

DO PATENT HOLDUP AND ROYALTY STACKING LEAD TO SYSTEMATICALLY EXCESSIVE ROYALTIES?

*Einer Elhauge**

ABSTRACT

Some recent literature has concluded that patent remedies result in systematically excessive royalties because of holdup and stacking problems. This article shows that this literature is mistaken. The royalty rates predicted by the holdup models are often (plausibly most of the time) below the true optimal rate. Further, those predicted royalty rates are overstated because of incorrect assumptions about constant demand, one-shot bargaining, and informational symmetry. Although this literature concludes that overcompensation problems are exacerbated by doctrines measuring damages using past negotiated royalties, in fact such doctrines exacerbate undercompensation problems. Undercompensation problems are further increased to the extent that juries cannot measure damages with perfect accuracy, a problem that persists even if damages are just as likely to be overestimated as underestimated. Nor do the royalty rates predicted by the holdup model apply if there is competition in the downstream product market or upstream market for inventions. Royalty stacking does not lead to royalties that exceed the optimal rate, contrary to this literature, but in fact tends to produce royalties that are at or below the optimal rate.

JEL: K00; K10; K11; K20; K21; K29; K30; K39; K40; K41; K49; L40; L49; L50; L51; L59

I. INTRODUCTION

In some recent influential work, Professors Mark Lemley and Carl Shapiro have offered economic models that reach conclusions that—if true—are striking and indicate that our existing patent remedies systematically result in excessive royalties.¹ However, close examination reveals problems in their models that undermine the validity of their conclusions and indicate quite

* Petrie Professor of Law, Harvard Law School, Cambridge, MA, USA. E-mail: elhauge@law.harvard.edu. I am grateful for comments on prior drafts from Yochai Benkler, Julie Cohen, Tom Cotter, Aaron Edlin, John Golden, Andrei Goureev, Jim Krier, Anne Layne-Farrar, Mark Lemley, Clarisa Long, Doug Melamed, Ken Reinker, Ben Roin, Carl Shapiro, Steve Shavell, and George Triantis, and for research support from Qualcomm. The conclusions here are mine, not theirs.

¹ Mark Lemley & Carl Shapiro, *Patent Holdup and Royalty Stacking*, 85 TEX. L. REV. 1991 (2007); Carl Shapiro, *Injunctions, Hold-Up, and Patent Royalties* (August 2006), <http://faculty.haas.berkeley.edu/shapiro/royalties.pdf>.

the opposite: that current patent remedies often (arguably usually) result in royalty rates that are too low to sufficiently reward socially optimal invention.

The Lemley–Shapiro holdup model finds that, if a valuable product design feature arguably violates a patent, a patent holder can systematically extract royalties that exceed the optimal royalty rate by holding up the product maker for a share of the costs and time it would take to redesign the product if use of the patent were enjoined.² They conclude that this holdup problem persists even if the product maker learns about the patent before it actually designs the product.³ Although their holdup model is limited to cases where there is one patent holder and only one product maker, which necessarily implies a downstream product monopoly, they argue that the findings of their model provide a “very good guide” in cases where one of many downstream firms has a dominant market share.⁴ They also offer a model finding that the results are even worse if multiple patent-holders have claims to valuable features of a product design because that would result in “royalty stacking.”⁵ They further argue that these problems created by awarding injunctive relief for patent violations are worsened by doctrines that base damages on past negotiated royalties.⁶

Not surprisingly, Lemley and Shapiro conclude that their findings support changing current law on patent remedies, including staying injunctions during redesign, changing how damages are calculated, perhaps denying injunctions in cases raising holdup problems, and sometimes even changing antitrust law to allow users to collectively negotiate royalty rates with patent holders.⁷ Others have also relied on the Lemley–Shapiro conclusion that patent remedies overcompensate patent holders.⁸ Further, similar arguments by Lemley and Shapiro, joined by fifty other intellectual property professors, in Supreme Court briefing appear to have helped influence the four justice concurrence in the *eBay* case to conclude: “When the patented invention is but a small component of the product the companies seek to produce and the threat of an injunction is employed simply for undue leverage in negotiations, legal

² Lemley & Shapiro, *supra* note 1, at 1991, 2000–02, 2008, 2044; Shapiro, *supra* note 1, at 1, 12–17, 27–31.

³ Lemley & Shapiro, *supra* note 1, at 1991, 2003–05; Shapiro, *supra* note 1, at 21–22.

⁴ Lemley & Shapiro, *supra* note 1, at 2008.

⁵ *Id.* at 2046–48.

⁶ Lemley & Shapiro, *supra* note 1, at 1994, 2021–22; Shapiro, *supra* note 1, at 23–24, 32–33.

⁷ Lemley & Shapiro, *supra* note 1, at 1994, 2035–44; Shapiro, *supra* note 1, at 18–21; Mark Lemley, *Ten Things to Do About Patent Holdup of Standards (And One Not To)*, 48 B.C. L. REV. 149, 153–54, 161–62, 166–67 (2007); Carl Shapiro, *Patent Reform: Aligning Reward and Contribution*, in INNOVATION POLICY AND THE ECONOMY, 112–113 & nn. 5, 8, 120–22, 138–39 (Jaffe, Lerner & Sterns, eds. 2008).

⁸ See Robert Hahn & Hal Singer, *Assessing Bias in Patent Infringement Cases: A Review of International Trade Commission Decisions*, 21 Harv. J.L. Tech. 457, 467 (2008); Mark A. Lemley & Philip J. Weiser, *Should Property or Liability Rules Govern Information?*, 85 TEX. L. REV. 783, 795–800, 834 (2007).

damages may well be sufficient to compensate for the infringement and an injunction may not serve the public interest.”⁹ Many lower courts have relied on this proposition to deny injunctions against patent infringers.¹⁰

The model underlying this influential position thus merits close scrutiny, which this article undertakes. I begin my analysis, in Part I, by showing that the Lemley–Shapiro model understates the true optimal benchmark for royalties, which means that even their own predicted royalties frequently (plausibly most of the time) fall short of optimal royalties. I next establish that their model overstates predicted royalties from the threat of injunctive relief under any of the following more realistic assumptions: (i) that firms negotiate a series of patents when they make a multi-component product, (ii) that firms using the patents have information about their operations that patent holders lack; or (iii) that demand is not constant. I further prove that their model mistakenly concludes that measuring damages using past negotiated royalties increases overcompensation, when in fact it increases undercompensation. I also demonstrate that royalties are even more likely to be undercompensatory to the extent juries are inaccurate in measuring damages, whether their inaccuracies are systematic or balanced, which is another important reason not to shift from injunctive relief to damages. I then show, in Part II, why even if their holdup model were correct when there is both an upstream patent monopoly and a downstream product monopoly, it does not apply when either market level is competitive. Finally, Part III establishes that Lemley and Shapiro are wrong to conclude that royalty stacking produces royalties that exceed the optimal rate; instead, it tends to produce royalties that are at or below the optimal rate.

None of the analysis that follows denies that there are some cases in which extreme holdup problems could result in royalties that exceed the socially optimal rate. But the analysis that follows will disprove the Lemley–Shapiro claim that this result is systematic and will indicate many reasons to think we should have greater concerns that royalties tend to be undercompensatory.

II. PROBLEMS WITH THE LEMLEY–SHAPIRO PATENT HOLDUP MODEL

The Lemley–Shapiro holdup model is beautifully clear and straightforward. It predicts the royalty rate a patent holder and product maker will negotiate under various conditions, and then compares these predicted royalty rates with what they specify as the optimal royalty rate, finding that their predicted royalty rates systematically exceed the optimal. Although I reach different conclusions, I am indebted to the lucidity of their model, which I build on throughout my analysis below.

⁹ *eBay Inc. v. MercExchange, LLC*, 547 U.S. 388, 396–97 (2006) (J. Kennedy, joined by Stevens, Souter & J.J. Breyer, concurring).

¹⁰ See John M. Golden, *Commentary*, “Patent Trolls” and Patent Remedies, 85 *TEX. L. REV.* 2111, 2113, note 12 (2007).

Section A describes the Lemley–Shapiro holdup model. Section B demonstrates a central problem that pervades all their analysis—the benchmark they specify as optimal is in fact suboptimal—and shows that the correct optimal benchmark will often (plausibly most of the time) be above their predicted royalty rates. Section C then proves various reasons why their predicted royalty rates overstate the true royalty rates, which makes it even more likely that actual royalty rates are undercompensatory. Section D shows that calculating damages based on prior negotiated royalties will make royalty rates even more undercompensatory (not as they conclude, more overcompensatory), and that this problem is worsened to the extent juries do not adjudicate patent value with perfect accuracy, even if they are equally likely to overvalue or undervalue a patent. Section E finishes by observing that, even if their analysis were entirely correct, it would not justify their conclusion that patent holders who license should be treated worse than those who compete in downstream markets.

A. The Lemley–Shapiro Holdup Model

Lemley and Shapiro’s basic model addresses negotiations between a single patent holder and a single downstream product maker.¹¹ Lemley and Shapiro begin with the premise that the optimal benchmark royalty is $\theta\beta v$, where θ is the odds the patent will be found valid and infringed, v is the per-unit value of the patented feature (compared with the best non-infringing alternative that could be used in the product design), and β is a number between 0 and 1 reflecting the fraction of the joint gains from trade that the patent holder gets in bargaining with the downstream monopolist.¹² β does not reflect any notion that the patented invention confers only some share of value v , nor does it indicate the relative contributions of the patent holder and downstream monopolist to creating some joint value v . Rather, the value v is assumed to come solely from the patented feature, and β just reflects the downstream monopolist’s bargaining power to get a share of the joint gains of trade (if a license is negotiated) by threatening not to buy a license.

Lemley and Shapiro first analyze surprise patents, those unknown to downstream monopolist D at the time it designed its product to incorporate the patented feature.¹³ They distinguish between “weak” and “strong” patents. For weak patents, θ is so low that, pending patent litigation, D would without a license make more expected profits by selling the product without redesigning it, because the odds are sufficiently low that the patent

¹¹ Shapiro, *supra* note 1, at 25; Lemley & Shapiro, *supra* note 1, at 2005, 2008.

¹² Lemley & Shapiro, *supra* note 1, at 1996–2000; Shapiro, *supra* note 1, at 6–9. Lemley and Shapiro actually use capital V and B , whereas the Shapiro article uses small v and the Greek letter β . Throughout this paper, where there are these differences in notation, I will use the notation used in the Shapiro article, which forms the mathematical basis for the conclusions in the Lemley and Shapiro piece.

¹³ Lemley & Shapiro, *supra* note 1, at 1995; Shapiro, *supra* note 1, at 11.

holder will win and be able to threaten an injunction that forces D to cease operations. For strong patents, θ is high enough that, pending the patent litigation, D would still sell the product without a license, but would find it more profitable to redesign the product before it knows the outcome of the litigation, in order to avoid the risk of being enjoined from operating and losing profits during the lag time before redesign is possible. A patent is weak when $\theta < \theta^*$, and strong when $\theta > \theta^*$, where $\theta^* = F/[\beta((m - v)XL + F)]$, F is the fixed cost of redesign, m is the per-unit profit margin of D , X is the number of units that D sells per unit time (assumed not to vary with whether the patented feature is used), and L is the lag time for redesign as a fraction of the remaining patent lifetime.¹⁴ They assume for most of their analysis that, if the patent holder wins, per-unit damages will equal βv , which is the royalty rate that would be negotiated if the patent were known to be valid and there were no holdup problems.¹⁵

For weak surprise patents, they predict that (if litigation costs and bargaining power are symmetrical), the parties would, given the threat points, negotiate a license with a royalty rate¹⁶ of

$$\theta\beta v + \frac{\theta\beta F}{X} + \theta\beta(m - v)L. \quad (1)$$

The first term in the formula reflects expected damages plus the expected post-trial royalty rate, leaving aside holdup problems. The second term reflects the expected value of the ability of the patent holder, if it wins an injunction that bars D from continued production, to hold up D for a share of the costs of redesign. The third term reflects the expected value of being able to use the same injunction to hold up D for a share of the lost profits during the lag time associated with redesign.

