

VIRGINIA LAW REVIEW

VOLUME 106

DECEMBER 2020

NUMBER 8

ARTICLES

DAMAGES FOR PRIVILEGED HARM

*Stephen Yelderman**

The law often permits us to impose substantial harm on others without incurring liability. Once liability is triggered, compensatory damages require a defendant to pay for the harm caused by his wrongful conduct. Calculating these damages requires consideration of the harm that the defendant could have caused without incurring liability in the first place. This harm is “privileged,” in the sense that the defendant would have been free to impose it in a counterfactual universe in which he complied with the substantive law. Having transgressed that law, he is responsible for damages. But the question is whether these damages should be reduced to account for the harm he could have imposed without owing damages at all.

The treatment of privileged harm is fundamental to the calculation of compensatory damages. Nonetheless, it has received little scholarly attention and has been the subject of conflicting decisions in the courts. In some areas of law, damages are routinely reduced to account for privileged harm; in others, this credit is given only sporadically, or not at all. Critically, there is not yet any sound theoretical explanation for why the rule ought to be different in one set of cases than another.

This Article begins by exploring the effects of crediting or not crediting privileged harm. It then relates the treatment of privileged harm to

* Professor of Law, Notre Dame Law School. For very helpful feedback on prior drafts, I thank A.J. Bellia, Sam Bray, Bernard Chao, Thomas Cotter, John Duffy, John Golden, Bert Huang, Dan Kelly, Andrew Kull, Douglas Laycock, Saul Levmore, Mark McKenna, Collen Murphy, Jeff Pojanowski, Henry Smith, Jay Tidmarsh, and Stephen Waddams.

several well-known questions of remedial design. Finally, it proposes several general principles that a court or policymaker might follow in determining whether to reduce damages to account for privileged harm.

INTRODUCTION..... 1570

I. BACKGROUND..... 1575

 A. *The Case from Basic Principles* 1575

 B. *Crediting Privileged Harm* 1578

 C. *Limits and Complications*..... 1584

II. UNDERSTANDING THE EFFECTS OF PRIVILEGED HARM 1587

 A. *Damages Functions With and Without Privileged Harm* 1587

 B. *Marginal Incentives of Prospective Defendants*..... 1591

 C. *The Likelihood of Enforcement* 1594

 D. *Privileged Harm and Punitive Damages* 1595

 E. *The Magnitude of Privileged Harm*..... 1599

III. GENERAL PRINCIPLES..... 1602

 A. *Incentives for Defendants* 1603

 B. *Incentives and Compensation for Plaintiffs*..... 1608

 C. *Managing Complexity* 1611

IV. SPECIAL CASES 1614

 A. *The “No Remedy” Problem*..... 1614

 B. *Non-Conduct Elements* 1618

 C. *Unearned Privileges*..... 1620

 D. *Implausible Counterfactuals* 1623

 E. *Strategic Decision Making by Litigants* 1627

CONCLUSION..... 1629

INTRODUCTION

Compensatory damages are meant to restore a plaintiff to the position she would have enjoyed absent the defendant’s wrong.¹ Their amount is typically figured by way of a counterfactual. We ask the factfinder, “How

¹ See Restatement (Second) of Torts § 903 cmt. a (Am. L. Inst. 1979) (“When there has been harm only to the pecuniary interests of a person, compensatory damages are designed to place him in a position substantially equivalent in a pecuniary way to that which he would have occupied had no tort been committed.”); 1 Theodore Sedgwick, *A Treatise on the Measure of Damages* § 30, at 25 (9th ed. 1920) (“In all cases . . . of civil injury and breach of contract, the declared object of awarding damages is . . . to put the plaintiff in the same position, so far as money can do it, as he would have been if the contract had been performed or the tort not committed.”) (footnotes omitted).

much worse off is the plaintiff compared to the position she would have occupied in a hypothetical universe in which the defendant did not wrong her at all?" In theory, damages equal to this amount will "redress the concrete loss that the plaintiff has suffered by reason of the defendant's wrongful conduct."²

In many cases, it is simple and true to imagine this alternative, wrong-free universe as one in which the defendant does no harm to the plaintiff. For example, rather than negligently swerving into oncoming traffic, the defendant simply stays in his lane. In a case like that, the defendant's role in the counterfactual universe is so trivial that he can be imagined out of it completely. Instead of asking what the plaintiff's condition would have been if the defendant had driven safely, we could just as well ask what the plaintiff's condition would have been if the defendant had not driven at all. This works because the world in which the defendant commits no tort is also a world in which the defendant imposes no harm.

But sometimes things are more complicated. In some cases, the truth is that the defendant would have imposed *some* harm on the plaintiff even if he had acted lawfully. This is possible because the law often leaves certain injuries to lie where they fall. As Oliver Wendell Holmes put it, some harm is "privileged," in the sense the defendant was free to impose it, even intentionally, only subject to conditions set by the substantive law.³ Holmes was never known for timid examples:

[A] man has a right to set up a shop in a small village which can support but one of the kind, although he expects and intends to ruin a deserving widow who is established there already. He has a right to build a house upon his land in such a position as to spoil the view from a far more valuable house hard by. He has a right to give honest answers to inquiries about a servant, although he intends thereby to prevent his

² *Cooper Indus., Inc. v. Leatherman Tool Grp., Inc.*, 532 U.S. 424, 432 (2001).

³ Oliver Wendell Homes, Jr., *Privilege, Malice, and Intent*, 8 Harv. L. Rev. 1, 3–4 (1894). Note that the term "privilege" later took on a somewhat narrower, more technical meaning in tort law: "conduct which, under ordinary circumstances, would subject the actor to liability, [but that] under particular circumstances, does not subject him thereto." Restatement (First) of Torts § 10 (Am. L. Inst. 1934); see also Francis H. Bohlen, *Incomplete Privilege To Inflict Intentional Invasions of Interests of Property and Personality*, 39 Harv. L. Rev. 307, 308 (1926) (consciously reappropriating the term "privilege" to refer to excuses and justifications). Throughout this Article, I use the term "privileged" in the broader sense that Holmes used it—to describe harm that may be imposed on another without incurring legal liability.

getting a place. . . . In these instances, the justification is that the defendant is privileged . . . to inflict the damage complained of.⁴

In cases involving damages, the defendant has, by hypothesis, *not* complied with the strictures of the substantive law, making the plaintiff's injuries actionable rather than privileged. But the possibility of privileged harm forces us to confront a choice in counterfactuals. When measuring compensatory damages, do we work from a baseline in which the defendant imposes no harm on the plaintiff, or one in which the defendant imposes only privileged harm on the plaintiff?

The difference between these two conceptions can have significant consequences for the defendant's ultimate liability, and courts have come down both ways. To pick just one example among many, consider the landmark due process case of *Carey v. Piphus*.⁵ The plaintiff in that case, Piphus, was a freshman at a Chicago high school who had been summarily suspended after a principal observed him smoking what appeared to be marijuana. Piphus brought suit under 42 U.S.C. § 1983, alleging that this suspension without pre-deprivation process violated the Fourteenth Amendment.⁶ The school district argued that, whether or not the afforded process was constitutionally sufficient, Piphus would be unable to prove any actual damages, since Piphus deserved his suspension either way.⁷ At heart, this was an argument rooted in privileged harm. Piphus sought the value of the school days lost to his unconstitutional suspension; the school district sought to reduce its damages by invoking a hypothetical suspension it could have imposed constitutionally.

In a terse section of its opinion, the Supreme Court accepted the school district's argument for privileged harm. If it was really true that the school could have suspended Piphus in a counterfactual world in which it also afforded him due process, then "the failure to accord procedural due process could not properly be viewed as the cause of the suspension[]." ⁸ To hold otherwise, the Court suggested, would afford Piphus a windfall rather than compensation.⁹

Carey's damages rule has had far-reaching consequences for plaintiffs invoking their federal constitutional right to procedural due process. But

⁴ Holmes, *supra* note 3, at 3.

⁵ 435 U.S. 247 (1978).

⁶ *Id.* at 248–50.

⁷ *Id.* at 260.

⁸ *Id.*

⁹ *Id.*

some courts have blunted its practical import by treating privileged harm differently for claims arising out of state law. Following *Carey*, a terminated government employee alleging that her firing violated the U.S. Constitution will typically receive zero compensatory damages unless she can show that the constitutionally required procedure would have actually prevented her termination.¹⁰ But when that same employee alleges that she was denied the procedural protections promised by her *employment contract*, a number of state courts have refused to credit privileged harm. For example, in *Piacitelli v. Southern Utah State College*,¹¹ the Utah Supreme Court explicitly declined to follow *Carey*, holding that the State was liable for back pay until the contractually specified process had been substantially performed. If credit were given for privileged harm, the Court reasoned, “the employer could discharge an employee summarily and then omit or delay the contractual termination procedures with impunity so long as it was in possession of evidence which, when ultimately provided, would justify the discharge.”¹² The State was thus liable for the privileged harm, even if it *could have* terminated the employee by following the contractually specified procedures in the first instance.¹³

As we will see, variants of this question arise in many areas of law, from torts to contracts, intellectual property, antitrust, defamation, and constitutional litigation. Perhaps surprisingly, a search across these disparate fields reveals there is not yet a clear theoretical or consistent doctrinal answer to what seems like a very basic remedial question. As a pair, *Carey* and *Piacitelli* capture the dilemma succinctly. On the one hand, a court must consider the possibility of privileged harm to measure the injury actually caused by the defendant’s wrongful conduct; to do otherwise seems to afford the plaintiff a windfall rather than compensation. On the other hand, reducing damages to account for

¹⁰ See *infra* notes 45–48 and accompanying text.

¹¹ 636 P.2d 1063 (Utah 1981).

¹² *Id.* at 1069.

¹³ See *id.* A number of Western states have followed Utah’s lead. See, e.g., *Hom v. State*, 459 N.W.2d 823, 826 (N.D. 1990) (adopting same rule for North Dakota); *Bowler v. Bd. of Trs.*, 617 P.2d 841, 849 (Idaho 1980) (“[I]n order to prevail, appellant must allege and prove either that his employment contract was breached by the board or that he was unjustifiably discharged.”); *Brown v. Ford, Bacon & Davis, Utah, Inc.*, 850 F.2d 631, 633–34 (10th Cir. 1988) (applying *Piacitelli* in case arising under Utah law). But see *Nzomo v. Vt. State Colls.*, 411 A.2d 1366, 1367–68 (Vt. 1980) (applying *Carey* rule in case involving contractual claims).

privileged harm risks leaving violations unpunished, creating no particular incentive for the defendant to comply with the substantive law. *Carey* and *Piacitelli* are outliers in acknowledging these policy concerns so explicitly, but the same basic tradeoff lurks beneath the surface in countless compensatory damages cases.

This Article explores when and how damages should be reduced to account for privileged harm. It begins with an introduction to prior doctrinal treatment of this question—an apparent morass of inconsistent rules and unreasoned conclusions.¹⁴ It then models the effects of privileged harm on marginal incentives, both for prospective plaintiffs and prospective defendants.¹⁵ This analysis reveals that the treatment of privileged harm closely tracks a well-known question of remedial design: whether damages ought to operate as prices or sanctions.¹⁶ At its core, a decision to deny credit for privileged harm is a decision to erect sanctions-like damages around the threshold of liability, with increased marginal deterrence for defendants and wealth transfer to plaintiffs following as a result.¹⁷ In this way, there is a previously unrecognized commonality between the question of how to measure compensatory damages and the question of whether to impose punitive damages.¹⁸

Building on this theoretical account, this Article turns to the specifics of how courts should handle arguments for privileged harm. In addition to the first-order question of whether damages should operate as prices or sanctions,¹⁹ there are a number of practical, second-order concerns that may influence the decision to account for privileged harm. In identifiable categories of cases, the magnitude of any privileged harm will be small enough that the question can be safely ignored in the interest of remedial simplicity.²⁰ In other cases, incautious crediting of privileged harm has the potential to leave entire classes of plaintiffs with no remedy.²¹ As these and other examples will illustrate, the treatment of privileged harm might look like a simple calculation detail, but in fact it can significantly affect the ability of the substantive law to achieve its ends. For that reason, it is a question better confronted than avoided.

¹⁴ See *infra* Part I.

¹⁵ See *infra* Sections II.A, II.B, II.C.

¹⁶ See Robert Cooter, *Prices and Sanctions*, 84 *Colum. L. Rev.* 1523, 1523 (1984).

¹⁷ See *infra* Sections II.B, II.C.

¹⁸ See *infra* Section II.D.

¹⁹ See *infra* Section III.A.

²⁰ See *infra* Section III.C.

²¹ See *infra* Section IV.A.

I. BACKGROUND

The case for crediting privileged harm is intuitive. And, indeed, courts often *do* reduce compensatory damages to account for privileged harm. But in other cases they don't. And rarely do courts provide an explanation for this disparate treatment.

This Part introduces the basic legal principles supporting credit for privileged harm. It then shows how these principles alone are insufficient to answer many of the damages questions confronting courts in practice.

A. The Case from Basic Principles

A straightforward case for crediting privileged harm can be made from a few foundational principles. Privileged harm is possible because of the maxim of *damnum absque injuriâ*—literally translated as “loss without injury,” but standing for the proposition that not every material loss creates an actionable legal claim. As the Supreme Court explained in 1938, “It is an ancient maxim, that a damage . . . without an injury in this sense . . . does not lay the foundation of an action; because, if the act complained of does not violate any of his legal rights, it is obvious, that he has no cause to complain.”²² But even in 1938 this was very old news. Nearly a century earlier, Herbert Broom’s classic treatise on legal maxims had explained:

[I]t frequently happens, in the ordinary proceedings of life, that a man may *lawfully* use his own property so as to cause damage to his neighbour, provided it be not *injuriousum* In cases of this nature a loss or damage is indeed sustained by the plaintiff, but it results from an act done by another free and responsible being, which is neither unjust or illegal.²³

To put it simply, we are not absolutely liable for all the ways our various acts and omissions may work to the detriment of others. We are generally free, for example, to build houses that obscure our neighbors’ views of the ocean and even to divert water that floods their property, so

²² Ala. Power Co. v. Ickes, 302 U.S. 464, 479 (1938); see also 1 Sedgwick, *supra* note 1, § 32, at 27–28 (“[L]egal relief is [not] to be had for every species of loss that individuals sustain by the acts of others. . . . It is only legal injury that sets its machinery in operation; and this is meant by the maxim that *damnum absque injuriâ* . . .”).

²³ Herbert Broom, *A Selection of Legal Maxims, Classified and Illustrated* 139 (Phila., T. & J.W. Johnson & Co., 5th ed. 1864).

long as we act without negligence or intention to harm.²⁴ These actions might impose very real losses on others. But without a legal wrong, they create no legally cognizable injury.

When a defendant does commit a legal wrong, law seeks to provide a remedy.²⁵ But the distinction between literal loss and legal injury persists at this stage as well. Damages are available only to compensate for injuries in the legal sense—that is, those losses that were the result of the defendant’s wrongful act or omission.²⁶ A plaintiff might have to bear significant losses without compensation, even as against a defendant who committed serious wrongs.²⁷ Unless the plaintiff’s losses were caused by the defendant’s wrongful act, they are not injuries in the legal sense, and cannot be the basis for compensatory damages.

The traditional method for testing this causal relationship is to invoke a counterfactual, asking what *would have* happened if the defendant had committed no legal wrong. As the Supreme Court explained in 1867, “The injured party is to be placed, as near as may be, in the situation he would have occupied if the wrong had not been committed.”²⁸ So in a contracts case, we conjure a world in which the defendant performed on his

²⁴ See Restatement (Second) of Torts § 833 cmt. b (Am. L. Inst. 1979). For these and hundreds of other examples, see generally Edward P. Weeks, *The Doctrine of Damnum Absque Injuriâ: Considered in Its Relation to the Law of Torts* (S.F., Sumner Whitney & Co. 1879).

²⁵ As the corollary maxim goes, “For every wrong a remedy.” See *Parker v. Griswold*, 17 Conn. 288, 303 (1845); *Broom*, supra note 23, at 137.

²⁶ See *Birdsall v. Coolidge*, 93 U.S. 64, 64 (1876) (“[D]amages shall be the result of the injury alleged and proved . . .”); *Wicker v. Hoppock*, 73 U.S. (6 Wall.) 94, 99 (1867) (“[W]hen a wrong has been done, and the law gives a remedy, the compensation shall be equal to the injury.”).

²⁷ For example, consider a shipowner who negligently fails to install life preservers. When a sailor is swept overboard and drowns, his estate has obviously suffered a most serious loss. But unless the estate can show that the drowning was the result of the negligence, it will recover no damages. *N.Y. Cent. R.R. Co. v. Grimstad*, 264 F. 334, 335 (2d Cir. 1920) (“But there is nothing whatever to show that the decedent was not drowned because he did not know how to swim, nor anything to show that, if there had been a life buoy on board . . . it would have prevented him from drowning.”); see also *Ford v. Trident Fisheries Co.*, 122 N.E. 389, 390 (Mass. 1919) (noting that, even if lifeboats were negligently installed, there could be no liability without evidence that a different installation would have saved the drowning sailor). In recent decades, some jurisdictions have relaxed the proof necessary to show that the plaintiff’s injury was the result of the defendant’s wrongful act. See Prosser and Keeton on the Law of Torts § 41, at 270–71 (W. Page Keeton, Dan B. Dobbs, Robert E. Keeton & David G. Owen eds., 5th ed. 1984); *Zuchowicz v. United States*, 140 F.3d 381, 388–91 (2d Cir. 1998) (explaining this doctrinal development). But even with this relaxed standard of proof, a causal link between the wrong and the injury remains necessary to the plaintiff’s claim. See id.

