Faculty Scholarship

4-2001

Of Two Wrongs that Make a Right: Two Paradoxes of the Evidence Law and Their Combined Economic Justification

Alex Stein
alex.stein@brooklaw.edu

Essay

Of Two Wrongs That Make a Right:
Two Paradoxes of the Evidence Law
and Their Combined Economic Justification

Alex Stein[*]

This essay offers a new rationale for the standard of proof requirements in civil trials. The civil proof doctrine, as traditionally understood, presents two economic paradoxes. First, it focuses on accuracy ex post by requiring judges to reconstruct the relevant events as they unfolded in reality, including the actual damage to the plaintiff, based on the information available at the trial. This retroactive (ex post) accuracy is economically inefficient. For deterrence purposes, only information that had been available to the defendant prior to taking the litigated action (ex ante information) matters. Moreover, accuracy ex post is an investment-dependent opportunity rather than a static good. As such, it fosters a secondary market for competitive adversarial investments in information, which might adversely affect the primary market, that is, the market for goods, services, risks, and precautions. Thus, when prospective litigants are rationally unwilling to commit themselves to the required investments in information, inefficiencies are bound to occur. In such cases, each party will account for the event that he will be wrongfully defeated in the future trial because his opponent's investment in information outscored his. This prospect will foil transactions that are otherwise efficient and chill many other socially beneficial activities.

Second, if the doctrine is nonetheless committed to accuracy ex post, then it should require judges to determine the ultimate probability of the

plaintiff's case by multiplying the probabilities of the relevant entitlement, breach, and damage. Yet, the doctrine refuses to apply the multiplication principle and thus reduces the total number of correct verdicts, instead of maximizing it. On these grounds, the controlling civil proof doctrine has been criticized as economically unsound.

Under the new rationale offered by this essay, the two alleged wrongs make a right because in combination they generate a synergetic mechanism that aligns, to the extent feasible, the ex ante and the ex post probabilities of transgression. This alignment is attained by the combined, but not conjunctive, functioning of the two probabilities: the probability of the litigated entitlement and the ex post probability of the entitlement's breach. The entitlement's probability dominates the defendant's ex ante information, thus adjusting the ex ante probability of breach. This adjustment is achieved due to the visibility element, uniformly featured by legal entitlements: under the definition of virtually any entitlement, the entitlement must both exist and be reasonably ascertainable ex ante, that is, at the time and in the circumstances of its breach. The ex post probability of breach has a different function, namely, to substantiate the allegation that the defendant has actually violated the entitlement. This combined framework secures the appropriate alignment between the ex post and the ex ante probabilities of breach. The plaintiff will prevail at trial only when each probability is preponderant, that is, greater than 0.5, so prospective defendants can safely rely on the probabilities of breach that exist ex ante. Indeed, because the entitlement's probability functions as a misalignment-corrector for the ex post probability of breach, and not as its conjunctive companion, the two probabilities must not be multiplied in determining the ultimate probability of the plaintiff's case.

This doctrinal mechanism also saves litigation expenses. The plaintiffs' litigation effort is substantially alleviated by the doctrinal refusal to apply the multiplication principle. This entails greater hardship for defendants, but they receive offsetting benefits: the doctrine allows potential transgressors—before they become defendants—to rely on their ex ante information. The doctrine thus encourages potential transgressors not to incur uneconomic expenses by acquiring further information that might become available at their subsequent trials.

This essay also supports the existing award system, under which, as a matter of general rule, the winner takes all. Due to the existing constraints in law enforcement, the essay prefers this system over that of probabilistic awards. Under the existing law enforcement constraints, trial awards function as incentives for reducing the number of both underusers and overusers of the adjudication facility. This rationale turns trial awards into a discrete component of civil litigation. The doctrinal mechanism exposed in this essay consequently separates the breach-related and the

damage-related proof requirements. Probability of the litigated damage must therefore also be removed from the multiplication formula, so the conjunction paradox disappears altogether. As demonstrated by the essay, this rationale also necessitates an adjustment in punitive damages. The essay therefore offers the required adjustment.

The essay establishes the above arguments by using the following methodology: it constructs a simple Bayesian model of civil litigation, from which it derives the ideal proof requirement (the "firstbest"); subsequently, the essay compares that requirement with the positive law in a way that accounts for the existing constraints in law enforcement (the "secondbest"). The essay demonstrates that the ideal proof requirement is economically unfeasible under these constraints, which forces the legal system to develop an adequate surrogate. This surrogate is identified by the essay as embedded in positive law.

I.   Introduction

Many people say about instant coffee "the good thing about it is that one doesn't have to drink it." Yet, many such people drink instant coffee. Moreover, they do that on perfectly rational grounds: ground coffee generally tastes better than instant coffee, but it takes more time to prepare it, so the time spared by drinking instant instead of ground coffee can be used for promotion of other objectives that override the person's craving for ground coffee. This explanation, however, contains a puzzle with respect to the transitivity of such people's preferences: those objectives that override their craving for ground coffee might also prevail over their qualified desire for instant coffee, so they could spare even more time and thus further promote their more important objectives if they decided not to drink coffee at all. This puzzle is, of course, unreal. Every person who drinks instant coffee while still preferring ground coffee is, in fact, unwilling to sacrifice her desire for coffee for the sake of other objectives. Any such person is determined to drink coffee, ground or instant. She is prepared to sacrifice her desire for coffee only when its satisfaction tangibly diverts her from the pursuit of some more important objective. In the case of instant coffee, however, there is virtually no diversion because the coffee is instant. Instant coffee is thus rationally consumed by people as a cost-efficient surrogate of ground coffee. Consequently, it becomes those people's second-best.

Law is permeated with second-bests. Typically, the law endorses them when the first-bests are either unattainable or attainable only at a prohibitive price. Descending from the first- to the second-best thus facilitates many of the law's utilitarian tradeoffs. The familiar tradeoff between the ideal deterrence system and its operational costs is, perhaps, one of the most striking examples of that phenomenon. This tradeoff

spawns second-bests that function as economically affordable substitutes for the first-best principles of deterrence. These second-bests are the law's "instant coffee."

This essay focuses on one of the principal offsprings of this important tradeoff: the basic proof requirements for civil trials. According to its common understanding, which can be traced back to Bentham's first principle of procedure,[1] this doctrine promises to maximize the veracity of judicial[2] verdicts and thus to facilitate the implementation of the law's substantive rules that apply to contracts, torts, and so forth. However, as demonstrated by the contemporary scholarly literature, the doctrine fails to deliver on that fundamental promise. First, it reduces, rather than maximizes, the total number of correct verdicts. Second, its prescriptions for factfinding are both costly in implementation and vastly irrelevant to deterrence. These two fundamental flaws are embedded in the doctrine's internal conditions. For these reasons, the doctrine has been accused of being paradoxical or anomalous.

These are very serious accusations that entail far-reaching implications for the law. These accusations therefore merit careful evaluation. As explained below, arguments supporting these accusations appear impeccable. These arguments would, indeed, be irrefutable if we were to accept their basic premise with respect to the doctrine's underlying ambition. According to that premise, the doctrine aims at attaining the first-best; specifically, it aims to maximize the amount of correct verdicts, so that the greatest possible number of wrongdoers may be exposed to liability and the greatest possible number of faultless defendants may be found not liable. This premise, however, is not necessarily correct. The doctrine may well be directed towards achieving another goal that was marked as second-best. Specifically, it may furnish the second-best surrogate for optimal deterrence that may have been chosen for perfectly good reasons. In such a case, the doctrine would have to be understood differently and the accusations leveled against it would be unfounded. Indeed, if such an understanding of the doctrine were available, then the doctrine would adequately respond to these accusations and its alleged paradoxes and anomalies would disappear. But is there a theory that could justify the controlling proof doctrine as a second-best?

This essay provides a positive response to that question by offering the required justificatory theory. This is done by examining two fundamental

---

1. *See* Jeremy Bentham, *The Principles of Judicial Procedure, in* 2 WORKS OF JEREMY BENTHAM 1, 6 (John Bowring ed., 1838-43) ("Of the adjective branch of law, the only defensible object . . . is the maximization of the execution and effect given to the substantive branch of the law."). For an illuminating discussion of Bentham's theory of procedure, see Gerald J. Postema, *The Principle of Utility and the Law of Procedure: Bentham's Theory of Adjudication*, 11 GA. L. REV. 1393 (1977).

2. The terms "judges," "judicial," etc. refer also to jurors.

paradoxes that academic literature has detected in the evidence law. The first is widely known as "the conjunction paradox"; the other is yet to be labeled, so I will call it "the accuracy-maximization paradox" for reasons that will soon become apparent.[3]   When examined separately, each of those paradoxes uncovers an important truth about its underlying doctrinal arrangement: although intuitively appealing, this arrangement is economically unsound.   I will demonstrate, however, that the *combined effect* of the two allegedly paradoxical doctrines is a defensible economic rationale, previously unnoticed.   Under this new rationale, ideal deterrence is unattainable because its attainment is bound to be foiled by adjudicative and other enforcement constraints; consequently, ideal deterrence is substituted for by its economically attractive surrogate.

Part II describes the alleged paradoxes.   The doctrine's surrogate approach is expounded in Parts III and IV.   Part V sketches the implications of the preceding discussion for the economics of punitive damages. Economic analysis of punitive damages recommends them *(inter alia)* as a tool for intensifying deterrence in cases where wrongdoers can escape liability and are consequently underdeterred.   As revealed in Part IV, the existing evidence doctrine already contains an award-upgrading remedy for liability-escaping cases. This implicit remedy must therefore be taken into account in punitive damage decisions.   If it is ignored, punitive damages will exceed their optimal level.   Part VI summarizes the. essay's principal conclusions.

II.   The Two Paradoxes

A.   The Conjunction Paradox

Take an ordinary civil lawsuit in which the plaintiff must establish three independent elements:

(1)   entitlement-related facts (denoted E);

(2)   breach-related facts (denoted B); and

(3)   damage-related facts (denoted D).[4]

---

3. Perhaps I employ here a rather loose notion of "paradox," for each of the discussed doctrinal arrangements can be held paradoxical only if one assumes, as the present essay does, that the law's overriding objective is to maximize utility. Also: for those who would not consider the discussed doctrinal arrangements intuitively appealing, "anomaly" or "fallacy" would be a better label. *See* R.M. SAINSBURY, PARADOXES 1 (2d ed., 1995) (paradox is established only when an apparently unacceptable conclusion is derived by apparently acceptable reasoning from apparently acceptable premises).

4. My assumption that the three elements are mutually independent has been made for the sake of convenience (in order to deal with monadic probabilities alone). Nothing would change if we treated E, B, and D as co-dependent in their occurrence: the same paradox would emerge if we ascribed a 0.75 value to each of the dyadic probabilities—$P(E|B,D)$, $P(B|E,D)$, and $P(D|E,B)$.

