), recovery may be disallowed under the rules governing causation.[66]

Insofar as these exclusions reduce defendants' expected liability, they will reduce the deterrent effect. Indeed, the need to make up for such exclusions from "compensatory" damages has itself been cited as one possible rationale for punitive awards.[67] What is less often noted, however, is that these exclusionary doctrines are often applied in a way that makes them highly responsive to changes in a defendant's underlying behavior. And this responsiveness creates an offsetting effect that *increases* the law's deterrent effect, for all of the reasons discussed earlier in this article.

For example, in cases where the defendant's conduct merely increased the risk of a probabilistic injury, it is often said that a defendant will be liable only if it is "more likely than not" that the defendant's negligence actually caused the plaintiff's injury. In economic analyses, it is sometimes assumed that this rule makes defendants liable if their negligence increased the probability of an injury by more than 50%.[68] If such a 50% cutoff could be applied perfectly, then defendants' expected liability would drop instantaneously to zero once their behavior improved to the point where their contribution to the risk fell below that threshold. In other words, a perfectly applied 50% rule would produce effects very similar to those of the perfectly applied negligence standard discussed earlier.[69] But if there is instead any uncertainty in the application of the 50% cutoff — for example, if defendants cannot know in advance precisely how much of the probability a judge or jury will ascribe to their particular behavior — then the expected punishment will fall more gradually, as defendants who improve their behavior face an increasing likelihood that their behavior will be

---

66. *See* 4 HARPER ET AL., *supra* note 63, § 20.2, at 93-101, 107-10. For economic analyses of this rule, see William M. Landes & Richard A. Posner, *Causation in Tort Law: An Economic Approach,* 12 J. LEGAL STUD. 109 (1983); Rosenberg, *supra* note 30; Steven Shavell, *Uncertainty over Causation and the Determination of Civil Liability,* 28 J.L. & ECON. 587 (1985).

67. *See, e.g.,* Chapman & Trebilcock, *supra* note 3, at 768-69; Thomas C. Galligan, Jr., *Augmented Awards: The Efficient Evolution of Punitive Damages,* 51 LA. L. REV. 3, 39 (1990); *see also* Polinsky & Shavell, *supra* note 3, at 896 ("To the extent that [certain losses are excluded from compensatory damages], an argument can be made that the level of damages should be *higher* than that called for by our multiplier formula . . . ." (emphasis added)). Polinsky and Shavell ultimately recommend that these excluded losses not be used to justify punitive awards, but only because the cost of measuring them will often be too high, and because (when the cost is not too high) a superior solution would be to revise the rules so that those losses were no longer excluded. *See id.* at 939-41.

68. *See, e.g.,* Shavell, *supra* note 66, at 588.

69. *See supra* text accompanying note 52.

found to fall below the relevant cutoff.[70] In other words, defend-
ants who improve their behavior are more likely to be exonerated
on causal grounds from full responsibility for all of the losses they
may have caused.

More generally, defendants who improve their behavior may
also be more likely to receive the benefit of the doubt under most
of the other doctrines governing the measurement of damages. In
some cases, the categorical exclusion of certain kinds of damages
may be relaxed against defendants who behaved in a particularly
egregious way.[71] In other cases, the responsiveness may come less
from an explicit legal rule and more from biases in the application
of a vague legal standard, such as the requirement that losses be
proven with *reasonable* certainty,[72] or that the losses have been *rea-
sonably* foreseeable to the defendant.[73] Defendants who behaved
well may be more likely to get the benefit of the doubt from the
judge (or jury) in these matters, while defendants who behaved
badly may be treated more harshly.[74]

The reason this is important, of course, is that if a limit on the
recoverability of damages is applied in a way that itself responds to
changes in a defendant's behavior, this will increase the law's deter-
rent effect. That is, even if the absolute level of damages is still less
than the harm caused by the violation (because some losses are still
being excluded from the measure of damages), there can still be an

---

70. For cases suggesting that the actual rule is applied quite flexibly, see 4 HARPER ET
AL., *supra* note 63, § 20.2, at 93-101.

71. *See, e.g.,* Dold v. Outrigger Hotel, 501 P.2d 368, 372 (Haw. 1972) (damages for emo-
tional distress not normally recoverable for breach of contract, but held recoverable in tort if
the defendant breached "in a wanton or reckless manner"), *overruled by* Francis v. Lee
Enters., Inc., 971 P.2d 707 (Haw. 1999). For discussions of other similar doctrines, see
George M. Cohen, *The Fault Lines in Contract Damages,* 80 VA. L. REV. 1225 (1994); and
Patricia H. Marschall, *Willfulness: A Crucial Factor in Choosing Remedies for Breach of Con-
tract,* 24 ARIZ. L. REV. 733 (1982).

72. Harper, James, and Gray note an increasing liberality in allowing plaintiffs to recover
without proving their loss with literal certainty. They also report that "the tendency is great-
est where the nature and impact of defendant's act is such as to make likely the kind of harm
that plaintiff is claiming, or to be *especially offensive in light of public policy considerations.*"
4 HARPER ET AL., *supra* note 63, § 25.3, at 510 (emphasis added) (footnote omitted).

73. *See* L.L. Fuller & William R. Perdue, Jr., *The Reliance Interest in Contract Damages*
(pt. 1), 46 YALE L.J. 52, 85 (1936) ("As in the case of all 'reasonable man' standards, there is
an element of circularity about the test of foreseeability. 'For what items of damages should
the court hold the defaulting promisor? Those which he should as a reasonable man have
foreseen. But what should he have foreseen as a reasonable man? Those items of damages
for which the court feels he ought to pay.'").