For strong surprise patents (again assuming symmetrical litigation costs and bargaining power), they predict that the negotiated royalty rate¹⁷ will instead be

$$\theta\beta v + \frac{\beta F}{X}. \quad (2)$$

Here the first term once again reflects expected damages and post-trial non-holdup royalties. The second term again reflects the holdup costs of redesign, but this time without any discount for θ because redesign would occur with certainty absent a negotiated license. There is no third term for

¹⁴ Shapiro, *supra* note 1, at 6–7, 13–14.

¹⁵ Lemley & Shapiro, *supra* note 1, at 2000–01; Shapiro, *supra* note 1, at 10, 13–14.

¹⁶ Shapiro, *supra* note 1, at 13–14, 31. *See also* Lemley & Shapiro, *supra* note 1, at 2001 (using the same formula to calculate the percentage overcharge, where C is set to F/vX).

¹⁷ Shapiro, *supra* note 1, at 14, 31. *See also* Lemley & Shapiro, *supra* note 1, at 2002 (using the same formula to calculate the percentage overcharge).

holdup due to lag time costs because they assume that any redesign would take less time than litigation. Thus, lag time costs would never be incurred given that, by assumption, any redesign would be complete by the time any injunction issues.¹⁸

For non-surprise patents, that is patents known to *D* when it is originally designing its product, they reason that this will add a useful option to *D* that leads to pre-design license only if the payoff from originally designing its product without the patented feature exceeds the payoffs from the above options of just using the feature and responding when sued by either selling without redesign or selling with redesign.¹⁹ They further conclude that, because the only threat *D* can make at this stage is not to use the patented feature, the royalty rate for a pre-design license will reflect a split of the value of using the patent feature, without any discount for the probability the patent is not valid or infringed, meaning a royalty rate of²⁰

$$\beta v. \quad (3)$$

For surprise patents, whether weak or strong, given that the first term of both their predicted royalty rates is $\theta\beta v$, which they assume is the optimal rate, and the remaining terms are positive, their predicted royalty rates always exceed their assumed optimal rate.²¹ For non-surprise patents, they predict either (i) the same royalty rates as for surprise patents or, (ii) if a pre-design license is negotiated, a royalty rate of βv , which again necessarily exceeds their presumed optimal rate of $\theta\beta v$ for any patent that is less than 100 percent likely to be held valid and infringed. These results lead them to conclude that patent law “systematically over-rewards” patent holders who license components of a larger product.²²

Finally, they relax the assumption that per-unit damages equal βv . They conclude that if courts follow precedent setting damages based on prior negotiated royalty rates, then in a fulfilled expectations model damages will exceed βv .²³ For strong surprise patents,²⁴ for example, they conclude damages will instead be

$$\beta v + \frac{\beta F}{X(1-T)}. \quad (4)$$

This leads them to conclude that measuring damages by prior negotiated royalties will exacerbate the tendency for royalties to exceed the optimal royalty rate.

¹⁸ Shapiro, *supra* note 1, at 7–8.

¹⁹ Lemley & Shapiro, *supra* note 1, at 2004–05; Shapiro, *supra* note 1, at 21–22.

²⁰ Shapiro, *supra* note 1, at 21–22.

²¹ For $\theta > \theta^{**} = \beta\theta^*$, Shapiro concludes the overcompensation is greater for the weak patents than for the strong patents. Shapiro, *supra* note 1, at 15.

²² Shapiro, *supra* note 1, at 1; Lemley & Shapiro, *supra* note, at 2044.

²³ Lemley & Shapiro, *supra* note 1, at 2021–22; Shapiro, *supra* note 1, at 23–24, 32–33.

²⁴ See Shapiro, *supra* note 1, at 33.

B. The Lemley–Shapiro Benchmark Uses a Suboptimal Royalty Rate

1. The Optimal Benchmark

The Lemley–Shapiro analysis critically depends on their assumption that the optimal benchmark royalty is $\theta\beta v$, which they base on the claim that such a royalty rate “provides an efficient reward to innovators.”²⁵ In fact, their recommended benchmark bears no relation to the reward necessary to efficiently incentivize invention. Indeed, given the premises, any royalty rate below $v\theta$ would underincentivize many socially desirable inventions.

Take the following simple example. Suppose that an invention is 100 percent likely to have a patentable value of \$1 billion. If $\beta = 50$ percent, as they generally posit is plausible, then the innovator gets \$500 million, with the downstream firm getting the other \$500 million in value. Now suppose that the creation of that invention requires a \$750 million investment. Then their benchmark royalty rate would deter the firm from ever creating that invention, and \$250 million in social value would be lost. The same is true for inventions that require an investment anywhere between \$500 million and \$1 billion; all of them would be efficient investments to make, and yet would all be inefficiently deterred by the benchmark royalty rate.

More generally, an investment in invention that costs I would be efficient whenever the expected value it creates exceeds the cost of the investment, that is, whenever $\theta v X > I$. A royalty rate of $\theta\beta v$ would provide a return of only $\theta\beta v X$. It will thus deter efficient investments whenever $\theta v X > I > \theta\beta v X$. In other words, the Lemley–Shapiro benchmark royalty rate will deter an investment that costs less than the value of the investment whenever the investment costs more than β times the expected value of the investment. This is not surprising because using the Lemley–Shapiro benchmark royalty effectively assumes it is instead normatively appropriate for the downstream firm to expropriate part of the value $(1 - \beta)$ of the upstream invention by using its monopsonistic purchasing power. Any such expropriation by a non-inventor will necessarily deter efficient investments to some extent.

Indeed, in other writings, Shapiro indicates that the optimal royalty rate would be $v\theta$.²⁶ Others have suggested the same.²⁷ Nonetheless, in their patent holdup and royalty stacking work, Lemley and Shapiro argue that $\theta\beta v$ is the natural benchmark. Their main argument is that this is the royalty

²⁵ *Id.* at 9. Golden correctly criticizes Lemley and Shapiro’s failure to justify their benchmark. See Golden, *supra* note 10, at 2115–16, 2137–40. But this article goes beyond such critique to explain why their benchmark is too high and what the correct benchmark would be.

²⁶ Joseph Farrell & Carl Shapiro, *How Strong Are Weak Patents?* 4, 8–10 (January 2007), forthcoming in AM. ECON. REV., <http://faculty.haas.berkeley.edu/shapiro/weak.pdf>.

²⁷ In addition to the article co-authored by Professor Farrell in the last footnote, see Joseph Farrell & Robert P. Merges, *Incentives to Challenge and Defend Patents*, 19 BERKELEY TECH. L.J. 943, 958 (2004).

that would result from negotiation without the sort of holdup problems they analyze.²⁸ But it is not at all clear why the natural benchmark should be based on negotiations where a downstream monopoly happens to exist, rather than on more typical cases of a competitive downstream market where (as Part II shows) the natural royalty rate would be θv . Indeed, the only reason that this natural level gets reduced by β in the Lemley–Shapiro model is that in their model the downstream firm exploits its downstream monopoly to “hold up” the patent holder for much of the value of its patent. Thus, if one eliminated holdup on both sides, then θv would be the natural benchmark. Further, Shapiro has himself argued in other works (joined by Professor Farrell) that, when the downstream firm is oligopolistic, then without any holdup problems the royalty will exceed even θv ,²⁹ leaving it unclear why that should not be the “natural” benchmark under the negotiation-mimicking standard posited by Lemley and Shapiro.

Nor can one justify the Lemley–Shapiro benchmark on the ground that it is the benchmark intended by patent law, because patent law also gives the injunctive rights and damage remedies that Lemley and Shapiro argue lead to excessive royalties.³⁰ One needs a standard external to patent law if one is going to critique the royalties produced by patent law. Lemley and Shapiro ultimately defend their benchmark on the ground that they are trying to avoid holdups that create “inefficiency.”³¹ I agree with that standard here, but it means that one must ascertain the optimal benchmark using efficiency standards.

In a subsequent paper, Lemley and Shapiro rely on a separate article by Shapiro to offer efficiency justifications for their benchmark of $\theta\beta v$. This separate article argues that, although allowing the patent holder to fully recover the social value of its invention is the natural starting point, it is not optimal under various conditions, two of which Lemley and Shapiro rely on in their analysis.³²

First, they argue that an individual patent holder should not get the full value of a product when many inventions contribute to a product’s value.³³ But that is not an argument that can apply to their model because it assumes that the patented feature has a positive value that cannot be achieved by *any* non-infringing alternative.³⁴ It is also an argument limited to a subset of

²⁸ See Lemley & Shapiro, *supra* note 1, at 1999–2000; see also Mark A. Lemley & Carl Shapiro, *Reply: Patent Holdup and Royalty Stacking*, 85 TEX. L. REV. 2163, 2164–65 (2007) (hereinafter Lemley & Shapiro, *Reply to Golden*).

²⁹ See Farrell & Shapiro, *supra* note 26. Below I explain some reasons to doubt this conclusion.

³⁰ Nor does it work to focus on what the law provides without injunctive remedies, because the holdup power of the downstream monopolist is enforced by injunctive rights to prevent others from invading its own monopoly facility.

³¹ Lemley & Shapiro, *Reply to Golden*, *supra* note 28, at 2164.

³² *Id.* at 2166–67; Shapiro, *Patent Reform*, *supra* note 7, at 114–26.

³³ Lemley & Shapiro, *Reply to Golden*, *supra* note 28, at 2166–67; Shapiro, *Patent Reform*, *supra* note 7, at 122–25.

³⁴ Shapiro, *supra* note 1, at 7; Lemley & Shapiro, *supra* note 1, at 1996.

patents, and thus cannot justify their sweeping conclusion that patent remedies provide systematic overcompensation. Moreover, to the extent this argument is valid, it provides no justification for the downstream monopolist appropriating part of the patent value. Instead, as Shapiro acknowledges, the literature shows only that, when a combination of patents creates a synergetic value over and above the separate value provided by each patent, that synergetic value should be spread among the patent holders.³⁵ Nothing in this literature provides any reason to think it would be appropriate to discount all the patent holders' returns by β , a measure of a bargaining power in cases when a patent holder happens to face a downstream monopolist.

Second, Lemley and Shapiro argue that a royalty of θv would be overcompensatory in cases where the downstream firm independently discovered the same invention, citing prior literature for the proposition that, when multiple firms compete to be the first innovator, allowing patent holders to recover the full social value of their inventions leads to excessive research and development.³⁶ However, a royalty of θv does *not* award the full social value to the patent holder, for two reasons: (i) The patent holder does not receive any of the social value reaped after the patent term expires. The Lemley–Shapiro model fails to take this into account because it simply ignores the effects of limited patent terms.³⁷ (ii) The patent holder does not receive any of the additional consumer surplus that results from its invention because v is set by the value of the invention to the marginal consumer, and thus does not reflect any higher value that inframarginal consumers place on the invention. The Lemley–Shapiro model excludes this consumer surplus effect by assuming that product output is constant and that all consumers place equal value v on the patented feature. However, this consumer surplus effect is an important factor in real markets where those artificial assumptions do not hold.

Because the prior literature did not exclude these sources of social value that are not reaped by the patent holder, it did not at all conclude that, under current patent law, patent races lead to excessive research and development. To the contrary, it concluded that, because the social value exceeds the value reaped by the inventor, “whether there is too much or too little research is ambiguous.”³⁸ Indeed, it concluded that, even if patent terms

³⁵ Shapiro, *Patent Reform*, *supra* note 7, at 122–25.