Under the ordinary civil proof standard, the plaintiff will win the case if each of those elements (E, B, and D) is more probable than not.[5] Evidence adduced by the plaintiff is clear and convincing, so judges infer from it that the probability of each of the above elements equals 0.75. Consequently, they rule for the plaintiff, a verdict that has a strong intuitive appeal. This verdict also squarely aligns with the law of the land, but not with the laws of probability. Under the controlling evidence doctrine, the plaintiff is entitled to recover because his case is deemed "more probable than not";[6] under the probability laws, however, the overall probability of the plaintiff's case is only 0.42, so his case is more improbable than probable. The product rule for conjunctive probabilities, also known as the multiplication principle, prescribes that $P(E\&B\&D) = P(E) \cdot P(B) \cdot P(D)$,[7] which gives us 0.42. Hence, the plaintiff should lose, not win, however counterintuitive that may appear.[8] By refusing to apply

---

5. *See, e.g.*, 2 JOHN W. STRONG ET AL., MCCORMICK ON EVIDENCE § 339 (5th ed., 1999) (stating that, in most civil cases, the plaintiff must prove his case by a preponderance of the evidence).

6. *See, e.g.*, ELEVENTH CIRCUIT PATTERN JURY INSTRUCTIONS CIVIL, Basic Jury Instruction 6.2 (2000) ("In this case each party asserting a claim or a defense has the responsibility to prove every essential part of his contention by a 'preponderance of the evidence.' . . . Where more than one claim is involved, and when more than one defense is asserted, you should consider each claim and each defense separately; but in deciding whether any fact has been proved by a preponderance of the evidence, you may consider the testimony of all of the witnesses, regardless of who may have called them, and all of the exhibits received in evidence, regardless of who may have produced them. If the proof fails to establish any essential part of a claim or contention by a preponderance of the evidence you should find against the party making that claim or contention."); FIFTH CIRCUIT PATTERN JURY INSTRUCTIONS CIVIL, Cautionary Instruction 2.20 (1999) ("In this case, the plaintiff must prove every essential part of his claim by a preponderance of the evidence. . . . In deciding whether any fact has been proven by a preponderance of the evidence, you may, unless otherwise instructed, consider the testimony of all witnesses, regardless of who may have called them, and all exhibits received in evidence, regardless of who may have produced them. If the proof fails to establish any essential part of the plaintiff's claim by a preponderance of the evidence, you should find for the defendant as to that claim."); STEPHEN M. LACHS ET AL., 1 CALIFORNIA JURY INSTRUCTIONS CIVIL 49 (Paul G. Breckenridge, Jr. ed., 8th ed. 1994) ("Plaintiff has the burden of proving by a preponderance of the evidence all of the facts necessary to establish: [The essential elements of [each separate] [the] claim. The essential elements of [the] [each separate] claim [is] [are] set forth elsewhere in these instructions. In addition to these essential elements, plaintiff has the burden of proving by a preponderance of the evidence all of the facts necessary to establish the nature and extent of the [damages] [injuries] claimed to have been suffered, the elements of plaintiff's damage and the amount thereof.") (brackets in original); LEON D. LAZER ET AL., COMM'N ON PATTERN JURY INSTRUCTIONS, ASS'N OF SUPREME COURT JUSTICES, 1A NEW YORK PATTERN JURY INSTRUCTIONS—CIVIL 59 (3d ed. 2000) (instructing that a party needs to "prevail on *an issue* on which he or she has the burden of proof") (emphasis added); *see also* Fisher v. Vassar Coll., 114 F.3d 1332, 1344 (2d Cir. 1997) (stating that in the usual case there must be sufficient evidence to support a finding in plaintiff's favor on *every element of the claim by a preponderance*); Four Corners Helicopters, Inc. v. Turbomeca, S.A., 979 F.2d 1434, 1438 (10th Cir. 1992) (same); Sandoval v. Hagan, 7 F. Supp. 2d 1234, 1245 (M.D. Ala. 1998) (same).

7. *See* L. JONATHAN COHEN, AN INTRODUCTION TO THE PHILOSOPHY OF INDUCTION AND PROBABILITY 18-19 (1989) (explaining that the multiplication principle can be used to determine the probability of the occurrence of conjunctive events).

8. *See* L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE 58-67 (1977) (explaining that, even when a plaintiff establishes each *individual* element of her claim by a preponderance of the evidence, application of the product rule may bring her case below that threshold).

the product rule in this and similar settings, the law magnifies the number of adjudicative errors in the long run of cases. Thus, out of 100 cases with probabilities identical to ours, the law would produce only 42 correct decisions and 58 erroneous verdicts. Perhaps, it seems, we should allow the probability laws to reshape our legal intuitions and doctrines.[9]

Before proceeding to the next paradox, it is important to notice that the doctrinal refusal to apply the product rule is restricted to elemental facts alone (such as formation of a contract, its breach by the defendant, and the ensuing damage), as opposed to *intermediary facts* (such as those that specify the terms of a transaction offered by the plaintiff and those that specify the defendant's acceptance of those terms). Elemental facts—or "operative facts," in Hohfeld's terms[10]—highlight the lawsuit's constitutive elements, as determined by the controlling substantive law. Intermediary facts are facts that establish elemental facts: in order to establish an elemental fact, one usually needs two or more intermediary facts. This distinction must be kept clear because the law's probabilistic separation is maintained only between elemental facts: if judges were to reason mathematically, they would have to apply the product rule in combining the probabilities of intermediary facts that are necessary for determining an elemental fact. Informally, this requirement is already embedded in the standard-of-proof doctrine: as famously acknowledged in the law, an inferential chain can never be stronger than its weakest link.[11]

B. The Accuracy-Maximization Paradox

In civil cases, the controlling level of proof is that of "preponderance of the evidence": the party whose evidence preponderates over that of her opponent must prevail. This has been termed "the $P > 0.5$ rule."[12] The $P > 0.5$ rule refers to the probability of the lawsuit as determined at trial. As such, it has been credited for its capacity to minimize the overall amount of adjudicative errors: any other rule of decision would produce

---

9. *See, e.g.*, Maya Bar-Hillel, *Probabilistic Analysis in Legal Factfinding*, 56 ACTA PSYCHOLOGIA 267 (1984); David Kaye, *The Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34 (1979); Bernard Robertson & G.A. Vignaux, *Probability–The Logic of the Law*, 13 OXFORD J. LEGAL STUD. 457 (1993) (all favoring application of probability rules in judicial factfinding); Ferdinand Schoeman, *Cohen on Inductive Probability and the Law of Evidence*, 54 PHIL. SCI. 76, 80-82 (1987).

10. WESLEY NEWCOMB HOHFELD, FUNDAMENTAL LEGAL CONCEPTIONS AS APPLIED IN JUDICIAL REASONING 32-35 (Walter Wheeler Cook ed., Greenwood Press, 1978) (1919).

11. *See* 1A JOHN H. WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 41 (Peter Tillers ed., 1983) (explaining that the weakest link in a chain of inferences determines the strength of the conclusion that can be drawn from those inferences).

12. For a brief restatement of this rule and its justifications, both old and contemporary, see Alex Stein, *Allocating the Burden of Proof in Sales Litigation: The Law, Its Rationale, a New Theory, And Its Failure*, 50 U. MIAMI L. REV. 335, 340-43 (1996) (discussing the economic rationales for the $P > 0.5$ rule).

fewer correct decisions than this rule does.   Proof of this creditworthy
capacity is straightforward.   By focusing on the amount, rather than on
substantive consequences, of judicial errors, the error-minimizing objective
treats every error as a fixed disutility unit (u).   Consequently, there is no
difference between errors that harm defendants (false positives) and errors
that harm plaintiffs (false negatives).   There is also no difference between
errors that occur in high-stakes, as opposed to low-stakes, litigation: all
errors are treated as equally bad.   The utility principle therefore ordains
that a party whose case has probability P should win the case whenever
$Pu > (1-P)u$, that is, whenever $P > 0.5$.[13]

In short, judges will maximize the total number of correct decisions
by treating their best chances to arrive at the factually correct result as
decisive.   This   maximization   of   correct   decisions   indeed   appears
economically beneficial: greater accuracy in factfinding gives more space

---

13. *See* David Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked
Statistical Evidence and Multiple Causation*, 1982 AM. B. FOUND. RES. J. 487 (1982); David Kaye,
Naked Statistical Evidence, 89 YALE L.J. 601 (1980) (reviewing MICHAEL FINKELSTEIN,
QUANTITATIVE METHODS IN LAW: STUDIES IN THE APPLICATION OF MATHEMATICAL PROBABILITY
AND STATISTICS TO LEGAL PROBLEMS (1978)).

When the disutilities that false positives and false negatives respectively produce are equal—that
is when P=0.5—plaintiffs should lose. This decision rule eliminates the enforcement costs that would
be incurred if plaintiffs were allowed to recover. Furthermore, by allowing plaintiffs to recover when
P=0.5, the legal system would raise the number of unmeritorious lawsuits, thus incurring greater
litigation costs. *See, e.g.*, Ralph K. Winter, Jr., *The Jury and the Risk of Nonpersuasion*, 5 LAW &
SOC'Y REV. 335, 337 (1971). Arguably, this decision rule is optimal also because "taking" is generally
more harmful than "not giving." This widespread perception has been justified by the diminishing
utility of income. RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 604 (5th ed. 1998). But
perhaps it has to do more with the "endowment effect" that causes people to put greater value on
psychologically vested, as contrasted with psychologically unvested, rights. *See* Amos Tversky &
Daniel Kahneman, *Rational Choice and the Framing of Decisions*, 59 J. BUS. S251, S260 (1986),
available at http://www.jstor.org.

A different conclusion would be arrived at if the law's postulated goal were to minimize large
errors, estimated in disutility units as u^2, as opposed to simply u (by plausibly invoking the diminishing-
utility-of-income assumption). On that theory, the $P > 0.5$ rule should give way to the probability-based
recovery regime that splits the stakes in accord with the probabilities favoring each party's case. The
total amount of disutility that would be incurred under this regime would equal to:

$$p_1(p_2u)^2 + p_2(p_1u)^2 = p_1p_2u^2,$$

with p_1 and p_2 denoting, respectively, the probabilities favoring plaintiffs' and defendants' allegations.
Under the $P > 0.5$ rule, the total disutility amount would equal either p_1u^2 (when defendants win) or p_2u^2
(when plaintiffs recover). Because both p_1 and p_2 range between 0 and 1:

$$p_1p_2u^2 \le p_1u^2; \; p_1p_2u^2 \le p_2u^2.$$

Probability-based recovery consequently becomes superior to the $P > 0.5$ rule. *See* Neil Orloff & Jery
Stedinger, *A Framework for Evaluating the Preponderance-of-the-Evidence Standard*, 131 U. PA. L.
REV. 1159 (1983). The same policy should apply in recurrent transgression cases. *See infra* notes 59-
60 and accompanying text.

to the controlling substantive law. It implies that liability would be allocated more accurately, so there would be less under-deterrence (because more wrongdoers would be held liable) and less overdeterrence (because fewer faultless defendants would be held liable).[14] The $P > 0.5$ rule therefore has a strong intuitive appeal.