74. This aspect of legal bias was often noted by the legal realists. *See, e.g.,* Ralph S.
Bauer, *The Degree of Moral Fault as Affecting Defendant's Liability,* 81 U. PA. L. REV. 586
(1933); Fuller & Perdue, *supra* note 73, at 77. For more recent evidence consistent with this
view, see Stephan Landsman et al., *Be Careful What You Wish For: The Paradoxical Effects
of Bifurcating Claims for Punitive Damages,* 1998 WIS. L. REV. 297, 334-35.

increased marginal incentive if the damages respond sufficiently to *changes* in a defendant's behavior. Here, too, it is the rate of change in the expected liability, rather than the absolute level of liability, that matters for deterrence.

An example will serve to illustrate. In the toxic waste scenario, we originally assumed that each leak would cause $6 million in damages if no extra precautions were taken, and $5 million if the precautions were taken. Suppose now that many of these damages are difficult to measure, or are difficult to attribute to any particular defendant. But suppose, too, that more of these losses will be attributed to defendants who did not take any extra precautions than to defendants who did take precautions, simply because the former are less likely to get the benefit of any doubt. For concreteness, suppose that defendants who did not take the precautions will be assessed $5 million in damages each time they are caught (as compared to the $6 million in damages their activity actually caused), while defendants who did take precautions will be charged with only $1 million in damages (as compared to the $5 million they actually caused). In other words, suppose that both types of defendants get away with paying less than the full social costs of their behavior, but that this shortfall is greatest for defendants who took the extra precautions.

In such a case, even if the probability of punishment is 25% for both defendants (and thus is completely *un*responsive to changes in their behavior), defendants will still have an incentive to take the optimal level of care. It might seem as though their incentives in this case would have to favor underdeterrence, since neither defendant is being charged with the full social costs of its behavior, and since both defendants are discounting these already-low penalties by the same 25% probability of punishment. However, in this example defendants who do not take the extra precautions will face an expected liability of $12,500 ($.01 \times .25 \times \$5,000,000$), while defendants who do take the precautions will see their expected liability fall to $2,500 ($.01 \times .25 \times \$1,000,000$). This means that the precautions can save defendants $10,000 in expected liability ($12,500 − $2,500), which just equals the social value of the precautions and thus achieves optimal deterrence.

The lesson here is the same as it was in preceding subsections of this article. That is, it is not simply the absolute level of expected liability that matters, but rather the *rate* at which that expected liability declines with improvements in a defendant's behavior. It will be relatively rare for that decline to come from changes in the mere

probability of detection, but it should be quite common — the rule, rather than the exception — for improvements in a defendant's behavior to reduce the probability of prosecution, the probability of conviction, and/or the likely damage award. Declines in any of these factors will cause a divergence between the optimal case-by-case multiplier and the optimal constant multiplier (or the optimal constant fine). As a result, satisfaction of the traditional multiplier principle will be sufficient but not necessary for optimal deterrence — and this conclusion, too, should be considered the rule rather than the exception.

IV.   Choosing Among the Alternatives

We are now in a position to compare the strengths and weaknesses of case-by-case multipliers, constant multipliers, and the other alternative penalty systems discussed above. While any of these systems could, in theory, lead to optimal levels of deterrence, they differ in many other respects that could make them more or less desirable. In this final Part of the article, I consider differences based on their relative ease of administration; their effect on other economically relevant variables, such as the level of overall activity or the optimal allocation of risk; and their possible symbolic or expressive effects.

A.   *Ease of Calculation*

One virtue of the traditional multiplier principle is that it is conceptually simple and, therefore, easy to explain to a judge or jury. The case-by-case multiplier requires only two pieces of information: the actual harm caused by the defendant's behavior, and the actual probability of punishment that the defendant faced. And while it will rarely be possible to measure either of these elements precisely, in many cases it should be possible to come to a rough estimate.[75] Moreover, all that really matters under this approach is that the average or *expected* penalty equal the actual harm divided by the probability of punishment, so it may not even matter if juries err as long as their errors are not systematically biased in either direction.[76]

By contrast, many of the alternative remedies are more difficult to calculate. For example, to determine the optimal constant multi-

---

75. *See* Polinsky & Shavell, *supra* note 3, at 891-93. For somewhat more skeptical views, based on the psychology of jury decision-making, see Sunstein et al., *supra* note 4, at 2111-12; Viscusi, *supra* note 4, at 327-32.

76. *See* Polinsky & Shavell, *supra* note 3, at 892.

plier, the decisionmaker must know several factors in addition to the actual harm caused by the defendant's behavior and the actual probability of punishment. In addition, the decisionmaker must be able to estimate the actual harm that would have been present, and the probability of punishment the defendant would have faced, if the defendant had instead committed a slightly more or slightly less serious violation. In particular, the decisionmaker must be able to measure each set of factors with sufficient precision to estimate the *difference* between the two violations, in order to calculate the rate of *change* in the relevant variables. Finally, once these estimates have been made, they must then be combined through a formula that is more complicated than simply multiplying by one over the probability of punishment.[77] Thus, the optimal constant multiplier will always be harder to calculate than the optimal case-by-case multiplier.

