

Antitrust limits to patent settlements

Carl Shapiro*

Patents, patent litigation, and patent settlements increasingly influence competition. Settlements of patent disputes come in many forms, including licensing and cross-licensing agreements, patent pools, mergers, and joint ventures. While frequently procompetitive, such settlements can stifle competition and harm consumers. I propose a specific antitrust rule limiting such settlements: a settlement must leave consumers at least as well off as they would have been from ongoing patent litigation. After establishing that profitable settlements satisfying this constraint generally exist, I show how this antitrust rule can be used to evaluate three types of settlements: mergers, patent pools, and negotiated entry dates.

1. Introduction

■ Intellectual property rights are widely recognized as critical assets in many industries. Large companies like IBM, Intel, Hewlett-Packard, and Microsoft regard their patents and copyrights, along with their other intangible assets such as know-how, as central elements giving them competitive advantage. Likewise, many start-up companies seeking funding from venture capitalists recognize that patent or copyright protection is essential if they are to attract financing and recoup their R&D expenditures. Put simply, patents and copyrights are often the crown jewels in a high-tech company's collection of assets.¹

Patents are playing an increasing role in determining competitive outcomes. The increased propensity of companies to file for patents has been well documented. See, for example, Kortum and Lerner (1998), Grindley and Teece (1997), Cohen, Nelson, and Walsh (2000), and Hall and Ziedonis (2001). Gallini (2002) reports that the number of new U.S. patents granted doubled from 1985 to 1999.

The increasing importance and number of patents and copyrights is leading inevitably to more and more intellectual property disputes between rightsholders and alleged infringers. Looking over the period from 1978 to 1995, Lanjouw and Schankerman (2002, p. 1) report that "The number of patent suits filed rose by almost tenfold, with much of this increase occurring during the 1990s." Making matters even more complex, many products can potentially infringe *multiple patents*. As described in Shapiro (2001), more and more companies are facing a *patent thicket* requiring them to obtain multiple licenses to bring their products safely to market.

* University of California at Berkeley; shapiro@haas.berkeley.edu.

I am grateful to a number of colleagues who have provided comments on this work: Jeremy Bulow, Joseph Farrell, Richard Gilbert, John Hayes, Mark Lemley, Jennifer Reinganum, Marc Schildkraut, Joel Schrag, Suzanne Scotchmer, Hal Varian, Robert Willig, and especially Michael Katz. I also thank conference participants at Toulouse (2001) and at Berkeley (2001), and two referees for their suggestions.

¹ For simplicity, in this article I shall typically refer to "patents" rather than "intellectual property" more generally. Apart from differences in the legal treatment of different forms of intellectual property, virtually everything in this article could be applied equally to copyrights, or even trade secrets, as well as patents.

The need to negotiate licenses or other settlements of intellectual property disputes is made even greater because of the danger of hidden or submarine patents, which make it all too easy for a company unintentionally to infringe on a patent that was not yet issued when the company's product was designed.² Likewise, the need to resolve intellectual property disputes is arguably made yet greater to the extent that the U.S. Patent and Trademark Office has issued "bad" patents, i.e., patents on technology that does not in fact meet the novelty requirement. Many critics have charged that the Patent and Trademark Office has had a poor understanding of prior art, especially in the software area, and has improperly issued a number of patents. Bad or not, there is no dispute that the number of patents being issued, and the number of patent disputes, is growing dramatically. In short, a compelling case can be made that intellectual property disputes are of increasing importance in determining just which firms can compete in which markets, and on what terms.

The lion's share of patent disputes are settled rather than litigated to a resolution in court. Lanjouw and Schankerman (2002) find that some 95% of patent lawsuits are settled prior to a court judgment. In fact, a wide range of commercial arrangements involving intellectual property can be regarded as settlements of intellectual property disputes, either literally (in the sense that litigation has been initiated and is dropped once an agreement is reached) or effectively (because negotiation takes place in the shadow of possible litigation). Virtually every patent license can be viewed as a settlement of a patent dispute: the royalty rate presumably reflects the two parties' strengths or weaknesses in patent litigation in conjunction with the licensee's ability to invent around the patent. The same is true of cross-licenses, where net payments reflect the strength of each party's patent portfolio along with its commercial exposure to the other's patents. Mergers and joint ventures are yet more ways to settle patent disputes.