Mitchell Polinsky and Steven Shavell have recently developed a model that further emphasizes the advantages of the $P > 0.5$ rule.[15] In this model,

$g =$ the expected gain from a potentially harmful action contemplated by a person;

$p =$ probability that the person will be detected as a possible violator of the relevant legal standard;

$f =$ a pecuniary sanction (say, a fine that derives from the amount of harm associated with the contemplated action) that would be imposed on the person by the court if it finds him liable;

$\epsilon_1 =$ probability of the type-I error (false positive), that is, probability that a nonliable defendant will be mistakenly found liable;

$\epsilon_2 =$ probability of the type-II error (false negative), that is, probability that a liable defendant will be mistakenly found not liable.

For the sake of simplicity, this model assumes that the relevant actor is neutral towards risk, a standard assumption also made by the present essay. The model further assumes that the probability of the person's (right or wrong) detection as a possible violator (p) is constant. This assumption is also made for the sake of simplicity: detection errors are bound to occur, and their actual rate has no impact on the lesson to be learned from the model. If detection errors could be reduced, the incidence of adjudication errors would obviously decline.

Based on the above assumptions, the person would take the contemplated action when

$$g - p(1-\epsilon_2)f \; > \; -p\epsilon_1 f.$$

The left side of this inequality represents the difference between the person's gain and expected loss that arise from the harmful action. The person's expected loss equals the fine, multiplied by the person's probability of being detected and subsequently found liable. The right side of the inequality is a negative sum that represents the person's initial

---

14. *See, e.g.*, Louis Kaplow & Steven Shavell, *Accuracy in the Determination of Liability*, 37 J.L. & ECON. 1, 10-11 (1994) (arguing that greater accuracy produces better deterrence, which avoids costlier law-enforcement efforts).

15. *See* A. Mitchell Polinsky & Steven Shavell, *The Economic Theory of Public Enforcement of Law*, 38 J. ECON. LITERATURE 45, 60-62 (2000).

position: even if the person does not take the contemplated action, he may be erroneously detected as a violator and subsequently found liable. This initial loss is predicated on the existence of errors in the legal system, and its magnitude would depend on the size of the fine and on the incidence of both detection and judgmental false positives. The person, therefore, would rationally take the contemplated action if its outcome improves this initial position. Consequently, the action would be taken even if the left side of the inequality yields a negative sum, as long as that sum is less than $p\epsilon_1 f$. This implies that deterrence will be reduced not only by false negatives, but also—albeit less trivially—by false positives.

Formally, the potentially harmful action will be taken if
$$g > p(1-\epsilon_2)f-p\epsilon_1 f,$$
that is, if
$$g > p\,f(1-\epsilon_2-\epsilon_1).$$
This formula makes clear that both types of error ought to be reduced in order to optimize deterrence and thereby augment social welfare. False negatives dilute deterrence and thereby reduce social welfare by lowering the expected fine for potential violators. False positives dilute deterrence less trivially: they do so by eroding the difference between the fine expected from violating the law and from not violating it.[16] For obvious reasons, false positives also chill socially desirable activities. For example, if courts too often fail to distinguish between negligent and careful truck-drivers, drivers who are both rational and careful may then decide not to drive their trucks. The lesson to be learned from this analysis is straightforward: the optimal proof requirement is one that maximizes the total number of correct decisions, an outcome that the $P > 0.5$ rule attains.

From a deeper economic perspective, however, the $P > 0.5$ rule seems to be misdirected: it turns out that the law has chosen the wrong object for maximization. Facts pertaining to a litigated event are reconstructed by judges on the basis of evidence that becomes available at trial, that is, after the event (ex post). By using that evidence, judges must determine the probability of the plaintiff's allegation that the defendant violated the litigated entitlement. Two sets of information, not one, thus appear to be at work, and they also differ from each other: information available to judges ex post is not the same information that was available ex ante to prospective transgressors. The two sets of information could be aligned if individuals had enough incentives for improving their earlier decisions by acquiring more information (e.g., through experts or legal advice). This would make the individuals better informed about the consequences of their actions in terms of both harm and liability. Accuracy ex post promotes this

---

16. *Id.* This point was originally made by I.P.L. Png, Note, *Optimal Subsidies and Damages in the Presence of Judicial Error*, 6 INT'L REV. L. & ECON. 101 (1986).

objective by inducing individuals to acquire information ex ante: failure to do so increases the risk of the individual's liability. However, as noted by Louis Kaplow, "individuals will be induced to learn information ex ante only if their benefits from adjusting behavior exceed the cost of the information."[17] This condition is unlikely to be satisfied when the two sets of information are substantially misaligned. In numerous cases, the required alignment would be exorbitantly expensive, if not altogether unattainable.[18] Moreover, normative postulation that individuals ought to be induced to acquire more information ex ante does not necessarily establish that accuracy ex post is the best inducement. A direct requirement that individuals obtain the best information that can practicably be obtained prior to acting in a potentially harmful way might provide a better inducement. This direct inducement is already embedded in the negligence doctrine and in many other liability rules that are based on objective, rather than subjective, standards of conduct.[19]

Therefore, if the law is determined to use its civil liability mechanisms, such as torts and contracts, for setting optimal deterrence for economic actors, then it should focus on the ex ante probability of trans-gression as determined by a reasonable economic actor in deciding whether to take the (subsequently) litigated action. The ex post probability of litigated events is a poor proxy for the actor-centered probability. In such circumstances, efforts directed at attaining accuracy ex post therefore amount to a waste of resources. Moreover, by inducing excessive invest-ments in the acquisition of information, accuracy ex post produces various effects that are detrimental to efficiency.[20]

Take, for example, contractor C, who builds a fence for house-owner H and guarantees that the fence will withstand bad weather. At that point in time, C properly estimates the risk that the fence will not withstand bad weather and fall down as amounting to 5%. This estimation turns out to be wrong ex post because there is something that C does not know and cannot plausibly ascertain. A year later, the fence falls down after being exposed to severe winds and floods. H sues C for damages, claiming that the fence did not withstand the weather. C contends that the fence fell down because someone must have damaged its foundations during the year.

---

17. Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307, 317 (1994).

18. *Id.* at 330-31.

19. *See, e.g.*, W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 32, at 185 (5th ed. 1984) (noting that negligence standards require individuals to seek information prior to acting in a potentially hazardous way).

20. *See* Kaplow, *supra* note 17, at 365-66. For an earlier analysis of this issue that points to similar and several other problems, see George P. Fletcher, *Paradoxes in Legal Thought*, 85 COLUM. L. REV. 1263, 1284-88 (1985).

At the ensuing trial, an expert witness testifies that the damage-through-bad-weather scenario advanced by H is more likely than C's competing scenario because in the past year the weather was particularly bad. The expert further clarifies that H's scenario has a 60% likelihood. The court finds this testimony credible. By applying the preponderance-of-the-evidence standard, it holds that the damage was caused by bad weather.

From the deterrence perspective, the ex post probability of the damage-through-bad-weather scenario could not be rationally accounted for by C at the time he built the fence. The court and the parties therefore wasted resources in their combined efforts to determine that probability. The court's holding in favor of H's scenario is doubtlessly more accurate than the alternative. Such decisions would indeed produce more correct than incorrect verdicts in the long run of cases (subject to the required removal of the conjunction paradox). Thus, out of 100 similar cases 60 would be decided correctly and only 40 incorrectly. Yet, the court's holding will induce no changes in behavior because any rational contractor situated similarly to C would estimate the relevant risk as amounting to only 5%. In accordance with this estimate, the contractor will not make an increased investment in the construction to further prevent it from falling down. He will not make such an investment even if he is aware of the $P > 0.5$ rule and its accuracy-enhancing effect. This awareness will have no impact on the contractor's behavior because he is unaware of the information that might become available at his trial. Under these conditions, the contractor will rationally assume that any ex post information will be distributed randomly across cases. He will, of course, speculate that he may have underestimated the risk that the fence will fall down due to bad weather. He will also speculate that the court might make a finding to that effect, based on the information that may become available at his trial. These speculations, however, will be treated by him as equally plausible as speculations that go in the opposite direction. Being indifferent between the conflicting speculations, the contractor will cancel them out. Accordingly, he will base his decision on the information available at the time of the decision.[21]

---

21. Negligence regimes mitigate this problem by conditioning the defendant's liability for a damaging occurrence on the foreseeability of that occurrence. If the occurrence was not foreseeable, the defendant will not be liable for the damage. Judges must determine this issue on the basis of information that was reasonably available to the defendant prior to taking the litigated action. Information that surfaced after the event is largely irrelevant to this issue, so judges must not use it. *See, e.g.*, KEETON, *supra* note 19, § 31 at 170 ("The actor's conduct must be judged in the light of the possibilities apparent to him at the time."); Fletcher, *supra* note 20, at 1285 (same).

This mitigation of the problem is far from sufficient because foreseeability is an indivisible legal standard that judges apply only ex post. Application of this standard is dichotomous: a damaging occurrence is either foreseeable or not. There is no such thing, for example, as a 5% foreseeability, so in the present case C would have to consider whether a 5% risk to the fence makes the potential

In the long run, this decision should also prove perfectly rational. Contractor *C* must account for the possibility that the ex ante and the ex post information may not be similar. Yet, under the given facts, there is no reason to believe that the discrepancy between the two types of information will be systematically slanted in a particular way. *C*'s assessment of the risk may turn out to have been too optimistic in some cases and too pessimistic in others, but there is no reason to believe that the former cases—to which the present example belongs—will outnumber the latter. Therefore, amongst the fences constructed by *C*, only 5% will actually collapse due to bad weather, so he must not be held responsible for more than 5% of the fences. Holding *C* responsible for more than 5% of the fences would produce excessive deterrence.

Arguably, houseowner *H* should still prevail in his lawsuit against *C* under the scenario that his fence belongs to the above 5%. After all, this scenario is more likely than not. Holding *C* liable would therefore not expand his general liability for fallen fences beyond 5%. By contrast, holding *C* not liable in this and similar cases would bring his liability below 5%, in which case the deterrence would be insufficient. This argument is correct only if one assumes that future information will not be slanted in any particular direction. However, this is too crude an assumption to proceed upon if one takes into account the creation and the operation of a secondary market for trial information—a factor that I consider in the next few paragraphs. But even if one takes this argument as essentially correct, one must still consider the actions of the other 99 houseowners who commissioned their fences from *C*. Any such person will sue *C* in the event that her fence falls down. Consequently, 100 lawsuits will be adjudicated; 200 experts will submit rival opinions, over which 200 lawyers will conduct adversarial warfare; and, finally, 100 judges (and possibly 1,200 jurors) will have to work hard on identifying the fences that collapsed due to bad weather. This massive investment in factfinding can only be justified if it is economically necessary.