The same objection can be raised against many of the other alternatives discussed above. For example, calculating the optimal constant to add or subtract from a compensatory award, or the optimal amount by which to adjust the substantive legal standard, is just as complex as calculating the optimal constant multiplier.[78] Calculating the optimal constant fine is slightly simpler conceptually, for this requires knowledge of only (a) the rate at which the probability of punishment changes in response to changes in defendants' behavior, and (b) the rate at which the expected social costs change in response to changes in defendants' behavior.[79] But information about these hypothetical rates of change will often be hard to come by, compared to information about the actual social harm and the actual probability of punishment. As a result, the case-by-case multiplier will surely rank highest in ease of calculation.

---

77. If x represents the defendant's level of precautions, $H(x)$ represents the expected social harm at any particular level of precautions, and $P(x)$ represents the probability of punishment at any particular level of precautions, then the optimal case-by-case multiplier is given by a simple fraction: $1/P(x)$. To calculate the optimal constant multiplier, we need to let x^* represent the socially optimal level of precautions, and $H'(x^*)$ and $P'(x^*)$ represent the *rate* at which the social harm and the probability of punishment change with slight deviations from the optimal level of precautions. The optimal constant multiplier can then be written as the following complex formula:

$$\frac{H'(x^*)}{H'(x^*)\ P(x^*) + H(x^*)\ P'(x^*)}$$

See Craswell & Calfee, *supra* note 13, at 294. Obviously, this complex formula would be more difficult to present to a judge or jury.

78. *See* Polinsky & Rubinfeld, *supra* note 40, at 159-60, app.

79. Using the notation introduced *supra* in note 77, the optimal constant fine can be represented by the simple fraction $H'(x^*)/P'(x^*)$. *See* Craswell & Calfee, *supra* note 13, at 297.

B. *Institutional Responsibility*

Another difference is that a case-by-case multiplier, by its very nature, must be calculated separately for each defendant, so all the calculations must be made by a judge or jury. By contrast, a constant multiplier could be set once by a legislature or administrative agency, leaving judges and juries with only the task (in each individual case) of measuring the losses to which the constant multiplier would be applied. If a constant fine were used, the judge or jury would not even need to calculate the actual losses in each case, thus shifting even more of the work to a centralized body.

The ability to make such calculations centrally *could* be an advantage. Obviously, it is sometimes cheaper to have calculations made once by a centralized body, rather than making them anew in every case. A centralized body may also be able to assemble more expertise than a judge or a jury — for example, it could commission statistical studies of the probability of punishment and the rate at which it changes. A centralized body can also benefit from the ability to "fine tune" a constant multiplier or a constant fine (or an adjustment to the substantive legal standard), raising or lowering it until the desired level of deterrence is achieved. Finally, in some contexts (or from some perspectives) a centralized body might be seen as more "democratic" or politically accountable.

On the other hand, there are also drawbacks to a centralized decision process, which may make these advantages moot. A centralized body may be more prone to political "capture" by groups favorable to plaintiffs or defendants.[80] Also, even a well-motivated central body must face the practical problem of defining, in advance, the exact class of cases to which any particular fine or multiplier would be applied. After all, the constant multiplier that is optimal for malpractice cases is unlikely to be the same as the one that is optimal for products liability cases; it may also be different for some malpractice cases than for others. Thus, any centralized solution must either (a) define a separate penalty for many different categories of cases, with the attendant difficulty of defining the boundaries of each; or (b) rely on a smaller number of relatively crude categories, recognizing that the penalty selected for each category will not be ideal for every case within that category.

By contrast, one advantage of a decentralized system is that it is unnecessary to define such categories in advance. As Polinsky and

---

80. The legislatively enacted caps on punitive damage awards, discussed *supra* in the text following note 41, may be an example of this.

Shavell pointed out in their analysis of the case-by-case multiplier, courts can estimate the probability of punishment facing any particular defendant without having to assign that defendant to some a priori category: they can define that probability based on whatever is known about this individual defendant's characteristics.[81] To be sure, this does not mean that judges or juries will necessarily get the probability right, for (as the preceding subsection discussed) even case-by-case calculations cannot be based on anything more than a rough estimate of the relevant variables. But since judges and juries decide cases after the fact, they will at least be spared the cost of trying to define relevant categories *in advance,* which is an unavoidable cost of any centralized calculation.

In short, there may be cases where the appropriate penalty can best be calculated centrally, and in those cases the traditional case-by-case multiplier will be inferior (insofar as ease of calculation is concerned). In many cases, though, the costs of centralization will be too high, so both the case-by-case multiplier and the possible alternatives will have to be estimated in every case by individual judges or juries. In that event, the advantage in terms of ease of calculation will usually rest with the traditional case-by-case multiplier, for the reasons discussed in the preceding subsection.

C. *Optimal Levels of Activity*

In addition, whenever it is important to optimize defendants' levels of activity, the case-by-case multiplier may well be superior. As Steven Shavell first emphasized, many social costs can be reduced by carrying on an activity more carefully *or* by reducing the frequency of the activity itself.[82] For example, drivers can reduce the number of auto accidents by driving more carefully, but they can also reduce the number of accidents by using their cars less fre-

---

81. *See* Polinsky & Shavell, *supra* note 3, at 893. More precisely, the probability can be estimated based on (1) all facts that were known to the defendant at the time it chose its level of care, together with (2) any other facts that did not become known until later, but only if those facts were just as likely to raise the probability as to lower it. The qualification is important because it would not be correct, under the traditional multiplier principle, to calculate the probability of punishment for a narrow category defined as "all defendants who behaved in the following way *and* who happened to leave evidence that allowed their victim to bring suit and prevail at trial." The probability of punishment for defendants who are defined in this way is, of course, exactly one, so calculating the multiplier on this basis would be the same as employing no multiplier at all.