In this article I do not distinguish between settlements that take place after patent litigation commences and those that take place before the filing of a patent lawsuit. Both types of settlements take place in the shadow of an ultimate court ruling on patent validity and/or infringement, and both types of settlements raise the same antitrust issues.³

Given the importance of patents and their licensing to innovation, and given the many commercial arrangements that are effectively settlements of intellectual property disputes, the legal rules governing the resolution of such disputes are of first-order importance. In a very real sense, the rules governing settlements affect what is truly meant by the patent grant itself. In fact, in many fast-moving industries, the rules governing patent litigation and settlements are arguably far more important to patentees than the single variable on which economists have traditionally focused, namely patent length.⁴

My focus in this article is on one particular class of legal rules that govern intellectual property rights: the *antitrust limits* imposed on patent settlements. The need for some basic antitrust limits should be obvious. Suppose that firm A has a fairly weak and fairly narrow patent that it is asserting against its sole rival, firm B. The two firms are competing vigorously, with firm B evidently not deterred from competing just because firm A has sued it for infringement. Firm B believes that the patent is likely to be found invalid if tested in court. Even if the patent is valid, firm B is optimistic that it will not be found to infringe the patent, and in any event firm B believes it could easily design around the patent and remain a very effective rival to firm A. Now imagine that firms A and B agree to merge to resolve their patent dispute. If the merger would be judged anticompetitive in the absence of the patent, there is no reason to believe that this one weak patent should reverse

² Recent reforms to disclose some patent applications before the issuance of the patent should alleviate, but not eliminate, this problem. The ability of those applying for patents to revise their patent applications over time tends to exacerbate the problem.

³ I also make no sharp distinction between patent cases involving validity and those involving infringement; the main difference here is that a finding of invalidity typically will enable more competition than a finding of noninfringement, a factor that I include in my analysis.

⁴ The importance of patent litigation, and thus rules governing settlements, varies across industries and by type of patent. For a study of how the frequency of court cases varies with characteristics of patents and their owners, see Lanjouw and Schankerman (2001).

that conclusion. Or imagine that firms A and B agree to a settlement under which firm B pays significant per-unit royalties to firm A and firm A makes a fixed payment to firm B. As shown in Katz and Shapiro (1985), such settlements can replicate the cartel or monopoly outcome. Even a clearly invalid patent could be used as cover for a cartel-like agreement.

Precisely because patent settlements can be anticompetitive, and because settling parties may have an incentive to insert anticompetitive provisions into their agreements, antitrust interest in the settlements of intellectual property disputes is very high. The 1995 Department of Justice/Federal Trade Commission *Antitrust Guidelines for the Licensing of Intellectual Property* describe generally how the agencies will analyze various commercial agreements involving intellectual property. In 1997, a clear statement of concern about settlements was voiced by then Assistant Attorney General Joel Klein, who even floated the idea of having companies notify antitrust officials of their settlements of patent disputes.⁵ More recently, senior antitrust enforcement officials have spoken repeatedly about intellectual property rights and their role in the new economy.⁶ In spring 2002, the FTC and the DOJ held an extensive series of hearings on competition and intellectual property, including sessions focusing on the settlement of patent disputes.

Beyond these general statements, we can observe a number of cases in which the Justice Department and the Federal Trade Commission have investigated and/or challenged settlements that they regarded as anticompetitive, or expressed views on the antitrust limits on patent settlements. I shall refer to several such cases below, but pause here to note two examples. After Digital Equipment Corporation sued Intel for patent infringement, Intel settled with Digital by purchasing certain assets from Digital and entering into a supply agreement with Digital. The FTC issued a complaint and required modifications of the settlement agreement.⁷ When six companies (Toshiba, Hitachi, Matsushita, Mitsubishi, Time Warner, and JVC) sought to jointly license their patents necessary to the production of DVD discs and players, the DOJ issued a business review letter approving their patent pool subject to certain conditions.⁸