²⁸ *Wicker*, 73 U.S. at 99.

contract;²⁹ in a torts case, we imagine that the defendant had breached no legal duty.³⁰ We then compare the plaintiff's actual condition to the condition she would have been in but for the unlawful act or omission. The difference between these two is the measure of the harm caused by the legal wrong—the very object of compensatory damages.³¹

A straightforward application of these principles suggests that damages should, as a general rule, be reduced to account for privileged harm. If compensatory damages are meant to restore the plaintiff to the position she would have occupied but for the wrong, then the counterfactual world should be defined by the defendant abiding the law—not by the defendant imposing zero harm.³² In fact, to equate the absence of legal injury with the absence of harm would be to ignore the very distinction taught by *damnum absque injuriâ*. Just as the defendant was free in the real world to impose some harm without becoming liable, he should be imagined as free to impose that same harm in the counterfactual world. Acknowledging the possibility of harm without legal injury leads directly to credit for privileged harm.

²⁹ See Restatement (Second) of Contracts § 344 (Am. L. Inst. 1981) (defining “expectation interest” as the plaintiff’s interest in “being put in as good a position as he would have been in had the contract been performed”). Alternatively, in some situations contracts law will vindicate the plaintiff’s “reliance interest” by “put[ting him] in as good a position as he would have been in had the contract not been made.” *Id.*

³⁰ See Restatement (Second) of Torts § 903 cmt. a (Am. L. Inst. 1979) (“When there has been harm only to the pecuniary interests of a person, compensatory damages are designed to place him in a position substantially equivalent in a pecuniary way to that which he would have occupied had no tort been committed.”).

³¹ See 1 Sedgwick, *supra* note 1, § 30, at 25 (“[T]he declared object of awarding damages is . . . to put the plaintiff in the same position, so far as money can do it, as he would have been if the contract had been performed or the tort not committed.”); *Livingstone v. Rawyards Coal Co.*, [1880] 5 App. Cas. 25 (HL) 39 (appeal taken from Scot.) (defining compensatory damages as “that sum of money which will put the party who has been injured, or who has suffered, in the same position as he would have been in if he had not sustained the wrong for which he is now getting his compensation”); *Cooper Indus., Inc. v. Leatherman Tool Grp., Inc.*, 532 U.S. 424, 432 (2001) (stating that the goal of compensatory damages is to redress the loss “suffered by reason of the defendant’s wrongful conduct”). To be sure, there are strong objections that the but-for inquiry is not a true test of causation. See generally Michael S. Moore, *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics* (2009). Nonetheless, this test is a common heuristic used in law, despite its imperfection. See, e.g., *Burrage v. United States*, 571 U.S. 204, 211–13 (2014) (“This but-for requirement is part of the common understanding of cause.”).

³² Michael G. Pratt, *What Would the Defendant Have Done But For the Wrong?*, 40 *Oxford J. Legal Stud.* 28, 42–43 (2020) (“[T]he court must make a correction to the actual world so that the defendant in the corrected, counterfactual world treats the claimant lawfully.”).

In the abstract, then, the answer looks easy. The basic goal of compensatory damages—returning the plaintiff to the position she would have enjoyed absent the defendant’s wrong—suggests that damages should be reduced to account for privileged harm. This credit reflects the background principle that we are liable only for the harm caused by our legal wrongs.

B. Crediting Privileged Harm

These foundational principles may explain why courts often assume that credit for privileged harm is available as of right. Indeed, within many areas of law, this credit is a routine component of damages calculation. This Section provides three examples of situations in which damages are regularly reduced to account for privileged harm: one from the common law of torts, one from contemporary public law, and one from intellectual property law.

To begin with the common law, the Restatement (Second) of Torts considers credit for privileged harm to be unobjectionable, provided there is a “reasonable and rational basis” for apportioning damages between the tortious and innocent causes.³³ Unfortunately, the Restatement treats the topic quite tersely, with nothing more than a passing sentence and a single illustration based on a New Jersey case from 1902.³⁴

That 1902 New Jersey case—*Jenkins v. Pennsylvania Railroad Co.*³⁵—does, however, present a textbook case of privileged harm. The plaintiff in *Jenkins* alleged that the defendant railroad had negligently discharged smoke and noxious vapors “in greater quantities than were required for the legitimate and proper use and operation of the railroad,” damaging the plaintiff’s furnishings and rendering his home unfit for habitation.³⁶ The defendant contended that it emitted no more smoke than was necessary. The damages question put to the jury was “how much of the damage was the result of carelessness in firing” and “how much was necessarily incident to the careful operation of the railroad.”³⁷ Any harm

³³ Restatement (Second) of Torts § 433A cmt. d, e, illus. 7 (Am. L. Inst. 1965) (“Apportionment may also be made where a part of the harm caused would clearly have resulted from the innocent conduct of the defendant himself, and the extent of the harm has been aggravated by his tortious conduct.”).

³⁴ *Id.* at illus. 7 (citing *Jenkins v. Pa. R.R. Co.*, 51 A. 704, 705–06 (N.J. 1902)).

³⁵ 51 A. 704.

³⁶ *Id.* at 705.

³⁷ *Id.*

that would have been caused by ordinary railroad operations was non-recoverable as *damnum absque injuriâ*.

On appeal, the court considered it unquestionable that the railroad should not have to pay for harm that it would have caused by running its engines properly. The only question was who should bear the burden of apportioning the harm between the negligent and non-negligent emissions. As the court noted, “It was as impossible for the plaintiff to adduce evidence separating the unnecessary from the necessary damage, as for the defendant to split up each smoke cloud into two, label one, ‘Necessary,’ and the other, ‘Unnecessary,’ and send them separately to the plaintiff’s premises.”³⁸ Fearing that too much strictness of proof would leave the plaintiff’s injuries entirely uncompensated, the court reasoned that the jury should be permitted to form its best estimate from the evidence available under the circumstances.³⁹ Though the primary question on appeal was burden of proof, *Jenkins* is quite clear that a defendant should be entitled to reduce his damages to account for privileged harm, at least where there is adequate evidence to do so.⁴⁰

More recently, the Supreme Court itself permitted credit for privileged harm in the constitutional tort case of *Carey v. Piphus*.⁴¹ As noted above, the Court quickly accepted the defendant’s premise that the plaintiff could not recover compensatory damages for claims of inadequate process unless he could show that process would have made a difference in outcome. That portion of the opinion is worth reading in full, if only to savor its brevity:

In this case, the Court of Appeals held that if petitioners can prove on remand that “respondents [Piphus] would have been suspended even if a proper hearing had been held,” then respondents will not be entitled to recover damages to compensate them for injuries caused by the suspensions. The court thought that in such a case, the failure to accord

³⁸ *Id.*

³⁹ *Id.*

⁴⁰ For a similar example sounding in nuisance, see *Varjabedian v. City of Madera*, 572 P.2d 43 (Cal. 1977). In *Varjabedian*, the plaintiffs alleged that the odors wafting from a neighboring sewage treatment plant constituted a nuisance and sought damages based on their reduced property values. *Id.* at 46. The defendant plant operator sought to reduce its damages by arguing that sewage treatment plants reduce the value of neighboring properties whether they emit unlawful odors or not. *Id.* at 48. The Supreme Court of California agreed with the defendant in principle, but concluded that the jury’s damages assessment was based entirely on the actionable odors. *Id.* at 48–49.

⁴¹ 435 U.S. 247 (1978).

procedural due process could not properly be viewed as the cause of the suspensions. The court suggested that in such circumstances, an award of damages for injuries caused by the suspensions would constitute a windfall, rather than compensation, to respondents. We do not understand the parties to disagree with this conclusion. Nor do we.⁴²

Those three short words—“nor do we”—have had far-reaching effects on the law of damages for cases involving government defendants. Since *Carey*, the black-letter rule in Section 1983 suits has been that a procedural deprivation gives rise to compensatory damages only if the lack of procedure actually caused the plaintiff harm.⁴³ A government defendant that has denied a plaintiff due process can often take its damages down to zero by conjuring a hypothetical world in which it imposed the same harm on the plaintiff while checking all the procedural boxes.⁴⁴

This rule leaves a procedural due process plaintiff with two options for proving non-zero damages. First, she can allege some injury, such as mental and emotional distress, that came from the lack of process itself.⁴⁵ (Post-*Carey*, a few plaintiffs have succeeded in this way, though usually the emotional harm from being denied process pales in comparison to the value of the underlying interest that was actually taken.⁴⁶) If that path is unpromising, the plaintiff’s only alternative is to prove that affording her

⁴² *Id.* at 260 (internal quotation marks and citations omitted).

⁴³ *Id.* at 263.

⁴⁴ The High Court of Australia recently adopted a similar rule in a suit for false imprisonment brought by a convict whose “periodic detention” was revoked in favor of full-time imprisonment without following the required procedures. *Lewis v Australian Capital Territory* [2020] HCA 26 (5 August 2020) 1. While accepting that the plaintiff was indeed falsely imprisoned, a majority of Court found he could not show any compensatory damages, since the relevant agency had grounds to revoke his periodic detention and hypothetically *could have* done so while affording the plaintiff all the process he was due. See *id.* at 71 (Edelman, J.) (“The correct method of framing the counterfactual is therefore to ask whether [the plaintiff] would lawfully have been subject to the same imprisonment but for the decision of the Board made in denial of procedural fairness. The answer to that question is ‘yes.’”).

⁴⁵ *Carey*, 435 U.S. at 264.

⁴⁶ See, e.g., *Laje v. R.E. Thomason Gen. Hosp.*, 665 F.2d 724, 730 (5th Cir. 1982) (affirming \$20,000 award for mental anguish and emotional distress but vacating award for \$32,400 loss of income); *Contract Design Grp., Inc. v. Wayne State Univ.*, 635 F. App’x 222, 233–34 (6th Cir. 2015) (affirming \$100,000 of emotional distress caused by deficient termination of contractor); see also *Burt v. Abel*, 585 F.2d 613, 616 (4th Cir. 1978) (remanding to allow plaintiff to prove actual injury caused by deprivation of procedure); *Wilson v. Taylor*, 658 F.2d 1021, 1032 (5th Cir. 1981) (remanding, with expressed skepticism, to permit plaintiff to present evidence of damage resulting from deprivation of process).

the appropriate process would have actually prevented her deprivation.⁴⁷ In the bulk of cases, a plaintiff alleging procedural impropriety must come prepared to litigate the substantive merits.

The liberal availability of credit for privileged harm has imposed a serious obstacle to damages in procedural due process cases. In the wake of *Carey*, plaintiffs across the country alleging improper dismissal from their government employment have seen their damages awards vacated or significantly reduced because of the possibility that they could have been afforded the appropriate process and terminated all the same.⁴⁸ Likewise, plaintiffs alleging deprivations of process relating to their detention in solitary confinement,⁴⁹ involuntary commitment to psychiatric facilities,⁵⁰ interstate extradition,⁵¹ termination of public

⁴⁷ See *Knudson v. City of Ellensburg*, 832 F.2d 1142, 1149 (9th Cir. 1987) (finding plaintiff was entitled to disability benefits on the merits); *Stein v. Bd. of N.Y.*, 792 F.2d 13, 18–19 (2d Cir. 1986).

⁴⁸ See *Kendall v. Bd. of Educ.*, 627 F.2d 1, 6 (6th Cir. 1980) (remanding for determination of whether plaintiff would have been terminated if a proper hearing had been held); *Wilson*, 658 F.2d at 1035 (denying backpay to police officer discharged without proper procedure); *Conley v. Bd. of Trs.*, 707 F.2d 175, 182 (5th Cir. 1983) (“[I]f it be proved that plaintiffs would have been terminated even if proper procedures had been used, that proof of actual injury must flow from the loss of the procedural rights.”); *Patkus v. Sangamon-Cass Consortium*, 769 F.2d 1251, 1265 (7th Cir. 1985) (reducing due process claim to nominal damages after defendants demonstrated employee would have been discharged even if “proper pre-deprivation procedures had been employed”); *Stein*, 792 F.2d at 18–19; *Fraternal Ord. of Police Lodge No. 5 v. Tucker*, 868 F.2d 74, 81 (3d Cir. 1989); *Brewer v. Chauvin*, 938 F.2d 860, 864–65 (8th Cir. 1991) (en banc) (vacating backpay award); *Hopkins v. Saunders*, 199 F.3d 968, 979–80 (8th Cir. 1999); *McClure v. Indep. Sch. Dist. No. 16*, 228 F.3d 1205, 1213 (10th Cir. 2000); *Nzomo v. Vt. State Colls.*, 411 A.2d 1366, 1367–68 (Vt. 1980) (affirming denial of reinstatement, back pay, and moving expenses because failure to afford process did not affect ultimate termination decision); *Bd. Of Educ. v. Crawford*, 395 A.2d 835, 842 (Md. 1979); *District of Columbia v. Gray*, 452 A.2d 962, 965 (D.C. 1982) (vacating award of backpay and reinstatement; remanding to determine whether employee could have been terminated with process); *County of Dallas v. Wiland*, 216 S.W.3d 344, 357 (Tex. 2007); see also *McGhee v. Draper*, 639 F.2d 639, 645 (10th Cir. 1981) (“The burden remains on the plaintiff to prove the link between the due process violation and any particular consequence that would support an award of damages.”); *D’Iorio v. County of Delaware*, 592 F.2d 681, 690–92 & n.16 (3d Cir. 1978) (invoking *Pullman* abstention, in part because “even if it were determined in this case that only the County Council was authorized to discharge [the plaintiff, his] claim to back pay would appear to be undermined seriously by proof that the County Council would have discharged him using proper procedures”).

⁴⁹ See *Patterson v. Coughlin*, 905 F.2d 564, 568 (2d Cir. 1990).

⁵⁰ See *Warren v. Pataki*, 823 F.3d 125, 141–42 (2d Cir. 2016). The parties in *Warren* apparently referred to the argument for privileged harm as the “no harm, no foul” defense. See *id.* at 141.

⁵¹ See *Brown v. Nutsch*, 619 F.2d 758, 764 & n.9 (8th Cir. 1980).

benefits,⁵² property condemnation,⁵³ suspension of taxicab licenses,⁵⁴ and garnishment of tax refunds⁵⁵ have run headlong into arguments for privileged harm. In general, if the State can show that it could have inflicted the same harm by affording process, it will owe only nominal damages as a result of denying process.⁵⁶

Privileged harm plays a similarly central role in intellectual property law. For example, a patent holder seeking lost profits must contend with the argument that a defendant who sold an infringing product could have made the same sales with a non-infringing product. As the Federal Circuit has explained:

[A] fair and accurate reconstruction of the “but for” market also must take into account, where relevant, alternative actions the infringer foreseeably would have undertaken had he not infringed. Without the infringing product, a rational would-be infringer is likely to offer an acceptable noninfringing alternative, if available, to compete with the patent owner rather than leave the market altogether.⁵⁷

In other words, lost profits must take into account the competitive injury the defendant *could have* inflicted without running afoul of the patent laws. Sales that could have been lost to non-infringing substitutes are not cognizable for purposes of patent damages because they are a quintessentially privileged harm.

The typical way that privileged harm affects lost profits calculations is by introducing hypothetical competing products into the damages

⁵² See *Alexander v. Polk*, 750 F.2d 250, 264 (3d Cir. 1984) (affording government defendant the opportunity to make privileged harm arguments on remand).

⁵³ See *Brody v. Village of Port Chester*, 345 F.3d 103, 121 (2d Cir. 2003).

⁵⁴ See *Nnebe v. Daus*, 306 F. Supp. 3d 552, 558 & n.1 (S.D.N.Y. 2018).

⁵⁵ See *Watts v. Wing*, 308 A.D.2d 391, 391 (N.Y. App. Div. 2003) (accepting the “general proposition that plaintiffs would not be entitled to recover damages to compensate them for their losses resulting from the government’s violation of their due process rights if such violation was not the cause of the plaintiffs’ losses”).

⁵⁶ To be clear, *Carey*’s acceptance of privileged harm is a damages rule, *not* a holding about the elements of a procedural due process claim. A plaintiff who was denied process but deserved the outcome may nonetheless recover nominal damages. See *Carey v. Piphus*, 435 U.S. 247, 266 (1978); *Zinermon v. Burch*, 494 U.S. 113, 126 n.11 (1990). In other contexts, lack of proof that the plaintiff would have obtained a different result causes the claim for liability to fail entirely. See, e.g., *Texas v. Lesage*, 528 U.S. 18, 21 (1999) (per curiam).

⁵⁷ *Grain Processing Corp. v. Am. Maize-Prods. Co.*, 185 F.3d 1341, 1350–51 (Fed. Cir. 1999).

counterfactual.⁵⁸ But some patent defendants have pushed arguments for privileged harm in more creative directions. For example, in a long-running dispute involving disposable cameras, Jazz Photo argued that the “first sale” doctrine permitted it to replace the film and resell cameras that had previously been manufactured and sold by the patent holder, Fuji. That argument lost on the merits: at the time, the first-sale defense applied only to goods previously sold in the United States,⁵⁹ and ninety percent of Jazz Photo’s used cameras came from foreign markets.⁶⁰ But the plaintiff’s triumph on liability proved less lucrative than one might have thought. For damages purposes, Jazz Photo simply conjured an imaginary world in which it *would have* qualified for the first-sale defense by purchasing only cameras previously sold in the United States.⁶¹ Privileged harm took lost profits entirely off the table.

Crediting privileged harm does not necessarily leave the plaintiff without a remedy. In some cases, the credit simply reduces the magnitude of compensatory damages.⁶² And even in a case in which credit for privileged harm eliminates compensatory damages entirely, the plaintiff might still receive nominal damages, punitive damages, and attorneys’ fees.⁶³ Nonetheless, crediting privileged harm can significantly reduce the stakes of a given case—and appears to reliably do so in a number of areas of law.