For a more general discussion of the choice between categorical and individualized judgments, see Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992).

82. *See* Steven Shavell, *Strict Liability Versus Negligence*, 9 J. LEGAL STUD. 1, 2-3 (1980).

quently. Thus, optimizing the level of activity can also be a legitimate goal of deterrence policy.

Significantly, negligence regimes often do not give defendants any direct incentive to constrain their levels of activity. Negligence regimes can easily condition liability on whether a driver was exercising appropriate care, and this gives drivers an incentive to choose their care appropriately. But such regimes rarely condition liability on whether a defendant drove with unnecessary frequency (probably because of the difficulty of determining a "reasonable" frequency), so they give drivers no incentive to limit their total amount of driving. In terms of the analysis used in this article, the probability of being held liable (for any given accident) usually is not responsive at all to any change in the number of miles the defendant drove.

If enforcement is perfect, strict liability regimes can improve defendants' incentives to adjust their levels of activity. In theory, strict liability forces defendants to internalize all of the social costs caused by their activity, and thus gives them an incentive to think about *every* way of reducing those costs, including engaging in the activity less frequently.[83] Even under a regime of strict liability, however, imperfect enforcement will dilute a defendant's incentives. That is, if a driver (or a firm employing a fleet of drivers) should be liable for every accident that results, but if there is only a 25% chance that any given accident will lead to a successful suit, the firm's liability will reflect only 25% of the costs its activity imposes, thus giving it too weak an incentive to reduce its amount of driving. In such a case, the traditional multiplier can restore the optimal incentives by multiplying the firm's expected liability by four, thus making the firm again bear the full social costs of its driving activity.

It might seem, then, that whenever the level of the defendant's activity can affect social costs, the traditional multiplier would always be the better way of achieving optimal deterrence. However, this conclusion is subject to three important limitations. First, the traditional multiplier only produces this effect in regimes of "strict liability," as that term was defined earlier.[84] In "negligence" regimes, defendants will still escape responsibility for the social costs of their activities whenever they are found to have complied with the legal standard, so even a traditional multiplier will not optimize

---

83. *See id.*
84. *See supra* text accompanying note 56.

levels of activity. And since most legal regimes really are "negligence" regimes (under the definition used here), this argument for the traditional multiplier is of limited applicability.

Second, even under regimes of "strict liability," there are a few cases in which increases in the level of activity could increase the probability of punishment per violation. For example, if one of a company's trucks is detected leaking toxic chemicals, the authorities may then decide to inspect all the other trucks owned by that firm. If so, the per-violation probability of punishment will indeed respond to changes in the overall quantity of driving, because increased driving brings a greater probability that at least one of the firm's trucks will be detected leaking chemicals, thus triggering an inspection of all the trucks. While this form of interdependence is probably not very common, it sometimes occurs in the investigation of criminal activities (tax fraud, racketeering, drug rings, etc.), where detection of one offense triggers close investigation of all of a defendant's activities.

Third, even when the traditional multiplier is superior in its effect on levels of activity, it may give rise to other problems that outweigh this benefit. After all, the level of activity is only one of the dimensions of social cost with which the law should be concerned. As the remaining subsections will discuss, there are other dimensions along which a case-by-case multiplier might sometimes be inferior to the alternative routes to optimal deterrence.

D. *Optimal Levels of Risk*

As noted earlier, the optimal case-by-case multiplier will typically be greater than (and will never be any less than) the optimal constant multiplier. This could make the case-by-case multiplier inferior from the standpoint of optimal levels of risk, quite independently of any effect the multiplier might have on levels of deterrence. If either defendants or their victims are risk averse, there will be benefits from keeping the total amount of risk in the system at a minimum. In a world of imperfect enforcement, however, a case-by-case multiplier increases the risk in the system because it presents defendants with the risk of paying even larger penalties. This could be a drawback in either of two situations.

First, as Polinsky and Shavell have noted, the effect on risk is a drawback if defendants are risk averse and full insurance is unavail-

able.[85] If defendants are risk averse, a 25% chance of having to pay $20 million in damages (as the traditional multiplier would recommend) will press on them more than four times as heavily as a 25% chance of having to pay only $5 million. The larger award thus could produce too much of a deterrent effect; and the larger award could also reduce total welfare directly, by forcing defendants to bear the disutility of this increased risk. However, the significance of these drawbacks are limited by the fact that they apply only to defendants who are risk averse, and even then only if they cannot get liability insurance (which would transfer the risk to a risk neutral insurance company). Most damage awards are legally insurable.[86] Even when they are not, many defendants are publicly held corporations who can usually be presumed to be risk neutral, or capable of self insurance through diversification of their shareholders' portfolios.[87]