Within the area of antitrust limits on settlements, the current article is fairly ambitious. In Section 2 I discuss generally the benefits and costs of patent settlements and explain more fully why antitrust limits on such settlements are unquestionably needed to prevent abuse of the settlement process. Then, in Section 3, I propose and develop a general rule for evaluating proposed settlements, namely a requirement that the proposed settlement generate at least as much surplus for consumers as they would have enjoyed had the settlement not been reached and the dispute instead been resolved through litigation. My proposed rule is designed to fully respect intellectual property rights while emphasizing that such rights are inherently uncertain or imperfect, at least until they have successfully survived a challenge in court. My proposed rule is also intended to enable a wide range of settlements that can enhance efficiency and promote competition without depriving rightsholders of their legitimate returns to invention. I prove a very general result showing that in virtually all cases, settlements exist that are better for consumers as well as the settling parties in comparison with ongoing litigation.

The balance of the article then applies the general rule to three different settlement types. For each type of settlement, I develop some basic theory and describe some actual settlements in this category where antitrust issues have arisen. Section 4 handles mergers. Section 5 covers negotiated entry dates. Section 6 covers patent pools. Section 7 summarizes my conclusions and outlines some ways in which my analysis can be extended.

⁵ See Klein (1997). Currently, companies entering into certain exclusive licensing arrangements must notify the government under the Hart-Scott-Rodino Act, as some exclusive licensing arrangements have been interpreted as involving the acquisition of assets.

⁶ See especially former FTC Chairman Pitofsky's speech, "Antitrust and Intellectual Property: Unresolved Issues at the Heart of the New Economy," March 2, 2001 (available at www.ftc.gov/speeches/pitofsky/ipf301.htm). For broader discussions of antitrust and intellectual property, see Gilbert and Shapiro (1997) and Gilbert and Tom (2001).

⁷ See the FTC's Complaint in the Matter of Digital Equipment Corporation (available at www.ftc.gov/os/1998/9807/9810040cmp.htm) and the Agreement Containing Consent Order (available at www.ftc.gov/os/1998/9804/9810040_agr.htm).

⁸ See Justice Department Business Review Letter of June 10, 1999 (available at www.usdoj.gov:80/atr/public/busreview/2485.htm).

2. Benefits, costs, and dangers of patent settlements

■ **Benefits and costs of settlements.** Settlements of litigation generally are recognized to provide a number of private and social benefits.⁹ Private benefits include the avoidance of litigation costs and the resolution of uncertainty. Social benefits include savings on court costs and/or reduction of congestion in the court system. Social costs can include the lack of resolution of a legal issue with applicability beyond the individual case at hand. Generally speaking, the courts have strongly favored settlements, in large part to reduce congestion and delays in the court system.

Unlike many other settlements of litigation, settlements of patent litigation between rivals by their very nature implicate competition and thus tend to have effects on third parties, most notably (but not only) customers of the litigating parties. Patent settlements certainly can enable the settling parties to compete more effectively with others, as when two firms with complementary patents agree to a cross-license enabling each of them to make higher-quality products or achieve lower production costs. But patent settlements can also enable the settling parties to restrict competition between themselves, to the detriment of consumers. Consumers may suffer from lost rivalry, both during the interim period while patent litigation would have continued, and perhaps in the longer term as well, at least until the patent expires. Settlements can deprive consumers not only of competition between the settling parties, but from other firms as well if an invalid patent is never actually challenged.

□ **Unconstrained settlements and their dangers.** Patent settlements present an especially tricky area for antitrust because of the undisputed procompetitive benefits that can result from a wide range of settlements, including the vast majority of patent licenses. Drawing the line between “price-fixing agreements” and “procompetitive licensing arrangements” is not a simple matter. But the need for such a line should not be in dispute.

Suppose that two rivals are permitted to settle their patent dispute with no antitrust limits. Of course, they still must successfully reach an agreement to settle their dispute, and this may be difficult for the usual reasons that negotiations break down, including potentially asymmetric information (more specifically, optimism on both sides about their prospects in litigation). But for now let us assume that the two firms bargain efficiently and thus reach a settlement that maximizes joint profits. What do such settlements look like?