⁵⁸ See, e.g., *Micro Chem., Inc. v. Lextron, Inc.*, 318 F.3d 1119, 1123 (Fed. Cir. 2003) (evaluating whether alternative product was actually available); *Apple Inc. v. Samsung Elecs. Co.*, 786 F.3d 983, 1004 (Fed. Cir. 2015) (discussing whether Samsung’s proposed alternatives would have been adequate substitutes), *rev’d on other grounds*, *Samsung Elecs. Co. v. Apple Inc.*, 137 S. Ct. 429 (2016).

⁵⁹ See *Jazz Photo Corp. v. Int’l Trade Comm’n*, 264 F.3d 1094, 1110–11 (Fed. Cir. 2001). The Federal Circuit’s rule that only U.S. sales trigger the first-sale doctrine was overturned in *Impression Products, Inc. v. Lexmark International, Inc.*, 137 S. Ct. 1523, 1530–31 (2017).

⁶⁰ See *Fuji Photo Film Co. v. Jazz Photo Corp.*, 249 F. Supp. 2d 434, 451–52 (D.N.J. 2003), *aff’d*, 394 F.3d 1368 (Fed. Cir. 2005).

⁶¹ *Fuji Photo Film Co.*, 249 F. Supp. 2d at 455.

⁶² For example, a patent holder may still receive a lost profits award, even if that award is reduced to account for potential sales of non-infringing substitutes. See *Grain Processing Corp.*, 185 F.3d at 1350–51.

⁶³ Cf. *Carlson v. Green*, 446 U.S. 14, 22 n.9 (1980) (“[A]fter *Carey* punitive damages may be the only significant remedy available in some § 1983 actions where constitutional rights are maliciously violated but the victim cannot prove compensable injury.”).

C. Limits and Complications

In other cases, however, courts have either rejected or overlooked credit for privileged harm. A few examples will be helpful to illustrate why the question of privileged harm is nowhere as simple as it might first appear.

First, as mentioned above, a number of state courts have refused to credit privileged harm when government employees allege they were denied the procedural protections promised by their employment contracts. In *Piacitelli v. Southern Utah State College*,⁶⁴ for example, the Utah Supreme Court explicitly declined to apply *Carey*'s damages rule to the plaintiff's contract claim. The state was denied any credit for privileged harm and, as a result, was liable for back pay until the contractually specified process had been substantially performed.⁶⁵

Though they are not irreconcilable, the contrasting holdings of *Piacitelli* and *Carey* are at least counterintuitive. When the State denies procedural protections promised by a contract, the State is made to pay damages until it jumps through the procedural hoops. But when that same State denies the procedural protections promised by the U.S. Constitution, damages are constrained by a kind of "no harm, no foul" principle. Given our legal culture's high regard for constitutional rights and generally forgiving attitude towards broken contractual promises, it would not be surprising if these two claims received different remedial rules. But the damages rules break exactly opposite to the direction one would expect.⁶⁶

⁶⁴ 636 P.2d 1063 (Utah 1981).

⁶⁵ *Id.* at 1069–70.

⁶⁶ *Piacitelli*'s strict treatment of government actors who fail to follow the required procedures seems to be more consistent with historical practice than *Carey*'s forgiving approach. For roughly three hundred years, procedural errors automatically triggered a loss of credit for privileged harm as a result of the *trespass ab initio* doctrine. See Prosser and Keeton on the Law of Torts, *supra* note 27, § 25, at 150–52 (explaining the doctrine and tracing it to the *Six Carpenters Case* of 1610) (citing (1610) 77 Eng. Rep. 695). For example, if a sheriff lawfully seized goods but failed to get the required appraisal before selling them, he became liable in trespass not just for the loss caused by the lack of appraisal, but for the full value of the goods. See, e.g., *Wilson v. Ellis*, 28 Pa. 238, 240 (1857) ("For where the law has given an authority, it will protect persons from the abuse of the authority, by leaving the abuser in the same situation as though he had acted without any authority . . ."). At bottom, this was a sweeping rejection of credit for privileged harm. The doctrine was heavily criticized in the early twentieth century, recommended for abandonment in the Second Restatement of Torts, and is apparently dead today. See Oliver Wendell Holmes Jr., *The Path of the Law*, 10 Harv. L. Rev. 457, 469 (1897) (criticizing the *trespass ab initio* doctrine as a rule that "simply persists from blind imitation of the past"); Restatement (Second) of Torts § 214 cmt. e (Am. L. Inst. 1965) (calling for abandonment of the doctrine's "peculiar and anomalous fiction");

Frequently, the treatment of privileged harm receives only passing attention. For example, the damages considered by the Supreme Court in *Bigelow v. RKO Radio Pictures*⁶⁷ appear to have been entirely rooted in privileged harm. The plaintiffs, owners of independent movie theaters, brought suit under the Sherman and Clayton Acts alleging that the major distributors had conspired to give their own theaters exclusive access to first-run films.⁶⁸ But, as Justice Frankfurter observed in dissent, the distributors could have reserved first-run films for their own theaters without a conspiracy, and such a preference for their own affiliates would not have offended the antitrust laws.⁶⁹ In his view, the plaintiffs had failed to show any damage beyond what the defendants were equally capable of imposing through lawful, non-conspiratorial action. Accounting for privileged harm would thus take the plaintiffs' damages to zero.⁷⁰

Two things are notable about *Bigelow*. First, though all of the plaintiffs' damages appeared to be rooted in privileged harm, seven Justices voted to reinstate the \$120,000 verdict (automatically trebled to \$360,000) without mounting any response to Justice Frankfurter's argument for privileged harm.⁷¹ And second, this appears to have truly been Justice Frankfurter's argument—the defendants' briefing made no mention of privileged harm, either in the Seventh Circuit or the Supreme Court.⁷² In this way, *Bigelow* is an example of another puzzling phenomenon. In many cases, arguments for privileged harm seem to leap from the facts, but were apparently not advanced by the litigants.⁷³ In a case like this, it is no surprise that the court did not award credit for privileged harm. The real mystery is why the defendant failed to seek it.⁷⁴

Foreign Car Ctr. v. Essex Process Serv., 821 N.E.2d 483, 488 (Mass. App. Ct. 2005) (“While older cases support the position of [plaintiff], more recent cases . . . suggest that where minor errors occur, the plaintiff has to show some damage.”).

⁶⁷ 327 U.S. 251 (1946).

⁶⁸ *Id.* at 253–55.

⁶⁹ *Id.* at 267.

⁷⁰ *See id.*

⁷¹ *Id.* at 254, 266.

⁷² Because of the age of this case, we cannot determine whether arguments for privileged harm were advanced in the district court.

⁷³ In some cases, defendants might be making a strategic decision not to advance any damages arguments at all, instead focusing the factfinder's attention on the issue of liability. But this explanation is incomplete. Some defendants—including the defendants in *Bigelow*—contest damages vigorously, yet still do not make arguments about privileged harm. *Id.* at 254.

⁷⁴ *See Berkey Photo, Inc. v. Eastman Kodak Co.*, 603 F.2d 263, 296–98 (2d Cir. 1979) (holding that damages in Sherman Act § 2 suit should be calculated by comparing prices charged by anticompetitive monopolist to those that would be charged by a “pristine

In other cases, by contrast, defendants have pushed arguments for privileged harm uncomfortably far. For example, consider the arguments around lost profits in another long-running patent dispute, this one between Cardiac Pacemakers and St. Jude Medical. The defendant, St. Jude, had acquired a company called Ventritex, which manufactured implantable cardiac defibrillators under a license granted by the patent holder, Cardiac.⁷⁵ As it happened, St. Jude's acquisition of Ventritex triggered a "change of control" provision in the license agreement, terminating the license and leaving St. Jude vulnerable to a new infringement suit by Cardiac.⁷⁶ When Cardiac tried to claim lost profits, St. Jude invoked a but-for world in which it had never acquired Ventritex in the first place. Lost profits should be completely unavailable, St. Jude argued, because Ventritex *could have* remained an independent company, kept its license, and imposed all the same harm on Cardiac.⁷⁷

Quite reasonably, Cardiac protested that this took privileged harm too far. In its view, lost profits should be reduced only by the availability of alternative product designs, not alternative corporate structures. While recognizing this as an "unusual" case, the district court discerned nothing in existing Federal Circuit precedent that would limit privileged harm to substitute products.⁷⁸ Over Cardiac's repeated objections, St. Jude made these arguments for privileged harm to the jury, which in turn found that Cardiac was not entitled to any lost profits.⁷⁹ The case then made repeated trips to the Federal Circuit, but that court never had occasion to rule on these boundary-pushing arguments for privileged harm.⁸⁰

monopolist"); 3 Phillip E. Areeda & Herbert Hovenkamp, *Antitrust Law: An Analysis of Antitrust Principles and Their Application* ¶ 657b, at 116 (2d ed. 2002) (noting that the rule of *Berkey Photo* may lead to little or no damages in some cases).

⁷⁵ See *Cardiac Pacemakers, Inc. v. St. Jude Med., Inc.*, 418 F. Supp. 2d 1021, 1037 (S.D. Ind. 2006), rev'd in part on other grounds, 576 F.3d 1348 (Fed. Cir. 2009).

⁷⁶ *Id.*

⁷⁷ *Id.* at 1036–37.

⁷⁸ *Id.* at 1037–38.

⁷⁹ See *Cardiac Pacemakers, Inc. v. St. Jude Med., Inc.*, No. 96-1718, 2002 WL 1801525, at *2, *76–77 (S.D. Ind. July 5, 2002) (original denial of post-trial objection to the relevant jury instruction); see also *Cardiac Pacemakers*, 418 F. Supp. 2d at 1038–39 (similar ruling on remand).

⁸⁰ The history of *Cardiac Pacemakers, Inc. v. St. Jude Medical, Inc.* is complex. See 381 F.3d 1371 (Fed. Cir. 2004); 144 F. App'x 106 (Fed. Cir. 2005); *In re Cardiac Pacemakers, Inc.*, 183 F. App'x 967 (Fed. Cir. 2006); 303 F. App'x 884 (Fed. Cir. 2008); 315 F. App'x 273 (Fed. Cir. 2009); 576 F.3d 1348 (Fed. Cir. 2009). Despite this large volume of appellate activity, the Federal Circuit never reached the argument for privileged harm.

Perhaps that is a good thing. These cases force questions that existing theory is simply not equipped to answer. For example, is there any reason to treat privileged harm differently depending on whether the claim sounds in constitutional tort (*Carey*) or contract (*Piacitelli*)? If the answer is that credit for privileged harm must be denied in order to give meaning to contractual promises, then shouldn't the same strict treatment be given to a patent infringer (like *St. Jude*) who lost its license under the express terms of a contract? And do those who offend the antitrust laws really have to pay for privileged harm not just once but three times over? Or did the *Bigelow* defendants simply commit a devastating strategic error, failing to make a winning argument that could have taken their damages to zero?

For its valor, the maxim of *damnum absque injuriâ* is not up to the task of such line drawing. It is clear that courts sometimes award credit for privileged harm and other times do not. But further work is necessary to explain how this distinction is and ought to be drawn.

II. UNDERSTANDING THE EFFECTS OF PRIVILEGED HARM

This Part begins by exploring the effects of reducing or not reducing damages to account for privileged harm. As this discussion will show, treatment of privileged harm can have significant consequences, both for prospective defendants' incentives to comply with the law and for prospective plaintiffs' incentives to sue. In this way, the denial of credit for privileged harm has many similarities with the imposition of punitive damages. Finally, the magnitude of privileged harm can swing widely from case to case, with a counterintuitive dependence on the strictness of the underlying substantive law that will be explored in greater detail below.

A. *Damages Functions With and Without Privileged Harm*

To illustrate the effects of crediting or not crediting privileged harm, let us consider a simple case of nuisance. Suppose a prospective defendant, Polluter, must determine how much of a particular substance to emit. We will assume that Polluter's emissions will impose some degree of harm on Neighbor, and that the harm suffered by Neighbor increases as a function of Polluter's emissions level.

Suppose that the law permits Polluter to release some low level of emissions and impose some degree of harm on Neighbor without

incurring liability at all. But once Polluter's emissions rise above a certain threshold, Polluter becomes liable to Neighbor. Call this liability-triggering threshold "L."

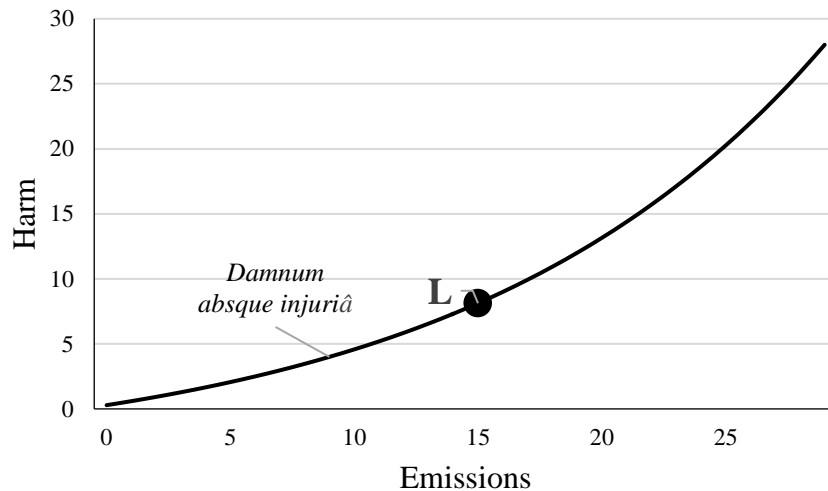


Chart 1

The solid black line of Chart 1 illustrates the relationship between Polluter's emissions and Neighbor's harm.⁸¹ If Polluter releases no emissions, Neighbor suffers no harm. As Polluter selects higher emissions levels, Neighbor begins to suffer increasing amounts of harm.

Under a regime of absolute liability, the damages owed to Neighbor would simply track the harm function depicted by the solid line, causing Polluter to internalize all the harm caused to Neighbor. But by hypothesis, the relevant law permits some amount of harm without injury (*dammun absque injuriâ*), provided Polluter does not cross the liability-triggering threshold L. In the region to the left of L, Polluter does harm but owes no damages.

But what should happen when Polluter *does* cross the threshold L, resulting in liability under the substantive law? There are two possibilities. Start with this one:

⁸¹ The curve shown here is merely an example. As we will soon see, the shape of this curve can affect the desirability of crediting privilege harm. See *infra* Charts 5–7 and accompanying text.

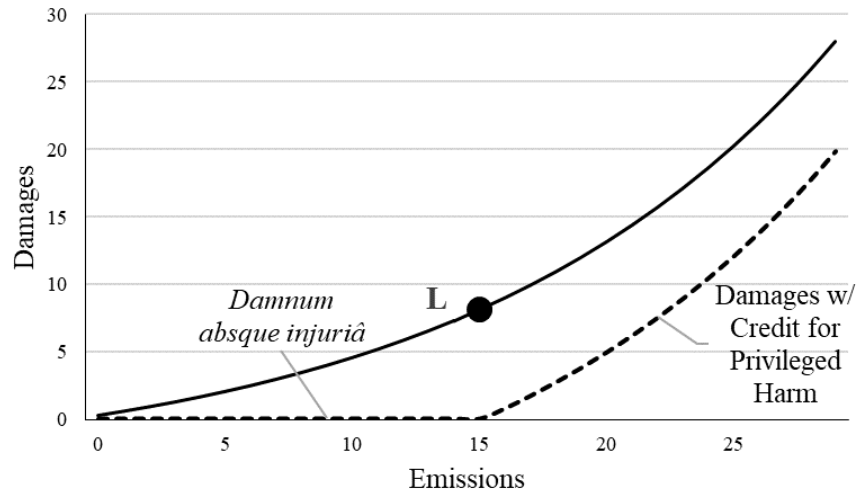


Chart 2

Chart 2 illustrates the defendant's damages function if credit is given for privileged harm. As before, the solid black line illustrates the harm caused by the defendant's emissions. The dotted line represents the damages owed by the defendant as a function of his chosen emissions level. To the left of L, the defendant owes no damages. When the defendant's emissions reach the level shown by point L, he becomes liable under the substantive law. Nonetheless, under the rule shown here, the defendant can argue that all the harm he caused *up to* level L is non-compensable as *damnum absque injuriâ*. The defendant's damages therefore start at zero and become non-zero only as the defendant imposes some additional harm beyond what the substantive law permitted him to impose without liability.

Now consider the effect of denying credit for privileged harm:

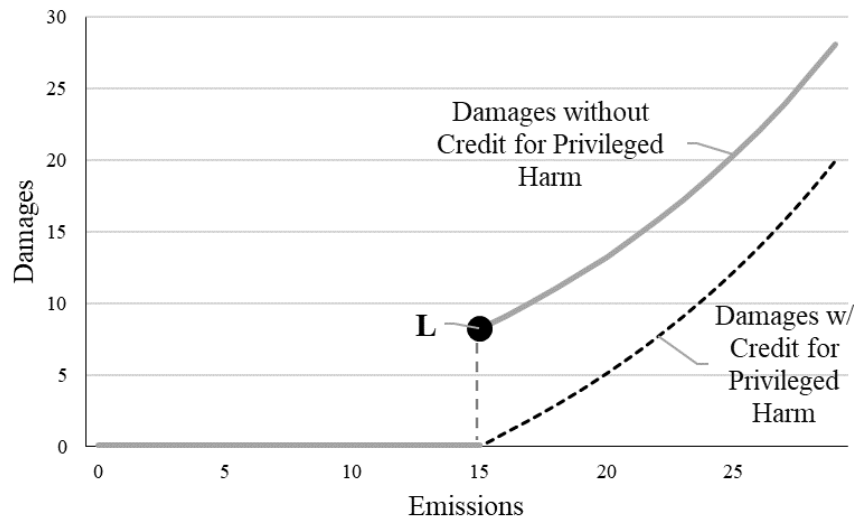


Chart 3

In Chart 3, the dotted black line continues to represent the defendant's damages function if credit is given for privileged harm. The new, gray line depicts the damages owed by the defendant if credit is *not* given for privileged harm. As before, at emissions levels below L, the defendant owes no damages for any harm those emissions might cause. But now, the defendant immediately starts to owe non-zero damages at the threshold of liability itself. Without a credit for privileged harm, triggering liability makes the defendant responsible for all the harm caused by his emissions up to that point, even though this harm would have been non-compensable if the defendant had kept his emissions within the legal limits.