Even when insurance is readily available, however, the effect on risk may still be a drawback if *victims* are risk averse, at least when the defendant and its victims stand in a market relationship (such as seller-customer or employer-employee). To be sure, there is sometimes no need for any liability at all when the defendant and its victims stand in a market relationship, if the victims are well informed about the risks and so the defendant will have adequate market incentives to improve its behavior.[88] In some markets, though, potential victims may not be quite so well-informed, in which case market incentives alone may be inadequate for optimal deterrence. This, at least, is the standard argument for any form of liability for deterrence purposes when defendants and their victims stand in a market relationship.[89] And if the enforcement of such liability is imperfect, the same arguments that would call for a dam-

---

85. *See* Polinsky & Shavell, *supra* note 3, at 886-87. For more technical economic analyses, see Louis A. Kaplow, *The Optimal Probability and Magnitude of Fines for Acts That Definitely Are Undesirable,* 12 INTL. REV. L. & ECON. 3, 6-8 (1992); A. Mitchell Polinsky & Steven Shavell, *The Optimal Tradeoff Between the Probability and Magnitude of Fines,* 69 AM. ECON. REV. 880, 884-85 (1979); Steven Shavell, *On Liability and Insurance,* 13 BELL J. ECON. 120, 124-26 (1982).

86. Even punitive damage awards are insurable in some states. *See* Polinsky & Shavell, *supra* note 3, at 931 n.193.

87. *See id.* at 887 n.44. Obviously, this conclusion would not apply to the extent that the risk in question was systematic (e.g., the risk of a change in legal rules that would increase the expected liability of *all* corporations).

88. *See id.* at 935-36.

89. For formal economic models, see, e.g., Shavell, *supra* note 82, at 14-17; Michael Spence, *Consumer Misperceptions, Product Failure and Producer Liability,* 44 REV. ECON. STUD. 561 (1977).

age multiplier in other contexts would appear to call for a multiplier in market relationships as well.[90]

The important point about market relationships, though, is that defendants' expected liability will ultimately have to be borne by consumers in the form of higher prices.[91] For example, if product defects cause $5 million in damages but defendants are held liable only 25% of the time, a premium will have to be added to the product's price to cover the defendant's expected liability of $1.25 million (.25 × $5,000,000). If the legal system then adopts the traditional case-by-case multiplier, thus making defendants pay $20 million every time they are caught (4 × $5,000,000), their expected liability will rise to $5 million (.25 × 4 × $5,000,000), so the price premium will have to be four times as large. In other words, the good news is that those consumers who are lucky enough to bring a successful lawsuit will have their recovery increased by a factor of four, but the bad news is that all consumers will have to pay an up-front price that includes a premium that is four times as large. In effect, the introduction of a multiplier turns the liability component of the price into a lottery ticket, with a bigger price up front supporting the chance of a bigger payoff at the end.

If customers are risk neutral, they will be indifferent toward this lottery; and if they are risk-preferring then they might actually like this sort of a gamble. It is more plausible, though, to assume that customers are risk averse, at least with respect to this sort of contingency. Those who would like some additional gamble can always go to the race track or play the state lottery, and it would be odd to posit a taste for gambling that could only be satisfied by wagering on a particular product defect. For most customers, then, the introduction of this extra lottery element will make the product less attractive, causing an additional welfare loss.[92]

Moreover, this effect (unlike the effect on risk averse *defendants*) will not disappear with the purchase of insurance. While full insurance can protect parties from many forms of risk, in this context consumers could be protected only by purchasing "reverse" insurance of a sort that is rarely available. That is, consumers

---

90. Polinsky & Shavell, *supra* note 3, at 935; *see also id.* at 938 (making the same argument for punitive damages in breach of contract cases). Breach of contract cases, by definition, always involve defendants and victims who were in a market relationship.

91. For a review of the literature on this point, see Richard Craswell, *Passing On the Costs of Legal Rules: Efficiency and Distribution in Buyer-Seller Relationships*, 43 STAN. L. REV. 361 (1991).

92. I develop this point at more length in Richard Craswell, *Damage Multipliers in Market Relationships*, 25 J. LEGAL STUD. 463 (1994).

would need a policy in which the insurance company paid the consumer up front (rather than the consumer paying an insurance premium), in return for which the consumer would assign to the insurance company his or her right to the punitive portion of any damage award. In theory, the up-front payment by the insurance company would exactly compensate the consumer for that part of the higher product price that reflected the punitive portion of the defendant's expected liability. Such a policy would thus have the same effect as restoring consumers to the premultiplier regime, in which they paid a lower price and were limited to compensatory damages if they sued. But since this sort of reverse insurance is rarely available, most consumers will have no way to insulate themselves from the risk-increasing effect of the larger, case-by-case multiplier.[93]

Of course, even when this risk-increasing effect is a drawback, that may not be sufficient reason to reject the traditional multiplier. If damage multipliers are used only rarely, the effect on total risk-bearing costs is not likely to be very large.[94] Moreover, the effect on risk averse parties is only one of the relevant consequences the law must consider, and a slight negative effect on risk averse consumers might be outweighed by greater improvements in (say) the ease of administration, or the effects on defendants' levels of activity. My point here is simply that *all* of these effects must be considered before any overall decision is reached — and that in evaluating this totality of effects, the effect on risk averse consumers will usually count as a negative.

E. *Practical Constraints on the Maximum Penalty*

The traditional, case-by-case multiplier may also be less desirable if defendants have limited assets, and would be unable to pay a fine as high as that required by the multiplier principle. This constraint is particularly likely to be a problem when the probability of punishment is small, so the penalty required under a case-by-case multiplier would be large. For example, if a violation causes $5 mil-

---

93. One possible way of providing such insurance might be a market in which consumers could sell their right to recovery in advance, at the same time they purchased the product. For a discussion of how this market might work, see Robert Cooter, *Towards a Market in Unmatured Tort Claims,* 75 VA. L. REV. 383 (1989). As Cooter points out, though, these transactions are illegal under current law. *See id.* at 383 & n.1.