It is immediately evident that such settlements could be used to eliminate competition that would have arisen had the patentholder lost. By eliminating such competition, monopoly profits can be enjoyed, even if the patent was very weak or even worthless. There are many ways that such settlements could be structured: (1) the patentholder could acquire the challenger, with the purchase price set in some mutually agreeable fashion to split the gains from trade, including the gains from eliminating competition; (2) the patentholder could make a fixed payment to the challenger in exchange for the challenger’s agreement not to compete, either at all or in certain product areas, geographic areas, or during some specified time period; (3) the two companies could enter into a joint venture or other cooperative arrangement (such as a supply agreement or co-marketing setup) whereby they both participate in the market without directly competing against each other; or (4) the challenger could agree to pay certain per-unit royalties to the patentholder in conjunction with a fixed payment running from the patentholder to the challenger. The only requirements for such profit-maximizing settlements are (i) that they preserve the monopoly power that the patentee would have had in the absence of the challenger, and (ii) that each party find it individually rational to accept the settlement rather than continue to litigate.

A hallmark of these anticompetitive agreements is that the patentholder agrees to share its monopoly profits with the challenger in order to induce the challenger to give up its fight. In the merger context this is clear: the challenger is paid the acquisition price. A bald payment not to compete is even more explicit (and more difficult to justify). A joint marketing program could also explicitly share the monopoly profits with the challenger.

⁹ For a general discussion of the benefits and costs of settlement versus litigation, see Cooter and Rubinfeld (1989).

Clearly, such agreements will tend to be contrary to the interests of consumers: without limits on patent settlements, consumers will receive only the surplus available facing a monopolist. Such settlements can deprive consumers of the advantages that competition, or at least its prospect, would have offered to them, during the interim period before the resolution of the patent dispute and subsequently, if the patent would have been declared invalid or not infringed, or had the challenger found a practical way to invent around the patent. Thus, antitrust oversight of patent settlement is clearly warranted. In principle, all manner of settlements are subject to antitrust scrutiny; in practice, the FTC and the DOJ (as well as private plaintiffs) are far more likely to become aware of settlements that occur after patent lawsuits are filed than those that arise in the context of private patent infringement negotiations.¹⁰

□ **Patents as partial property rights.** It is important to bear in mind that the monopoly profits that can be (jointly) achieved through unconstrained settlement do not merely represent the rights granted to the patentee by virtue of having obtained the patent in the first place. As emphasized by Ayres and Klemperer (1999), a patent is best viewed as a *probabilistic* property right. What the patent grant actually gives the patentholder is the right to sue to prevent others from infringing the patent. Nothing in the patent grant guarantees that the patent will be declared valid, or that the defendant in the patent suit will be found to have infringed. In other words, all real patents are *less strong* than the idealized patent grant usually imagined in economic theory.

A real patent may prove to be less valuable than the idealized patent in several distinct ways: (1) The real patent may be found invalid, either in whole or in part; (2) the real patent may be found not to be infringed by a given product sold or process employed by another firm; (3) the real patent may be relatively easy for others to invent around rather than pay royalties or be forced to cease production; (4) the real patentee may be unable to find infringers or obtain a preliminary injunction to prevent infringement; (5) enforcing the real patent entails litigation costs and delays; or (6) the real patentee may not be able to receive as a judgment all of its lost profits that result from infringement, e.g., if the infringer's assets are insufficient to satisfy the award. In short, a real patent does not give the patentee "the right to exclude" but rather the more limited "right to *try* to exclude" by asserting its patent in court.

In this article, I shall take as given the bundle of uncertain and imperfect rights that we call a "patent." These rights are typically far less valuable than would be idealized "ironclad" patent rights. In my view, the patentholder is not "entitled" to obtain the same level of profits, or the same rights to exclude rivals, as would the owner of the fictionalized ironclad patent. Therefore, the patentholder is not "entitled" to negotiate a monopoly outcome, just because the patentholder *asserts* that its patent is valid and infringed by a particular rival. Rather, the patentholder's rights are calibrated according to the likelihood that the patentholder would win the patent litigation, and the extent of exclusion that such a victory would permit. Generally, these rights are not as strong or as valuable as the rights of a full-fledged monopolist owning an ironclad, blocking patent.