As this example illustrates, the difference between crediting and not crediting privileged harm is essentially a matter of starting points. If damages are reduced to account for privileged harm, the defendant's damages begin from zero. If they are not, the defendant's damages take off from some non-zero amount reflecting the harm that defendant has already caused up to that point. To the right of the liability-triggering threshold, the two damages functions move in parallel, each tracking the incremental harm caused by the defendant's additional emissions. The distance between the two curves is constant: anywhere above L, it is equal

to the one-time credit given (or not) at the liability-triggering threshold itself.

B. Marginal Incentives of Prospective Defendants

Let us now consider the marginal incentives of a prospective defendant like Polluter facing either of the two damages functions depicted above. We will assume that, in the absence of damages, Polluter would obtain some benefit by increasing his emissions (such as greater factory output or mitigation costs avoided), though we will not make any assumptions about the magnitude of these benefits. The question is how these two damages functions will affect Polluter's choice of an emissions level.

In selecting a level of emissions, Polluter will compare the marginal benefit he receives by emitting an additional unit to the marginal damages he faces. Marginal damages are given by the derivative of the damages function—the slope of the curves depicted in Chart 3. When the curve is flat, each additional unit of emissions is cheap. For example, in the no-liability region to the left of L, moving from 10 units of emissions to 11 units of emissions imposes no additional damages. If these additional emissions benefit Polluter in some way (as we assume they do), he will therefore choose to set his emissions above 10. But when the curve is steep, each additional unit of emissions is expensive for Polluter, which will tend to discourage additional emissions. For example, far to the right of L, moving from 25 to 26 units of emissions results in significantly more liability. Without knowing more about Polluter's marginal benefits in that zone, we cannot say whether it will be profitable for him to take this incremental step. But, in general, the steeper the damages function at a particular point, the greater the deterrence of polluters considering whether to increase their emissions past that level.

Comparing the slope of the two damages functions depicted in Chart 3, they turn out to be remarkably similar. At all points to the left of L, their slope is zero. Moreover, at all points to the right of L, their slope is also identical, as each traces the incremental harm imposed by the defendant as a result of emissions above the liability-triggering threshold. Across much of the range of potential emissions levels, Polluter will have the same marginal incentives, whether or not he expects to receive a credit for privileged harm.

An important exception, however, occurs right around the threshold of liability itself. Here, the two damages functions have dramatically different slopes. If credit is awarded for privileged harm, the defendant's

damages liability turns on smoothly, gradually increasing from zero. But if credit is not awarded for privileged harm, the defendant experiences a sharp, one-time jump in his expected damages. Even a slight incursion into the realm of liability can trigger a wall of damages, making the marginal cost of those additional emissions extremely high.⁸²

This difference in treatment around the liability-triggering threshold can have important consequences for how Polluter will set his emissions levels. To illustrate, suppose Polluter obtains a constant \$1 of incremental benefit for each additional unit of pollution. This means he will start his emissions at 0, and increase from there until he encounters a point on the damages curve where the slope becomes greater than \$1/unit. If credit is given for privileged harm (that is, his damages track the dotted line in Chart 3), Polluter will find it profitable to go well past the legal limit. Based on the shape of the curve shown in Chart 3, it appears that he can increase his emissions to somewhere around 20 units before encountering marginal damages sufficient to deter additional pollution. But if credit is *not* given for privileged harm (that is, his damages track the *gray* line in Chart 3), Polluter will encounter marginal deterrence much greater than \$1/unit at the moment he becomes liable under the substantive law. In this example, the treatment of privileged harm will thus determine whether Polluter stays within legal limits or continues to increase his emissions to substantially higher levels.

To be clear, this is just an illustration, and one can construct examples in which Polluter will comply with or violate the substantive law regardless of the treatment afforded to privileged harm. The more general point is that, compared to awarding credit for privileged harm, denying credit for privileged harm will increase marginal deterrence around the threshold of substantive liability. Depending on how profitable

⁸² An analogous “cliff effect” can be found in the treatment of causation in negligence cases. See Jennifer Arlen, *Tort Damages*, in 2 *Encyclopedia of Law and Economics* 682, 685 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000) (“[Under a negligence regime,] an injurer’s expected liability increases dramatically from zero to the expected damage award if he reduces his care-taking from due care to less than due care. . . . This result does not hold, however, if the application of ‘but for’ causation effectively eliminates the discontinuity in the injurer’s expected liability function.”). The desirability (or not) of this discontinuity at the liability stage has been explored at some length. See, e.g., Mark F. Grady, *A New Positive Economic Theory of Negligence*, 92 *Yale L.J.* 799 (1983); Marcel Kahan, *Causation and Incentives To Take Care Under the Negligence Rule*, 18 *J. Legal Stud.* 427 (1989); Stephen Marks, *Discontinuities, Causation, and Grady’s Uncertainty Theorem*, 23 *J. Legal Stud.* 287 (1994); Keith N. Hylton & Haizhen Lin, *Negligence, Causation, and Incentives for Care*, 35 *Int’l Rev. L. & Econ.* 80, 81–82 (2013).

lawbreaking is, this single-point wall of damages may create incentives for some prospective defendants to stay within legal limits.

So which rule is better? The answer is that it depends. The choice between crediting and not crediting privileged harm turns out to be closely related to a distinction well-known in the law and economics literature: the choice between prices and sanctions. As Robert Cooter describes in his seminal work on the subject, the hallmark of a sanction is “an abrupt jump in an individual’s costs when he passes from the permitted zone into the forbidden zone where behavior is sanctioned.”⁸³ The purpose of a sanction is to discourage people from crossing that line at all.⁸⁴ This is precisely the effect of denying credit for privileged harm. By contrast, the hallmark of a price is a continuous internalization of the harm caused by a person’s actions.⁸⁵ Prices convey no judgment about prohibited actions. The defendant simply pays the cost of that which he is permitted to do.⁸⁶ Once again, this is precisely the effect of awarding credit for privileged harm. Beyond the threshold of liability, the defendant simply internalizes the incremental harm he imposes on the plaintiff.

Neither prices nor sanctions are universally preferable—one of Cooter’s insights is that the choice between prices and sanctions ought to depend on the rule maker’s access to various forms of information. When it is cheaper for officials to observe the external costs of the conduct at issue, prices are preferable. But when it is cheaper for officials to observe the optimal standard of care, sanctions are preferable.⁸⁷ Cooter provides a number examples of how a rule maker may sometimes find herself in the former position and sometimes find herself in the latter position.⁸⁸ From the perspective of marginal deterrence, therefore, the optimal treatment of privileged harm may be contingent on the substantive law at issue, turning on whether price-like or sanctions-like damages are better suited for regulating the underlying conduct. So even if marginal deterrence of prospective defendants were the only relevant consideration, we should not expect a single rule to answer all questions about the treatment of privileged harm.

⁸³ See Cooter, *supra* note 16, at 1523.

⁸⁴ *Id.* at 1524 (“A sanction is a detriment imposed for doing what is *forbidden* . . .”).

⁸⁵ *Id.* at 1527–28.

⁸⁶ *Id.*

⁸⁷ *Id.* at 1533.

⁸⁸ *Id.* at 1533–37.

C. The Likelihood of Enforcement

Apart from these effects on defendant's incentives, treatment of privileged harm can also have significant consequences for whether a plaintiff will enforce her rights. Because privileged harm will sometimes constitute the bulk of a plaintiff's potential damages, the economic feasibility of suit can turn dramatically on whether this harm is included or excluded from damages calculations.

The impact of privileged harm is generally greatest near the threshold of liability itself. When there is only a small difference between what the defendant did and what would have been permitted, it will often be difficult for a plaintiff to show that she has suffered much in the way of incremental harm.⁸⁹ In that case, privileged harm may make up the bulk of her potential damages, and excluding it may deprive the plaintiff of the economic incentive (or ability) to bring suit at all. When this occurs, it may be infeasible for a plaintiff to enforce her rights, except in cases where the defendant has engaged in conduct far exceeding the legal threshold of liability.⁹⁰

On the other hand, denying credit for privileged harm can work as a kind of enforcement bounty, moving the threshold of a profitable suit back towards the legal threshold of liability. To illustrate, let us refer again to Chart 3. The defendant becomes liable when his emissions exceed 15 units, so the plaintiff will win on the merits for any emission level above that. But assume that it costs the plaintiff \$10 to bring a lawsuit. With a credit for privileged harm, the plaintiff will not find it profitable to sue unless the defendant's emissions exceed 25 units, since her prospective damages are less than \$10 at all points below that. But if the defendant is denied credit for privileged harm, the plaintiff will find it profitable to sue whenever the defendant's emissions exceed 17 units.

⁸⁹ Another situation in which it can be difficult for plaintiffs to prove incremental harm—even for conduct well past the threshold of liability—arises when the damages function is quite flat to the right of the threshold. That special case is taken up in Section IV.A.

⁹⁰ This lack of enforcement interest could, in turn, have a compounding effect on the price-like operation of damages discussed in the prior section. A defendant contemplating a small incursion into the plaintiff's legally protected interest will expect to owe only small damages, and the deterrent effect of those already small damages will be discounted further to reflect the unlikelihood of a plaintiff bringing an unprofitable lawsuit. Note, however, that some plaintiffs may be motivated to bring suits for reasons besides compensatory damages: to obtain punitive damages, to establish ownership, and so on. Fee shifting and statutory damages can also be used to make it profitable for plaintiffs to enforce their rights when compensatory damages are small. These possibilities are discussed at various points to follow.

Moving from the “privileged harm” curve to the “no credit” curve causes the plaintiff’s prospective recovery to exceed her cost of suit at lower emissions levels.⁹¹

To be clear, this is an illustration, not a universal claim. One can construct examples where it is always profitable for a plaintiff to sue (or is never profitable to sue) regardless of how the court treats privileged harm. All else being equal, though, a rule denying credit for privileged harm increases the plaintiff’s expected recovery, which may make it profitable to bring more suits at the margin.

So is more vigorous enforcement of the plaintiff’s rights a good thing or a bad thing? Again, the answer will depend on context. In some cases, increased private enforcement of the underlying substantive law may be for the better. But it is also possible to have too much enforcement. Indeed, the consensus in the law and economics literature is that there is no general rule about whether private plaintiffs will bring too many suits or too few. As Steven Shavell sums it, “[T]he private incentive to bring suit is fundamentally misaligned with the socially optimal incentive to do so,” and, unfortunately, “the deviation between them could be in either direction.”⁹² In the face of this ambiguity, the same effect could be either a reason to award credit for privileged harm or a reason to deny it, depending on whether a substantive regime is likely to yield too much or too little private enforcement near the threshold of liability.

D. Privileged Harm and Punitive Damages

To summarize the observations so far, denying credit for privileged harm can have two important consequences. First, it can produce sanctions-like damages, increasing marginal deterrence for prospective defendants to comply with the substantive law. Second, it can increase plaintiffs’ expected recoveries, thereby resulting in more vigorous enforcement by plaintiffs in marginal cases.

⁹¹ This example also reveals a subtlety in the choice between damages and sanctions discussed in the prior Section. Once the plaintiff’s costs of suit are factored in, denying credit for privileged harm may not be sufficient to set damages in a sanctions mode around the threshold of liability. In Chart 3, the threshold of liability is emissions level 15, and the damages formally owed by the defendant jump significantly at that point. But depending on the plaintiff’s costs of suit—and whether or not other remedial tools like attorneys’ fees are available—the defendant may or may not expect to face sanctions-like damages until an emissions level somewhat higher than that.

⁹² Steven Shavell, *Foundations of Economic Analysis of Law* 391 (2004).

These two features map neatly onto two of the primary rationales for awarding punitive damages: deterring wrongful conduct and encouraging private enforcement of the law.⁹³ Given this, one might wonder if damages for privileged harm are, at root, really just another form of punitive damages.

Tempting as this comparison may be, the similarities with punitive damages only go so far. In fact, there are two important features that distinguish compensation for privileged harm from punitive damages. The first is doctrinal. Typically, punitive damages operate well past the liability-triggering threshold, targeting conduct that is reckless, egregious, or otherwise exceptional.⁹⁴ Privileged harm, by contrast, is potentially at play in any case involving compensatory damages, and is most significant in cases involving conduct close to the threshold of liability. So while both tools can be used to increase deterrence and enforcement, they tend to affect two very different kinds of cases.

⁹³ See Dan B. Dobbs & Caprice L. Roberts, *Law of Remedies* § 3.11(3), at 323–25 (3d ed. 2018); *Exxon Shipping Co. v. Baker*, 554 U.S. 471, 491–95 (2008) (discussing goals of deterrence and creating incentives to sue). Likewise, the Supreme Court has explained that the purposes of awarding treble damages under the Sherman Act are: (1) to “penaliz[e] wrongdoers and deter[] wrongdoing” and (2) to “counterbalanc[e] the difficulty of maintaining a private suit.” See *Brunswick Corp. v. Pueblo Bowl-O-Mat, Inc.*, 429 U.S. 477, 485–86 & n.10 (1977) (internal quotation marks and citation omitted). Under the heading of “counterbalancing the difficulty of maintaining a private suit,” one could arguably include the risk that wrongdoing may go undetected. See *Mathias v. Accor Econ. Lodging, Inc.*, 347 F.3d 672, 677 (7th Cir. 2003) (“The award of punitive damages in this case thus serves the additional purpose of limiting the defendant’s ability to profit from its fraud by escaping detection and (private) prosecution. If a tortfeasor is ‘caught’ only half the time he commits torts, then when he is caught he should be punished twice as heavily in order to make up for the times he gets away.”). But see A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 *Harv. L. Rev.* 869, 898 (1998) (“Courts sometimes allude to the possibility of escaping liability, but they rarely recognize its importance with respect to deterrence.”).

⁹⁴ See, e.g., *Exxon Shipping Co.*, 554 U.S. at 493 (“The prevailing rule in American courts also limits punitive damages to cases of . . . ‘enormity,’ where a defendant’s conduct is ‘outrageous,’ owing to ‘gross negligence,’ ‘willful, wanton, and reckless indifference for the rights of others,’ or behavior even more deplorable.” (citations omitted)); *Restatement (Second) of Torts* § 908 (Am. L. Inst. 1979) (“Punitive damages are . . . awarded against a person to punish him for his outrageous conduct”); *Halo Elecs., Inc. v. Pulse Elecs., Inc.*, 136 S. Ct. 1923, 1932 (2016) (holding that enhanced damages under the Patent Act “are not to be meted out in a typical infringement case, but are instead designed as a ‘punitive’ or ‘vindictive’ sanction for egregious infringement behavior”). But see 15 U.S.C. § 15(a) (2018) (providing for automatic trebling of damages in suits brought under the Clayton Act); 18 U.S.C. § 1964(c) (2018) (same for RICO).

The second distinction is functional. If punitive damages are imposed as a multiplier on compensatory damages,⁹⁵ they will have two effects on the defendant's damages function. First, they will create a discontinuity at the punitive damages threshold. Second, they will increase the slope of the damages function from that point onward. Privileged harm can do something similar to the first, but can do nothing like the latter. To illustrate, consider a situation in which both kinds of sanctions are imposed:

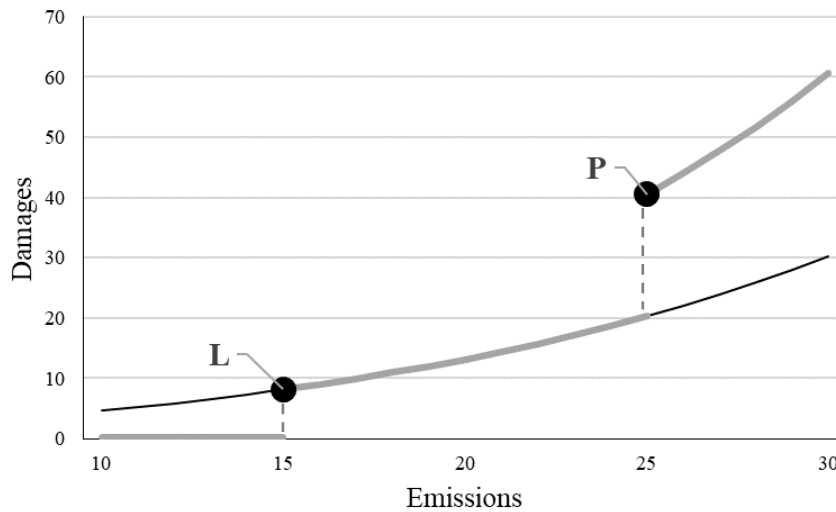


Chart 4

⁹⁵ Judges and lawyers often think of punitive damages as a multiplier on compensatory damages, likely because this ratio plays a central role in the due process review of punitive damages awards. See *State Farm Mut. Auto. Ins. Co. v. Campbell*, 538 U.S. 408, 416–18 (2003). However, we do not know whether the deciders—lay jurors—actually conceive of punitive damages in this way. They might, for example, tend to award punitive damages as a lump sum when a defendant crosses some threshold of egregiousness, and without tying that figure to the degree of actual injury. If that is the case, punitive damages would produce the discontinuity shown below, but not the change in slope to the right of P. At least some regimes, however, explicitly peg the level of enhanced damages to the underlying compensatory damages, and these regimes would produce a damages function like the one in Chart 4. See, e.g., 35 U.S.C. § 284 (2018) (providing for discretionary trebling of patent damages); *Halo Elecs.*, 136 S. Ct. at 1932 (describing the kind of egregious conduct that would merit such damages).