But is it really necessary from an economic point of view? Why not devise a cheaper mechanism that would allow every houseowner—damaged and undamaged alike—to recover from *C* a sum that equals, say, 5% of the

---

damage fully foreseeable. Application of this standard also depends on how judges will treat the relevant risk, an ex post factor of which *C* is informed only partially at the time of contemplating his action. *See generally* Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992). Furthermore, the foreseeability standard applies only to the type, rather than size, of the relevant harm, so it does not affect the quantum-of-damage decisions. *See* KEETON, *supra* note 19, at 290-93. The size of the future damage and the corresponding extent of legal liability therefore remain unascertainable ex ante. These factors can only be approximated prior to the event. *See infra* notes 26-33 and accompanying text.

houseowner's payment for the fence plus 5% of the external damage that a falling fence produces in the average case?[22]   Under this framework, proof of each houseowner's payment to C would be cheap. Determination of the expected average damage figure might be costly, but because this knowledge needs to be acquired only once, its costs would be spread across many cases. This mechanism would therefore be much cheaper than conventional adjudication, while its effect on deterrence would be at least as good.[23]

This mechanism's impact on general efficiency would, in fact, be better than that of conventional adjudication.   Accuracy ex post is expensive: it would require substantial investments in information and other litigation efforts on the part of both C and the houseowners. Both C and the houseowners may be *antecedently* unwilling to commit themselves to such investments.   Instead of making that commitment, each party will consider his prospect of being defeated at trial due to erroneous fact-finding.   Each party will consider it not only as a general prospect. He will also consider the event that he will be defeated at trial because his opponent invested more in the litigation than he did.   Before entering into each fence-building transaction, C will therefore account for his prospect of paying more in liability damages.   The houseowners, in turn, will consider their prospect of obtaining less compensation.   Aware of his increased liability risk, C will attempt to charge more for his fences. Aware of their risk of receiving less in money damages, the houseowners will only be willing to pay less.   Consequently, transactions that could otherwise be efficient will not be made.[24]   Adjudicative accuracy would thus produce anomalous effects that sharply contrast with its strong intuitive appeal. Hence, it is paradoxical.

To sum up, accuracy ex post is an investment-dependent opportunity rather than a static good.   Under this standard, an individual who obtains more evidence to credibly support her case has a greater chance to prevail at trial. This standard therefore fosters a secondary market for competitive adversarial investments in information.   To obtain accuracy ex post, litigants must appropriately invest in the determination of facts (both before and during trial), and the level of their investments will depend on competition.   This creates a secondary market, that is, a market for evidence and associated litigation efforts.   This market will affect the

---

22. By removing further liability from C, this mechanism would induce those houseowners who are averse towards risk to appropriately insure their fences immediately after their construction.

23. *See* Kaplow, *supra* note 17, at 330-31.

24. Theoretically, C and his clients may enter into a contractual arrangement that would replace conventional adjudication with the mechanism I suggested.   However, asymmetric information, coordination problems, and high transaction costs would usually frustrate such arrangements.

individuals' conduct in the primary market, that is, in the market for goods, services, risks, and precautions. The secondary market may, indeed, exert an adverse effect on the primary market. Thus, when prospective litigants are rationally unwilling to commit themselves to the required investments, inefficiencies are bound to occur. In such cases, each party will account for the event that he will be wrongfully defeated in the future trial because his opponent's investment in information and other litigation efforts outscored his. This prospect will foil transactions that are otherwise efficient and chill many other socially beneficial activities. Take, for example, a careful truck-driver who rationally considers his prospect of being sued for damages allegedly caused by his negligence. If he is rationally unwilling to commit himself to a substantial investment in information and other litigation expenses, he will have to consider the possibility of being wrongfully defeated by a claimant whose litigation investment outscores his. The truck-driver may consequently decide not to drive his truck. If the truck-driver works for a company that assumes full liability for damages associated with his negligent driving, then the company will have to account for a similar prospect. The company will consequently attempt to charge more for its services, which may foil some of its transactions. Accuracy ex post is socially beneficial only when its attainment is relatively inexpensive; but when its attainment is inexpensive, it will not substantially differ from accuracy ex ante.[25] Therefore, the law should focus on accuracy ex ante rather than on accuracy ex post.

From the same perspective, the law creates a distortion by allowing plaintiffs to recover the whole damage[26] instead of its expected average value, as determined by using the ex ante probability of transgression. In contemplating a potentially damaging action, an economic actor compares her damage-avoidance costs with the expected, rather than the actual, damage, that is, with the prospective damage, as reduced by the probability of its (rationally estimated) non-occurrence. In the average case, this estimation corresponds to the average level of harm associated with the type of actions to which the contemplated action belongs. Basing the trial award system on the expected average level of harm thus taxes the wrongdoers for their externalities in an efficient Pigouvian way. Under that system, damages to be paid by a wrongdoer equal a tax that would have been imposed on him by the state if he were to pay in advance for his

---

25. As already mentioned, individuals should indeed be encouraged to expand the informational base upon which they decide to act, if the expansion is relatively inexpensive.

26. By "damage" I mean any remedy that can be plausibly reduced to money, so my discussion equally applies to specific-performance lawsuits brought under the contract law, as well as to any injunctive relief (in which case, the remedied expectation, reliance, or proprietary interest would be treated as damage).

harmful externalities.[27] In reality, some wrongdoers inflict harm that is
greater than the average; some inflict harm that goes below its average
amount (including zero harm); and others inflict the average harm. For
example, wrongdoer A may inflict damage that amounts to 10 and
wrongdoers B and C may respectively inflict damages that equal 20 and 30.
For the sake of simplicity, let it be assumed that the ex ante probability of
inflicting each damage is 0.5 and that none of the (then) potential
wrongdoers can estimate in advance the magnitude of the future damage.
In that setup, the total amount of the expected average damage equals 30.
An advance externality tax imposed on A, B, and C would divide this
damage equally, so each of them would pay 10 (the expected average
damage in each case would equal $0.5 \cdot 20 = 10$). None of them, therefore,
would need to be forced to pay more than 10 in money damages. This
compensation amount would be efficient because in the absence of case-
specific information that could be available ex ante, A, B, and C should not
be forced into avoiding the damage in their individual cases if the cost of
avoiding it exceeds this amount.[28] Forcing them to do so would be mani-
festly inefficient. Although their individual victims have sustained damages
in the amounts of 10, 20, and 30, respectively, those victims should seek
their redress elsewhere, especially because the compensation case made out
by each of the victims is merely probable rather than certain. Ideally, a
Pigouvian tax should be collected from every potential wrongdoer, so the
total amount of the tax would allow the state to fully compensate the
victims (or to promote other socially useful policies).[29] The law,
however, prefers the winner-takes-all principle over this approach. The
underlying motivation of the winner-takes-all principle is transparent: the
$P > 0.5$ rule reputedly maximizes the number of correct verdicts; by
contrast, any probabilistic recovery rule would yield zero correct decisions,
which would frustrate the law's accuracy-maximizing objective. This
objective, however, is misdirected. An efficiency-oriented legal system
should value adjudicative accuracy only when it is instrumental to
efficiency and not for its own sake.[30]

---

27. *See generally* William J. Baumol, *On Taxation and the Control of Externalities*, 62 AM. ECON.
REV., June 1972, at 307 (delineating Pigou's tax criteria and identifying their operational limitations).
See also Henry E. Smith, *Ambiguous Quality Changes from Taxes and Legal Rules*, 67 U. CHI. L.
REV. 647, 679-96 (2000) (comparing Pigouvian taxes with liability rules as instruments for inducing
efficient conduct).

28. Pollution cases would be different because the aggregated pollution level is a pivotal factor.
The latest pollution unit that takes the pollution above its tolerable level may consequently be treated
as more harmful than the previous pollution units. Pollution and similar cases therefore deserve special
treatment. *See* POSNER, *supra* note 13, at 412.

29. *See generally* Christopher H. Schroeder, *Corrective Justice and Liability for Increasing Risks*,
37 UCLA L. REV. 439 (1990) (advancing a corrective justice theory that imposes compensation duty
upon risk-creators even in cases in which no actual damage was inflicted and that channels the
compensation money into a general compensation fund).

30. *See generally* Kaplow, *supra* note 17 (examining the value of adjudicative accuracy from an

In short, accuracy in the adjudicative assessment of the plaintiff's damage is beneficial only when it aligns with the ex ante assessment of the damage by a rational economic actor. If such an actor is unable to cost-efficiently assess the actual damage that would result from her contemplated action, then she will base her action on the average damage. In such cases, spending resources on determining the plaintiff's actual damage would be wasteful.[31] Such cases are far from exceptional: the specifics of the plaintiff's damage are typically unknown ex ante to his (then) prospective injurer.[32] The present level of investment in the adjudicative determination of plaintiffs' damages is therefore socially wasteful.[33]

The difficulties presented by the two paradoxes are probably not insurmountable. By using different economic[34] criteria, an arguable case can be constructed in favor of both the non-multiplication principle[35] and the $P > 0.5$ rule.[36] This essay pursues a different strategy. It accepts the normative criteria that denounce each of these evidentiary arrangements *individually* but argues that the arrangements should be examined *jointly*. Accordingly, this essay evaluates the combined effect of the two evidentiary arrangements by using these normative criteria as a frame of reference. The motivation behind this endeavor must now be apparent: the $P > 0.5$ rule reputedly aims to maximize accuracy, an objective that the non-multiplication principle systematically frustrates. It is therefore

---

economic perspective).

31. *Id.* at 313-16.

32. *See* Louis Kaplow, *Accuracy in Adjudication, in* 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 1, 4 (1998). In contract law, the prospective injurer's antecedent knowledge can be expanded by using penalty-default rules: arguably, the limitation of liability famously set in *Hadley v. Baxendale*, 156 Eng. Rep. 145 (1854), is one such rule. *See* Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 YALE L.J. 87 (1989).

33. *See* Kaplow, *supra* note 32, at 4.

34. For non-economic justifications of the doctrine's suppression of the multiplication principle, see Ronald J. Allen, *The Nature of Juridical Proof*, 13 CARDOZO L. REV. 373, 409-20 (1991) (justifying the suppression with the demand that each party produce a single coherent account of the events, which precludes defendants from relying on disjunctive scenarios that are mutually inconsistent); Alex Stein, *The Refoundation of Evidence Law*, 9 CAN. J.L. & JURIS. 279, 336-38 (1996) (justifying the suppression by equality between plaintiffs and defendants).

35. *See, e.g.*, Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439, 450-54 (1986); Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1512-14 (1999) (all justifying the suppression of the multiplication principle using reasons consistent with utility); Alex Stein, *An Essay on Uncertainty and Fact-Finding in Civil Litigation, with Special Reference to Contract Cases*, 48 U. TORONTO L.J. 299, 311 n.27 (1998); *see also* Saul Levmore, *Conjunction and Aggregation* (2000) (unpublished manuscript on file with the *Texas Law Review*) (arguing that group decisionmaking by jurors heightens the standard proof requirement to a level that justifies suppression of the multiplication principle).

36. *See, e.g.*, James Brook, *Inevitable Errors: The Preponderance of the Evidence Standard in Civil Litigation*, 18 TULSA L.J. 79 (1982) (justifying the $P > 0.5$ rule on both economic and equality-based grounds); Stein, *supra* note 34, at 333-36 (same).

possible that the law suffers from schizophrenia. Alternatively, the ostensibly bizarre interaction between the two doctrines may, after all, produce economically defensible outcomes. The story unfolded by this essay has a happy end: in what follows, I will demonstrate that the second possibility is more plausible than the first.