94. For empirical evidence consistent with this view, at least where punitive damages are concerned, see Jonathan M. Karpoff & John R. Lott, Jr., *Punitive Damages: Their Determinants, Effects on Firm Value, and the Impact of Supreme Court and Congressional Attempts to Limit Awards,* 52 J.L. & ECON. (forthcoming 1999).

lion worth of damages but brings only a 25% probability of punishment, the optimal case-by-case multiplier would require a fine of $20 million (4 × $5,000,000). But if many defendants are thinly capitalized firms, the threat of an $20 million penalty may have no more effect than the threat of a $9 or $10 million penalty, if $9 or $10 million is the most they could possibly pay. Indeed, in some cases the adoption of larger penalties could even *reduce* deterrence, by giving potential defendants an incentive to operate with even less capital than they otherwise might. This would make it even more likely that these defendants would not have to pay the full legal penalty (because they would not have enough money to pay it), and so would further reduce the law's deterrent effect.[95]

One possible solution to this problem is to (1) raise the penalty as high as possible, then (2) make up for any remaining underdeterrence by improving the enforcement system to raise the probability of punishment.[96] In the above example, if the maximum fine that could possibly be collected is only $10 million, optimal incentives could still be achieved if the enforcement system were improved to raise the probability of punishment from 25% to 50%, since that would make the expected penalty equal the expected social harm (.50 × $10,000,000 = $5,000,000). However, improving the enforcement system has costs of its own that must also be taken into account. As a result, this solution usually leads to a compromise in which the probability of punishment is raised to some extent (at some cost), but it is not raised high enough to satisfy the multiplier principle.[97]

What is less often noted, though, is that the alternative routes to optimal deterrence may provide another solution to this problem. As we have seen, alternatives such as a constant multiplier or a constant fine typically require smaller penalties than those required by the case-by-case multiplier. As a result, the fines or damage awards required under these alternatives are less likely to run up against

---

95. For an economic model of this effect, see James Boyd & Daniel E. Ingberman, *Do Punitive Damages Promote Deterrence?*, 19 INTL. REV. L. & ECON. 47 (1999). For an analysis of the effect of limited capital on deterrence generally, see S. Shavell, *The Judgment-Proof Problem*, 6 INTL. REV. L. & ECON. 45 (1986).

96. Polinsky & Shavell, *supra* note 3, allude to this possibility at 922 n.167.

97. For mathematical analyses of these trade-offs, see A. Mitchell Polinsky & Steven Shavell, *A Note on Optimal Fines When Wealth Varies Among Individuals*, 81 AM. ECON. REV. 618 (1991); A. Mitchell Polinsky & Steven Shavell, *supra* note 85, 69 AM. ECON. REV. 883 (1979); A. Mitchell Polinsky & Steven Shavell, *The Optimal Use of Fines and Imprisonment*, 24 J. PUB. ECON. 89 (1984) [hereinafter Polinsky & Shavell, *Fines and Imprisonment*]. As the second of these articles indicates, another possible solution to the problem of defendants with limited assets — albeit a solution with costs of its own — is to resort to nonmonetary sanctions such as imprisonment. *See* Polinsky & Shavell, *Fines and Imprisonment, supra.*

the constraints imposed by defendants' limited assets. Where limited assets are a problem, then, these alternatives may actually be superior to the traditional multiplier.

Moreover, there are other factors besides limited assets that may constrain the maximum penalty. Jurors (even judges or prosecutors) are sometimes reluctant to impose penalties that seem "too large," thus constraining the maximum penalty through jury nullification (or through the exercise of prosecutorial discretion).[98] Whenever this is likely, the effect will be the same as if the maximum penalty were constrained for any other reason. In particular, if this constraint is below the size of the penalty required by the multiplier principle, the case-by-case multiplier will lead to underdeterrence. Here, too, there may be an advantage to one of the alternatives that can achieve optimal deterrence with smaller penalties.

F. *Litigation Costs*

Because of their smaller size, the alternatives to the traditional multiplier may also reduce the amount spent on litigation in any particular case.[99] This is so for two reasons: the total stakes will be lower, and not as much will turn on establishing any particular probability of punishment.

First, a traditional multiplier raises the stakes involved in litigating other issues, such as the amount of damages to which the multiplier will be applied (or the underlying issue of liability itself). Under a system with no multiplier at all, defendants obviously will have some incentive to try to convince the court that their behavior did not do very much harm, or that they should not be found liable at all. But if case-by-case multipliers are used, every reduction in the measure of damages will be three times as valuable to the defendant if the multiplier is three, or five times or ten times as valuable (if the multiplier is five or ten). To be sure, the measure of damages and the question of underlying liability will also take on greater importance under any regime using a constant multiplier, as long as that multiplier is greater than one. But since the optimal constant multiplier will almost always be less than the optimal

---

98. An economic model with some of these features — specifically, a model in which finders of fact implicitly raise the burden of proof when higher sanctions are sought — is presented in James Andreoni, *Reasonable Doubt and the Optimal Magnitude of Fines: Should the Punishment Fit the Crime?*, 22 RAND J. ECON. 385 (1991).