Chart 4 illustrates how a rule denying credit for privileged harm would interact with the imposition of punitive damages. The thin black line indicates the harm actually caused by Polluter's emissions—that is, the damages Polluter would owe under a rule of absolute liability. The thick gray line represents the damages owed by a prospective defendant in a jurisdiction that both denies credit for privileged harm and imposes punitive damages.

In this hypothetical, Polluter becomes liable under the substantive law at point L, somewhere just past emissions level 15. As before, denying credit for privileged harm creates a discontinuity in Polluter's damages function around the threshold of liability, as he instantly becomes responsible for all the harm caused by his conduct. Then, in the range immediately above L, Polluter's damages track the harm caused by his incremental emissions. Depending on the actual harm caused by additional emissions, this may or may not provide much in the way of marginal deterrence. This continues until Polluter's emissions approach the threshold for punitive damages, shown by point P. Around that point, Polluter's damages experience a *second* discontinuity as the punitive damages multiplier kicks in. This constant multiplier increases the slope of Polluter's damages function from that point onward.⁹⁶

From the perspective of marginal incentives, therefore, privileged harm and punitive damages share the feature of imposing a discontinuity on the prospective defendant's damages function. Immediately below the threshold of liability, Polluter faces significant marginal damages—just as he does immediately below the threshold of punitive damages. But treatment of privileged harm is a one-time shot, with an effect on marginal deterrence only at the threshold of liability itself. Punitive damages, by contrast, have the ability to increase marginal deterrence across a wider range of conduct.

From the perspective of incentive to sue, the differences between the two tools are more dependent on circumstances. If, as a doctrinal matter, punitive damages are not available until the defendant's conduct becomes egregious, then the two remedial tools create enforcement incentives for two different kinds of cases. Denying credit for privileged harm moves the frontier of a profitable suit closer to the threshold of liability, while

⁹⁶ In the example shown here, punitive damages are equal to compensatory damages, and thus have the effect of doubling the defendant's total damages at any given point to the right of P.

punitive damages increase incentives for litigants to pursue cases against more egregious wrongdoers. But, to be clear, this distinction is a product of punitive damages doctrine, rather than some universal truth. If enhanced damages are imposed at the threshold of liability itself (as they are in antitrust and RICO cases⁹⁷), this particular distinction disappears.

E. The Magnitude of Privileged Harm

The comparison to punitive damages raises a final descriptive question: just how significant is privileged harm for purposes of bottom-line damages calculations? In cases like *Bigelow v. RKO Radio Pictures* and *Carey v. Piphus*, privileged harm appears to constitute the bulk of the plaintiffs' potentially cognizable damages. But there are also identifiable categories of cases in which privileged harm will be quite small. This section will briefly explore the factors that determine the magnitude of privileged harm.

As discussed above, the difference between crediting and not crediting privileged harm comes down to a one-time offset awarded at the threshold of liability itself. Below that threshold, the defendant's damages are zero either way; above that threshold, the two damages functions move in parallel. The treatment of privileged harm simply determines whether compensatory damages start smoothly from zero or begin with a discontinuous jump.

The magnitude of privileged harm is equal to the distance between these two parallel damages functions. And the size of this potential credit is affected by the substantive law itself. To see this, let us return to the simple form of the Neighbor/Polluter example illustrated in Chart 3:

⁹⁷ See 15 U.S.C. § 15(a) (2018) (antitrust); 18 U.S.C. § 1964(c) (2018) (RICO); see also *Smith v. Wade*, 461 U.S. 30, 53–54 (1983) (noting that some common law torts trigger eligibility for compensatory damages and punitive damages at the same threshold).

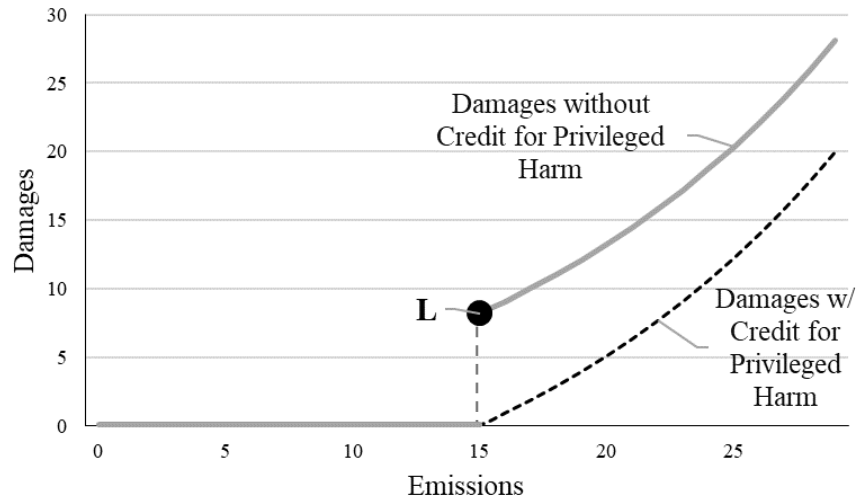


Chart 3 (reprise)

In this example, the underlying substantive law permits Polluter to release a substantial amount of emissions and impose a significant degree of harm without becoming liable to Neighbor at all. As shown in Chart 3, the liability threshold is around 15 units of emissions, which causes Neighbor to suffer about \$9 of harm. This figure—the maximum amount of harm Polluter could impose without liability—establishes the spread between the two damages functions at all points above the threshold of liability.

By making the underlying substantive law stricter, however, we can significantly reduce the magnitude of the credit for privileged harm. For example, if we tweak our hypothetical pollution law so that liability is triggered by emissions at level 5 instead of 15, the significance of privileged harm is greatly diminished:

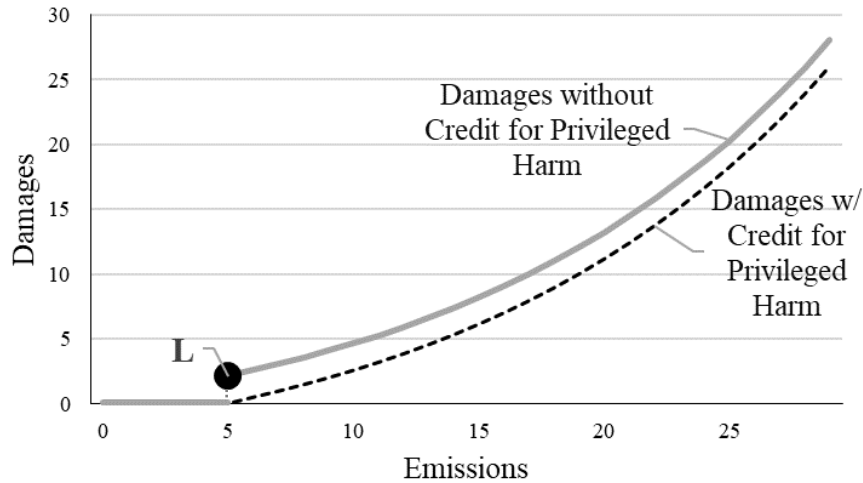
*Chart 5*

Chart 5 illustrates the same damages functions as before, but operating on a substantive law that triggers liability at a much lower threshold. Now, the maximum amount of harm that Polluter can impose without incurring liability is only about \$2. This establishes a much smaller spread between the two damages functions, rendering the treatment of privileged harm much less important as a result of the stricter substantive law.

The point generalizes. The greater the harm that a defendant can impose without triggering liability, the larger the consequences of crediting or not crediting privileged harm. The converse holds as well, and can be taken to its limit: when the law imposes absolute liability, credit for privileged harm is worthless, since the amount of harm the defendant was privileged to impose is stuck at zero.

Moreover, because the magnitude of the credit for privileged harm is fixed at the threshold of liability, its significance as a share of the defendant's overall liability will diminish as the defendant moves farther past the liability-triggering threshold. So long as the underlying harm function remains upward sloping (that is, incremental emissions continue to impose additional harm), the defendant's total compensatory damages will continue to increase as he moves further into the region of liability. But all the while, the magnitude of the privileged harm remains the same. The importance of this issue therefore fades as the defendant's conduct goes farther and farther beyond the threshold of liability.

These observations can be synthesized into two conditions that must be present for privileged harm to be significant in a particular case. First, the underlying substantive law must permit the defendant to impose meaningful harm on the plaintiff without triggering liability. If the underlying substantive law compensates even low-level injuries, privileged harm will be small in absolute terms. Second, the defendant must not have caused significant harm beyond those permitted at the threshold of liability. If the defendant has gone well past the threshold of liability (and done much more harm as a result), privileged harm will necessarily be small as a percentage of total damages. The most important cases for privileged harm, therefore, will involve both (a) a substantive law that permits lots of uncompensated harm and (b) a defendant operating close to the threshold of liability.

Once these conditions are recognized, it makes sense that the doctrine of privileged harm has been most extensively developed in cases of employment law and intellectual property.⁹⁸ Wrongful termination cases are excellent candidates for litigation of this issue, since employers are generally permitted to impose a substantial harm on their employees—layoffs—without owing compensation. Likewise, because the intellectual property laws are a tailored exception to a general policy of free competition,⁹⁹ infringers often would have had other ways to harm a plaintiff's business without triggering liability at all. The substantive law's tolerance of large amounts of uncompensated harm makes treatment of privileged harm a central question.

III. GENERAL PRINCIPLES

With this descriptive account in mind, we now turn to normative questions about when and how credit for privileged harm should be afforded. This part evaluates the *when* question through the lens of three basic and persistent considerations: (1) the need to establish appropriate incentives for prospective defendants; (2) the need to establish appropriate incentives and provide adequate compensation for plaintiffs; and (3) the need to manage remedial complexity. Part IV will then turn to the *how* question, introducing additional concerns that may arise in particular cases.

⁹⁸ See *supra* Section I.B.

⁹⁹ See *Bonito Boats, Inc. v. Thunder Craft Boats, Inc.*, 489 U.S. 141, 150–51 (1989).

A. Incentives for Defendants

First, the treatment of privileged harm can affect the marginal incentives of prospective defendants. As discussed above, requiring defendants to pay for privileged harm creates sanctions-like damages around the threshold of liability.¹⁰⁰ By crossing that threshold, a defendant will face a sharp discontinuity in his damages function, going from owing nothing to owing an amount equal to all the harm that he could have imposed through lawful conduct. A rule that denies credit for privileged harm may thus discourage prospective defendants from making small incursions into the legally protected interests of others.

This particular feature is not universally desirable. In some cases, it may be quite fitting that minor incursions will trigger only minor damages. Indeed, unless there is some reason for wanting to deter prospective defendants from crossing into the zone of liability, sanctions-like damages could result in excessive precautions.¹⁰¹ But when this deterrence is desirable, privileged harm can be used to impose a one-time, sanctions-like discontinuity around the threshold of the liability. This interest in marginal deterrence could justify damages that might otherwise appear excessive.

Because the preferability of prices versus sanctions depends on the circumstances,¹⁰² we should likewise not expect a single, trans-substantive rule that will tell us when to credit damages for privileged harm and when not to. While that may be unsatisfying, it is important to recognize the question of privileged harm as the policy question that it ultimately is. Indeed, at least some of the confusion in existing doctrine can be attributed to a failure to confront these tradeoffs head on. For example, consider the Utah Supreme Court's efforts in *Piacitelli v. Southern Utah State College* to explain why it was treating privileged harm so differently than the U.S. Supreme Court had in *Carey v. Piphus*. First, the *Piacitelli* court noted that *Carey* and its progeny involved constitutional torts brought under Section 1983, while it was reviewing damages in a contract case.¹⁰³ That is obviously valid as a doctrinal distinction, though it fails to explain why the damages rule should be

¹⁰⁰ See *supra* Section II.B.

¹⁰¹ Cf. Polinsky & Shavell, *supra* note 93, at 879–82 (noting potential for punitive damages to have a similar effect); Grady, *supra* note 82, at 811 (describing similar effect in negligence law).

¹⁰² See Cooter, *supra* note 16, at 1533.

¹⁰³ See 636 P.2d 1063, 1068–69 (Utah 1981).

different for one kind of claim versus the other. More to the point, the court then noted the sanctioning effect of denying credit for privileged harm:

This result comports with what we deem to be sound policy for contractual employer-employee relations. It will encourage employers to comply promptly with their contractual termination procedures, and if they fail to do so will impose the monetary consequences on the party at fault. If the rule were otherwise, the employer could discharge an employee summarily and then omit or delay the contractual termination procedures with impunity so long as it was in possession of evidence which, when ultimately provided, would justify the discharge.¹⁰⁴

This is a perfectly sound explanation for why Utah, as a policy matter, might want to deny credit for privileged harm in cases of contractually improper discharge. But the same reasoning would apply in equal force in *Carey*, or in any of the other cases in which courts *do* reduce damages to account for privileged harm. Denying credit for privileged harm in cases like *Carey*, for example, would likewise “encourage employers to comply promptly with [the Constitution], and if they fail to do so will impose the monetary consequences on the party at fault.”¹⁰⁵ Instead, the U.S. Supreme Court adopted a rule which, to borrow the language of the Utah Supreme Court, would permit a government employer to “discharge an employee summarily and then omit or delay the [constitutionally required] procedures with impunity so long as it was in possession of evidence which, when ultimately provided, would justify the discharge.”¹⁰⁶ Right as it may be, the argument proves too much.

Carey and *Piacitelli* can potentially be reconciled, however, by appreciating the tradeoffs at play in the treatment of privileged harm. For example, one could argue that there is less need to pile deterrence onto government employers contemplating constitutional violations, since the Section 1983 regime in some cases permits plaintiffs to recover punitive damages and attorneys’ fees.¹⁰⁷ By contrast, the same employer

¹⁰⁴ *Id.* at 1069.

¹⁰⁵ *Id.*

¹⁰⁶ *Id.*

¹⁰⁷ At least one court following *Carey*’s rule has made this reasoning explicit. See *Brewer v. Chauvin*, 938 F.2d 860, 864 (8th Cir. 1991) (“[T]here is no reason to fear that public employers will rush to deny employees their . . . rights. First, an award of attorney fees is proper, even when only nominal damages are awarded Second, punitive damages may be awarded to a plaintiff recovering only nominal damages”).

contemplating a contractually deficient firing will face neither the enhanced remedies nor the public stigma that comes with committing a constitutional tort. As a result, the risk of small incursions into the plaintiff's legally protected interests might be much greater in cases involving contractual claims than in cases involving constitutional ones. Given that, a rule of crediting privileged harm for constitutional torts but not for breaches of contract might be perfectly sensible.

But one could plausibly argue the exact opposite: that there is more need to deter constitutionally deficient firings than contractually deficient firings. For example, one might argue that attorneys' fees and punitive damages are available in constitutional tort cases *exactly because* those violations are more serious than garden-variety breaches of contract. By those lights, enhanced remedies like punitive damages and attorneys' fees should be seen as complements rather than substitutes for the sanction of denying credit for privileged harm—all of the above will assure that even “minor” violations of the Constitution are appropriately deterred. So, far from defending the distinction, one could just as sensibly argue that the *Carey* progeny and the *Piacitelli* progeny have gotten the rule exactly backwards.

The fact that these conclusions are debatable only demonstrates why treatment of privileged harm ought to be made explicit. Behind what might seem like a straightforward damages issue lurks subtle predictions about the likelihood and seriousness of small incursions just across the threshold of liability. Rather than assuming that a single approach will be appropriate in all cases, courts should consider whether it is preferable to impose sanctions-like or price-like damages around the threshold of liability for the particular substantive rule at issue.

Though they may or may not think in those specific terms, this is a distinction judges routinely make in other contexts. For example, in deciding whether to enjoin a defendant from future violations of the plaintiff's rights, a court will often be determining whether the plaintiff's legal entitlement should be subject only to a price-like obligation to pay damages or instead backed by the stiffer threat of contempt.¹⁰⁸ Judges will also be familiar with other indicia that sanctions are appropriate, such as

¹⁰⁸ To be clear, this will not be the only question at play when considering a plaintiff's request for an injunction—a number of other factors may also be relevant to the court's exercise of equitable discretion.

whether a violation would be publicly viewed as shameful, or whether it would be considered more blameworthy if done intentionally.¹⁰⁹

And this question need not be posed on a blank slate. In many areas of law, doctrinal signposts already indicate whether damages are intended to work as prices or sanctions for conduct just past the threshold of liability. For example, the Clayton Act provides that private plaintiffs injured by antitrust violations “shall” be awarded treble damages and attorneys’ fees.¹¹⁰ There is no ambiguity in the Clayton Act or its legislative history that violating the antitrust laws is meant to be a sanctionable offense.¹¹¹ The same goes for statutes like the Racketeer Influenced and Corrupt Organizations (“RICO”) Act and the False Claims Act, both of which impose automatic damage multipliers the moment a defendant becomes liable.¹¹² Plainly, the damages resulting from antitrust conspiracies, racketeering, and defrauding the government are not to meant to be just another cost of doing business.¹¹³ A strict rule denying credit for privileged harm would therefore be consistent with the greater remedial framework.¹¹⁴

¹⁰⁹ See Cooter, *supra* note 16, at 1538 n.33.

¹¹⁰ 15 U.S.C. § 15(a) (2018).

¹¹¹ See *Brunswick Corp. v. Pueblo Bowl-O-Mat, Inc.*, 429 U.S. 477, 485–86 & n.10 (1977) (observing punitive nature of antitrust damages and collecting legislative history); see also *Blue Shield of Va. v. McCready*, 457 U.S. 465, 472 (1982) (“[I]n enacting § 4: Congress sought to create a private enforcement mechanism that would deter violators and deprive them of the fruits of their illegal actions”); *Perma Life Mufflers, Inc. v. Int’l Parts Corp.*, 392 U.S. 134, 139 (1968) (“The plaintiff who reaps the reward of treble damages may be no less morally reprehensible than the defendant, but the law encourages his suit to further the overriding public policy in favor of competition.”). But see *Berkey Photo, Inc. v. Eastman Kodak Co.*, 603 F.2d 263, 297–98 (2d Cir. 1979) (awarding credit for privileged harm in a Sherman Act case on the theory that even an anticompetitive monopolist should not “forfeit [the] legitimately acquired advantage” of its monopoly).