3. Proposed principle: settlement cannot harm consumers

■ Given the obvious incentive to use settlements to replicate the monopoly outcome, and given that the patent grant is not the same as an ironclad right to monopoly profits, antitrust limits on settlements are clearly needed. At the same time, a prohibition on settling patent disputes cannot make sense: as noted earlier, virtually every patent license can be viewed as the settlement of a patent dispute, and settlements generally can provide many benefits not only to the settling parties but to consumers as well. Since many settlements are procompetitive, in the sense that consumers are better off under the settlement than they would be from ongoing litigation, overly strict antitrust limits, not to mention a ban on settlements, would clearly be counterproductive.

¹⁰ More stringent antitrust limits on patent settlements will tend to push competitors to resolve their patent disputes quietly and privately rather than going to court, since embarking on patent litigation will increase the chance that any subsequent settlement they reach will face an antitrust challenge.

So, we must face the complex question of how to draw the line between acceptable and unacceptable patent settlements from an antitrust perspective. In this article, I propose and explore in some depth the following simple antitrust rule: a patent settlement cannot lead to lower expected consumer surplus than would have arisen from ongoing litigation. Effectively, consumers have a “property right” to the level of competition that would have prevailed, on average, had the two parties litigated the patent dispute to a resolution in the courts. So long as consumers’ rights to this level of competition/benefits are respected, the two parties are permitted to negotiate more profitable arrangements that they each prefer to litigation.¹¹

I believe that this standard has much to commend itself. First, it balances the rights of patentees with consumer interests. Antitrust enforcement (such as merger review) often uses a consumer-welfare standard rather than a total-surplus standard. Clearly, a short-run consumer-surplus standard is not sensible when intellectual property rights are involved: declaring all extant intellectual property rights invalid could well maximize short-run consumer surplus, but at the obvious expense of longer-term innovation and consumer interests. Indeed, it is hard to articulate an alternative standard that encourages innovation in the long run, promotes efficient commercial arrangements in the short run, and still protects consumers from cartel-like settlements.

Second, my proposed standard fully respects the property rights granted to patentholders. Since patents involve “partial” or “probabilistic” property rights, as discussed above, patentholders are not entitled to the same level of profits that would result from an ironclad patent covering the same patent claims. Put differently, competition that would take place under the shadow of patent litigation is considered entirely legitimate, even though it may wind up constituting infringement. Lurking behind this view are two broad assumptions worth making explicit: (1) I take as given the intellectual property rights regime, with its necessary imperfections, such as the granting of patents that will later be found invalid and the chance that the holder of a valid patent may not be able to obtain an injunction to stop what turns out to be actual infringement; and (2) I take as given the damages regime associated with patent infringement, including both the rules for calculating damages and the fact that patentees may not be able to fully collect on damage awards in some cases.

My proposed standard for patent settlements is also consistent with how antitrust policy and law treat other forms of collaboration among competitors. A proposed merger, for example, is usually judged to be procompetitive if consumers are better off under the proposed merger than they would be in the absence of the merger. The same standard is used for joint ventures and co-marketing arrangements between direct rivals.¹² Likewise, under the DOJ/FTC *Antitrust Guidelines for the Licensing of Intellectual Property*, licenses are generally regarded as procompetitive if they do not restrict competition that would have taken place in the absence of the license. While the FTC and the DOJ have not published specific guidelines regarding settlements, their enforcement actions involving settlements are consistent with my standard. For example, the FTC has challenged drug settlements on the grounds that they delayed generic entry in comparison with ongoing litigation.

An alternative antitrust rule, less favorable to settling parties, would look not only at the effects on consumers of their overall agreement, as I am proposing here, but also at the effects of specific *provisions* in the agreement. Under a full-blown “less-restrictive alternative” approach, an agreement would be declared anticompetitive if an alternative agreement could have been fashioned, perhaps by removing or modifying certain provisions in the original agreement, to accomplish the same legitimate ends while leading to higher consumer surplus.

Clearly, taken to an extreme such an approach would *not* in fact respect intellectual property rights. Consider, for example, the case of two holders of complementary, blocking patents who agree to place their patents into a pool and license them jointly at an agreed-upon royalty rate. As shown in Section 6 below, such a pool can easily be in consumers’ interests, in comparison with,

¹¹ Meurer (1989) imposes exogenous antitrust limits on settlements of patent litigation. If these limits on the joint profits from settling are stringent, a mutually beneficial and permissible settlement may not exist.