III. The First-Best Proof Requirement

This part of the essay elicits the first-best proof requirement from the optimal liability standard, already identified in Part II. This is done by using a simple Bayesian probability model. Part IV compares the resulting proof requirement with the combined effect of the non-multiplication principle and the P>0.5 rule. This discussion develops a defensible economic rationale for the two ostensibly paradoxical arrangements. The significance of this endeavor is straightforward: it is important to know whether the existing proof requirements require reform. Making adjustments in legal proof requirements is basically costless, so any improvement in efficiency that could thereby be attained would be attained for free.[37]

> In a paradigmatic civil trial, the following probabilities are at work:
> (1) probability of the litigated entitlement;
> (2) probability of the entitlement's breach as a potential cause of the litigated damage; and
> (3) probability of the litigated damage as resulting from the entitlement's breach.

Under the normative criteria set forth in Part II, each of those probabilities is an ex ante probability, reconstructed by judges on the basis of information that was cost-efficiently available to the defendant prior to her embarkment on the litigated action. Also: damage focused upon by this model is the average damage figure, rather than the damage actually sustained by plaintiffs. As already explained, damages actually sustained are usually not ascertainable ex ante at a cost that prospective injurers can reasonably bear. The actual (ex post) damage figure thus becomes largely irrelevant for deterrence. Defendants will consequently be held liable for the risk engendered by their actions. From the deterrence perspective, any person externalizing risk should assume liability even when he ultimately causes no damage to another person. This precept, however, belongs only to the ideal-type model, isolated, for methodological reasons, from the existing drawbacks in law enforcement that I consider later in the essay. The existing enforcement constraints, dealt with in Part IV, generally prevent imposition of liability for bare risk. Finally, my discussion

---

37. *See* Kaplow, *supra* note 17, at 358-59.

proceeds on the standard assumption that prospective injurers and other relevant actors are risk-neutral.

The plaintiff's holding of the litigated entitlement is not factually dependent on the entitlement's breach and the ensuing damage. The entitlement's probability is therefore denoted as $P(E)$. The entitlement's breach depends on the entitlement's existence, so probability of breach is dyadic. This probability is accordingly denoted as $P(B|E)$. The litigated damage depends on the existence of the entitlement and on the causally relevant breach. Probability that this damage will be[38] sustained by the plaintiff is thus denoted as $P(D|E,B)$. The magnitude of the plaintiff's average damage is marked as d.

Using the Bayes Rule, we can now formulate the amount of the defendant's liability that would be optimal under our normative criteria. First, this amount equals

$$P(E) \cdot P(B|E) \cdot P(D|E,B) \cdot d. \tag{1}$$

Second, legal rules that determine entitlements and corresponding breaches presuppose that the combination of the two is ordinarily harmful. Any other assumption would make those rules economically redundant (at best). Consequently, as a matter of legal standard, we are entitled to assume that *some* damage is ordinarily associated with the relevant entitlement and its breach. Because the magnitude of the actual damage is unknown prior to its infliction, we are also entitled to assume that it equals the average damage. If so, because we want to focus precisely on the average level of harm, we can make an important simplifying assumption: $P(D|E,B)=1$. Making this assumption would not be possible if we were to focus on the actual, rather than average (and consequently standard) level of harm. The optimal liability amount thus transforms into

$$P(E) \cdot P(B|E) \cdot d. \tag{2}$$

Let us now have a close look at $P(B|E)$ (probability of breach, given the entitlement). This probability equals $P(B) \cdot P(E|B)/P(E)$.[39] In other words, it equals the prior probability of breach multiplied by the relevancy quotient;[40] the latter is determined by the probability of the entitlement,

---

38. Future tense is employed here in order to emphasize the forward-looking character of the liability standard.

39. Under the Bayesian "inversion theorem," this derivation is straightforward: $P(E\&B)=P(E) \cdot P(B|E)=P(B) \cdot P(E|B)$; hence, $P(B|E)=P(B) \cdot P(E|B)/P(E)$. *See* WILLIAM KNEALE, PROBABILITY AND INDUCTION 127-32 (1949).

40. *See* RICHARD JEFFREY, PROBABILITY AND THE ART OF JUDGMENT 109 (1992). More often, this quotient carries the name of "likelihood ratio." *See* BERNARD ROBERTSON & G.A. VIGNAUX, INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM 16-20 (1995).

given breach, divided by the general probability of the entitlement. Note that in this formulation, P(B) is a probability of breach unupdated by the relevancy quotient P(E|B)/P(E). Evidence pertaining to both breach and the underlying entitlement will, however, be considered in determining P(B), so this probability will be affected by the entitlement-related evidence to the extent relevant to the prospect of breach. In determining P(B), the prospective injurer must therefore ask herself the following question about her contemplated action: *How likely is it that the action belongs to the category of actions that inflict the damage protected against by the relevant entitlement?* The same question must be addressed by a judge who subsequently steps into the injurer's shoes. As will soon become clear, P(B) is a pivotal factor in my analysis. Its features are important and I allude to them later in the essay.

My formulation of the optimal liability amount can be simplified further: P(E|B)=1. Indeed, it is certain that an entitlement must exist whenever its breach occurs, so that P(B|E)=P(B)/P(E). Incidentally, this formulation gives us an interesting insight: the greater the general probability of the litigated entitlement, the lesser the impact of the entitlement-related evidence on the posterior probability of breach. For example, when the entitlement is certain, it becomes logically irrelevant to the issue of breach: when P(E)=1, the prior probability of breach remains unmodified. When the entitlement is recurrent, it transforms the prior probability of breach into an only slightly higher posterior probability. But when the entitlement is rare and thus has a relatively low probability of occurrence and detection, it raises the posterior probability of breach quite substantially. This has an intuitive explanation: breaches are avoided more easily and more often when entitlements are more recurrent and consequently more familiar and provable. Breach of a rare and correspondingly less familiar entitlement is usually more onerous to avoid and more tempting to commit. Acquisition of information pertaining to a non-recurrent entitlement is, indeed, costlier than with standard entitlements. The relative frequency of transgressions thus clearly depends on the informational (and other) costs of avoiding those transgressions. It should also be noted that because P(B|E)≤1, P(B) can never be greater than P(E). Empirically, P(E) is always greater than P(B): legal entitlements are more recurrent than their respective breaches.

This point is of some importance to the discussion taking place in Part IV, and I will return to it there.[41] Here, we focus on a different issue: our formulation of the optimal proof standard is yet to be completed. By substituting P(B|E) with P(B)/P(E) in formula (2), we receive

---

41. *See infra* note 49 and accompanying text.

$$P(B) \cdot d \qquad\qquad (3)$$

as representing the optimal liability amount. Correspondingly, the controlling proof requirement (to be employed as a damage award quotient) would be any value of $P(B)$.

The resulting proof requirement thus features three economically attractive factors:

(1) The ex ante probability of breach, $P(B)$, is a function of the prospective injurer's rational response to the question "How likely is it that my contemplated action belongs to the category of actions that inflict the damage protected against by the relevant entitlement?";

(2) d is the average damage associated ex ante with the contemplated action, not the damage actually incurred ex post;

(3) $P(B) \cdot d$ is the expected average damage that sets the optimal level of precaution for prospective injurers.

This combination of the ex ante probability of breach and the average damage is obviously attractive: it generates optimal deterrence at minimal informational costs and precludes excessive investment in adjudicative accuracy.

This outcome is remarkable for two additional reasons. First, it requires judges to determine a non-conjunctive monadic probability of breach alone. The multiplication principle consequently does not apply. This important refinement is plainly necessary for maintaining the proper deterrence standard. The refinement is not merely formal: it has to do with the functional significance of the elemental facts. Deterrence would go far below the desired level if elemental-fact probabilities (all below 1) were to be multiplied by each other. Second, consonantly with exempting plaintiffs from establishing damages personally sustained, this formulation imposes no restrictions upon *locus standi*. Apparently, a non-damaged plaintiff can come to court and recover damages in the name of deterrence. This is a well-familiar problem of all deterrence-driven arrangements: such arrangements, as affecting relevant individuals, are concerned primarily with imposing duties, not with conferring entitlements. In principle, therefore, violations of deterrence-driven duties must be prosecuted by the state. The state must collect a Pigouvian tax equal to the amount of $P(B) \cdot d$ from any person who externalizes risk. Liable defendants would consequently pay the trial awards to the state (in the form of taxes or fines) rather than to the actual victims of violations.[42] If so, then why not allow any plaintiff to sue any defendant on behalf of the state?

---

42. *See* POSNER, *supra* note 13, at 209.

This familiar problem has a familiar solution. Any deterrence system must be supplemented by an efficient enforcement mechanism that will make the system work. The state-prosecution mechanism is manifestly inefficient for remedying contract, tort, and similar violations: costs involved in searching for and prosecuting such violations are enormous. Thus, when a person externalizes risk, the risk is often invisible until it materializes into a tangible damage. The state cannot therefore exercise efficient control over impositions of bare risk. The state is also unable to police many, if not most, damaging occurrences. The state-prosecution mechanism is consequently applied only in special regulatory areas, such as criminal law, that involve particularly harmful (and therefore regulated) activities. As a result of this enforcement constraint, contract, tort, and similar violations are prosecuted by self-interested individuals upon whom the state has conferred a license to sue. This privatization would be doomed to failure if the licensed individuals had no incentives to sue. Therefore, by subsidizing adjudication and by allowing such individuals to recover for their damages, the state provides them a sufficient incentive to sue.[43] Furthermore, by giving the license to sue *only* to those individuals who claim to have sustained personal damages, the state fosters zealous advocacy that enhances the integrity of adjudication through incentives to gather more information and to appeal against wrong decisions.[44] These enforcement-related factors must be accounted for in devising liability principles that depend on adjudication. The P(B)·d principle must therefore be treated as a substantive economic ideal that would have to be adjusted and thereby compromised in order to accommodate enforcement and other real-life constraints. The need to provide plaintiffs with the right incentives to sue is one of those constraints. Other constraints derive from the ex post character of adjudication. Adjudication always follows, rather than precedes, the litigated event. Facts necessary for adjudication are therefore always reconstructed ex post. Determination of the ex ante probability of transgression thus becomes costly. Furthermore, in a privatized litigation, private facts are easier to prove than social facts. The plaintiff's private damage is, therefore, generally easier to establish than the average damage figure. The legal system's resort to the second-best consequently becomes inevitable.

With these conclusions in mind, I will now compare the normative with the positive: the optimal and the existing proof requirements will now be measured against each other. This comparison reveals that the legal

---

43. From that perspective, "private law," as conventionally understood, should thus be perceived as a privatized public law.

44. *See* Mathias Dewatripont & Jean Tirole, *Advocates*, 107 J. POLIT. ECON. 1 (1999) (analyzing the social benefits that zealous advocacy provides).

system employs the existing proof requirement as an economically justified surrogate for the optimal requirement. The existing proof requirement is consequently justified as an economically necessary second-best.