99. The effect on *total* litigation costs, by increasing or decreasing the number of suits that are filed, was discussed in *supra* text accompanying notes 49-50.

case-by-case multiplier, the incentive to spend on litigation should at least be lower if a constant multiplier is used (and similarly for a constant fine, or for an adjustment to the substantive standard with no adjustment in liability).

Second, a case-by-case multiplier also gives parties an incentive to spend money litigating the size of the multiplier itself. Consider, for example, a defendant who caused $5 million worth of damages, and for whom the probability of punishment was somewhere around 20%. If the court finds that the probability of punishment was exactly 20%, the case-by-case multiplier will be set at five and the defendant will have to pay $25 million in damages (5 × $5,000,000). But if the defendant persuades the court that the probability of punishment was really 25%, the case-by-case multiplier will then be set at four, and the damage award will be reduced to $20 million (4 × $5,000,000). In other words, the defendant can save $5 million in liability ($25,000,000 − $20,000,000) just by altering the court's perception of the probability by five percentage points. This is the same as the amount the defendant could gain by establishing that it was not liable at all, thus reducing its liability from $5,000,000 to 0, in a regime that did not use any multipliers. For many litigants, then, it will pay to spend just as much litigating the probability of punishment (in a system with case-by-case multipliers) as they would spend contesting liability itself (in a system with purely compensatory damages).

Of course, if there is a chance they could alter the court's finding on probability even more — say, by raising it from 20% to 33%, thus reducing the multiplier from five to three — they will have an incentive to spend even more. By contrast, this incentive will be eliminated in any system employing a constant multiplier or a constant fine. For all of these reasons, then, the amount spent litigating each case should be significantly higher under the traditional case-by-case multiplier.

G. *Symbolic or Expressive Effects*

Finally, the alternatives to the case-by-case multiplier may also have symbolic or expressive advantages. Recall that, when the probability of punishment declines with improvements in a defendant's behavior (as it usually will), the case-by-case multiplier has to be largest for those defendants who behaved relatively well (to make up for their low probability of punishment), and smallest for defendants who behaved relatively badly. For instance, in one of the examples discussed earlier, defendants who took extra precau-

tions saw their probability of punishment fall from 25% to 10%, thus increasing the case-by-case multiplier from four to ten.[100] The actual damages in that example were $6 million without the extra precautions and $5 million with the precautions. With a case-by-case multiplier, however, defendants who took the extra precautions had to pay $50 million every time they were caught (10 × $5,000,000), while defendants who did not take the precautions had to pay only $24 million (4 × $6,000,000).

A possible objection is that this inverts the "fair" or "just" relation between wrongfulness of behavior and severity of punishment, by punishing those who behave well more harshly than those who behave badly. The objection is not that this will distort such parties' incentives, because incentives depend on the *expected* levels of punishment, and the expected punishment is more severe for defendants who do not take precautions (.25 × 4 × $6,000,000 = $6,000,000) than for defendants who do (.10 × 10 × $5,000,000 = $5,000,000). But it is sometimes argued that justice places independent constraints on the penalties that can be meted out by the state — independent, that is, of any utilitarian or deterrence-related goals — and that the regime described here would run afoul of those constraints.[101] Indeed, the Supreme Court has suggested (though without addressing this precise issue) that some proportionality between the size of the penalty and the wrongfulness of the defendant's conduct may even be constitutionally required.[102]

Moreover, even those who object to deontological constraints on the size of the permissible penalties might still worry that bad consequences would follow from a regime that punished mild offenses more severely than egregious ones. It is sometimes said that one function of law is to educate its citizens and to instill appropriate attitudes concerning right and wrong.[103] Presumably, this expressive or educative function includes expressing appropriate at-

---

100. *See supra* text following note 14.

101. *Cf.* IMMANUEL KANT, THE METAPHYSICAL ELEMENTS OF JUSTICE 100 (John Ladd trans., Bobbs-Merril Co. 1965) (1916) ("The law concerning punishment is a categorical imperative, and woe to him who rummages around in the winding paths of a theory of happiness [i.e., utility] looking for some advantage to be gained. . . ."). For modern discussions of this position, describing in more detail its conflict with the multiplier principle, see Chapman & Trebilcock, *supra* note 3, at 779-98; Alan H. Goldman, *The Paradox of Punishment,* 9 PHIL. & PUB. AFF. 42 (1979).

102. *See* BMW of N. Am., Inc. v. Gore, 517 U.S. 559, 575-76, 580-81 (1996).

103. Paul H. Robinson & John M. Darley, *The Utility of Desert,* 91 Nw. U. L. REV. 453, 472-73 (1997). For further development of this idea in connection with punitive damages, see Marc Galanter & David Luban, *Poetic Justice: Punitive Damages and Legal Pluralism,* 42 AM. U. L. REV. 1393, 1430-40 (1993).

titudes about relative *degrees* of wrongfulness. But the regime described above could send the wrong signal in this regard, as it could suggest that defendants who took the extra precautions (and therefore had to pay $50 million) had behaved worse than the defendants who did not take the extra precautions (and who only had to pay $24 million).[104] If citizens' attitudes are shaped by this incorrect signal, that could affect the citizens' willingness to reduce their own pollution (or even their willingness to comply with laws in general), thus raising the cost of achieving any given level of deterrence.