³⁶ Lemley & Shapiro, *Reply to Golden*, *supra* note 28, at 2166–67; Shapiro, *Patent Reform*, *supra* note 7, at 115–17.

³⁷ See Golden, *supra* note 10, at 2136.

³⁸ Dasgupta & Stiglitz, *Uncertainty, Industrial Structure, and the Speed of R&D*, 11 BELL J. ECON. 1, 18 (1980); Tandon, *Rivalry and the Excessive Allocation of Resources to Research*, 14 BELL J. ECON. 152, 156–57 (1983) (patent races might result in underinvestment or overinvestment in research, depending on the ratio of patent reward to social value); SUZANNE SCOTCHMER, *INNOVATION AND INCENTIVES* 112, note 7 (MIT Press 2005) (summarizing the economic literature as concluding that: “Patent races may lead to deficient incentives or excessive incentives”); *Id.* at 102 (using the Tandon model to conclude that the “optimal number of participants [in research] can be either larger or smaller than the

were infinite, “there is no clear presumption whether . . . there will be excessive or inadequate research.”³⁹ Instead, the answer turns on the demand elasticity because that determines the consumer surplus and thus the ratio of private profits to social returns under an infinite patent term.⁴⁰ The prior literature also concluded that, even “where, with an infinite-lived patent, there is excessive expenditure on R&D, there is an optimal patent life . . . which will guarantee that the market will undertake the correct amount of research.”⁴¹ Thus, if we assume, as makes sense to isolate the remedial issues at hand, that substantive patent law on issues such as patent length has been optimally set, then this literature supports awarding patent holders the full θv rather than discounting that amount by β .

Further, the patent race literature concludes that “for *small inventions the market always provides inadequate research*.”⁴² Because the Lemley–Shapiro analysis focuses precisely on small inventions that are components of a larger product,⁴³ this conclusion runs directly counter to their analysis. Instead, this literature indicates that patent races for small inventions are likely to mean that a royalty of θv underincentivizes invention, rather than overrewarding it.

Perhaps Lemley and Shapiro have in mind cases where independent invention is very easy, requiring little or no investment, so that it would not be deterred by the prospect of losing royalties.⁴⁴ But Lemley and Shapiro do not limit their analysis or conclusions to cases where easy independent invention is proven. Nor is such easy independent discovery established merely by showing a “surprise” patent because firms might not know that others were racing to make the same invention, might be surprised because they deliberately avoided a patent search that could have exposed them to the risk of treble damages for willful infringement, or might be influenced by prior inventions (that might reach them indirectly) even if they never learned those inventions were patented. Further, Lemley and Shapiro extend their analysis to non-surprise patents, for which the independent invention may simply reflect reverse engineering or inventing around an existing patent.

In any event, to the extent independent invention was extremely easy, that should already be taken into account by θ because, if the invention was sufficiently obvious that anyone easily could have independently discovered it,

equilibrium of the race, . . . , depending on the patent value” and how close it comes to the full social value).

³⁹ Dasgupta & Stiglitz, *supra* note 38, at 21.

⁴⁰ *Id.* at 26.

⁴¹ *Id.* at 21.

⁴² *Id.* at 19 (emphasis in original).

⁴³ See Shapiro, *supra* note 1, at 6 (“We are interested in situations in which m [the profit margin] is relatively large in comparison with v [the value of the patented feature]”).

⁴⁴ Such a view seems suggested in other work by Shapiro. See Farrell, Hayes, Shapiro & Sullivan, *Standard Setting, Patents, and Hold-Up* 17 (December 2007), forthcoming ANTITRUST L.J., <http://faculty.haas.berkeley.edu/shapiro/standards2007.pdf>.

then a patent should be denied for being obvious or lacking novelty.⁴⁵ If those patent doctrines are inadequate, that inadequacy must independently be established, which Lemley and Shapiro do not purport to do since their analysis assumes the soundness of substantive patent law and that it is a normatively “attractive property that the patent holder’s reward is proportional to patent strength, i.e., to the probability [θ] that the patent holder in fact owns a valid right covering an innovation that the downstream firm is using.”⁴⁶

Even if we thought that substantive patent law provided the wrong patent terms or had poor doctrines to deal with independent invention cases, that would not provide any support for the Lemley–Shapiro benchmark. The reason is that β is a measure of bargaining power between a patent holder and downstream monopolist that lacks any relation to the level that optimizes royalties given patent races. Rather, if substantive patent law is inadequate, it would justify reform through a general change to the substantive doctrine to shorten patent terms or deny patent protection in the set of cases where current doctrine is shown to be overinclusive. It would not justify limiting remedies for a class of patent litigants that is not defined to match the problem using a discount factor that bears no relation to the problem. If we instead assume that substantive patent law is correct—in order to focus the analysis on the remedial issue at hand, as Lemley and Shapiro did in their papers—then the proper benchmark is $v\theta$.

2. Lemley and Shapiro’s Predicted Rates Do Not Systematically Exceed the True Optimal Benchmark

Using the proper benchmark of $v\theta$ disproves the Lemley–Shapiro claim that patent law results in royalty rates that are systematically excessive, even if we assumed they were right about the royalty rates that result under current patent law. Recall that Lemley and Shapiro predict a royalty rate for strong surprise patents of $\theta\beta v + \beta F/X$. This predicted royalty exceeds the correct optimal rate of θv only if $F > \theta v X(1 - \beta)/\beta$. Assuming the neutral case where bargaining power is symmetrical, then $\beta = 0.5$, which implies a royalty rate that will exceed the true optimal benchmark only when $F > \theta v X$. In short, applying the correct benchmark to their predicted royalty rates under neutral bargaining indicates a royalty overcharge for a strong patent only when the fixed costs of a redesign exceed the expected value of the patent, taking into account the odds that the patent claim will be found invalid. When the fixed costs do not exceed the expected patent value, then Lemley and Shapiro’s predicted royalty rates will be *undercompensatory*.

Lemley and Shapiro’s predicted royalty rate for weak surprise patents will exceed the correct optimal rate only if $\theta\beta v + \theta\beta F/X + \theta\beta(m - v)L > \theta v$, which is true only when $F + (m - v)LX > vX(1 - \beta)/\beta$. Again assuming the

⁴⁵ See 35 U.S.C. §§ 101, 103(a).

⁴⁶ See Lemley & Shapiro, *supra* note 1, at 1999–2000.

neutral case where $\beta = 0.5$, the royalty rate will exceed the optimal rate only if $F + (m - v)LX > vX$. That is, their predicted royalty rate for a weak surprise patent will exceed the true optimal rate under neutral bargaining only when the value of the lost profits from the lag time to redesign plus the fixed cost of a redesign exceeds the value of the patent without any discount for its possible invalidity. When that condition does not hold, their predicted royalty rate will be *below* the optimal.

Finally, Lemley and Shapiro's predicted royalty rate for non-surprise patents that lead to pre-design licenses will exceed the correct optimal rate only if $\beta v > \theta v$, which is when $\beta > \theta$. Thus, their predicted rate for early negotiated licenses will exceed the true optimal rate only when the share of joint gains a patent holder can get in bargaining exceeds the odds its patent will be held valid and infringed. If we assume the neutral case where $\beta = 0.5$, this will be true only when the odds of patent validity are less than 50 percent. When this is not the case, then Lemley and Shapiro's predicted royalty rate will be *undercompensatory*.

The above analysis proves Theorem 1.

Theorem 1. *Making the same assumptions as the Lemley–Shapiro model, then with symmetrical litigation costs and bargaining power, royalties will be lower than the optimal benchmark of $v\theta$ if*

- (1) $F + (m - v)LX < vX$ for weak surprise patents;
- (2) $F < \theta vX$ for strong surprise patents;
- (3) $\theta > 0.5$, for pre-design licenses.

In short, even if Lemley and Shapiro's predicted royalty rates were accurate, they would be incorrect to conclude that the current system "systematically over-rewards" patent holders who license components in a larger product.⁴⁷ For strong surprise patents, such a conclusion depends on redesign costs being higher than expected patent value. For weak surprise patents, it depends on redesign and lag costs exceeding patent value. For non-surprise patents that lead to pre-design licenses, it depends on the patent holder's bargaining power exceeding the odds of patent validity. None of those requisite conditions can be assumed to hold systematically, and when they do not hold, the predicted royalty rates will be below the optimal rate.

To the extent we can make inferences about empirical likelihoods, they tend to suggest that Lemley and Shapiro's predicted royalty rates are likely to be suboptimal. For example, to reach their conclusions, their model assumes the costs of initially designing the product are precisely the same whether or not the technology is included.⁴⁸ Given this assumption, it

⁴⁷ Shapiro, *supra* note 1, at 1; Lemley & Shapiro, *supra* note 1, at 2044.

⁴⁸ Shapiro, *supra* note 1, at 7.

seems unlikely that the fixed costs of a redesign will be high enough to exceed the expected value of the patent. This assumption also makes it unlikely that those fixed costs plus the lag time lost profits will be higher than the value of the patent without any discount for its likely validity.

It also seems likely that $\theta > 0.5$ in the set of cases analyzed by the Lemley–Shapiro model because their model assumes that the patented feature has a positive value that cannot be achieved by *any* “non-infringing alternative.”⁴⁹ By definition, this means that the technology chosen by *D* must be infringing because it achieves that positive value. The fact that the patented feature describes a technology that has a positive value that no other alternative can achieve also suggests the sort of novelty that makes it likely that the odds that the patent will be held valid should exceed 50 percent. The fact that θ likely exceeds 0.5 in the cases they model is important because the above shows that, when this is the case, the downstream firm can (unless patent search costs are insuperable) always assure it pays a royalty rate that does *not* exceed the true optimal rate. It need simply search the patent records to avoid surprise and then negotiate a license before designing anything.

C. The Lemley–Shapiro Model Overstates Predicted Royalties

The Lemley–Shapiro model overstates predicted royalties for various reasons. First, their model overstates predicted royalties by assuming a one-shot game; if one assumes, instead, a repeated game with multiple patent-holders, predicted royalty rates are lower. Second, even in a one-shot game, royalties will be lower than Lemley and Shapiro’s predicted royalty rates if there is asymmetric information or if fairness norms permit commitments to refuse to pay excessive royalties. Third, even with the above problems, their assumption of inelastic output is unrealistic and inflates predicted royalties.

1. *The Lemley–Shapiro Model Overstates Predicted Royalties by Assuming a One-Shot Game*

The royalty rates predicted by the Lemley–Shapiro model depend on their assumption of a single-shot game. If the negotiations over patent royalties are repeated between a downstream firm and patent holders, then the Lemley–Shapiro rates set only an upper bound on the range of royalty rates. Indeed, in a repeated game, *D* should be able to pay patent holders a rate that is fairly low, as long as the profits foregone from refusing to deal with any patent holder are offset by the profits gained with other patent holders from adhering to a strategy of paying the low rate.

⁴⁹ *Id.* at 6. The odds of patent validity under these assumptions cannot be inferred from the empirical odds at which current patent claims are held valid and infringed because current patents involve many patents for which this strong assumption is untrue. Nor can one simply drop those assumptions because they are integral to establishing the patent holder’s bargaining power under their model.