IV. The Second-Best: Justifying the Existing Proof Requirement

Under existing doctrine, a lawsuit will succeed if each of its following three elements is more probable than not:
  (1) $P(E)$ = probability of the litigated entitlement;
  (2) $P^*(B)$ = probability of the entitlement's breach as a potential cause of the litigated damage; and
  (3) $P(D)$ = probability of the litigated damage as resulting from the entitlement's breach.
Within this framework, $P^*(B)$ is the ex post probability of breach that must be determined by judges on the evidence presented at trial. This probability is marked with an asterisk in order to be distinguished from the ex ante probability of breach $P(B)$.

The doctrinal and the optimal proof requirements are similar in one important respect: both of them do not apply the product rule. Under both requirements, probabilities of the lawsuit's elemental facts need not be multiplied. If the product rule were to apply, it would seriously undercut deterrence by reducing the liability amount anticipated by prospective transgressors. This similarity between the two proof requirements is, however, only the beginning of the story because the requirements also differ from each other in four important respects. Specifically,
  (1) $P(B)$ and $P^*(B)$ rest upon different informational platforms (ex ante and ex post) and are consequently misaligned;
  (2) Under the doctrinal requirement, both $P(E)$ and $P(D)$ must be more probable than not, in addition to $P^*(B)$, whereas under the optimal requirement, $P(B)$ is the only material probability;
  (3) The optimal requirement favors probabilistic recovery for expected damages, whereas the doctrinal requirement follows the winner-takes-all principle;
  (4) The optimal requirement averages the litigated damage and thereby saves resources associated with ascertaining the actual damage in each case, whereas the doctrinal requirement insists upon proof of the actual damage.
Differences (1) and (2) both relate to the proof of the litigated breach, while differences (3) and (4) pertain to the litigated award and its appropriate amount. The doctrinal $P(D) > 0.5$ requirement must therefore be understood as part of difference (3). This splits our discussion into two parts: we must compare the positive and the normative sets of breach related and award-related requirements. Thus we must first compare $P(B)$ and the combination of $P^*(B)$ and $P(E)$; subsequently, we must compare

the positive and the normative award standards. This distinction is not merely analytical. We have already observed that the optimal award ideal cannot be implemented under the existing enforcement constraints. This observation requires us to treat the award-related requirements as a separate issue. For this reason, we can already see why P(D) must not be processed by the product rule.

A. Proof of Breach: Aligning the Ex Post and the Ex Ante Probabilities

As already explained, P(B) and P*(B) rest upon different informational bases (ex ante and ex post) that may overlap each other only occasionally. Those probabilities are consequently misaligned. Technically, the two probabilities are misaligned even when P*(B) equals P(B) because they stand on different informational platforms.[45] This problem is immanent to adjudication. Because adjudication always occurs after the litigated event, it can secure only an imperfect alignment between P*(B) and P(B). The degree of the alignment's perfection depends, of course, on the resources expended on adjudication. Those resources must not exceed the optimal level, and it is transparent that trials attempting to perfectly reconstruct P(B) would violate this precept (if they would ever succeed). The law must therefore look for an inexpensive way of aligning the two probabilities.

---

45. By adopting Keynes's terminology, we can also observe here that uncertain reasoning has two separate dimensions, "probability" and "weight," neither of which may be neglected. As explained by Keynes,

As the relevant evidence at our disposal increases, the magnitude of the probability of the argument may either decrease or increase, according as the new knowledge strengthens the unfavourable or the favourable evidence; but *something* seems to have increased in either case,—we have a more substantial basis upon which to rest our conclusion. I express this by saying that an accession of new evidence increases the *weight* of an argument. New evidence will sometimes decrease the probability of an argument, but it will always increase its "weight."

. . . .

[W]eight, to speak metaphorically, measures the *sum* of the favourable and unfavourable evidence . . . probability measures the *difference*.

JOHN M. KEYNES, A TREATISE ON PROBABILITY 71, 71, 77 (Harper & Row 1962) (1921) (emphasis in original).

This issue was further developed by L. Jonathan Cohen in *Twelve Questions About Keynes's Concept of Weight*, 37 BRIT. J. PHIL. SCI. 263 (1986). For its implications on judicial factfinding, see L. Jonathan Cohen, *The Role of Evidential Weight in Criminal Proof*, 66 B.U. L. REV. 635 (1986); D. H. Kaye, *Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?*, 66 B.U. L. REV. 657 (1986); Stein, *supra* note 34, at 296-322; and Stein, *supra* note 35, at 304-15. The probability-weight distinction also attracted Charles Peirce's attention. He wrote that "to express the proper state of belief, not *one* number but *two* are requisite, the first depending on the inferred probability, the second on the amount of knowledge on which that probability is based." Charles Hartshorne & Paul Weiss (eds.), 2 COLLECTED ESSAYS OF CHARLES SANDERS PEIRCE 421 (1934) (emphasis in original). Louis Kaplow made an essentially similar distinction by separating burden of proof from accuracy. *See* Kaplow, *supra* note 17, at 356-58.

This task is achieved with the help of the $P(E) > 0.5$ requirement. This requirement focuses upon probability of the litigated entitlement that does not combine with $P^*(B)$ in order to form a joint probability product $P^*(B) \cdot P(E|B)$. That probability is nevertheless necessary for delivering a verdict in the plaintiff's favor. Why is it so?

An economically plausible answer to this question is that $P(E)$ functions as a tool for aligning $P^*(B)$ and $P(B)$. Legal entitlements almost uniformly feature the *visibility element*: under the definition of virtually any entitlement, information about its existence is reasonably available to potential violators. This element permeates the whole body of private law. It is embedded in the publicity requirements that attach to property rights,[46] in the mutual disclosure requirements that attach to contractual entitlements,[47] and in the duty of care requirement as limiting liability in torts.[48] Therefore, by establishing $P(E)$ as more probable than not, the plaintiff also establishes that the litigated entitlement was reasonably ascertainable at the time and in the circumstances of its breach. When $P^*(B)$ is similarly established, the plaintiff's combined proof implies (on a balance of probabilities) that the entitlement was breached in circumstances that made it reasonably ascertainable ex ante. Specifically, the plaintiff's proof entails that the defendant knew (or should have known) about the plaintiff's entitlement at the time he violated it. Because the defendant's action falls into the violation category, the defendant—who knew (or should have known) about the violated entitlement—was also aware (or should have been aware) of the action's transgressing potential.

The key question therefore is whether the entitlement—as defined by the controlling substantive law—extends to cases in which the ex ante probability of breach amounts to what it does in the case at hand. If the answer is negative, then the entitlement is not established under its own definition, so the defendant prevails. If the answer is positive, then the entitlement's probability should acquire a preemptive status. Because the ex ante probability of breach does not exonerate the transgressor under the entitlement's conditions, this probability becomes immaterial. The entitlement's probability will thus dominate the ex ante probability of

---

46. *See, e.g.*, Jason S. Johnston, *Legal Formalities, in* 2 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 524-27 (Peter Newman ed., 1998).

47. *See, e.g.*, Avery W. Katz, *Contract Formation and Interpretation, in* 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 425, 429-30 (Peter Newman ed., 1998).

48. Strict liability is not really an exception to this observation, as there are virtually no cases in which a reasonably unascertainable entitlement has received strict liability protection. Strict liability is generally imposed in connection with unusual and abnormal dangers over which potential wrongdoers exercise control. *See* KEETON, *supra* note 19, at 537, 545-59. The resulting entitlements are consequently ascertainable by potential wrongdoers. Duties of care that judges impose by applying the negligence standard are not unexceptionably visible ex ante. The evidence doctrine, however, should take no blame for such deviations.

breach as an incentive for potential transgressors. Aware of the
entitlement's nature and probability, a potential transgressor must introduce
an appropriate adjustment in his ex ante probability of breach. This
adjustment will substitute any initial level of his ex ante probability of
breach with the level of the entitlement's probability. Hence, even when
the defendant's alleged act of transgression is determined ex post as more
probable than not, the plaintiff will not prevail at the trial. The plaintiff
will prevail only if the entitlement's probability—that is, the upgraded ex
ante probability of breach—is also preponderantly probable. The plaintiff's
breach-allegation can therefore be properly established by any evidential
means, including those that became known only after the event. This
entitlement-focused reasoning aligns $P^*(B)$ and $P(B)$ with the maximal
feasibility that can be attained ex post without incurring exorbitant
adjudication costs.

By employing $P(E)$ as a probability-aligning tool that brings $P^*(B)$ and
$P(B)$ closer to each other, the law also removes it from the standard
multiplication formula for conjunctive probabilities. This removal is
justified by the law's deterrence policy. In accordance with that policy, the
law is primarily interested in the ex ante probability of breach; and because
this probability cannot be determined at trial at an affordable cost, the law
substitutes for it a suitable surrogate. Under this framework, $P(E)$ serves
as a misalignment-corrector for $P^*(B)$, not as its conjunctive companion.[49]

To make things less abstract, let us now return to my houseowner-
contractor hypothetical.[50] In that hypothetical, the ex post probability of
H's claim that the fence collapsed due to severe winds and floods amounts
to 0.6. Antecedently, the risk of this damage amounted to 0.05. This
setup unfolds into two different scenarios, depending on the exact nature
of C's liability towards H. If C gave H an unqualified warranty that the
fence would not fall down due to bad weather, the pivotal issue would then
be C's refusal to compensate H. If the fence was indeed destroyed by
winds and floods, this refusal would constitute a breach of the warranty.
The ex ante probability of this breach must be determined only after the
fence fell down. From that point in time and onwards, the ex ante and the
ex post probabilities of the breach overlapped each other. This alignment
occurs because the two probabilities have a common evidential base: after
the fence's collapse, evidence supporting H's claim was available to both
parties not only at trial, but also before trial.

---

49. As mentioned earlier, $P(E)$ is always greater than $P(B)$ or $P^*(B)$ (legal entitlements exist more
frequently than breaches). Consequently, the plaintiff will be granted recovery when
$P(E) > P^*(B) > 0.5$, which further secures the defendant's reliance on the ex ante probability of breach.

50. *See supra* text accompanying note 20.

Another liability scenario, in which C would be liable only if the fence was ill-constructed or otherwise defective, is less trivial. In that scenario, which is reminiscent of product liability, H could be entitled to C's precautions even when the risk of damage equals 5%. If this entitlement is properly established, it would tell C, both before and after the event, "H is entitled to your precautions also against risks as low as 5%." By its very definition, this entitlement was visible ex ante. It required C either to take the appropriate precautions or to let the damage happen and then pay for it. This entitlement would then have been conferred on H for presumptively good reasons: damage associated with falling fences can be serious and the costs of its prevention are usually not exorbitant.[51] The entitlement therefore overrides the originally low chance of the fence's collapse. This low chance did not erode H's entitlement to C's precautions, as could happen, for example, under the negligence regime.