¹² For a more general discussion of the limits on collaboration, see DOJ/FTC (2000).

say, independent licensing programs by the two firms, since independent licensing runs into the problem of Cournot complements. Just as clearly, however, consumers would be better off (in the short run) if the two firms agreed to a royalty-free cross-license, as each could then compete independently with no licensing cost burden. But compelling such a cross-license over the patent pool could well deprive both firms of a return on their R&D that led to their patent. I do not expect to resolve this debate here, which arises in other areas of antitrust; I merely note that the standard I explore here may not be universally accepted.

□ **The need to assess patent strength.** I would like to highlight one key practical problem with the approach advocated and analyzed here: typically, to compare consumer surplus under a settlement with consumer surplus from ongoing litigation requires an informed judgment as to the strength of the patent(s) at issue. If the patent is very strong, i.e., very likely to be found valid and infringed and difficult to invent around, the challenger is unlikely to offer much independent competition to the patentholder if litigation proceeds. Alternatively, if the patent is very weak, ongoing litigation is likely to lead to greater competition and greater consumer benefits. Except in special cases where inferences about patent strength are possible based on the commercial decisions made by the two parties,¹³ there does not appear to be any way around the need to assess patent strength directly if one is trying to determine whether a settlement benefits consumers.

I treat patent strength as a parameter outside the scope of my economic analysis. This is reasonable, since patent validity and patent breadth ultimately are technical issues. However, by treating patent strength as exogenous I cannot address various tricky and deep issues, such as signalling and self-selection, that arise when the two parties differ in their assessments of patent strength. Asymmetric information, specifically optimism on each side about its chances in court, is generally considered the primary reason why disputes are litigated rather than settled. Empirically, the fact that 95% of filed patent cases are settled (and many, many more patent disputes are settled without any case being filed) suggests that the extent of asymmetric information may often be modest. In any event, the current article is a necessary precursor to an analysis of the antitrust limits on patent settlements that integrates oligopoly theory and competitive effects (treated here) with asymmetric information and signalling (not covered here).¹⁴

□ **Pareto optimality and gains from settlement.** I now prove a general result showing that there are invariably gains from settling a patent dispute, even ignoring the savings associated with reduced litigation costs and uncertainty.

Consider two firms that are actual or potential competitors who are engaged in a patent dispute. Firm 1 we will call the *patentholder* or the *incumbent*. Firm 2 we will call the *alleged infringer*, the *challenger*, or the *entrant*. There are a number of possible outcomes of their patent dispute, if it is litigated to completion; we index these states of the world by $w = 1, \dots, N$. For example, one state might be that the patent is declared invalid. A number of other states might correspond to various levels of cost required for the challenger to invent around the patent. This framework is quite general: a finding of noninfringement would correspond to a zero cost of inventing around the patent; a finding that some of the broader claims in the patent are invalid while narrower claims are valid would correspond to lower, but still positive, costs of inventing around the patent.

Denote the probability of state w by θ_w ; unless otherwise noted, I shall assume that both sides share these probability assessments. As discussed above, in practice the two firms may well differ in their beliefs about the likelihoods of the various states of the world. Indeed, it is relative optimism that tends to lead to litigation rather than settlement. The proper treatment of asymmetric information, including possible updating of beliefs based on settlement offers, is beyond the scope of this article. Relaxing this assumption leads to a rich area for further research.

¹³ For example, the willingness of the challenger to indemnify its customers for infringement, or to compete vigorously during the pendency of the patent litigation, signals that the challenger considers the patent weak. Likewise, a large cash payment from the patentholder to the challenger to exit the market suggests that the patent is vulnerable.

¹⁴ For example, modelling asymmetric information, an important ingredient for models of equilibrium litigation, appears necessary to study how antitrust limits affect the likelihood of settlement versus litigation.

The general antitrust rule explored in this article is that a settlement should be permitted if it leaves consumers at least as well off as they would be, in expected value, from ongoing litigation and resolution of the patent dispute in court. Denote consumer surplus in state w by S_w . This is a reduced form that incorporates whatever duopoly (or oligopoly) solution concept applies in the various states of the world. If resolution of the patent dispute takes time, then S_w includes consumer surplus during the interim period while the dispute is litigated. To the extent that information is revealed gradually, and the firms' behavior can adjust (e.g., to preliminary court rulings), consumer surplus will vary somewhat with the state of the world even before that state is fully revealed.