Assume that we have a repeated game in which downstream monopolist D faces a new patent holder P_i every period with no determinate last patent holder. D adopts the following strategy. It offers the lowest royalty rate r^* that satisfies the condition that the profits D could make from using P_i 's patent at the Lemley–Shapiro rate are lower than the additional profits D would make from paying all subsequent P_i , a rate of r^* rather than the Lemley–Shapiro rate.⁵⁰ D 's strategy is to adhere to r^* unless it has paid more than r^* in the past, in which case it will offer the single-shot royalty rates predicted by Lemley and Shapiro. A strategy of paying no more than r^* to each patent holder is credible because, if any patent holder rejects r^* , D will make more profits by refusing to use the patented feature than by using it at the Lemley–Shapiro rate.⁵¹ In contrast, no patent holder can credibly commit to refusing to take r^* because each plays only one period, and taking r^* exceeds zero, the payoff each would otherwise earn. Thus, paying r^* is a subgame-perfect Nash equilibrium. This proves Theorem 2.

Theorem 2. *If one alters the Lemley–Shapiro model to assume a repeated game in which the downstream firm D faces multiple patent-holders P_i , then there is a subgame-perfect Nash equilibrium in which D pays a royalty rate of r^* that is lower than the Lemley–Shapiro rates for each class of patent. This lower royalty rate of r^* equals the lowest royalty rate for which it is true that the profits D could make from using the individual P_i 's patent at the Lemley–Shapiro rate are exceeded by the additional profits D would make from paying all subsequent P_i a rate of r^* rather than the Lemley–Shapiro rate.*

Thus, in a repeated game with multiple patent-holders, the royalty rate is likely to be significantly lower than the royalty rates Lemley and Shapiro predict. Another way to think of this is that, when the downstream firm faces multiple patent-holders, bargaining power is likely to favor the downstream firm rather than be symmetrical, so that β can be well below 0.5, which increases the likelihood of undercompensatory royalties.

There are several reasons why actual royalty rates are more likely to reflect a repeated game with multiple patent-holders than a single-shot game. To begin with, the entire premise of the Lemley and Shapiro analysis is that producers face not just one patent holder, but multiple patents on various components of a single product. If one is going to consider this fact for the purpose of concluding that having a patent on one component of a larger product may create holdup and royalty stacking problems, then one should

⁵⁰ Profits would have the appropriate discounts to the extent the first set of profits comes in earlier. But even if there are serial negotiations, the profits may all come in the same time when the product is launched, in which case such discounting is unnecessary.

⁵¹ Avoiding the patented feature will require either designing the product not to include the patented feature or (in the case of a surprise patent) ceasing to sell the product pending redesign.

not limit the bargaining game to one patent holder, but rather should consider the whether the bargaining game is changed when the downstream firm faces a series of patent holders. Given such a series of patent holders, the producer will realize that failing to adhere to the above strategy will subject it to higher holdup costs for all the other patents on components of the product. Adhering to the strategy will accordingly be profit-maximizing for D , and thus quite credible to each patent holder.

Moreover, even if D only had a single product with one patented feature, such a product will typically have multiple generations, with no clear endpoint. D will know that if it fails to adhere to a strategy of sticking to low royalty rate r^* , then D will face similar holdup problems on subsequent generations, and thus D will adopt the strategy that maximizes its profits over all the generations. Further, the typical producer D will have multiple products and will know that failing to adhere to the strategy on this product will lower its expected profits on others.

To be sure, it might instead be the case that a single patent holder faces multiple downstream firms. But then the Lemley–Shapiro model is inapplicable for different reasons, as I discuss in Part II.

2. *Predicted Royalty Rates Will Be Lower With Either Asymmetric Information or Norm-Based Commitments to Reject Excessive Royalties*

Even in a single-shot game, a threat of lower royalty payments than Lemley and Shapiro's predicted royalty rates is likely to be believable to the patent holder because of asymmetric information and norm-based commitments.

Asymmetric Information. To reach their conclusions, Lemley and Shapiro assume that information about all the variables is common to both sides. In reality, the patent holder often will not know what course of action is most profitable for D because it is uncertain about the relevant variables. Lemley and Shapiro argue that informational uncertainty could cut in both directions if the uncertainty were equal on both sides.⁵² However, the information known to the patent holder that bears on the likelihood of patent validity will generally be publicly available,⁵³ because patent law requires public disclosures about the patents. In contrast, D will likely have lots of private information on its profit margins, its output, the value of the patent to it, the likelihood that it is infringing the patent, and the costs and lag time necessary for redesign,⁵⁴ which will be confidential information that it is hard for the patent holder to obtain and may even be protected by trade secrets. In short, the patent holder and D will have largely equal information on validity factors that bear on θ , whereas D will have an informational

⁵² See Lemley & Shapiro, *Reply to Golden*, *supra* note 28, at 2170.

⁵³ See Golden, *supra* note 10, at 2133.

⁵⁴ *Id.* at 2132.

advantage not only on infringement factors that bear on θ , but also on all the other variables in the equations.⁵⁵

The result is that D will benefit from an informational asymmetry, so that assertions by D that it would be profitable for it to avoid the patented feature will often seem credible to the patent holder even though D knows they are untrue. This will reduce predicted royalties below the rates predicted by Lemley and Shapiro. Lemley and Shapiro argue otherwise because they say that imperfect information is as likely to “make the patentee demand more money as less.”⁵⁶ But the problem for their position is that D will accept when the patent holder demands too little (because the patent holder underestimates how much it could get), but won’t accept when the patent holder demands too much (because D knows it would be paying too much). Thus, the actual negotiated royalties will be lower than they predict.

Norm-Based Commitments. Although economic models generally assume that threats cannot be credible if they would be unprofitable to carry out, this assumption often turns out to be untrue. Empirical studies of the ultimatum game, for example, have shown that, if an offeror does not offer a fair split of a joint surplus, offerees will often reject the split even though rejection means they get nothing.⁵⁷ This violates the conditions for subgame perfection, because the offeree is carrying out a threat to take unprofitable action, but in fact people do it frequently. Further, these studies indicate that the threat of such an unprofitable rejection induces offerors to make fairer initial offers, and that when offerees adopt an explicit strategy in advance, they are more likely to reject unfair splits. Perhaps the reason has to do with biological hard-wiring explicable by evolutionary mechanisms,⁵⁸ but whatever the reason, this suggests a threat to carry out unprofitable action, rather than cave to an unfair demand, will often be credible.

Fairness norms should indicate it is unfair to pay anything to the patent holder above the expected value of its patent. Thus, a threat to refuse to pay more than the optimal benchmark of $v\theta$ might well be credible even without a repeated game. This, again, will reduce predicted royalties below the rates suggested by Lemley and Shapiro. If parties believed that $\theta\beta v$ was the fair

⁵⁵ This is true even for litigation costs, because although both sides can equally estimate the costs of litigating patent validity, the user can better estimate the costs of litigating its infringement.

⁵⁶ Lemley & Shapiro, *Reply to Golden*, *supra* note 28, at 2170.

⁵⁷ JOSEPH HENRICH, ROBERT BOYD, SAMUEL BOWLES & COLIN CAMERER, *FOUNDATIONS OF HUMAN SOCIALITY: ECONOMIC EXPERIMENTS AND ETHNOGRAPHIC EVIDENCE FROM FIFTEEN SMALL-SCALE SOCIETIES* (Oxford University Press 2004); Hessel Oosterbeek, Randolph Sloof & Gijs van de Kuilen, *Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis*, 7 *EXPERIMENTAL ECONOMICS* 171 (2004).

⁵⁸ See Bjorn Wallace, *et al.*, *Heritability of Ultimatum Game Responder Behavior*, Proceedings of the National Academy of Sciences, vol. 104, no. 40, pp. 15631–34 (2 October 2007) (identical twin study showing that genes explain 40 percent of the variation of ultimatum game responder behavior).

benchmark, as Lemley and Shapiro argue, then they are likely to refuse royalties above that, making royalties even more undercompensatory. Norm-based commitments may also induce the patent holder to refuse to accept an r^* that is too low in the repeated game model described above, but the point here is simply that such norm-based commitments provide another reason to explain why it is unlikely that royalties will exceed expected patent value.

3. *The Lemley–Shapiro Model Overstates Royalties by Assuming Constant Output*

Even if we assumed a one-shot case with symmetric information and no fairness-based commitments, the Lemley–Shapiro model would overstate royalties because it assumes the downstream output X is constant and totally unaffected by whether D incorporates a patented feature that increases product value.⁵⁹

This assumption is highly dubious. It seems inconsistent with the assumption that having a patented feature with value v allows the downstream firm to raise prices by v ; if demand were constant, D could have raised prices by v without the patented feature. The notion that downstream demand is totally inelastic is also wrong because, even under Lemley and Shapiro's assumption that the downstream firm is a monopolist, it would not price on the inelastic portion of its demand curve.⁶⁰ More generally, it is unclear why firms would want to pay for patents that increased product value unless those value increases increased demand for the product. This assumption thus seems inconsistent with the very existence of a license for such a patent.

The dubious nature of this assumption is important because assuming constant demand inflates predicted royalties, given that it overstates the profits from licensing the patent. It does so because having the patented feature increases prices, which reduces demand, and increases marginal costs, which reduces the downstream's firm's willingness to produce. Both effects will reduce output with the patent, which in turn reduces total profits with the patent, the joint gains from trade, and thus royalty rates.

For example, assume that the downstream monopolist faces the linear demand function that Lemley and Shapiro assume in their royalty stacking analysis, $Q = A + v - P$, where Q is the quantity, A is a constant, and P is the downstream product price.⁶¹ Suppose further that royalties are paid on a lump sum basis. Then the appendix proves Theorem 3, showing that royalties will be lower than Lemley and Shapiro's predicted royalty rates in their holdup model.

⁵⁹ Shapiro, *supra* note 1, at 6–7.

⁶⁰ See EINER ELHAUGE, UNITED STATES ANTITRUST LAW AND ECONOMICS 189 (Foundation Press 2008) (explaining why “even an absolute monopolist would never set a price that leaves it on the inelastic portion of its demand curve”).

⁶¹ Lemley & Shapiro, *supra* note 1, at 2046.

Theorem 3. *If one alters the Lemley–Shapiro model to assume linear rather than constant demand, but otherwise uses all the same assumptions, then assuming lump-sum royalties are used, royalties will be lower than Lemley and Shapiro predict by*

- (1) $0.25\theta\beta v^2 + \theta\beta(0.5vX - 0.25v^2)L$ for weak surprise patents;
- (2) $0.25\theta\beta v^2$ for strong surprise patents;
- (3) $0.25\beta v^2$ for pre-design licenses.

If one instead assumes royalties are paid per unit, then Theorem 4 (also proven in the appendix) again proves that the amount of royalties will be lower than Lemley and Shapiro’s predicted royalty rates.

Theorem 4. *If one alters the Lemley–Shapiro model to assume linear rather than constant demand, but otherwise uses all the same assumptions, then assuming per-unit royalties are used, royalty rates will be lower than Lemley and Shapiro predict by*

- (1) $0.25\theta\beta(v-r)^2 + \theta\beta[0.5(v-r)X - 0.25(v-r)^2]L$ for weak surprise patents;
- (2) $0.25\theta\beta(v-r)^2$ for strong surprise patents;
- (3) $0.25\beta(v-r)^2$ for pre-design licenses.

These formulas might seem to suggest that the total amount of royalties paid is larger if per-unit royalties are used than if lump-sum royalties are used because the overstatement of royalties looks smaller in the former case, but as the appendix shows that is an artifact of the fact that X is higher with a lump-sum royalty than with a per-unit royalty.

Whether lump-sum or per-unit royalties are used, all of Lemley and Shapiro’s predicted royalty rates would be lower if we altered their bargaining model to relax the unrealistic assumption of constant demand. This makes it even less likely, even in a one-shot game, that accurately predicted royalties will be overcompensating when compared with the proper benchmark. Further, when the downstream monopolist faces multiple patent-holders, these lower royalty rates will lower the upper bounds on royalty rates under Theorem 2.