The key issue must therefore be formulated as follows: Did C take the adequate precautions to prevent the fence's collapse? H contends that C did not take them, and the ex post probability of this allegation equals 0.6. Admittedly, this probability is not an ex ante probability, but it is not a pure ex post probability either. This is so because H's entitlement is both well-established (because it is either uncontested or more probable than not) and visible ex ante, so C knew or ought to have known about it at the time when he built the fence. Because this entitlement extended to risks that are as low as 5%, it dominated C's ex ante information. This domination upgrades the ex ante probability of breach: this probability becomes equal to the high probability of H's entitlement.

By suppressing the product rule, this doctrinal framework also saves substantial litigation resources. If the product rule were to apply, plaintiffs would have to establish the discrete elements of their lawsuits at exceedingly high levels of probability. Thus, in the example exhibiting the conjunction paradox, the plaintiff would succeed only if the probabilities of the contested entitlement, breach, and damage equaled approximately 0.8 each, which would drive plaintiffs into substantial litigation expenses. By making it harder to defend against lawsuits, the doctrinal refusal to apply the product rule apparently increases the defendants' litigation expenses. This, however, is only partially true because the doctrine allows potential transgressors—before they become defendants—to rely on their ex ante information. Any potential transgressor can thus safely rely on the

---

51. Because the issue at hand is evidential, it must be assumed that the underlying substantive entitlement is supported by economically sound reasons. *See* Postema, *supra* note 1, at 1396-97 (presenting and discussing Bentham's precept that "we are to judge the adequacy of a system of judicial procedure not directly in terms of the Principle of Utility but rather in terms of the system's success (or likely success) in properly executing the substantive law, and only indirectly in terms of the system's utility").

information reasonably available to him at the time when he contemplates his action. The doctrine encourages him not to take on uneconomic expenses to acquire further information that might become available at his subsequent trial. Moreover, because the plaintiff's and the defendant's litigation efforts are co-dependent, by inducing plaintiffs to spend less on litigating facts, the doctrine also allows defendants to reduce their litigation expenditures.

Therefore, contrary to the accusations against the doctrine, the doctrine does not actually induce wasteful expenditures on accuracy ex post. These accusations are ill-founded because they ignore the doctrinal suppression of the product rule and its economic implications. The doctrine does not aim at maximizing the number of correct verdicts, nor does it aim at enhancing accuracy ex post. The doctrine effectively sets two conditions for a lawsuit's success: the plaintiff's allegation that the defendant violated the litigated entitlement must be more probable than not, and the ex ante probability of the alleged violation, as determined by the information available to the defendant at the time when he acted, must also be preponderantly probable.[52] This framework generates a suitable second-best surrogate for the ideal proof requirement.

The analysis conducted thus far justifies the probabilistic component of the doctrinal proof requirement as economically efficient. The winner-takes-all principle still awaits justification, which will be provided now.

B. Proof of Damage: Efficient Awards as Affected by Drawbacks in Enforcement

If there were no drawbacks in law enforcement, the law's formula for optimal awards would be $P(B) \cdot d$, with d denoting the average, rather than actual, damage. But law enforcement, as we know it, does suffer from serious drawbacks. These drawbacks must therefore be taken into account in devising the policy for litigation awards. Ideally, violations of deterrence-driven entitlements, including those that the law labels as "private," should be prosecuted by the state. This point has already been mentioned along with the cost problem that makes the state-prosecution mechanism unattractive: indiscriminate processing of all transgressions through this mechanism is unaffordably costly. This problem is responsible for the fundamental law enforcement drawback: *the state must privatize the license to sue and to subsidize adjudication*, which is bound to generate inefficiencies. These inefficiencies spawn from the divergence

---

52. This doctrinal mechanism depends on the visibility of legal entitlements: when an entitlement is invisible, the mechanism will not produce the desired results. In any such case, however, it is the entitlement itself, not the evidence doctrine, that should take the blame for producing anomalies.

between the private and the social incentives to use the adjudication ·facility.[53]

The privatized license to sue can be misused in two ways: it can be utilized both insufficiently and excessively. The license will be underused, thus failing to produce the desired social benefits, if adjudication is costly and accrues insufficient income to private entrepreneurs. Adjudication must therefore be subsidized, and it must also promise attractive awards to successful plaintiffs. But the license to sue may also be overused: a privately advantageous lawsuit may incur excessive social costs that outweigh the adjudication's social benefits. Corrective measures that minimize the number of both over-users and underusers of the license to sue are, therefore, in order. I will now demonstrate that the winner-takes-all principle is one such measure.[54]

What would happen with overusers and underusers of the license to sue, if the ideal-world formula for optimal awards—$P(B) \cdot d$—were adopted in our non-ideal litigation milieu as a replacement to the winner-takes-all principle? The answer is obvious: both overusers and underusers of the license to sue would increase in number. There would be more underusers because the awards awaiting successful plaintiffs under the ideal rule are considerably less attractive than under the winner-takes-all principle. Apart from that, it would usually be easier for a plaintiff to prove her personal damage than to establish the average damage figure. Under the ideal rule, plaintiffs would therefore also have to spend more on litigation, which would inevitably reduce the number of deserving plaintiffs. The deterrence and other economic objectives of the law would consequently be frustrated.

The overusers' community would also grow in its number: because any meaningful probability that attaches to a lawsuit would suffice for recovery, more unmeritorious lawsuits would be brought and expensively processed through adjudication. Because the multiplication principle for conjunctive probabilities is out of use, the number of such lawsuits would, in fact, increase most dramatically. The overuse problem would be further exacerbated by the externalized litigation expenses: the $P(B) \cdot d$ rule invites too many plaintiffs to impose litigation expenses on defendants and to eat away the adjudication subsidy as an almost free lunch. Indeed, under the American rule (which rejects the British "costs follow the event" principle), a party to litigation pays only her attorney's fee, which enables her to externalize part of the adjudication expenses upon her adversary.[55]

---

53. *See* Steven Shavell, *The Fundamental Divergence Between the Private and the Social Motive to Use the Legal System*, 26 J. LEGAL STUD. 575, 577-86 (1997).

54. For discussion of other corrective measures, see *id.* at 586-88.

55. *See id.* at 584. Note that under the $P(B) \cdot d$ rule, a low-probability lawsuit would not be considered "frivolous" for purposes of Rule 11 of the Federal Rules of Civil Procedure. The possibility of cost-shifting would therefore not deter the holders of such lawsuits.

Overusers of the license to sue would also fail to produce the deterrence benefits that the P(B)·d rule strives to attain. Because their only interest is to secure for themselves a profitable outcome of the venture, such plaintiffs would settle for any profit that *they*—not the social welfare principle—would consider reasonable.[56] They would thus largely accept cheap settlement offers made by defendants. Typically, such plaintiffs would be represented by attorneys acting on a contingent-fee basis and thus running portfolios of cases. Those attorneys would initiate settlements at a point in which their expected earnings from litigation equalize with their opportunity costs (calculated as steadily growing on the hourly-fee basis). Cheap settlements that leave wrongdoers underdeterred could consequently be expected to mushroom.[57]

Largely frivolous lawsuits initiated by the overusers would also generate some adverse selection impact or, alternatively, an additional adjudication-cost externality. Aware of their lawsuits' frivolous nature, and also cognizant of the ensuing judicial mistrust of *all* lawsuits that do not unequivocally signal credibility, the overusers would attempt to pool with owners of good lawsuits. The latter would consequently have to consider an additional investment in litigation (such as gathering of more evidence) in order to separate their lawsuits from the bad ones. Because a lawsuit's expected returns would already be cut down for reasons given above, this additional expenditure would turn some good lawsuits into non-profitable actions, so that their holders would consequently prefer not to initiate them in the first place. This market-for-lemons scenario[58] would, perhaps, be too morbid for the present context: intuitively, the additional separating investment to be made by owners of good lawsuits does not appear to be large enough to bring the (once profitable) basic investment in the lawsuit up to a level that wipes out the expected returns. Good lawsuits therefore seem unlikely to be driven away from the courts. But in this more optimistic scenario, the additional lawsuit-separating expenses

---

56. *See id.* at 581-97.

57. This point is well developed with regard to class action attorneys. *See* John C. Coffee, Jr., *The Unfaithful Champion: The Plaintiff as Monitor in Shareholder Litigation*, LAW & CONTEMP. PROBS. Summer 1985, at 5, 49 (advancing the idea that the significant costs that contingent-fee lawyers incur in litigating a case provide an incentive for early and cheap settlements); John C. Coffee, Jr., *Understanding the Plaintiff's Attorney: The Implications of Economic Theory for Private Enforcement of Law Through Class and Derivative Actions*, 86 COLUM. L. REV. 669, 690 (1986) (demonstrating that class counsel "have an incentive to settle prematurely and cheaply"). For a defense of frivolous lawsuits, see Charles M. Yablon, *The Good, the Bad, and the Frivolous Case: An Essay on Probability and Rule 11*, 44 UCLA L. REV. 65, 73-75 (1996) (defending frivolous lawsuits on the grounds that standard portfolio theory urges a lawyer to take high risks with the potential for enormous awards and to diversify risks within her caseload).

58. As famously set by George A. Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 QUART. J. ECON. 488 (1970).

would actually be incurred. Those expenses should therefore be counted as a harmful externality.

The winner-takes-all principle tackles these problems by branching into two rules:

  (1) the full-recovery rule, activated when the lawsuit's probability is greater than 0.5;

  (2) the zero-recovery rule that applies when the lawsuit's probability is not greater than 0.5.

The full-recovery rule makes litigation obviously more attractive for potential underusers of the license to sue. Under this rule, successful plaintiffs would recover more damages than are justified by deterrence. They would also have to prove only their personal damages, which would usually be less expensive than establishing an impersonal average damage as a general social fact. This would substantially increase the underusers' expected income from litigation.

The zero-recovery rule makes litigation considerably less attractive for potential overusers of the license to sue. Under this rule, low-probability lawsuits are driven away from the courts. The rule also requires plaintiffs to establish their personal damages, thus driving away personally un-damaged plaintiffs. Finally, by raising the stakes of the average litigation, in comparison with the ideal $P(B) \cdot d$ approach, this rule induces both plain-tiffs and defendants to spend more on litigation. Instead of making this increased investment, at least some unmeritorious plaintiffs (or, more realistically, their contingent-fee attorneys) might consequently prefer not to come to court.

These reasons do not only justify the winner-takes-all principle. They also turn the probability of damage into a separate non-conjunctive factor, thus removing it from the multiplication formula. Because damage awards have acquired a special function in law, probability of litigated damages also acquires a special non-conjunctive status. The task faced by prospective plaintiffs would become considerably more onerous if proba-bility of damage were to be processed by the product rule and multiplied by the probabilities of breach and entitlement. This high burden of proof would increase the number of underusers of the license to sue.

Before praising the winner-takes-all principle, however, we must look also on the other side of the coin. Under this principle, some transgressors (which we can label as "high-probability transgressors") are overdeterred, while other transgressors (which we can label as "low-probability transgressors") are underdeterred. The amount of overdeterrence (caused by the full-recovery rule) equals $D-P(B) \cdot d$, with D representing the plaintiff's award. The amount of under-deterrence (caused by the zero-recovery rule) equals $P(B) \cdot d$. This is the price exacted by the winner-takes-all principle.