¹¹² See 18 U.S.C. § 1964(c) (2018) (RICO private suit and treble damages provision); 31 U.S.C. § 3729(a)(1) (2012) (False Claims Act civil penalty and treble damages provision).

¹¹³ For this reason alone, the majority in *Bigelow* was right to reject Justice Frankfurter’s argument for privileged harm: Congress clearly intended antitrust damages to work as sanctions rather than prices. See *Bigelow v. RKO Radio Pictures, Inc.*, 327 U.S. 251, 254 (1946) (noting that damages were automatically trebled). For other examples of sanctions-like remedial provisions, see 12 U.S.C. § 2607 (2018) (real estate settlement anti-kickback provision); 15 U.S.C. § 1693f (2018) (consumer credit protection); Conn. Gen. Stat. § 52-560 (2019) (cutting or destroying a tree intended for use as a Christmas tree punishable by a payment to the injured party of five times the tree’s value).

¹¹⁴ Some readers may ask whether combining the privileged-harm sanction with statutory treble damages creates a risk of *too much* deterrence. While this is possible, one of Cooter’s observations about prices versus sanctions is that when damages operate as sanctions, prospective defendants’ conduct will be much more strongly affected by the standard of care

By contrast, other areas of substantive law lack these sanctions-like features or reserve them for the most egregious cases. For example, punitive damages are categorically unavailable in contract law—even if the parties undisputedly agreed to them in advance.¹¹⁵ The law and economics literature is also well-developed on this point: contract damages should be imposed as prices, not sanctions.¹¹⁶ It would thus be incongruous to deny credit for privileged harm, at least in ordinary contract cases.¹¹⁷ As another example, patent law features a treble damages provision,¹¹⁸ which might at first glance suggest an interest in sanctioning patent infringers. But for nearly two hundred years these enhanced damages have been reserved for “egregious cases of misconduct beyond typical infringement.”¹¹⁹ Congress deliberately created space between the threshold of liability and the threshold of punishment because the prior rule of punishing “[t]he defendant who acted in ignorance or good faith . . . was manifestly unjust.”¹²⁰ This suggests that patent damages are meant to work as prices (not sanctions) near the threshold of patent liability. As with breach of contract cases, the larger remedial framework implies that damages should generally be reduced to account for privileged harm.

To be sure, other areas of law lack such signposts. A court will not always have the benefit of statutory features, legislative history, or a scholarly consensus indicating that damages ought to operate as prices or sanctions for a particular class of conduct. And, in the end, these are only

rather than the level of sanctions. See Cooter, *supra* note 16, at 1532 (“[M]istakes in computing the level of the sanction or the frequency of its application are not crucial, because most people will conform in spite of these mistakes.”). It is therefore more important to ensure damages provide *some* sanction rather than to worry that they might provide too much. So while a legislature could make a different judgment, denying credit is more naturally consistent with a rule for trebling of compensatory damages where the statute is silent on the treatment of privileged harm.

¹¹⁵ See Restatement (Second) of Contracts §§ 355–56 (Am. L. Inst. 1981); 2 Simon Greenleaf, *A Treatise on the Law of Evidence* § 257, at 253 (16th ed. 1899).

¹¹⁶ See Cooter, *supra* note 16, at 1545–46.

¹¹⁷ Note that these are merely guideposts, however, and should not preclude a policy maker from creating different treatment in particular kinds of cases. For example, the Utah Supreme Court was not necessarily wrong to impose sanctions-like deterrence in *Piacitelli*. This rule should simply be recognized as an exception to the price-like approach generally taken to contractual damages.

¹¹⁸ See 35 U.S.C. § 284 (2018).

¹¹⁹ See *Halo Elecs., Inc. v. Pulse Elecs., Inc.*, 136 S. Ct. 1923, 1932, 1935 (2016) (summarizing history and synthesizing the modern test).

¹²⁰ *Seymour v. McCormick*, 57 U.S. (16 How.) 480, 488 (1854).

signposts. There is no guarantee that the existing regime is correct in its election of a price-like or sanctions-like damages regime. But the response to this ambiguity should be further study and conscientious doctrinal development, rather than masking an important question behind a supposedly uniform rule.

B. Incentives and Compensation for Plaintiffs

The treatment of privileged harm can potentially affect plaintiffs' behavior as well. These considerations turn out to be less persuasive—and stake a weaker claim to judicial competence—than the defendant-focused concerns just discussed.

As noted above, denying credit for privileged harm can encourage private enforcement in response to slighter incursions. All else equal, the frontier of a profitable suit moves closer to the threshold of liability when plaintiffs stand to recover privileged harm.¹²¹ That could be a good thing or a bad thing, depending on whether plaintiffs would otherwise enforce their rights too strictly or too leniently in cases close to the threshold of liability. It is possible to have either too much or too little enforcement, and there is no reason to assume that the private incentives to sue will err in one direction or the other.¹²²

To determine whether additional incentives for plaintiffs to enforce their rights would be beneficial, a court would need to know two things: (1) the magnitude of the public benefits promised by that particular kind of suit, and (2) the existing likelihood that plaintiffs holding those claims will enforce those rights. Unfortunately, neither factor is reliably amenable to judicial assessment. The magnitude of the public benefits flowing from private lawsuits is a famously elusive question.¹²³ And the private incentives to enforce legal rights are similarly opaque, particularly since unpursued infractions are by their very nature not usually observable by courts. As a result, a court deciding whether to deny privileged harm as a way of encouraging future lawsuits would be poorly positioned to assess the need for such an intervention.¹²⁴

¹²¹ See *supra* Section II.C.

¹²² See Shavell, *supra* note 92, at 391.

¹²³ See *id.*

¹²⁴ The matter is further complicated by the possibility that the public and private benefits of an enforcement action can vary from case to case. When the benefits of suits are heterogenous, the ideal policy might not be subsidized enforcement across the board, but rather incentives targeted at the subset of cases likely to yield the largest public benefits. See

Even in situations where it is clear that more vigorous enforcement would be a good thing, a court would also need to consider other remedial tools that could be used to accomplish a similar result. For example, attorneys' fees, statutory damages, and damage multipliers can likewise be used to encourage plaintiffs to enforce their rights even in cases where compensatory damages are small.¹²⁵ As compared to these alternatives, a distinguishing feature of denying credit for privileged harm is that it will primarily affect cases close to the threshold of liability.¹²⁶ If the goal is to encourage enforcement near the threshold of liability, but without affecting incentives to assert, settle, or litigate cases far beyond that threshold, a privileged harm bounty may be preferable to enhanced damages and attorneys' fees, which have the potential to affect a broader spectrum of cases.¹²⁷

But relying on privileged harm to induce additional private enforcement comes with its limitations. The implicit enforcement bounty is fixed by external factors and in some cases may be too small to be effective. Particularly when the underlying substantive law is already strict—that is, it holds defendants responsible for their conduct at low levels of harm—then the difference between crediting and not crediting privileged harm will not amount to very much.¹²⁸ So while denying credit for privileged harm may be a useful and elegant tool for inducing private enforcement in some cases, it is not a tool that will be universally available.

When it comes to determining whether plaintiffs require additional incentives to enforce their rights, doctrinal signposts are unlikely to be much help. For example, the fact that a substantive area of law already provides attorneys' fees to prevailing plaintiffs does not imply that a court should also award a privileged harm bounty. From a plaintiff-focused perspective, it is just as possible that attorneys' fees and a privileged harm

Stephen Yelder, *Do Patent Challenges Increase Competition?*, 83 U. Chi. L. Rev. 1943, 1996 (2016) (arguing that policies to encourage patent challenges should focus on those most likely to yield pro-competitive benefits).

¹²⁵ See Bert I. Huang, *Surprisingly Punitive Damages*, 100 Va. L. Rev. 1027, 1046–48 (2014) (noting that statutory damages may create incentives for plaintiffs to enforce their rights).

¹²⁶ See *supra* Section II.E.

¹²⁷ See Huang, *supra* note 125, at 1048, 1059 fig.1 (noting that statutory damages can inadvertently exceed the fixed subsidy necessary to induce suit).

¹²⁸ See *supra* Chart 5 and accompanying text.

bounty should be seen as mutually exclusive policy tools, and that affording both would lead to an excessive rate of private enforcement.

Given all this, arguments sounding in plaintiffs' incentives are better suited to legislative rather than judicial appraisal. This is not to say that privileged harm's consequences for plaintiffs are minor or irrelevant. But the ambiguous sign of these effects—and the lack of judicial competence to assess them—ultimately make them unhelpful for a court confronting a question of privileged harm.

But there is another plaintiff-focused effect that courts *are* competent (and often inclined) to consider: the possibility that compensation for privileged harm will constitute a “windfall” to the plaintiff. Judges and lawyers tend to incorporate the principle of *damnum absque injuriâ* into their conception of a plaintiff's baseline state. For example, consider again the Neighbor/Polluter hypothetical introduced above. The substantive law permits Polluter to impose about \$9 of harm a month without owing any compensation.¹²⁹ Against that backdrop, both parties might expect (and in fact regularly experience) the following outcome: Polluter emits, Neighbor loses \$9 of value, and Polluter owes nothing. So on the occasions on which Polluter's emissions exceed the legal threshold, Neighbor is only worse off by the amount that these unlawful emissions result in excess harm. And if damages are not reduced to account for privileged harm, Neighbor would actually come out *better* on the occasions that Polluter violates the law than on the occasions he complies with it. To a lawyer, at least, that sounds like a windfall.

A number of courts (including the Supreme Court in *Carey*) have invoked similar reasoning to reduce damages to account for privileged harm.¹³⁰ They are correct that these would constitute windfalls in the technical sense—that compensation for privileged harm would come as an “economic gain” that is “independent of work, planning, or other

¹²⁹ See *supra* Chart 2 and accompanying text.

¹³⁰ See *Carey v. Piphus*, 435 U.S. 247, 260 (1978) (“[A]n award of damages for injuries caused by the suspensions would constitute a windfall, rather than compensation”); *Hostrop v. Bd. of Junior Coll. Dist. No. 515*, 523 F.2d 569, 579 (7th Cir. 1975) (“[P]laintiff would be given a windfall at the expense of the taxpayers . . . [if] he is entitled to the same damages that would be recoverable if the contract had been terminated without just cause”); *Wheeler v. Mental Health & Mental Retardation Auth. of Harris Cnty.*, 752 F.2d 1063, 1071–72 (5th Cir. 1985); *Edwards v. Jewish Hosp. of St. Louis*, 855 F.2d 1345, 1352 (8th Cir. 1988) (“Under these circumstances, backpay would be a windfall to Edwards, even though he was the victim of intentional racial discrimination.”).

productive activities that society wishes to reward.”¹³¹ But even if windfalls are undesirable in general, we tolerate them routinely as a means to some other end.¹³² Punitive damages, for example, are even more obviously a windfall to the plaintiff who receives them, and yet are justified in the interest of deterring future egregious conduct.¹³³ In a similar way, the goal of establishing sanctions-like damages around the threshold of liability could justify leaving the plaintiff better off than she expected to be as a result of the defendant’s wrong.

C. Managing Complexity

Finally, in some cases, separately calculating the amount of privileged harm may not be worth the additional complication it brings. Accounting for privileged harm will often require a subtler counterfactual inquiry, raising decision costs and increasing the risk of error. Concerns of remedial complexity may sometimes justify a simplified damages calculation, even if it would be theoretically preferable to treat privileged harm in a more nuanced manner.

Because a number of case-specific considerations can influence the optimal relationship between accuracy and complexity, it is not possible to state a single, generally applicable rule for when this concern should trump the others. But drawing on the analysis above, we can at least identify a few categories of cases in which the comparative insignificance of privileged harm counsels in favor of simplification. To start, when the underlying liability regime is strict—requiring defendants to pay compensation even at low levels of harm—the magnitude of privileged harm will be small, no matter what else might be happening in the case.¹³⁴ As a result, there may be entire areas of law in which arguments for privileged harm can be categorically excluded with little threat to accuracy or policy goals.

¹³¹ Eric Kades, *Windfalls*, 108 *Yale L.J.* 1489, 1491 (1999). Of course, whether compensation for privileged harm is truly a “gain” depends on one’s choice of baseline. From the plaintiffs’ perspective, this compensation leaves him no better off than if the defendant had never harmed him in the first place.

¹³² See *id.* at 1526 (“Society gains more by tolerating double recoveries that serve deterrence than eliminating such windfalls at the cost of underdetering potential tortfeasors.”).

¹³³ See *Smith v. Wade*, 461 U.S. 30, 49 (1983) (“[D]eterrence of future egregious conduct is a primary purpose of . . . punitive damages”); *id.* at 58–59 (Rehnquist, J., dissenting) (“Despite these attempted justifications . . . [p]unitive damages are generally seen as a windfall to plaintiffs . . .”).

¹³⁴ See *supra* Section II.E.

Even in areas of law in which significant, uncompensated harm is tolerated, cases of particularly egregious conduct may present additional opportunities for simplification. Because the value of privileged harm is constant, the value of this credit will be small as a percentage of total damages in cases where the defendant has imposed a substantial amount of *unprivileged* harm on the plaintiff.¹³⁵ Under those conditions, privileged harm can be excluded from the damages model with only slight losses to accuracy.

These latter cases are harder to identify categorically. The most direct method—calculating the plaintiff’s total losses and comparing it to the losses the defendant was privileged to impose—relies on the very calculation we are hoping to avoid. Instead, some proxies may be helpful to identify cases where privileged harm is likely to constitute a small share of the defendant’s total damages. For example, cases involving egregious conduct far beyond the threshold of liability are good candidates for foreclosing arguments about privileged harm. Requiring the defendant to pay for privileged harm in these cases can simplify the remedial inquiry and deter future defendants from engaging in similar conduct. In cases of extreme violations, the twin goals of simplification and sanction may counsel against crediting privileged harm.¹³⁶

To be sure, it may not always be necessary for a court to manage remedial complexity in this way, since litigants will often have their own incentives to forgo arguments that are unlikely to make much difference. But sometimes one side or the other may have strategic reasons to increase the cost of litigation with arguments that will have only a small effect on the expected outcome at trial. This can be particularly powerful if the increased costs will be borne asymmetrically. For example, if a defendant can cheaply make a claim about privileged harm that is expensive for the plaintiff to rebut, it may be profitable to do so even if the claim itself has very low monetary value in isolation. Likewise, questions about privileged harm could be used to sow doubt about the plaintiff’s damages theory—to suggest that the evidence is unduly

¹³⁵ See *id.*

¹³⁶ In other areas of damages law, courts have imposed hard limits on calculation methodologies out of concerns for complexity. See, e.g., *Hanover Shoe, Inc. v. United Shoe Mach. Corp.*, 392 U.S. 481, 491–94 (1968) (prohibiting passing-on defense because proof would be too complex and uncertain).

speculative, and perhaps even cause the entire claim to fail.¹³⁷ A court can prevent certain strategic abuses of complexity by recognizing that privileged harm will not always be worth the candle.

* * *

In sum, a court determining whether to reduce damages for privileged harm should focus on two primary considerations: the need to deter prospective defendants and the need to manage remedial complexity. In confronting this question, the court will be implicitly setting damages around the threshold of liability to work as prices or sanctions, with attendant consequences for the marginal incentives of future defendants. At the same time, there are identifiable cases in which the magnitude of any privileged harm will be predictably small. In these situations, the court can dispense with the prices-versus-sanctions question and calculate damages in whichever way appears simplest.

Though treatment of privileged harm can have consequences for plaintiffs' incentives as well, the effects on this side are ambiguous, and their assessment will not typically be within the realm of judicial competence. (Plaintiff incentives might, however, serve as a legitimate basis for legislative intervention on the question of privileged harm.) Moreover, the plaintiff-focused consideration that has often seized judicial attention—that privileged harm come as a windfall—is less persuasive than it might first appear. Although it is certainly an acceptable starting position, the principle of avoiding super-compensatory damages should yield when necessary to deter defendant wrongdoing, as it regularly does in other contexts.

These high-level conclusions, however, are subject to a number of nuances and caveats in practice. The next Part addresses additional considerations that arise when implementing credit for privileged harm in specific cases.

¹³⁷ See, e.g., *Jenkins v. Pa. R.R. Co.*, 51 A. 704, 705–06 (N.J. 1902) (rejecting argument that plaintiff had failed to prove damages with sufficient accuracy when harm was caused by combination of tortious and non-tortious causes).

IV. SPECIAL CASES

A. The “No Remedy” Problem

The first complication we must consider is that sometimes more egregious conduct does not cause any additional harm to the plaintiff. When this occurs, damages for privileged harm will constitute the bulk of the compensatory damages potentially available to the plaintiff, no matter how badly a defendant might behave. This might be a reason to pause before reducing damages to account for privileged harm.