To be sure, each of these arguments has difficulties of its own. It is controversial (to say the least) whether we ought to accept purely deontological constraints on the size of permissible punishments. Moreover, the argument that inverted penalties will send the wrong moral message depends on a kind of misperception on the part of the audience of the signal. That is, if citizens realized that the regime described above was adopted solely for its deterrent virtues, and that the difference in penalties therefore expressed only the fact that the probability of punishment was different for the two defendants, there would then be no reason for citizens to draw an incorrect moral lesson. Indeed, if citizens realized that the *expected* penalty was actually harsher for defendants who did not take extra precautions, they might continue to draw the correct moral lesson, in which case the law's expressive or educative effect would be reinforced. The concern that citizens will draw the wrong moral lesson thus rests on an implicit assumption of "noise" or miscommunication between the message intended by the drafters of the policy and the message understood by the citizenry. And while such errors or misperceptions are no doubt common, it is notoriously difficult to predict the exact form they will take.

Still, the fact that such misperceptions are possible means that this concern cannot be dismissed out of hand. My only point here is that, to the extent this danger is real, it too can be avoided by using one of the alternatives to the traditional multiplier principle. Since these alternatives all allow the actual penalties (not just the expected penalties) to increase in severity with the egregiousness of the defendant's behavior, they would not pose any risk of the moral misperception at issue here.

---

104. This very objection is made by Galanter & Luban, *supra* note 103, at 1449-50.

V. CONCLUSIONS

As we have seen, the traditional multiplier of one over the probability of punishment can achieve optimal deterrence if it is recalculated on a case-by-case basis, to reflect the probability of punishment facing each individual defendant. Moreover, if it is applied in this way, the traditional multiplier will be optimal in a wide variety of circumstances, regardless of whether the premultiplier incentives favored under- or overdeterrence. Because a case-by-case multiplier can achieve optimal deterrence using a relatively simple formula under so broad a range of conditions, it is perhaps not surprising that the law review literature has focused almost exclusively on this method of correcting for imperfect enforcement.

In fact, though, the law uses a variety of other methods that do not fit the traditional multiplier principle. Sometimes the law uses a constant multiplier; sometimes it uses constant fines; and sometimes it uses adjustments to the substantive legal standard. Indeed, while all of these methods are relatively common, it is very difficult to find examples of a true case-by-case multiplier, in which defendants whose conduct faces a high probability of punishment are "rewarded" with a multiplier lower than that given to defendants whose conduct is less likely to be punished. Unfortunately, the legal literature has focused so much on the case-by-case multiplier principle that it has not even begun to address the pros and cons of these alternative (and much more common) systems of deterrence. This article is an attempt to begin to fill that gap. While its conclusions are necessarily tentative, several points can be made.

First, the traditional analysis is still perfectly valid whenever the probability of punishment is essentially unresponsive to changes in a defendant's behavior. This is most likely to be the case when the probability of punishment depends entirely on whether the defendant's offense is detected. In other words, the traditional analysis is strongest if, once the offense is detected, prosecution and conviction (or litigation and civil liability) are virtually sure to follow. If, in addition, the probability of detection depends purely on chance, rather than on a prosecutorial decision to allocate more resources to detecting serious offenses; and if any rules excluding certain losses from the fine or damage award are applied no more harshly against defendants who behaved badly than against defendants who behaved well, the expected punishment will then be completely independent of any improvements in a defendant's behavior. In such a case, there will be no difference between the optimal case-by-case multiplier and the optimal constant multiplier (or the optimal con-

stant fine), and all of them will have to satisfy the traditional multi-plier principle.

In all other cases, however, the probability of punishment will respond to improvements in defendants' behavior. In these cases, as we have seen, satisfaction of the multiplier principle will no longer be necessary for optimal deterrence (though it may still be sufficient). For example, the optimal constant multiplier will gener-ally be less than the level called for by the multiplier principle, per-haps even at or below compensatory levels. The optimal fine, too, could also be less than the traditional multiplier principle, though in some cases it could be greater. Moreover, if adjustments to the sub-stantive legal standard are also employed, the optimal fine or dam-age award could be even less. On the other hand, if the substantive standard is relaxed considerably, to the point where a defendant who behaves optimally faces no risk whatsoever of being found in violation of the standard, the penalties for those found in violation could then be substantially *raised* without interfering with optimal deterrence.

In any of these cases, the law faces a choice about which strat-egy to use to achieve optimal deterrence. This is the choice whose investigation I have tried to begin. Preliminarily, I can suggest that the case-by-case multiplier will work best whenever it is extremely important to optimize defendants' levels of activity as well as their levels of care, and/or if it is more efficient to have all the calcula-tions needed for deterrence made anew by a judge or jury in each individual case. On the other hand, one of the other strategies will probably be best if it is more efficient to have these calculations made by a central legislative or administrative body, and/or if there are practical constraints on the maximum penalty the law can assess, thus requiring penalties below the traditional multiplier. One of the alternative strategies may also be best if it is important to preserve a direct relationship between the harmfulness of a de-fendant's conduct and the size of the actual penalty that is imposed, either because such proportionality is constitutionally required, or because it is desirable in order to send the proper symbolic message.

I have stressed that these conclusions are tentative, and doubt-less they could be improved or refined through further analysis. Until we recognize that there is a choice to be made, however, no progress on these issues will even be possible. It is time to recog-nize that the multiplier principle is sufficient but not necessary for optimal deterrence.