D. Damage Measures That Use Prior Negotiated Royalties or Reflect Imperfect Assessments of Patent Value Will Make Predicted Royalties More Undercompensatory

1. Basing Damages on Prior Negotiated Royalties Makes Them More Undercompensatory

Lemley and Shapiro note that courts often set damages to equal prior negotiated royalty rates.⁶² They conclude this will tend to make damages

⁶² Lemley & Shapiro, *supra* note 1, at 2021–22; Shapiro, *supra* note 1, at 23–24, 32–33.

overcompensatory because damages will be based on prior negotiated rates that were themselves excessive, which will exacerbate the overcompensation problem.

This analysis is mistaken on several levels. First, there can be no prior negotiated royalty rates unless there are downstream rivals to have negotiated them, and the whole Lemley–Shapiro model assumes only one downstream firm. As noted below, their model does not apply if instead the downstream market is competitive.

Second, this conclusion misses the point that, under Lemley and Shapiro’s own analysis, the prior negotiated rate will be βv for pre-design licenses or even less for post-design licenses. Thus, even if we assume the complete accuracy of the Lemley–Shapiro model, putting aside the other reasons noted above to think it overstates predicted royalties, the damages awarded by juries under this formula will on average be less than βvX even after an adjudication has been made that determines that the patent was valid and infringed.⁶³ And of course no damages will be awarded if the adjudication determines otherwise. Because the former adjudications by definition occur θ percent of the time, this means that expected damages will be less than $\theta\beta vX$ even under the Lemley–Shapiro model.

This further undermines the negotiating ability of the patent owner and makes it even more likely that royalty rates will be undercompensating. As noted above, the true optimal total royalty would be θvX . The methodology used by courts means that they are not only discounting this by β , but awarding even less than that whenever post-design licenses are used. This greatly discourages the investments that might be necessary to create the valuable patent.

To illustrate, consider the following concrete example. Suppose we have a strong patent that did not lead to a pre-design license, that $\beta = 0.5$, $\theta = 0.4$, the total value (vX) of the patent (if valid) equals \$1 billion, and $F = \$200$ million. Because there is no prior negotiated license to refer to, the damage formula expected will be the royalty that would have been negotiated if the patent were known to be valid and infringed.⁶⁴ Assuming the Lemley–Shapiro model is accurate, this expected damage formula will be βvX . Thus, the strong patent under their model formula (2) will be licensed for a total of $\theta\beta vX + \beta F$, which here totals \$300 million. This means that, if a

⁶³ If we take into account the other reasons to think that the Lemley–Shapiro model overstates predicted royalties, like the fact that the predicted royalties are lower if D engages in repeat bargaining with patent holders, if there is informational asymmetry favoring D , if norms allow credible commitments not to pay excessive royalties, or if demand is not constant, then basing damages on past negotiated royalties would be even more undercompensatory.

⁶⁴ Lemley & Shapiro, *supra* note 1, at 2000, 2021; Shapiro, *supra* note 1, at 23.

court awarded the negotiated royalty rate as damages even after litigation established that the patent was in fact valid, a court would be awarding only 30 percent of the value of the patent.

Indeed, the problem is even worse, because after the first negotiation, the second patent negotiation would settle for $(\theta\beta vX + \beta F)\theta T + \theta\beta vX(1 - T) + \beta F$. The first term for expected damages will reflect the prior negotiated royalty because that is assumed to affect the measure of damages, the second term reflects the expected post-trial royalty, and the third term reflects a share of the joint gains of avoiding the fixed costs of redesign. This second negotiation will offer even less compensation than the first one whenever $1 - F/vX > \theta$. Absent surprise, that inequality holds whenever a strong patent leads to a post-design license, because under the Lemley–Shapiro model the firm would negotiate a post-design license rather than a pre-design license whenever $\beta v > \beta v\theta + \beta F/X$, which is true whenever $1 - F/vX > \theta$. Thus, absent surprise, whenever there is a post-design license on a strong patent, a rule that bases damages on prior negotiated royalties will mean that the second patent negotiation will settle for a rate lower than the first one.

To illustrate, assume again the facts of the above concrete example, adding the assumption that $T = 0.2$, which is realistic because patent litigation generally lasts no more than 20 percent of the life of a patent. Given the first patent license, the second patent negotiation should settle for $d\theta T + \beta v\theta X(1 - T) + \beta F$, where d is the damages the parties expect the court to award if the patent is held valid. Here, that equals \$300 million given the first patent license and a rule that bases damages on prior negotiated royalties. Therefore, the parties will negotiate a royalty rate of $0.08(\$300 \text{ million}) + \$160 \text{ million} + 0.5(\$200 \text{ million}) = \$284 \text{ million}$. And so on, with each round getting a bit smaller until they converge on the equilibrium, which should be reached when $d = (0.08)d + \$260 \text{ million}$, or $d = \$282.6 \text{ million}$. Further, the parties should realize this before they enter into the first negotiation. No firm will thus want to make the first deal at \$300 million because it knows its rivals will get better deals in subsequent negotiations. Thus, even the first negotiating user will insist on paying no more than \$282.6 million, as will all other users. This is well below the optimal royalty of \$400 million.

Lemley and Shapiro conclude otherwise using a self-fulfilling equilibrium.⁶⁵ Multiplying their court-ordered royalty rate by X to get the total damages, they reason that, if $\theta = 1$, then the expected negotiated royalty rate would be $r(1) = sT + \beta v(1 - T) + \beta F/X$, where s is the royalty per unit the court will award in damages. And since $r(1) = s$ in a self-fulfilling

⁶⁵ Shapiro, *supra* note 1, at 23–24, 32–33; Lemley & Shapiro, *supra* note 1, at 2022, note 79.

equilibrium, they conclude that for strong patents.⁶⁶

$$s = \beta v + \frac{\beta F}{X(1 - T)}. \quad (4)$$

But this analysis misses Lemley and Shapiro's own point that, in a pre-design negotiation, D will pay no more than βv .⁶⁷ Indeed, as I noted above, D will pay a royalty of βv whenever $\theta > 1 - F/vX$, and by definition if $\theta = 1$, it must be greater than $1 - F/vX$, unless $F = 0$, in which case the rate is βv under either formula. No one would thus negotiate for such a royalty unless they had not done the search necessary to discover the patent, and thus did not know to negotiate for a pre-design royalty. And given that, by hypothesis in Lemley and Shapiro's analysis, $\theta = 1$, it seems clear that anyone who did search for the patent would have been likely to discover it.

Rather than just assuming $\theta = 1$, we can describe the more general formula for the self-fulfilling equilibrium under the Lemley–Shapiro model as follows:

$$r(\theta) = \theta s T + \theta \beta v (1 - T) + \frac{\beta F}{X}.$$

Because $r(\theta) = s$, this means

$$s = \theta s T + \theta \beta v (1 - T) + \frac{\beta F}{X}.$$

Solving for s and multiplying by X , one gets the predicted damage amount:

$$d = sX = \frac{\theta \beta v X (1 - T) + \beta F}{1 - \theta T}.$$

One can confirm this theorem by plugging in the numbers from the above concrete example, where $\beta = 0.5$, $\theta = 0.4$, $vX = \$1$ billion, $F = \$200$ million, and $T = 0.2$, which gives us $d = \$282.6$ million. And one can complete the circle by noting that, if the parties expect $\$282.6$ million in damages, then the negotiated total royalty under the above-noted formula, $rX = \theta s X T + \theta \beta v X (1 - T) + \beta F$, comes to $\$282.6$ million, which is the same undercompensatory royalty I found above.

Lemley and Shapiro's self-fulfilling equilibrium for damages from a strong patent will exceed the correct self-fulfilling equilibrium whenever $\beta v X + \beta F / (1 - T) > [\theta \beta v X (1 - T) + \beta F] / (1 - \theta T)$, which can be

⁶⁶ Shapiro, *supra* note 1, at 33.

⁶⁷ *Id.* at 22; Lemley & Shapiro, *supra* note 1, at 2004–05.

rearranged as being true whenever $vX/F > T(\theta - 1)/[1 - T - \theta + \theta T]$. The right hand side of this inequality will be negative whenever $1 - T - \theta + \theta T > 0$, which can be rearranged as being true whenever $1 > T$, which by assumption is always true. Thus, Lemley and Shapiro's damage formula always overstates equilibrium damages from a strong patent, even if one otherwise assumes the correctness of their model.

The accurate damages equilibrium for a strong patent under the Lemley–Shapiro model will be lower than the βvX in damages that Lemley and Shapiro use to derive all their predicted royalties whenever $[\theta\beta vX(1 - T) + \beta F]/(1 - \theta T) < \beta vX$. This can be rearranged as being true whenever $\theta < 1 - F/vX$. As noted above, absent surprise, that inequality will be true whenever a firm chooses to negotiate a post-design license rather than a pre-design license for a strong patent. Therefore, unless prior negotiated royalties resulted from surprise, using the accurate damages equilibrium will always reduce the royalties one would predict for strong patents under the Lemley–Shapiro model, thus providing another reason why their analysis overstates predicted royalties.

The above thus proves Theorem 5.

Theorem 5. *If courts measure damages on the basis of prior negotiated royalty rates, and the Lemley–Shapiro model is correct, then in a self-fulfilling equilibrium, damages for a strong patent will equal*

$$\frac{\theta\beta vX(1 - T) + \beta F}{1 - \theta T}. \quad (4)$$

This will always be lower both than (i) the damages predicted by the Lemley–Shapiro model in such cases and (ii) the βvX they assumed for damages when predicting royalties if $\theta < 1 - F/vX$, which will always be true unless prior negotiated royalties resulted from surprise.

2. *Imperfect Adjudications of Patent Value Makes Negotiated Royalties More Undercompensatory*

The Lemley–Shapiro model assumes that each technology has some clear value v that courts and juries always ascertain with perfect accuracy. This is a highly dubious assumption. When many ideas must be combined to create something valuable, there is—even with perfect knowledge—no clear theoretical way to allocate value across those ideas other than firms' willingness to pay for those ideas. And the juries that award patent damages have far from perfect knowledge and perfect accuracy.

If courts instead systematically underestimate damages, then all the predicted royalty rates will be lower. Here, such systematic underestimation seems plausible given Lemley and Shapiro's own analysis, for it shows that even highly sophisticated scholars on patent law and economics can favor a

measure of damages (βvX) that systematically would underincentivize invention that has value v . Further, as noted above, courts often set damages equal to prior negotiated rates, and Theorem 5 proved that this method would produce damages even lower than βvX , which would thus be even more undercompensatory.

Even if courts were actually instructing juries to award damages equal to vX , other factors may cause juries to systematically underestimate damages. Hindsight bias may cause juries to underestimate v because patented inventions often seem more obvious after they have been created. Underestimations may also exceed overestimations if patent owners have information about the value of their patents that they either cannot communicate to courts or must bear large costs to communicate to courts.⁶⁸ This is likely often true with patents given their often complicated, highly technological nature, which mean that experts such as patent holders may have superior information about patent value that they cannot adequately convey to an inexpert court or jury. Further, underestimation of damages may be more likely for classes of damages that are particularly difficult to estimate because courts may be unwilling to bear the administrative costs of making such estimations.⁶⁹ Given the technological complexity of many patents and the difficulty in predicting the value and potential uses of the patent, courts in patent cases may bear large administrative burdens in estimating damages and thus may choose to ignore some aspects of damages to make the calculation more manageable. For example, courts and juries may choose to base patent damages on royalties negotiated for prior, less valuable, inventions that will undervalue the newer patent, rather than incur the difficulties of trying to assess the value of the patent without guidance from market forces. All these factors suggest it is likely that juries will systematically underestimate damages.