In the damages functions depicted above, more egregious conduct by the defendant reliably caused more harm to the plaintiff. This took some pressure off the privileged harm question. Even if a credit for privileged harm rendered damages small at first, we could rely on the upward-curving slope of the damages function to eventually restrain the defendant.¹³⁸

But suppose the damages function instead looks like this:

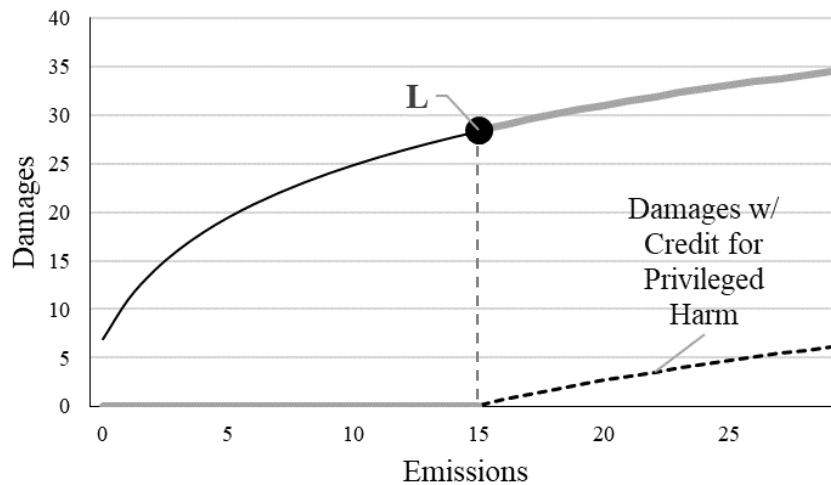


Chart 6

The thin black line in Chart 6 represents the harm actually imposed by the defendant’s conduct. The gray line depicts the damages owed by the defendant if credit is not given for privileged harm. As before, without

¹³⁸ See *supra* Section II.B.

the credit, damages jump from zero at the threshold of liability. Alternatively, if a credit for privileged harm is awarded, the defendant's damages (shown by a dotted black line) will take off smoothly from zero, and increase only as the defendant's more egregious conduct results in additional harm.

At the threshold of liability itself, the effect of crediting or not crediting privileged harm is the same here as it was in the examples illustrated above. What makes this example different from the others is that the damages function is very flat to the right of L—more egregious conduct does not impose much additional harm. When this occurs, crediting privileged harm will reliably leave the plaintiff with only small damages, even as the defendant's conduct moves well past the threshold of liability.

A damages function of this shape could potentially appear in almost any kind of case. There is no law of nature that requires the plaintiff's injury to continue getting worse as the defendant's conduct becomes more flagrant. But there are also certain causes of action where a flat damages curve can be expected with some regularity. For example, consider the common law tort of defamation, which allows a plaintiff to recover for the damage done by a false and defamatory statement or publication.¹³⁹ Because truth is always a defense,¹⁴⁰ the law of defamation lends itself to massive amount of privileged harm. The publication of any terrible thing that also happens to be true about a person would be unrecoverable as *damnum absque injuriâ*.¹⁴¹

Taken to its logical conclusion, this defense could stand in the way of just about any claim for defamation damages. Aside from the most virtuous of plaintiffs (and perhaps the most vicious of defamations), there are surely unflattering truths to be spoken about many plaintiffs that could do similar reputational harm. If defamation defendants are given credit for privileged harm, plaintiffs will recover only for the incremental harm their reputation suffered as a result of the actually spoken falsity over the unspoken truth. The difference might often be zero.

¹³⁹ See Restatement (Second) of Torts §§ 558 & 621 (Am. L. Inst. 1977).

¹⁴⁰ Id. § 581A.

¹⁴¹ A different common law tort—publicity given to private life—provides an eventual backstop to the widespread dissemination of truthful information. See id. § 652D. However, that tort has a few elements not found in defamation: the matter must be highly offensive to a reasonable person, and not of legitimate concern to the public. Id. Moreover, this latter tort requires *publicity*, not merely publication. See id. § 652D cmt. a (distinguishing the two).

Another set of cases with this feature are those involving violations of equal protection. When the substance of a claim is that the government has unjustifiably treated one group differently than another, there are often two potential solutions to the unequal treatment. First—the outcome universally preferred by equal protection plaintiffs—the government could treat the dispreferred group more favorably. Second, the government could treat the preferred group *less* favorably. Because the government could simply deny both groups the benefits in question without running afoul of equal protection, these claims have almost fatal levels of privileged harm baked into them from the beginning.¹⁴²

To make this general point more specific, consider a generation of Section 1983 claims brought against the City of Chicago for its policy of uniformly strip searching female—but not male—inmates without cause or suspicion.¹⁴³ If advanced as an equal protection claim, privileged harm would likely zero out all money damages to the victim of these searches. After all, the City could have simply expanded its strip-search policy to include male inmates too, and imposed all of the same privacy violations and traumas on the female inmates. It is only when the government's conduct violates some *other* prohibition—such as the Fourth Amendment—that such cases begin to produce non-privileged harm.¹⁴⁴

In cases like these, the all-or-nothing nature of privileged harm may call for special treatment. A rule that credits privileged harm may result in very small amounts of compensatory damages, not just in borderline cases but across the entire spectrum of potential conduct. Depending on the substantive regime, this could have far-reaching consequences, both for deterring defendants and giving plaintiffs incentives to sue.

¹⁴² This feature of equal protection claims can frustrate the design of equitable relief as well. See *Sessions v. Morales-Santana*, 137 S. Ct. 1678, 1698–1700 (2017) (holding that child citizenship statute violated equal protection but denying plaintiff a remedy). To be clear, none of this is meant to suggest that failures to afford equal protection result in only small amounts of harm. It may just be that the harm is largely structural or societal, and does not always map onto an individualized injury that can be measured on behalf of a particular plaintiff.

¹⁴³ See *Mary Beth G. v. City of Chicago*, 723 F.2d 1263, 1266–68 (7th Cir. 1983).

¹⁴⁴ See *id.* at 1268–73 (concluding that City's strip-search policy was unreasonable under the Fourth Amendment). But see *Florence v. Bd. of Chosen Freeholders of Burlington*, 566 U.S. 318, 322, 330 (2012) (permitting general inmate population strip searching without reasonable suspicion). Aside from the Fourth Amendment, another potential limit on privileged harm in these cases could come from the implausibility of the asserted counterfactual conduct. For example, if it would be extremely burdensome and therefore impractical to routinely strip search every man and woman arrested by the Chicago Police Department, the argument for privileged harm may be unavailable. This factually imposed limit on privileged harm is explored *infra* Section IV.D.

But before denying credit for privileged harm on these grounds, a court should first consider whether any other remedial tools besides compensatory damages are available to deter prospective defendants and ensure that plaintiffs have adequate incentives to enforce their rights. For example, if a plaintiff can potentially receive statutory damages, punitive damages, attorneys' fees, equitable relief, or a gains-based remedy, the underlying substantive law may retain some practical teeth, even if credit for privileged harm drastically reduces compensatory damages.¹⁴⁵

In the two examples introduced above, the availability of alternative remedies may justify different treatment. Government defendants who have denied plaintiffs equal protection face a number of potential consequences apart from compensatory damages, including attorneys' fees, injunctions, public shaming, and, in some cases, punitive damages.¹⁴⁶ By contrast, the First Amendment seriously constrains the remedial alternatives in defamation suits. Courts are reluctant to enjoin speech,¹⁴⁷ and neither presumed nor punitive damages are available unless the plaintiff can show with "clear and convincing proof that the defamatory falsehood was made with knowledge of its falsity or with reckless disregard for the truth."¹⁴⁸ As a result, in run-of-the-mill defamation cases, compensatory damages may very well be the only remedy on offer. This might justify awarding credit for privileged harm in equal protection cases but not in defamation suits, notwithstanding the similarly flat damages curves created by the two regimes.¹⁴⁹

¹⁴⁵ Cf. *Carlson v. Green*, 446 U.S. 14, 22 n.9 (1980) ("[A]fter *Carey* punitive damages may be the only significant remedy available in some § 1983 actions where constitutional rights are maliciously violated but the victim cannot prove compensable injury."). Another solution to this problem is to create an alternative head of damages entirely, one less closely tied to the losses the plaintiff actually suffered. See Jason NE Varuhas, *Lewis v Australian Capital Territory: Valuing Freedom*, 42 *Sydney L. Rev.* 123, 125, 133–34 (2020).

¹⁴⁶ See 42 U.S.C. § 1988(b) (2012) (permitting courts to award attorneys' fees to a prevailing party in constitutional torts and other cases); *Smith v. Wade*, 461 U.S. 30, 35–36 (1983) (recognizing availability of punitive damages in § 1983 suits). But see *City of Newport v. Fact Concerts, Inc.*, 453 U.S. 247, 271 (1981) (holding that municipalities enjoy absolute immunity against punitive damages in § 1983 suits).

¹⁴⁷ See, e.g., *Vance v. Universal Amusement Co.*, 445 U.S. 308, 315–16 (1980); *Willing v. Mazzocone*, 393 A.2d 1155, 1157 (Pa. 1978).

¹⁴⁸ See *Gertz v. Robert Welch, Inc.*, 418 U.S. 323, 342, 349–50 (1974) (applying standard set forth in *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964)).

¹⁴⁹ Indeed, this may explain why courts applying defamation law appear to stop so far short of crediting privileged harm. In many jurisdictions, a defendant who said false things cannot even reduce damages based on the truthful things he also said. See *Masson v. New Yorker Mag., Inc.*, 960 F.2d 896, 898–99 (9th Cir. 1992) (rejecting the "incremental harm" doctrine);

To be clear, these considerations do not compel one rule for privileged harm or the other. Courts should simply be cognizant that, in some specialized subset of cases, the treatment of privileged harm may determine whether the plaintiff receives any remedy at all.

B. Non-Conduct Elements

The next three complications relate to the counterfactual universe a defendant may conjure to establish privileged harm. As noted above, privileged harm is a kind of one-off credit that reduces damages equally across the spectrum of liability. The existence (and magnitude) of that credit will depend on a story about how the defendant would have behaved in a counterfactual world in which he did not violate the plaintiff's legally protected interests. Once a court has decided that credit for privileged harm is at least theoretically available, it will then confront questions about the nature of this counterfactual inquiry.

The first complication is really a limit, and it can be stated as a bright-line rule: the defendant's counterfactual argument for privileged harm must be rooted in alternative *conduct*. To put it differently, the possibility that the defendant could have imposed the same harm without satisfying non-conduct elements of the claim—such as mental state, conditions, or jurisdictional elements—should not be a basis for reducing damages.

To illustrate the mischief that would arise without this limitation, consider an example from patent law. Direct infringement is a strict liability offense. To establish liability, the patent holder need only prove that the defendant made, used, or sold the patented invention.¹⁵⁰ Other theories of liability, however, require proof of the defendant's mental state. To establish contributory infringement, for example, the patent holder must show that the defendant sold a material component of the patented invention with knowledge that it could be used to infringe that

Moldea v. N.Y. Times Co., 15 F.3d 1137, 1149–50 (D.C. Cir. 1994) (same); *Mandel v. Bos. Phoenix, Inc.*, 456 F.3d 198, 210 n.6 (1st Cir. 2006) (noting that whether Massachusetts recognizes the doctrine remains an open question); *Bustos v. A & E Television Networks*, 646 F.3d 762, 765–66 (10th Cir. 2011) (Gorsuch, J.) (criticizing incremental harm doctrine and noting that it has not been adopted in Colorado). Though these cases do not directly deal with the question of privileged harm, they certainly imply an answer: if a defamation defendant cannot reduce damages based on the true things he *did* say, it would be illogical to allow him to reduce damages based on other true things he hypothetically *could have* said.

¹⁵⁰ See 35 U.S.C. § 271(a) (2018).

particular patent.¹⁵¹ In the absence of such knowledge, there can be no contributory infringement, and the patent holder will fail to establish liability.

Once a patent holder shows that the defendant had the requisite knowledge and otherwise satisfies the elements of contributory liability, that defendant is subject to the same remedies as if he were a direct infringer—including damages for the profits the patent holder would have made but for the defendant's contributory infringement. But note that almost every defendant in this situation would seem to have a killer argument for privileged harm. Counterfactually, the defendant could have made all the same sales without knowledge of the patent, and those sales would have been just as harmful as the ones he made *with* knowledge of the patent. By conjuring a counterfactual involving changed mental state rather than changed conduct, the defendant could take lost profits off the table every time.

This loophole would not be limited to patent law. In any claim involving mental state, credit for privileged harm on that element could eviscerate compensatory damages:

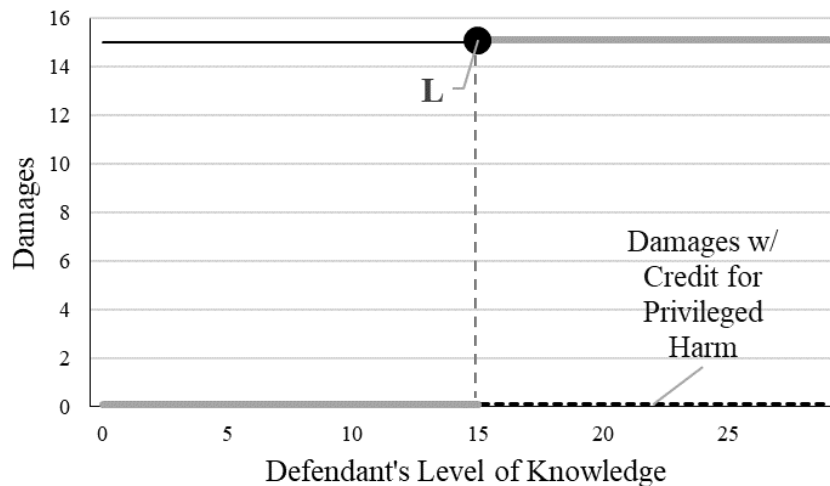


Chart 7

¹⁵¹ See *Global-Tech Appliances, Inc. v. SEB S.A.*, 563 U.S. 754, 765–66 (2011); 35 U.S.C. § 271(c) (2018).

Chart 7 illustrates why counterfactual mental states have the potential to generate such expansive privileged harm. A plaintiff's measurable harm is almost always a function of the defendant's conduct and rarely varies based on the defendant's mental state.¹⁵² Awarding credit for this kind of privileged harm could thus result in an extreme case of the "no remedy" problem discussed in the prior section—one in which the damages curve is completely flat past the threshold of liability. But unlike those special cases, this problem would be pervasive. Without a limit on the counterfactuals a defendant is permitted to conjure, compensatory damages would be effectively unavailable for most claims with a mens rea element, jurisdictional limitation, or non-conduct condition of liability.

The rule that solves this problem is straightforward: credit for privileged harm should be available only on the basis of conduct that the defendant might have counterfactually undertaken in lieu of his liability-triggering conduct. This principle is so deeply intuitive that it risks being taken for granted.¹⁵³ Nonetheless, it is essential that a defendant's argument for counterfactual harm be constrained in this way, lest compensatory damages become a mirage.

C. Unearned Privileges

Having limited arguments for privileged harm to alternative conduct, we must next consider the possibility that privileged harm should be available for some conduct elements but not others. In some cases, it may be desirable to give sanctions-like treatment to some elements of a

¹⁵² There are exceptions to this general observation. For example, an intentional tort may impose emotional or psychological injuries that an accident would not have. In cases like these, the defendant's mental state can exacerbate a plaintiff's injuries. See Guido Calabresi & A. Douglas Melamed, Property Rules, Liability Rules, and Inalienability: One View of the Cathedral, 85 Harv. L. Rev. 1089, 1126–27 n.71 (1972). Where the defendant's mental state is not just a condition of liability but an actual cause of the plaintiff's injuries, allowing the defendant to argue counterfactual mental states will not necessarily eliminate all damages.

¹⁵³ Indeed, I have been unable to find any cases in which a court has directly confronted this question. Consideration in secondary sources is surprisingly sparse as well. On the question of *liability* (not damages), the Restatement (Third) of Torts notes that a plaintiff must only prove but-for causation on conduct elements. Restatement (Third) of Torts: Physical and Emotional Harm § 26 cmt. h. (Am. L. Inst. 2016). Though the Restatement does not cite any authorities or explain the reasoning beyond that comment, the proposition that causal inquiries are typically concerned only with physical conduct appears to be well-established. See H.L.A. Hart & Tony Honoré, Causation in the Law 435–36 (2d ed. 1985).

plaintiff's claim, while nonetheless imposing price-like damages on the claim overall.

To see why, consider a hypothetical. Suppose the City Council enacts a zoning ordinance setting the maximum height of buildings at 50 feet and creates a private right of action for neighbors whose property values are adversely affected by buildings above that limit. Under this regime, a neighbor whose view is blocked by a 55-foot building can sue its builder for the loss of value to her property. Assume this ordinance establishes that ordinary violations are intended to trigger prices rather than sanctions. As a result, a builder who exceeds the height limit should be permitted to reduce damages by the amount of harm that a code-compliant, 50-foot building would have caused. The non-privileged harm is the measure of how much worse off the plaintiff is by adjoining a 55-foot building rather than the hypothetical 50-foot building the law would have permitted.

But now suppose that the City Council wants to encourage the construction of green buildings. As an inducement, it permits a slightly higher height limit—60 feet—for buildings that meet a demanding set of energy-efficiency standards. Neighbors of such buildings are out of luck if the resulting 60-foot building blocks their view or otherwise diminishes the value of their property.

Against this regulatory framework, it would be odd if the builder of a 55-foot, non-green building could use the new green-building provision to reduce or eliminate his damages. While it may be true that the defendant *could have* built a 55-foot green building and owed no damages, we cannot overlook the fact that he *didn't* actually meet the higher efficiency standards necessary to invoke that privilege. Credit for privileged harm on this element would moot the very incentives the City Council sought to create by treating green buildings more leniently. Instead of encouraging green construction, the Council's actions would provide no inducement at all, since a builder could claim the credit in court whether or not he complied with the more demanding standards in concrete.¹⁵⁴

At the same time, it is implausible that the new green-building initiative changed the fundamental nature of the rest of the zoning code. If price-

¹⁵⁴ Cf. Stephen Yelderman, *The Value of Accuracy in the Patent System*, 84 U. Chi. L. Rev. 1217, 1238–41 (2017) (noting that the incentives created by a prize system depend both on the expectation of receiving a prize when it is deserved and the expectation of not receiving a prize it when it is undeserved).

like damages were the appropriate remedy for excess height before, they remain so now. Thus, the builder of a 55-foot, non-green building should still be able to reduce damages based on the harm that a 50-foot building would have done. And, along similar lines, the builder of a 65-foot green building should be able to reduce damages based on the harm that a 60-foot green building would have done.