Intuitively, this price is worth incurring in dealing with unsystematic transgressions. When incurred sporadically, this price is unlikely to exceed the costs associated with applying the P(B)·d rule. But if the price is incurred systematically, as would happen in recurring transgression cases, then the opposite would be true. The winner-takes-all principle would defeat the deterrence objective of the law if a low-probability transgressor could repeat his conduct time and time again (theoretically, *ad infinitum*) without being penalized. The deterrence objective would also be defeated if a high probability transgressor were to gain nothing from taking precautions that reduce the transgression's probability but still keep it above the level of 0.5. In such settings, prominently exemplified by market-share liability in torts,[59] a strong case can be made for substituting the winner-takes-all principle with the P(B)·d rule.[60]

V. An Implication for Punitive Damages

The winner-takes-all principle has an important policy implication for punitive damages in torts. From the deterrence perspective, punitive damages are an important policy instrument: they can be utilized as an award-correcting tool by a legal system that suffers from drawbacks in law enforcement. Drawbacks in law enforcement dilute deterrence. Imposition of punitive damages can raise the level of deterrence back to the optimal and thereby counteract this dilution. Law-enforcement drawbacks would consequently be offset by punitive damages. This possibility is econ-omically attractive because elimination of law-enforcement drawbacks is usually expensive. In contrast, fixing punitive damages only requires an informed decision, so it is relatively cheap.

In making such decisions, legislators and judges must account for the punitive aspect of the winner-takes-all principle. By forcing defendants to pay increased compensation in order to increase the number of socially beneficial lawsuits and thereby counteract the potential underenforcement of the law, the winner-takes-all principle performs an essentially punitive function. Under this principle, the defendants' overpayment equals (1-P)·D, when P is the probability of the plaintiff's case and D is the relevant

---

59. *See* Sindell v. Abbott Labs., 607 P.2d 924, 933-35 (Cal. 1980).

60. *See* Saul Levmore, *Probabilistic Recoveries, Restitution, and Recurring Wrongs,* 19 J. LEGAL STUD. 691, 697-98 & 698 n.13 (1990). Arguments favoring this substitution can be put on a more principled footing. Many instances that involve low-probability transgressors are instances in which the transgression is *preponderantly* responsible for the factual uncertainty of the case. This is certainly true about all repeated-transgression cases. In such cases, transgressors should be made liable for the resulting *evidential damage,* measured by the plaintiff's lost chances to prevail at her trial with the help of the evidence of which she was wrongfully deprived. The evidential damage doctrine also optimizes the incentives for high probability transgressors. *See* Ariel Porat & Alex Stein, *Liability for Uncertainty: Making Evidential Damage Actionable,* 18 CARDOZO L. REV. 1891, 1940 (1997).

damage. This overpayment must be treated as a punitive ingredient of trial awards because it intensifies general deterrence at the expense of the individual defendant. If so, then $(1-P) \cdot D$ is the sum by which punitive damages must be reduced whenever they are explicitly imposed by courts.

Punitive damages must therefore be calculated in the following way.[61] In a trial for damage D, let q denote the defendant's ex ante probability of being subjected to law enforcement. The defendant's ex ante probability of escaping law enforcement would thus equal 1-q. The odds of escape vs. enforcement ("the enforcement odds") would accordingly be $(1-q)/q$. The total damage award, including both actual and punitive damages, would amount to $1/q \cdot D = D/q$ (1/q functions here as an award multiplier that compensates for the liability-escaping chances, thus bringing the award to the optimal level of deterrence). The award's punitive part would consequently equal $(D/q)-D$, that is, $[(1-q)/q] \cdot D$ (the non-punitive award multiplied by the enforcement odds). As already mentioned, the punitive ingredient of the full-recovery rule is $(1-P) \cdot D$. The optimal punitive award should therefore amount to $D \cdot [(1-q)/q-(1-P)]$ (the non-punitive award multiplied by the difference between the enforcement odds and the probability of non-breach).

VI. Conclusion

This essay has offered a new rationale for the existing proof requirements in civil trials. The civil proof doctrine, as traditionally understood, aims at maximizing the accuracy of judicial verdicts. Understood in this way, the doctrine exhibits two economic paradoxes. First, it focuses upon accuracy ex post by requiring judges to reconstruct the relevant events, as unfolded in reality, including the damage actually sustained by the plaintiff, on the basis of information that becomes available at trial. On the basis of that information judges are called upon to determine the probability of the plaintiff's allegation that the defendant violated the litigated entitlement and that the plaintiff has sustained the litigated damage as a result of the violation. This type of accuracy is economically inefficient. Accuracy ex post is a waste when information available at trial is not the same information that could be cost-efficiently obtained by the defendant before taking the litigated action. For deterrence purposes, it is only the latter type of information that matters. Moreover, accuracy ex post is an investment-dependent opportunity rather than a static good. As such, it fosters a secondary market for competitive adversarial

---

61. As developed by A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 870 (1998). *See also* A. Mitchell Polinsky & Steven Shavell, *Punitive Damages*, *in* 3 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 192, 193-94 (Peter Newman ed. 1998).

investments in information, which might adversely affect the primary market, that is, the market for goods, services, risks, and precautions. Thus, when prospective litigants are rationally unwilling to commit themselves to the required investments in information, inefficiencies are bound to occur. In such cases, each party will account for the event that he will be wrongfully defeated in the future trial because his opponent's investment in information outscored his. This prospect will foil trans-actions that are otherwise efficient and chill many other socially beneficial activities.

Second, if the doctrine is nonetheless committed to accuracy ex post, then it should require judges to determine the ultimate probability of the plaintiff's case by multiplying the probabilities that attach to the relevant entitlement, breach and damage. By refusing to apply the multiplication principle, the doctrine, in fact, reduces the total amount of correct verdicts, instead of maximizing it. The doctrine therefore seems to be doing the wrong thing in the wrong way, even though the thing and the way both have a strong intuitive appeal.

The traditional theory, however, is neither necessary nor most plausible: under the new theory offered by this essay, the two alleged wrongs make a right because in combination they generate a synergetic mechanism that aligns, to the extent feasible, the ex ante and the ex post probabilities of transgression. This alignment is attained by the combined, but not conjunctive, functioning of the two probabilities: the probability of the litigated entitlement and the ex post probability of the entitlement's breach. The entitlement's probability dominates the defendant's ex ante information, thus adjusting the ex ante probability of breach. This adjustment is achieved due to the visibility element, uniformly featured by legal entitlements: under the definition of virtually any entitlement, the entitlement must both exist and be reasonably ascertainable ex ante, that is, at the time and in the circumstances of its breach. The ex post probability of breach has a different function, namely, to substantiate the allegation that the defendant has actually violated the entitlement. This combined framework secures the appropriate alignment between the ex post and the ex ante probabilities of breach. The plaintiff will prevail at trial only when each probability is preponderant, so prospective defendants can safely rely on the probabilities of breach that exist ex ante. Indeed, because the entitlement's probability functions as a misalignment-corrector for the ex post probability of breach, and not as its conjunctive companion, the two probabilities must not be multiplied in determining the ultimate probability of the plaintiff's case.

This doctrinal mechanism also saves litigation expenses. The plaintiffs' litigation effort is substantially alleviated by the doctrinal refusal to apply the multiplication principle. This entails greater hardship for

defendants, but they receive offsetting benefits: the doctrine allows potential transgressors—before they become defendants—to rely on their ex ante information. The doctrine thus encourages potential transgressors not to undertake uneconomic expenses by acquiring further information that might become available at their subsequent trials.

The essay also supports the existing award system, under which, as a matter of general rule, the winner takes all. The essay justifies this system by the existing constraints in law enforcement. Under those constraints, the legal system employs the winner-takes-all principle to reduce the number of underusers and overusers of the adjudication facility, which is both privatized and subsidized by the state. This rationale turns trial awards into a discrete component of civil litigation. The doctrinal mechanism exposed in this essay consequently separates the breach-related from the damage-related proof requirements. Probability of the litigated damage must therefore also be removed from the multiplication formula. There is, consequently, no room for applying the multiplication principle: the conjunction paradox disappears because all the relevant probabilities are functionally non-conjunctive.

The essay has compared this doctrinal mechanism with the ideal proof requirement, as derived from the optimal deterrence standard. I have done this by using a simple Bayesian model that explores the interrelationship between three pivotal probability factors: probability of entitlement, probability of breach, and probability of damage. This formalization has proved useful: under the constructed first-best model, allocation of liability in civil trials ought to be determined by the ex ante probability of breach as a non-conjunctive monadic probability. Unfortunately, due to various constraints, determination of this probability cannot feasibly be attained in adjudication (ex post). The legal system must therefore search for a suitable surrogate for that probability. This essay has identified the required surrogate in the existing doctrinal mechanism. This mechanism is therefore justified as a second-best.

The winner-takes-all principle is the only doctrinal component that deviates from the ideal model. This finding is not surprising because the model was constructed as an ideal type and was thus isolated from the existing drawbacks in the administration of justice. As explained in the essay, these drawbacks call for corrective measures, and the winner-takes-all principle is one of those measures. This understanding of the winner-takes-all principle calls for an appropriate adjustment of punitive damages when these are awarded on economic grounds. The essay consequently offers the required adjustment.

Readers of this essay can also deduce from it a more general lesson. This essay unfolds an explanatory theory that instantiates the general economic theory of second-best. This general theory holds that finding an

economic distortion is not yet a good reason for eliminating it: the observed distortion may be counteracting another distortion that the relevant social mechanism, such as law, must accommodate because of some of its intrinsic constraints.[62]   This possibility calls for a comprehensive assessment of the mechanism's components, including the incentives that these components mutually produce and the constraints under which they operate. As demonstrated by this essay, the "second-best" methodology is particularly suitable for analyzing common-law doctrines, such as those that judges collectively develop to facilitate proof of facts in adjudication.   Those doctrines are presumptively efficient because they have been thought through by many judges as a means for resolving future cases and for shaping people's future behavior.[63] Therefore, prior to denouncing a common law doctrine that appears inefficient, one should examine its possible role in offsetting another ostensible distortion in a comprehensive second-best mechanism.[64]

---

62. *See* Smith, *supra* note 27, at 695 (citing Richard G. Lipsey & Kelvin Lancaster, *The General Theory of Second Best*, 24 REV. ECON. STUD. 11 (1956-57)) (offering a general economic theory of second-best).

63. *See* POSNER, *supra* note 13, at 27-28.

64. *See* POSNER, *supra* note 13, at 271, 274-75 (stating that while the "common law is to most lawyers a collection of disparate fields, each with its own history, vocabulary, and bewildering profusion of rules and doctrines," those doctrines "form a system for inducing people to behave efficiently," and although "[t]heir articulation in economic terms is beyond the capacity of most judges and lawyers, . . . their intuition is not").