Now suppose that courts correctly try to award vX in damages, and juries have no systematic bias, but rather under- and overestimate vX with equal likelihood. Even so, such symmetric errors have an asymmetric effect because the option of bargaining sets an effective cap at vX , the actual value of the patent. To see why, suppose we have some set of patented inventions, each of which has some value v_i that is known to the relevant firms, and that juries set damages for that set of inventions at v^* , the average of all the v_i . The jury-set damages will not be systematically high or low, but their effect on predicted royalties will be. To see why, take the simple case where there is no redesign fixed costs or lag time and the validity of the patent is clear

⁶⁸ See Henry Smith, *Property and Property Rules*, 79 N.Y.U. L. REV. 1719, 1729 (2004) (arguing that property rules are superior to liability rules because they reduce opportunism that results when owners cannot communicate to others the true value of the property).

⁶⁹ See Louis Kaplow & Steven Shavell, *Property Rules v. Liability Rules: An Economic Analysis*, 109 HARV. L. REV. 713, 731 (1996); Richard Epstein, *A Clear View of the Cathedral: The Dominance of Property Rules*, 106 YALE L.J. 2091, 2093 (1997).

(F and L both equal zero, and θ equals 1). When $v^* < v_i$, then, without a license, the downstream firm would just infringe and pay damages of v^*X , which will be undercompensatory because those damages are lower than v_iX . Thus, when it negotiates a license, the downstream firm would pay a royalty no more than average value v^* whenever the actual value exceeds the average value. When $v^* > v_i$, then the downstream firm will not use the invention without a prior license where it pays a royalty no higher than v_i . Thus, it will pay no more than actual value v_i for a license when the average value exceeds the actual value. The result is that even balanced errors in ascertaining damages will mean that the average royalty rate will be below the average value v^* , thus making damages more undercompensatory.⁷⁰ The analysis becomes more complicated when we take into account redesign costs, lag time, and probabilistic patents, but the above effect still creates a general tendency to make damages undercompensatory, which reduces predicted royalty rates in such cases.

Balanced errors might also have an asymmetric effect if the error directions tend to vary for different industries or doctrinal areas.⁷¹ Suppose, for example, juries tend to underestimate v in telecommunications, but overestimate v in software, and thus overall are equally likely to under- and overestimate. The effects of this sort of seemingly balanced errors will be asymmetric. In telecommunications, the underestimation means that (in the simple case) firms will infringe the patent and pay damages rather than pay anything above the underestimation of v , so that the innovator will receive less than v and thus have underincentives to innovate. In software, the overestimation will just mean that firms will license the patent and pay v . The software innovator will have proper incentives to innovate but this will not compensate for the underincentive in telecommunications. In short, the problem is that underestimations of damages lead non-inventors to appropriate patents for less than their value, thus discouraging investments in invention. Overestimations of damages, in contrast, effectively just return us to a property rule that induces a bargain where the patent holder can demand

⁷⁰ Kaplow and Shavell analyze a similar problem in the property versus liability rules discussion and reach the contrary conclusion that it will lead to inefficient takings only if the property owner has some idiosyncratic value that generally makes takings undesirable. See Kaplow & Shavell, *supra* note 69, at 759–63. Without such an idiosyncratic valuation, in their model any taking at underestimated damages would not be undesirable because the taker would gain the same value that the owner would lose. In my model, what makes takings at underestimated damages undesirable is not the idiosyncratic value the owner puts on its patent, but rather the resulting reduction in *ex ante* incentives to create the invention at all, which Kaplow and Shavell do not model. See generally Lucian Bebchuk, *Property Rights and Liability Rules: The Ex Ante View of the Cathedral*, 100 MICH. L. REV. 601 (2001) (observing that the property versus liability rules literature has only considered *ex post* effects, and thus wrongly failed to recognize that *ex post* distributive effects can reduce efficiency by altering *ex ante* incentives).

⁷¹ The unbalanced effect of balanced errors that vary for different industries or doctrinal areas does not appear to have been noticed in prior articles on property versus liability rules.

the full value of its patent, but users will not pay more than that value for it. Similar problems might arise if the error directions vary for different doctrinal areas in ways that are predictable by firms.

To illustrate, consider the following concrete example. Potential inventions *A* and *B* in, respectively, the telecommunications and software industries, each has a v of \$10. In the case of either invention *A* or *B*, it would be socially optimal for the inventing firm to invest up to \$10 per unit to create the invention. Suppose that, juries award damages that equal 100 percent of average patent value over all cases, but on average award 90 percent of patent value in telecommunications cases and 110 percent of it in software cases. For Patent *A*, expected damages would be \$9 per unit, thus in the simple case, firms would infringe rather than pay any royalty over \$9, which would underincentivize investment to create invention *A*. For Patent *B*, expected damages would be \$11 per unit, but no firm would be willing to pay more than the value of the patent to be able to use it, and thus the negotiated royalty will be no more than the \$10 per unit that equals the value of the patent.

Now suppose (unrealistically) that underestimations and overestimations are not only equally likely but also entirely unpredictable by firms. Even then the variation in damages creates problems. One reason is that some downstream firms with limited assets will have incentives to infringe. If the damage award is less than v , the firms will earn enough to offset the damages but invention will be discouraged by the underestimation of damages. If the damages award exceeds v , it will exceed the firms' limited assets and they will go out of business, and this will not offset the discouragement to invention from underestimations.⁷² Further, the variation in damages creates an additional risk that will increase risk-bearing costs, which are already great in invention markets given the uncertainty that invention will lead to valuable patentable products. Such an increase in risk-bearing costs will discourage invention.⁷³

To illustrate, consider again an invention with a v of \$10 per unit. Assume that, if this invention were patented and infringed, juries would assess per unit damages of \$6 with 25 percent chance, \$10 with a 50 percent chance, and \$14 with a 25 percent chance, for an average damage award of \$10 that equals the true value. Downstream firms only earn \$10 from each unit that incorporates the patented feature. Thus, if there is a downstream firm that sells, say, 1 million units and makes \$10 on each of those units, it will make \$10 million due to its use of the patent. If damages are properly set at \$10 per unit, then the firm will be able to pay \$10 million

⁷² See Kaplow & Shavell, *supra* note 69, at 739–41 (noting that limited defendant wealth can make damages undercompensatory and inefficient compared with property rules).

⁷³ In a similar vein, some prior commentators have argued that the reason to have property rules is that it creates greater certainty about outcomes. See, e.g., Epstein, *supra* note 69, at 2094; Carol Rose, *The Shadow of the Cathedral*, 106 YALE L.J. 2175, 2187 (1997); Henry Smith, *Exclusionary and Property Rules*, 90 VA. L. REV. 965, 984 (2004).

to the patent holder. But if damages are instead set at \$14 per unit—which under our assumptions occurs 25 percent of the time—then the downstream firm would be liable for \$14 million in damages. If the firm has limited assets, it may not be able to pay the additional \$4 million in damages above what it earned from the patent. Instead, the firm would only be able to pay the \$10 million in damages, or \$10 per unit. This means that for this firm, 75 percent of the time, it will be paying damages of \$10 per unit, because whenever the court rules that damages are \$14 per unit, the firm can only afford to pay \$10. In the remaining 25 percent of the time, it will be paying damages of \$6 per unit. Expected damages for this wealth-limited firm are thus only \$9. Therefore, because there will be some downstream firms with limited assets, the overall expected level of damages is less than \$10, which will provide suboptimal incentives for invention. In addition, the fact that damages range from \$6 to \$14 per unit means that patent holders face substantial variation in the damages that they receive rather than having the greater certainty that the use of a property rights would provide, thus increasing their risk-bearing costs.

E. The Lemley–Shapiro Model Wrongly Disfavors Licensors

Lemley and Shapiro stress that their conclusions are limited to cases where the patent owner licenses its patent and do not apply when the patent owner competes with the downstream firm that allegedly violates it patent.⁷⁴ Others have relied on their analysis to likewise conclude that injunctions raise no overcompensation problem if the patentee competes downstream.⁷⁵ Similar arguments also seemed to persuade the four justice concurrence in *eBay* that holdup problems create special reasons to deny injunctions when “firms use patents not as a basis for producing and selling goods but, instead, primarily for obtaining licensing fees.”⁷⁶ Many lower courts have likewise accepted this distinction in denying injunctions when the patent holder was not competing with the infringer.⁷⁷

However, to the extent there were any overcompensation problems, the Lemley–Shapiro analysis offers no convincing reason to believe that such problems would be lower for patent owners who compete in downstream markets. To the contrary, any misincentives would be even larger if the patent holder participated in the downstream market. There are two reasons.

First, in addition to any incentives to demand excessive royalties that the patent owner might have, it would also have anticompetitive incentives to

⁷⁴ See Lemley & Shapiro, *supra* note 1, at 2036; Shapiro, *supra* note, at 1, 24.

⁷⁵ See Hahn & Singer, *supra* note 8, at 467–68.

⁷⁶ See *eBay Inc.*, 547 U.S. at 396.

⁷⁷ See Golden, *supra* note 10, at 2113, note 12.

drive its rival out of the downstream market and take its profits. It would thus, if anything, demand a higher royalty rate than otherwise predicted by the Lemley–Shapiro model. For example, a product monopolist that holds a patent will not license it to a new entrant in the product market because the monopolist’s profits will decline if the market becomes a duopoly. Suppose that we assume the same demand function, $Q = A + v - P$, that (as noted above) Lemley and Shapiro assume elsewhere. Then standard monopoly models show that monopoly profits will equal $(A + v - C)^2/4$. If entry leads to Bertrand competition, then profits will drop to zero. If entry instead leads to Cournot competition, then each firm’s profits will drop to $(A + v - C)^2/9$, which is less than 50 percent of the monopolist’s profit.⁷⁸ A patent holder that has market power in the relevant market will accordingly have more profits to lose by licensing others, and thus will demand higher royalties.

Second, if the patent owner participates in the downstream market, its past damages claims would go beyond lost royalties to include lost profits. Such lost profits will result in higher damage awards because total profits per unit are greater than royalties per unit. This will, under the Lemley–Shapiro model, make it demand even more excessive royalties.

Therefore, even if the Lemley–Shapiro model were valid, it can offer no persuasive reason to treat patent owners worse because they do not practice their own patent. The Lemley–Shapiro model thus cannot justify a distinction based on whether patent owners license or practice their patents.

III. THE HOLDUP MODEL DOES NOT APPLY TO COMPETITIVE MARKETS

A. Competitive Downstream Product Markets

Lemley and Shapiro acknowledge that their holdup model does not apply to cases where multiple patent-licensees compete downstream.⁷⁹ However, they indicate that various factors cut in opposing directions from the bilateral monopoly case, from which might one might mistakenly infer that there is no strong reason to think the results differ very much.⁸⁰ In fact, there is every reason to think the results are totally different if the downstream market is competitive.

In a perfectly competitive downstream market, the downstream price equals cost, and equals the value of the product to the marginal purchaser. Because the situation will no longer present a bilateral bargaining game, Nash bargaining no longer applies. If one downstream firm threatens to do without the technology, the patent holder can simply license another

⁷⁸ See DENNIS CARLTON & JEFFREY PERLOFF, *MODERN INDUSTRIAL ORGANIZATION* 195–96 (4th ed. 2005).

⁷⁹ Shapiro, *supra* note 1, at 25; Lemley & Shapiro, *supra* note 1, at 2005, 2008.

⁸⁰ Lemley & Shapiro, *supra* note 1, at 2005–08.

downstream firm that can take the first firm's output away by offering a better product. Given competing downstream firms, the patent holder P can have them compete for the patent until they drive its price up to its expected value $v\theta$. In other words, $\beta = 1$ if the downstream market is competitive.