This example illustrates two generalizable points about privileged harm. First, to give the substantive law its intended effects, it will sometimes be necessary to permit arguments for privileged harm on some elements of a claim but not others. To avoid penalizing the creation of 51-foot buildings, credit for privileged harm must be given on the height element. But to preserve the desired incentives to construct green buildings, credit for privileged harm must be denied on the energy-efficiency element. In practice, this means partially limiting the counterfactual conduct a defendant can invoke, even within a regime that generally permits credit for privileged harm.

Second, this example reveals another reason why it may sometimes be necessary to deny credit for privileged harm within regimes that lack any punitive character. When a law is designed to encourage certain activity, privileged harm will need to be treated the same as if the law were seeking to punish the alternative activity. In the green building example, no one would say the purpose of the City Council amendment is to punish those who construct conventional buildings. And yet, giving effect to the Council's intended incentives structure requires treating non-green buildings in exactly that way.

These principles may explain another exception to the generally price-like damages found in patent law. Ordinarily, patent damages are reduced to account for privileged harm.¹⁵⁵ As a result, a patent defendant can reduce the patentee's lost profits by pointing to non-infringing alternatives that he could have sold instead of the infringing product. But that same defendant is not permitted to assert that, but for the infringement, he would have counterfactually gone out and invented a non-infringing substitute. Rather, the non-infringing alternative must have been actually available at the time of the infringement.¹⁵⁶ This is a

¹⁵⁵ See *supra* notes 57–61 and accompanying text.

¹⁵⁶ See *Grain Processing Corp. v. Am. Maize-Prods. Co.*, 185 F.3d 1341, 1353 (Fed. Cir. 1999) (non-infringing alternative must have been available for sale, not simply a theoretical possibility); *Micro Chem., Inc. v. Lextron, Inc.*, 318 F.3d 1119, 1123 (Fed. Cir. 2003) (finding

surprising rule, given that developing a non-infringing substitute often is an infringer's best option in the counterfactual world where the patented technology is unavailable. But refusing to consider inventive actions not taken makes sense in light of patent law's goal of stimulating subsequent invention.¹⁵⁷ If defendants could invoke theoretical alternative technologies without inventing them, there would be little incentive to incur the costs and risks of actually doing so.

In short, some privileges to harm are not enjoyed as of right, but rather have to be earned. Even if a defendant is entitled to credit for privileged harm in general, he should not be permitted to invoke specific privileges that were created for the purpose of inducing conduct he has not performed.

D. Implausible Counterfactuals

Another complication arises when a defendant asserts counterfactual conduct of dubious plausibility. Indeed, the availability of credit for privileged harm can push defendants to conjure a "just so" kind of counterfactual—one in which the defendant does the maximum allowable harm to the plaintiff without incurring liability under the substantive law. Should it matter if the defendant's argument for privileged harm relies on a counterfactual that seems unlikely, if not wholly implausible?

To illustrate the dilemma, let us return to the example of Polluter and Neighbor. Suppose Polluter has exceeded the legal emissions limit of 15 units by releasing 30 units of the regulated pollutant. To minimize his damages, Polluter will want to advance a counterfactual world in which he emits as much as possible without exceeding the legal limit—14.999 units. If this but-for scenario is used as the counterfactual baseline, Polluter will obtain the largest possible credit for privileged harm, and thus end up with the smallest possible damages figure.

But suppose there is strong evidence that it would be unprofitable for Polluter to operate the factory without emitting at least 20 units. Given this evidence, Neighbor might try to argue that credit for privileged harm is unavailable on the facts of this particular case. Specifically, Neighbor

an alternative product not available where new product required 984 hours to design and another 330 hours to test).

¹⁵⁷ See *State Indus., Inc. v. A.O. Smith Corp.*, 751 F.2d 1226, 1236 (Fed. Cir. 1985) ("One of the benefits of a patent system is its so-called 'negative incentive' to 'design around' a competitor's products . . ."); see also Joseph P. Fishman, *Creating Around Copyright*, 128 *Harv. L. Rev.* 1333, 1351–58 (2015) (collecting cases and scholarship discussing this goal).

will say that the only two courses of conduct available to Polluter were either (a) shutting the factory down (avoiding liability but imposing no harm) or (b) selecting an emissions level at 20 units or above (imposing harm but also triggering liability). Neither of these two possible scenarios would produce any privileged harm.

Situations like these raise a fundamental question about the nature of the counterfactual inquiry. Are we reducing damages to account for the defendant's abstract legal right to do some harm, or because the defendant really would have imposed that harm in the counterfactual universe in which he complied with the law? If it is the former, then treatment of privileged harm is a purely legal question, divorced from empirical predictions about how the defendant would have acted. If it is the latter, then the court must not only determine whether credit for privileged harm is available as a legal matter, but also make factual predictions about how the defendant would have acted in a world in which he respected the plaintiff's rights.

The answer is complicated. The task of measuring credit for privileged harm certainly includes a factual component. At a most basic level, the reason for conjuring a counterfactual universe is to test causation—to determine what the plaintiff's position would have been in the world in which the defendant had not committed the wrong.¹⁵⁸ Though the counterfactual universe is necessarily imaginary, its purpose is to tell us something about reality. At heart, it is a factual inquiry that should be rooted, as much as possible, in evidence from the observed universe.

But though this inquiry is fundamentally factual, it is also artificially limited in scope. As discussed above, the goal of the counterfactual inquiry is only to determine the effect of the defendant's *conduct*. We are not actually interested in whether the defendant would have subjectively chosen to perform the counterfactual actions that he later tries to claim. For that reason, evidence that the defendant was obstinate, or motivated by ill-will, and therefore never would have respected the plaintiff's legal rights, is irrelevant.¹⁵⁹ The counterfactual universe is to be built on facts from the observable universe, but not all facts are invited to join. (Indeed,

¹⁵⁸ See *supra* notes 26–31 and accompanying text.

¹⁵⁹ See Richard B. Wright, *Causation in Tort Law*, 73 *Calif. L. Rev.* 1735, 1806 (1985) (“[Courts] do not want to know, as part of the causal inquiry, whether the defendant would have driven at the legal speed if he had not driven at eighty miles per hour The causal issue is restricted to the narrow question of the actual effect of the excess speed given the other conditions (attentiveness, etc.) that actually existed.”).

if *every* fact came over to our imaginary universe, we would inadvertently recreate the actual world and be no closer to measuring damages than when we started.)

So how, exactly, are we to build this counterfactual universe, if some facts are to be brought over from the observed world, but others are to be replaced by fictional assumptions? The problem turns out to be deeply philosophical and can be explored at length without arriving at any fully satisfying answers.¹⁶⁰ In light of these difficulties, Richard Wright has suggested a practical, two-step process for isolating the effects of a defendant's wrongful conduct. First, we start the hypothetical inquiry by assuming that the defendant would have operated at precisely the legal limit.¹⁶¹ We then imagine how events would play out from there, including the probable reactions of the plaintiff and the defendant.¹⁶² This simulation is constrained, however, by the inviolable rule that in no event may the defendant transgress the plaintiff's rights.¹⁶³

In the emissions example, this approach would start by conjuring a hypothetical world in which Polluter sets his emissions at 14.999. We then predict how each party would react from this starting position. In this example, we know that Polluter's time at emissions level 14.999 would be fleeting, because—as the evidence shows—it would be unprofitable to operate the factory at anything below 20 units of emissions. Enjoined from selecting an emissions level above 15 units, Polluter would simply shut down the factory. Credit for privileged harm would thus be unavailable on the facts of this case.¹⁶⁴

Note that, in reaching this conclusion, we did not inquire about this particular polluter's intentions or mental state. Instead, technical and economic evidence demonstrated that privileged harm would not actually occur in the counterfactual universe in which Polluter could not exceed

¹⁶⁰ See Moore, *supra* note 31, at 387–89.

¹⁶¹ See Wright, *supra* note 159, at 1806.

¹⁶² See *id.* at 1806–07.

¹⁶³ For a careful explanation of this final constraint, see Pratt, *supra* note 32, at 45–49.

¹⁶⁴ Antitrust cases involving unlawful horizontal agreements will often fit this pattern as well. Though a single conspiring firm may have had the ability to impose some of the same harm on the defendant, it will often be unprofitable to do so without the cooperation of co-conspirators. See *Interstate Circuit, Inc. v. United States*, 306 U.S. 208, 222 (1939); 6 *Areeda & Hovenkamp*, *supra* note 74, ¶ 1415c, at 96. The exclusionary conduct thus immediately ceases in the imaginary world in which no conspiracy is possible. This may have been yet another basis for rejecting Justice Frankfurter's argument for privileged harm in *Bigelow*. See *Bigelow v. RKO Radio Pictures, Inc.*, 327 U.S. 251, 253–54 (1946) (describing allegations involving horizontal conspiracy among distributors).

the emissions level. The counterfactual inquiry is objective, not subjective.

This example also reveals why the privileged harm inquiry must, at bottom, be a factual one. Although we took it as a given that it would not be profitable to operate the factory without emitting at least 20 units of the pollutant, that “rule” almost certainly depends on the current state of markets and technology. A shift in demand or a new technological breakthrough could undermine the basic assumption of this exercise—for example by making it profitable to operate a factory with as few as 10 units of emissions. To accurately predict how a person in the defendant’s shoes might have behaved in a counterfactual world of legal compliance, we will often need to know quite a few facts.

An additional complication is that the state of the relevant markets and technology can themselves be affected by the counterfactual universe’s inviolable rule that in no event may the defendant transgress the plaintiff’s rights. Assuming that this imaginary edict applies to everyone, it seems likely that it would affect the prices of the factory’s inputs, outputs, or both, perhaps even making it profitable to operate at lower emissions levels.¹⁶⁵ It could also increase incentives to develop new pollution-mitigation technologies, which could make it possible to operate the factory with no harm to neighbors at all. Hypothetical developments like these could cause the counterfactual universe to deviate even further from observed reality.¹⁶⁶

In many cases, this thought experiment will end only at the point where the court’s indulgence for speculation reaches its limit. Plausible counterfactual claims about market reactions and technological developments may perish before the practical need to support arguments with real-world evidence. By hypothesis, we are calculating damages because, in the real world, the defendant violated the plaintiff’s rights. There is only so much we can do to project what might have happened in

¹⁶⁵ Cf. Herbert Hovenkamp, *Federal Antitrust Policy: The Law of Competition and Its Practice* § 17.4, at 889 (5th ed. 2016) (noting that marginal cost would normally be somewhat higher in a competitive market than in a cartelized market).

¹⁶⁶ Further complicating matters, in many areas of law, counterfactual changes in economic and technological conditions could in turn affect the demands of the law itself. See, e.g., Restatement (Second) of Torts § 298 (Am. L. Inst. 1979) (defining reasonable care for purposes of negligence); *id.* § 826 (defining unreasonable invasions of land for purposes of nuisance); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 452–54 (1984) (observing that copyright fair use defense turns on likelihood of future harm as well as the magnitude of a practice’s social benefits).

the counterfactual world in which the defendant behaved differently. Given that uncertainty naturally looms over this inquiry, the allocation of burden on this point can be outcome-determinative.¹⁶⁷

E. Strategic Decision Making by Litigants

A final complication is that defendants may sometimes make a strategic decision not to seek credit for privileged harm that they formally would be entitled to claim. In some circumstances, arguments for privileged harm might not only fail to persuade the factfinder that damages should be reduced, but also prejudice her against the defendant for other purposes as well. This may explain why privileged harm goes unmentioned in some cases where it seems to constitute a substantial portion of the damages at issue.

To be clear, this is only a theory as to why a particular argument sometimes goes unmade. But one can see why a defendant might fear that invoking privileged harm will do more harm than good. First, the argument has the potential to ring formal and legalistic: “Though I broke the law and did harm, I might have just as easily done some harm without breaking the law.” That statement alone might render the defendant unsympathetic. It also begs the question *why* the defendant broke the law. If the counterfactual lawful path was indeed so imminently available, the choice to take the unlawful path begins to look willful and perhaps deserving of sanction.¹⁶⁸

¹⁶⁷ See, e.g., *Brewer v. Chauvin*, 938 F.2d 860, 864 (8th Cir. 1991) (holding that a “public employer carries the burden of proving that the plaintiff would have been fired even if procedural due process had been observed”); *Thompson v. District of Columbia*, 832 F.3d 339, 346 (D.C. Cir. 2016) (finding that “[o]nce a plaintiff establishes that he was terminated without due process and demonstrates damages arising from that termination, the defendant is responsible for those damages unless the *defendant* shows they would have occurred regardless”); *Patterson v. Coughlin*, 905 F.2d 564, 568 (2d Cir. 1990) (holding that the “burden was on the plaintiff inmate to show that the challenged disciplinary action would not have been taken if he had been allowed to present a defense”); *Rite-Hite Corp. v. Kelley Co.*, 56 F.3d 1538, 1545 (Fed. Cir. 1995) (holding that a patent holder has burden of proving the profits it would have made but-for the infringement); *Grain Processing Corp. v. Am. Maize-Prods. Co.*, 185 F.3d 1341, 1353 (Fed. Cir. 1999) (holding that an infringer has burden of proving that alternatives not on the market were actually available for sale).

¹⁶⁸ Sometimes a defendant will have satisfying answers to this nagging question. Almost every defendant can invoke the strength of his conviction that he is not liable at all. (For example: “I did not redesign my product because of my certainty that the plaintiff’s infringement claims are frivolous, as I urge you to find.”) And some defendants will have particularly sympathetic explanations for why they persisted in an ultimately losing argument rather than changing their conduct to avoid liability. For example, in *WesternGeco LLC v. ION*

Second, in some cases the most powerful evidence of privileged harm will come from non-actionable injuries the defendant has caused in the past. For example, in the Neighbor/Polluter example discussed above, the best proof that Polluter's lawful emissions would do similar harm to Neighbor might be evidence about what Polluter has done in other situations that did not lead to liability. Introducing evidence of this prior uncompensated harm could easily backfire, perhaps even prompting the factfinder to find a way to increase the plaintiff's damages. Given these risks, Polluter might reasonably choose to forgo any argument about privileged harm.

But the risk of prejudicing the factfinder might not be the same in every case involving privileged harm. To the contrary, it seems likely that some arguments for privileged harm may present the defendant in a neutral or even positive light, notwithstanding the concerns raised above. For example, patent infringers appear to be quite comfortable presenting juries with alternative designs they might have offered to consumers, even introducing data revealing the frequency with which their non-infringing products have indeed taken sales from the patent holder.¹⁶⁹ By contrast, it is unfathomable that a modern defendant would make the argument found in *Jenkins v. Pennsylvania Railroad Co.*—in short, “we should owe no damages for negligently destroying the plaintiff's house, because our ordinary operations destroy similar houses all the time.”¹⁷⁰ One can imagine that line of argument not having its intended, damages-mitigating effect.

As a result, the practical availability of the credit for privileged harm may turn on case-specific factors that have little to do with the policy considerations discussed above. For example, the risk a defendant faces by arguing for privileged harm might turn on whether damages will be found by a judge or a jury, whether that factfinder will perceive the defendant's activities to be socially valuable, and on the sources of evidence available to prove the probable effects of counterfactual actions. There is no reason to think these factors will be predictive of, say, the need

Geophysical Corp., 138 S. Ct. 2129 (2018), the defendant's privileged harm option would have involved shutting down a U.S. factory and moving production overseas to avoid patent infringement. See *id.* at 2142 (Gorsuch, J., dissenting). In that case, it's hard to imagine a jury would punish the defendant for protecting its U.S. workforce by persisting in arguments that the plaintiff's patents were invalid and not infringed.

¹⁶⁹ See *supra* note 58 and accompanying text.

¹⁷⁰ *Jenkins v. Pa. R.R. Co.*, 51 A. 704, 705 (N.J. 1902).

for increased marginal deterrence around the threshold of liability, or the need to encourage more lawsuits. From the perspective of whether credit for privileged harm ought to be available, these concerns seem fortuitous at best.

This final claim is admittedly speculative. But these questions are amenable to experimental study. For example, it would be interesting to see how mock jurors respond to arguments for privileged harm, and particularly how their responses change based on the mechanism of the claimed counterfactual harm. If it turns out that jurors are indeed prejudiced by certain arguments for privileged harm, this could have far-reaching consequences for how we conceive of the prices-versus-sanctions question. It may be that, regardless of how we theorize them, compensatory damages regularly act as sanctions for the simple reason that jurors are unsympathetic to arguments for privileged harm in certain kinds of cases.

CONCLUSION

The treatment of privileged harm is rarely given explicit attention. Upon examination, the decision to credit or not credit privileged harm can have a significant effect on prospective defendants' marginal incentives near the threshold of liability. If credit is *not* given, prospective defendants will face a sanctions-like wall of damages at the moment they become liable. In some cases, that may be desirable. But in others it may not be, and failure to consider privileged harm when assessing damages will result in excess deterrence around that threshold.

This initial design choice becomes more complicated when it comes to the details of practical implementation. In some cases, damages for privileged harm may be the only effective remedy on offer to the plaintiff. Reducing damages to account for privileged harm may thus leave the plaintiff with no remedy at all. This effect can potentially appear in almost any kind of case. But there are particular areas of law where it can be expected to occur with some frequency. For cases like these, a sanctions-like rule might be preferable to a rule that leaves defendants completely undeterred and plaintiffs completely uncompensated.

In the other direction, practical obstacles may sometimes prevent defendants from claiming the credit for privileged harm to which they are at least theoretically entitled. Counterfactual developments are often quite difficult to prove, and courts routinely shift the burden to defendants making these kinds of damages-mitigating arguments. Moreover, even if

proof of counterfactual harm is actually available, defendants may sometimes hesitate to introduce that evidence in cases in which it will make them appear unsympathetic. If credit for privileged harm is practically inaccessible to defendants, compensatory damages may end up operating like sanctions even in regimes where they are formally designed to operate as prices.