With these definitions, the expected consumer surplus from ongoing litigation is

$$\bar{S} = \sum_{w=1}^N \theta_w S_w. \quad (1)$$

The profits earned by firm i in state of the world w (in the absence of any agreement between them) are denoted by π_{wi} with joint profits in state w given by $\pi_w = \pi_{w1} + \pi_{w2}$. The expected joint profits from ongoing litigation are

$$\bar{\pi} = \sum_{w=1}^N \theta_w \pi_w. \quad (2)$$

Now suppose that the two parties can write detailed contracts that specify their actions in each state of nature. Denote their specified actions by $x_w = (x_{1w}, x_{2w})$ in state w . Denote by $\pi(x_w)$ the joint profits corresponding to action x_w . Denote by $S(x_w)$ the consumer surplus associated with action x_w . If we require that any negotiated settlement satisfy the surplus constraint (1), the two parties will solve

$$\max_{x_w} \sum_{w=1}^N \theta_w \pi(x_w) \quad \text{subject to} \quad \sum_{w=1}^N \theta_w S(x_w) \geq \bar{S}. \quad (3)$$

Since the same θ_w are used in both the objective function and the constraint, the solution to this maximization necessarily involves solving the subproblem of

$$\max_x \pi(x) \quad \text{subject to} \quad S(x) \geq \bar{S} \quad (4)$$

and then replicating this solution across all states of nature. In other words, the firms should negotiate a set of actions that maximizes their joint profits subject to the surplus constraint and then follow this same set of actions regardless of how the patent litigation would have been resolved. Call this (constrained) optimal set of actions x^* .¹⁵ If the firms acted differently from x^* in any set of states of nature satisfying the consumer surplus constraint, they could do better by taking action x^* instead in each of those states of nature and leaving their actions in all other states of nature unchanged.

The key idea here is that an optimal settlement involves the two parties generating consumer surplus efficiently. It is inefficient for the two parties to vary their actions based on the various possible outcomes of their patent litigation just because of initial uncertainty about the strength of the patentholder's property rights. The implication is that there are always settlements available that raise joint profits and leave consumers whole, so long as the firms would have taken different actions in different states of nature under litigation. This is a very strong and general "gains from settlement" theorem.

¹⁵ I assume that the solution to (4) does not involve randomization. If it does, then x^* should just be reinterpreted as a randomized set of actions.

Proposition 1 (gains from settlement). If the actions of two rival firms engaged in a patent dispute would vary depending upon the outcome of their patent dispute, there is always a settlement available that raises their joint profits without harming consumers as a group. Under the profit-maximizing *ex ante* settlement that leaves consumers whole, the firms' *ex post* actions do not vary with the *ex post* property rights of the patentee.

By showing that there are always gains from settlement I do not mean to suggest that bargaining is always efficient. Nor do I intend to minimize various practical considerations that come into play when antitrust authorities attempt to evaluate the impact on consumers of patent settlements that are actually proposed. Most notably, as noted above, antitrust authorities may have difficulty determining the probabilities of the various states of the world, particularly inasmuch as this requires a technical assessment of one or more patents and their various claims.

Proposition 1 has an immediate implication for the value of uncertain patent rights when settlement is possible. To illustrate what is a more general point, suppose that there are only two outcomes of litigation: the patentholder "wins" or "loses," with the former state yielding higher profits for the patentholder and lower profits for the challenger. The probability that the patentholder "wins" will be called "patent strength." Now, we know (by definition) that the patentholder's payoff from litigation is linear in patent strength. We also know that the gains from the optimal permissible settlement are zero if the patent strength is either zero or unity, but positive for intermediate values of patent strength. Assuming that the patentholder receives some positive fraction of the gains from the optimal settlement, the patentholder's payoff from settling cannot be linear in patent strength, and it exceeds the linear interpolation of value in between a worthless patent and an ironclad patent.

Proposition 2 (value of patent is not linear in patent strength). Suppose that there are only two outcomes of the patent litigation. Call the probability of the outcome more favorable to the patentholder the "patent strength," θ . The patentholder's payoff under the optimal settlement is not linear in patent strength. A patent with strength θ is worth *more* than θ times that of a patent with strength of unity.