Assuming damages are properly set at v times X_i for any infringing seller, the expected damages for infringement will be $v\theta X_i$. Thus, if the patent owner tried to charge a royalty of more than $v\theta$, all the downstream firms would decline the license because they would incur expected losses from agreeing. Even if the downstream firms had already used the technology inadvertently, the patent owner could not charge more than $v\theta$ by trying to holdup the downstream firm for some of the costs of redesign, because if it did so the downstream firm would expect to lose money and prefer to exit the market.

Thus, the patent holder must pick the royalty that maximizes its profits subject to the constraint that the royalty must not be greater than $v\theta$. Assume a linear demand curve $Q = A - P$ without the invention, which because P equals marginal cost C in a competitive market, means $Q = A - C$. With an invention with value v and royalty rate r , the market quantity will be $Q = A + v - C - r$. The patent holder seeks to maximize its royalty times the market quantity, $rQ = Ar + vr - Cr - r^2$. Taking the derivative with respect to r and setting it at zero yields $r = (A - C + v)/2$. Because the monopoly price $P_m = (A - C)/2$, this means the profit-maximizing royalty rate is $P_m + v/2$ subject to the constraint that this rate is less than $v\theta$. Thus, the royalty rate will be $v\theta$ whenever $P_m + v/2 \geq v\theta$, which is the same as $P_m \geq (\theta - 1/2)v$. The royalty will thus always be $v\theta$ for any probabilistic patent with a likely validity of less than 50 percent, and also will be $v\theta$ for inventions whose value is not large compared with the price of the downstream product, which is precisely the set of cases Lemley and Shapiro analyze. This conclusion regarding the effects of a value-enhancing patent parallels the conclusions in prior literature dealing with cost-reducing inventions, which likewise found that if the downstream market were competitive and the invention was not drastic, then the royalty rate would equal the full per unit value of the cost-reduction, without any discount for β .⁸¹

The patent owner will thus maximize its profits by licensing the patent at a royalty rate infinitesimally below $v\theta$. If the patent owner charges any amount below $v\theta$, then all the downstream firms would maximize expected profits by accepting and using the technology. If any firms did not use the technology, then they would be driven out of the market by firms that did because the latter would offer a more desired product. If firms did use the

⁸¹ See W. KIP VISCUSI, JOHN M. VERNON & JOSEPH E. HARRINGTON, *ECONOMICS OF REGULATION AND ANTITRUST* 836 (The MIT Press, 2nd ed. 1998); Kamien & Tauman, *Fees Versus Royalties and the Private Value of a Patent*, 101 Q.J. ECON. 471, 477, 482-83 (1986).

technology, then their expected profits would be higher by paying any royalty up to $v\theta$ than by paying the expected damages.

The same is true if the market downstream is marked by recurring fixed costs or product differentiation, making models of “monopolistic competition” more appropriate. In such cases, each downstream firm will earn a profit margin over marginal costs, but earning that margin will be necessary to meet recurring fixed costs and stay in the market.⁸² Thus, the patent owner will once again maximize its profits by licensing the patent at a royalty rate that is infinitesimally below $v\theta$. All the downstream firms would pay up to that rate because, if they do not use the technology, they will be driven out the market by firms that do, and if they use it, they will pay less with a license than in expected damages. Nor will any firm agree to pay more than $v\theta$ because, if it did so, it would fail to make the expected margin necessary to cover recurring its fixed costs and stay in the market.

B. Downstream Product Markets with Dominant Firms or Oligopolies

Lemley and Shapiro argue that their holdup model provides “a very good guide” when one downstream firm earns “far greater revenues” than the other firms.⁸³ But they do not explain why they think so. Given that revenues just means gross sales, this scenario does not at all mean the downstream firm in question earns any abnormal returns. Having higher revenues than other firm does suggest a high market share, but that does not alone suffice to infer market power.⁸⁴ If they do mean to be considering cases where one of the downstream firms earns supracompetitive returns unavailable to the other firms, then one might try to extend the Lemley–Shapiro model, but all the problems noted above would apply.

In separate work, Professors Farrell and Shapiro find that patent holders will get excessive royalties in an oligopolistic market. But there the supposed mechanism does not turn on holdup problems at all. Instead, they reason that the reservation price of any downstream patent user will be higher than optimal because, if the user challenges a patent, a decision invalidating the patent benefits all users by relieving them of royalties, whereas a decision upholding the patent binds it but not other users, and results in a royalty that puts it at a disadvantage compared with licensees who agreed to royalty rates discounted for the *ex ante* likelihood of invalidation.⁸⁵ Because their model in that paper assumes that the patent holder can credibly commit to refusing to alter offered licensing terms even if a user litigates rather than accepts, their model predicts that the royalty will equal the inflated reservation

⁸² See Carlton & Perloff, *supra* note 78, at 200–34.

⁸³ Lemley & Shapiro, *supra* note 1, at 2008.

⁸⁴ See Elhauge, *supra* note 60, at 209–11.

⁸⁵ See Farrell & Shapiro, *supra* note 26; Shapiro, *Patent Reform*, *supra* note 7, at 119.

price of users.⁸⁶ If, instead, they assumed the sort of Nash bargaining that the Lemley–Shapiro model assumed, they would also have to take into account the fact that precisely the same litigation rules also depress the reservation price of the patent holder. The reason is that the patent holder can lose royalties to all users if it loses the litigation, but can gain royalties from only one user if it wins. Because the litigation rules shift both side’s reservation prices away from the middle, they have no clear effect on whether such a shift would produce a higher or lower royalty under Nash bargaining.

Farrell and Shapiro suggest that the patent holder can credibly commit to sticking to a profit-losing offer when it is rejected by one user because the patent holder’s decisions affect its profits from other users.⁸⁷ But if one takes that into account, then one must also take into account the parallel problem that any user knows that its decision on this patent will affect its profits on other patents, as I noted above in Section I.D.1. Given the reality of multiple patent-holders negotiating with multiple patent-users, litigation rules that both lower the reservation prices of patent holders and raise those of users are likely to have unclear overall effects on royalty rates.

In any event, even if the Farrell–Shapiro model offered a good basis for legal reform, it would not support any changes to the rules on injunctions or damages. Instead, it would suggest changing the litigation rules that produce the asymmetry they model. Farrell and Shapiro suggest determining patent validity before licensing, but other natural reforms might be to restore symmetry either by (i) changing the rule on collateral estoppel so that a decision of patent invalidity does not benefit non-challenging users, or (ii) joining all users to any patent challenge and allowing a renegotiation of royalty rates with all licensees after any decision upholding a patent.

C. Upstream Competition in Patent Markets

Finally, the upstream patent market might also be competitive. Suppose two patent holders develop separate patents that each can achieve precisely the same value v . Lemley and Shapiro say the benchmark royalty should be the difference between the value of the patent and the next best alternative.⁸⁸ But that standard would mean that they normatively favor a benchmark rate of zero in this case because there is no difference between the patent’s value and the value of the next best alternative. Such a benchmark would obviously be quite undercompensatory because, if the patent holders anticipated getting it, then neither would have invested to create the patent, and the social value v would have been lost.

Unfortunately, in such a case, a royalty rate of zero would be the predicted result if we assume Bertrand competition between the two patent

⁸⁶ See Farrell & Shapiro, *supra* note 26, at 5–6.

⁸⁷ *Id.* at 18.

⁸⁸ Lemley & Shapiro, *supra* note 1, at 2039, note 153.

holders, because such competition would drive the price for each patent down to its marginal cost, which is zero. Such Bertrand competition is much more plausible than in ordinary product markets because patent holders set prices, not output, for usage of their inventions, making Cournot and Stackelberg models inapplicable. Oligopolistic coordination among patent holders also seems less likely than coordination among ordinary product makers because patent licenses tend to be non-public, long-term, and reflect one-shot deals from the patent holder's perspective, all of which are factors that make coordination less likely. In any event, even if some oligopolistic coordination occurs, it will result in royalties that are lower than the monopoly price v so long as coordination is not perfect. This will thus still tend to make patent royalties undercompensatory.

More generally, each firm considering whether to invest in creating an invention will have to discount its expected profits by the odds that some other firm will discover a separate invention that achieves the same value in some other way. This effect is most devastating if Bertrand competition results but persists to some degree with oligopolistic coordination. This is another way in which royalties will tend to be undercompensatory and discourage the optimal level of research.

IV. THE ROYALTY STACKING THEORY IS INCORRECT

Lemley and Shapiro try to extend their model about one patent owner facing one downstream firm to cases where there are multiple patent-owners facing one downstream firm. They conclude that, in such cases, a "royalty stacking" problem will be created in which each patent owner charges more than the value of its product. But properly analyzed, their model instead shows precisely the opposite.

Lemley and Shapiro assume linear demand, where $X = A + V - P$, and V is the sum of all the v_i provided by each individual patent owner.⁸⁹ They then purport to show that, with royalty stacking, each royalty owner has incentives to charge a royalty r_i that is greater than v_i , so that the sum of all the royalties $R > V$. But their model depends on their assumption that the downstream firm has only two options, either (i) pay the royalty or (ii) use the technology without paying and litigate.⁹⁰ In reality, there is another simple option: simply decline to use the overpriced technologies at all.

Which option is more profitable? Under their own assumed linear demand function, a downstream monopolist will charge a price of $0.5(A + V + C + R)$, and thus earn a profit margin of $0.5(A + V + C + R) - C - R = 0.5(A + V - C - R)$. The firm's output will be $0.5(A + V - C - R)$. Thus its total profits will be equal to $0.25(A + V - C - R)^2$. The downstream firm that

⁸⁹ Lemley & Shapiro, *supra* note 1, at 2046.

⁹⁰ *Id.* at 2046–48.

simply declines to use the patented technologies will charge a price $0.5(A + C)$, and thus earn per-unit profits of $0.5(A + C) - C$, or $0.5(A - C)$. Its output will be $0.5(A - C)$, for total profits of $0.25(A - C)^2$.

Using the technologies will thus be profitable only if $0.25(A + V - C - R)^2 > 0.25(A - C)^2$, which is true only if $V > R$. Thus, if a downstream firm were presented with stacked royalties of $R > V$, as suggested by Lemley and Shapiro, then it would always be more profitable for the downstream firm to simply decline to use those overpriced technologies.

The same is true if we ask this question not for all the patent owners collectively, but for any one individually. If we are considering one individual patent, the value and cost from all the other licenses would simply be incorporated into A and C , and V and R would be replaced by v_i and r_i in the above formulas. The same analysis would thus apply as to whether it was more profitable to use a technology with value v_i when it is less than r_i . The conclusion would again be that it would always be more profitable for the monopolist to decline to use the technology whenever $r_i > v_i$. Thus, each royalty owner individually faces the same constraint that they do collectively. The royalty-stacking problem thus disappears.

Accordingly, a downstream monopolist will not find it profitable to pay for any license i unless $r_i \leq v_i$, and therefore it will be the case that $R = \sum r_i \leq \sum v_i = V$. Thus, any patent owner i will not offer a license at a rate greater than v_i because the downstream firm will reject the license. Indeed, given the series of patents being offered to the downstream firm, the downstream firm will (given Theorem 2), pay no more than r_i^* , which can be substantially less than v_i , because it knows that a threat to do so will be credible given that it is playing a multi-stage game while the patent holders are playing a single-shot game.

Further, when there are multiple patent-owners, individual owners might find it optimal to offer a license for less than v_i . In particular, if there are N patents and the value of each patent is the same, meaning $v_i = V/N$, then Lemley and Shapiro's model shows that the patent owners will choose the royalty rate $r_i = (A + V - C) / (N + 1)$ even if that is less than v_i .⁹¹ This royalty rate will be less than v_i whenever $(A + V - C) / (N + 1) < V/N$, which can be rearranged as $v_i > A - C$. Given standard monopoly models, the latter equals double the per-unit profit at the monopoly price if no patented inputs are used.

In short, royalty stacking will never lead to royalty rates that exceed the patent value because it would always be profitable for the downstream firm to refuse to use the technologies in such a case. But royalty stacking can lead patent holders to ask for royalty rates that are lower than the patent value if that value exceeds double the per-unit monopoly profit without using patents. Further, royalty stacking is likely to lead to royalty rates that are lower than patent values if the downstream firm has incentives to consider

⁹¹ See Lemley & Shapiro, *supra* note 1, at 2047.

multi-period effects of agreeing to higher royalties while the patent holders are playing a single-shot game. The combined effect is once again to make royalty rates more undercompensatory, not more overcompensatory.

V. CONCLUSION

Lemley and Shapiro have offered some illuminating models on patent holdup, damages, and royalty stacking. Correcting various problems with those models, however, leads to sharply different conclusions than the ones they reach. Correcting their assumed optimal benchmark shows that their predicted royalties are often (plausibly most of the time) below the optimal royalty. Correcting their assumptions about one-shot bargaining, informational symmetry, or constant demand each separately shows that their predicted royalty rates are too high. Their predicted royalty rates are even more overstated if we take into account the fact that courts often base damages on past negotiated royalties or that juries do not adjudicate damages with perfect accuracy, and the latter problem holds whether the jury errors are systematic or balanced. Because actual royalty rates will be lower than they predict for the above reasons, those reasons exacerbate the likelihood that actual royalty rates will be inefficiently low.

Further, their holdup model does not apply in cases where multiple patent-licensees compete downstream. In such cases, competition will likely drive royalties toward patent value. Nor does their holdup model apply in cases where multiple patent-owners compete upstream. In such cases, royalties will tend to be inefficiently low.

Finally, their conclusion that royalty stacking will make royalties overcompensatory is incorrect. Instead, a downstream firm will use a particular invention only if its royalty is no greater than its value. However, the existence of multiple patent-owners will sometimes lead individual patent owners to charge less than the value of their patents, thus increasing the likelihood that royalties are inefficiently low.

APPENDIX

Proof of Theorem 3. Let C be the marginal cost without the patent, P^* and X be the price and output, respectively, with the patented technology, and R_L be the lump-sum royalty. Then, with the patented feature, standard monopoly models indicate the downstream price $P^* = 0.5(A + v + C)$, and the downstream output $X = 0.5(A + v - C)$. D 's profits per unit will be $P^* - C$, making its total profits $(P^* - C)X - R_L$. Because $X = P^* - C$,⁹² this is the

⁹² We know this because $P^* - C = 0.5(A + v + C) - C = 0.5(A + v - C)$, which equals X .

same as $X^2 - R_L$. Without the patented feature, the downstream price will be $0.5(A + C) = P^* - 0.5v$, and the downstream output will be $0.5(A - C) = X - 0.5v$. D 's profits will equal $(P^* - 0.5v - C)(X - 0.5v) = (X - 0.5v)^2$. Thus, D 's profits will be higher with the patented invention by $X^2 - R_L - (X - 0.5v)^2 = vX - 0.25v^2 - R_L$. In contrast, under constant demand, using the patented invention increases D 's profits by $vX - R_L$. Thus, the increase in D 's profits from licensing is $0.25v^2$ lower with linear demand than with constant demand.

Using the same Nash bargaining approach used by Lemley and Shapiro, the royalty rate will equal the patent holder's disagreement payoff (if the parties do not agree to a license) plus β times the joint gains if they agree to a license. Take first the case of a non-surprise patent that leads to a pre-design license. The patent holder's disagreement payoff would have been zero if D had instead designed around its patent. If they negotiate a license, the patent holder gains profits of R_L , and D increases profits by $vX - 0.25v^2 - R_L$. The joint gains from trade are thus $vX - 0.25v^2$. The patent holder will accordingly bargain for a total payoff of $\beta vX - 0.25\beta v^2$, which is $0.25\beta v^2$ lower than the βvX royalty payment predicted by Lemley and Shapiro.

Now take the case of a strong surprise patent. The joint gains from trade (avoiding the fixed costs of redesign) will be the same as with constant demand because the fixed costs do not vary with demand. The patent holder's payoff from disagreement would be the combination of (1) the patent holder's expected damages plus (2) the expected post-judgment negotiated payment. Because Lemley and Shapiro assumed that damages reflect the rate that would have been negotiated for a valid patent without holdup problems, those damages will not be βvX but rather $\beta vX - 0.25\beta v^2$, making expected damages equal to $\theta[\beta vX - 0.25\beta v^2]T$. Likewise, if the patent holder wins the lawsuit, the negotiated post-judgment royalty rate would be $\beta v - 0.25\beta v^2/X$ because, given redesign, there is no holdup problem post-judgment.⁹³ Taking into account the litigation odds and post-trial length, the expected post-judgment payment will thus be $\theta[\beta vX - 0.25\beta v^2](1 - T)$. The combined payoff from disagreement to the patent holder will thus be $\theta\beta vX - 0.25\theta\beta v^2$. This is $0.25\theta\beta v^2$ lower than the disagreement payoff predicted by Lemley and Shapiro. Thus, changing their assumption of constant demand to linear demand makes royalties $0.25\theta\beta v^2$ lower than Lemley and Shapiro's predicted royalty rate.

For a weak surprise patent, the joint gains of trade from licensing in advance are avoiding litigation costs,⁹⁴ which I presume would be the same with constant or linear demand. As for the disagreement payoff, expected

⁹³ See Shapiro, *supra* note 1, at 30.

⁹⁴ Avoiding redesign and lost profits during the period of redesign are not joint gains from licensing in advance because those are equally obtainable by agreeing to a post-judgment license, and are taken into account when considering the royalty for such a license.

damages would be $0.25\theta\beta v^2 T$ lower than with constant demand. The expected post-judgment negotiated payment would be lowered by $0.25\theta\beta v^2(1 - T)$ because of the reduction in the profits gained from using the patented feature throughout the post-judgment period. The expected post-judgment negotiated payment would also reflect a share of gains from not shutting down during a redesign period, where those gains equal the length of redesign times the profits the downstream firm could make by selling its product without the patented feature. Under linear demand, those lost profits without the patented feature would be $(X - 0.5v)^2$. Under Lemley and Shapiro's model, they are $(m - v)X$, and since $m - v$ equals the profit margin without the patented feature, it equals $X - 0.5v$, making lost profits $(X - 0.5v)X$.⁹⁵ Taking the difference between the two shows that the lost profits during the redesign period are thus $0.5vX - 0.25v^2$ lower with linear demand than with constant demand. This will further reduce the expected post-judgment negotiated payment by $\theta\beta(0.5vX - 0.25v^2)L$. Thus, royalties will in total be $0.25\theta\beta v^2 + \theta\beta(0.5vX - 0.25v^2)L$ lower than Lemley and Shapiro's predicted royalty rate.

Proof of Theorem 4. Now, instead, assume a royalty is charged per unit produced. Then, with the patented feature, standard monopoly models indicate the downstream price $P^* = 0.5(A + v + C + r)$, and the downstream output $X = 0.5(A + v - C - r)$. D 's profits per unit will be $P^* - C - r$, making its total profits $(P^* - C - r)X$. Because $X = P^* - C - r$,⁹⁶ this is the same as X^2 . Without the patented feature, the downstream price will be $0.5(A + C) = P^* - 0.5v - 0.5r$, and the downstream output will be $0.5(A - C) = X - 0.5v + 0.5r$. D 's profits per unit without the patent will equal $P^* - 0.5v - 0.5r - C$, which equals $X - 0.5v + 0.5r$. Total profits are thus $[X - 0.5(v - r)]^2$ or $X^2 - X(v - r) + 0.25(v - r)^2$. D 's profits are thus $vX - rX - 0.25(v - r)^2$ higher with a patent. In contrast, under constant demand, using the patented invention increases D 's profits by $vX - rX$. Thus, the increase in D 's profits from licensing is $0.25\beta(v - r)^2$ lower with linear demand than with constant demand.

Take first the case of a non-surprise patent that leads to a pre-design license. The patent holder's disagreement payoff would be zero if D had instead designed around the patent. If they negotiate a license, the patent holder would gain profits of rX , and D would increase its profits by $vX - rX - 0.25(v - r)^2$. The joint gains from trade are thus $vX - 0.25(v - r)^2$. Given the Nash bargaining predicted by the Lemley-Shapiro model, the patent holder will bargain for a total royalty of $\beta vX - 0.25\beta(v - r)^2$. This

⁹⁵ I am indebted to John Golden for this point.

⁹⁶ We know this because $P^* - C - r = 0.5(A + v + C + r) - C - r = 0.5(A + v - c - r)$, which equals X .

is always $0.25\beta(v - r)^2$ lower than Lemley and Shapiro's predicted royalty rate.⁹⁷

Now take the case of a strong surprise patent. Once again, the joint gains from trade will be the same as with constant demand because the fixed costs do not vary with demand. The patent holder's payoff from disagreement would once again be the combination of expected damages and the expected post-judgment negotiated payment, which will equal θ times the rate that would be negotiated for a valid patent without holdup problems. Here, that rate is $\beta vX - 0.25\beta(v - r)^2$, making the patent holder's payoff from disagreement equal to $\theta\beta vX - 0.25\theta\beta(v - r)^2$. This is lower than the payoff predicted with constant demand by $0.25\theta\beta(v - r)^2$.

For a weak surprise patent, the joint gains of trade are avoiding litigation costs that are presumably constant and thus affected by linear demand. The portion of the disagreement payoff that reflects expected damages or the damage compensation portion of negotiated post-trial payments will be lower by $0.25\theta\beta(v - r)^2$ for the same reasons just noted above. The portion of post-trial payments reflecting the risk of having to lose profits while shut down for redesign will turn on profits without the patented feature. With linear demand, those profits are $[X - 0.5(v - r)]^2$, whereas with constant demand they would be $X^2 - 0.5(v - r)X$. The difference between those is $0.5(v - r)X - 0.25(v - r)^2$, which will further reduce the expected post-judgment negotiated payment by $\theta\beta[0.5(v - r)X - 0.25(v - r)^2]L$. Thus, royalties will in total be $0.25\theta\beta(v - r)^2 + \theta\beta[0.5(v - r)X - 0.25(v - r)^2]L$ lower than Lemley and Shapiro's predicted royalty rate.⁹⁸

Comparing the Results of Theorems 3 and 4. The above formulas might seem to suggest that the total amount of royalties paid is larger if per-unit royalties are used than if lump-sum royalties are used, given that the overstatement of royalties looks smaller in the former case. However, one can show that this appearance is an artifact of the fact that X is higher with a lump-sum royalty than with a per-unit royalty. Suppose that we call X the output with the patented feature under a lump sum payment; then the output with a per-unit royalty will be $X - 0.5r$. Thus, for pre-design licenses, the total amount of royalties paid will be $\beta v(X - 0.5r) - 0.25\beta(v - r)^2$ for the per-unit royalty, and $\beta vX - 0.25\beta v^2$ for the lump-sum royalty. The former will be always smaller than the latter. This lower royalty is consistent with the intuition that the joint gains from trade should be lower when one considers per-unit royalties because they lower output.

⁹⁷ Given that rX equals this payoff $\beta vX - 0.25\beta(v - r)^2$, one can solve the quadratic to conclude that the negotiated royalty rate with Nash bargaining will be $v - 2X/\beta + (2/\beta)[X^2 - (1 - \beta)\beta vX]^{0.5}$.

⁹⁸ One can solve both the strong and weak surprise patent equations for r , but the resulting quadratics are messy.