□ **Benefits of optimal settlement: price and quantity examples.** We can illustrate these gains from settlement more concretely when the firms are picking prices and quantities. In fact, for a given set of products, we know a great deal about the solution to equation (4), since it is the dual to the standard Ramsey pricing problem.

Suppose the incumbent firm produces a single product, the demand for which is denoted by $x = D(p)$, where p is price and x is output. Call the monopoly price p_M with corresponding consumer surplus S_M . To keep things simple, suppose that there are only two states of the world: either the incumbent wins the patent case and has a monopoly, or the challenger wins the patent case and a duopoly results. Let us suppose that the incumbent will win the patent case with probability θ , which we call *patent strength*. If the entrant wins the patent litigation, the duopoly price will be p_D , with corresponding consumer surplus S_D and joint profits π_D .

The expected consumer surplus if the parties do not settle is given by

$$\bar{S} = \theta S_M + (1 - \theta)S_D.$$

Call \bar{p} the price that generates surplus level \bar{S} . The optimal settlement subject to the consumer surplus constraint involves a price of \bar{p} in all states of the world. In other words, we must have

$$S(\bar{p}) = \theta S(p_M) + (1 - \theta)S(p_D). \quad (5)$$

Since $S'(p) = -D(p)$, we know that $S''(p) = -D'(p) > 0$, so the surplus function is convex in price. Therefore, equation (5) implies that $\bar{p} < \theta p_M + (1 - \theta)p_D$. In other words, consumers must receive a lower price on average from settlement than they would get from litigation. This result follows from the fact that consumer benefits grow disproportionately as price falls.

Proposition 3 (settlement must lower average price). If the two parties to a patent dispute sell a single homogeneous product, the price that prevails under their settlement must be *less* than the average price that would result from litigation, if the settlement is to be in consumers' interests.

We can measure the gains from settlement if we are prepared to make some stronger assumptions about costs, demand, and duopoly behavior. Suppose that demand is linear, $D(p) = A - p$, and marginal costs are a constant, c . The monopoly price is $p_M = (A + c)/2$.

Bertrand competition. If the firms are Bertrand competitors, then $p_D = c$. The consumer surplus function is $S(p) = (A - p)^2/2$. Consumer surplus under Bertrand pricing is $S_D = (A - c)^2/2$. Calculating consumer surplus under monopoly and under duopoly, the price \bar{p} must satisfy

$$\bar{p} = A - \frac{A - c}{2} \sqrt{4 - 3\theta}.$$

By construction, consumers are just as well off under the settlement as from litigation. By how much do profits, and thus total welfare, rise under settlement? With litigation, total expected profits are $\theta\pi_M = \theta(A - c)^2/4$, since there are no profits under duopoly. With settlement, profits are

$$\pi_S = (\bar{p} - c)D(\bar{p}) = \frac{(A - c)^2}{4} (2\sqrt{4 - 3\theta} + 3\theta - 4).$$

The extra profits (welfare) made possible by settlement as a fraction of the monopoly profits are given by

$$G \equiv \frac{\pi_S - \theta\pi_M}{\pi_M} = 2\sqrt{4 - 3\theta} + (3\theta - 4).$$

Naturally, these gains from settlement are zero if there is no uncertainty, i.e., if θ is zero or one. But for intermediate levels of θ , this measure of the gains from trade is nontrivial. For example, if $\theta = 1/2$, G equals .16. In other words, one-sixth of monopoly profits can be captured as extra profits from settlement without harming consumers.

If we assume that the patentholder and the challenger split the gains from trade associated with the settlement, the overall payoff to the patentholder is given by $\theta\pi_M + G\pi_M/2$, which equals $\pi_M(2\theta - 2 + \sqrt{4 - 3\theta})$. Note that this expression is concave in the strength of the patent, so the patentholder's payoff is concave in patent strength. A patent with strength one-half is worth *more* than half as much as an ironclad patent, as we know more generally from Proposition 2.

How would the optimal agreement be implemented? Since we are assuming Bertrand pricing competition with homogeneous products, the desired price \bar{p} can be achieved if the challenger pays a royalty equal to

$$\bar{r} = \bar{p} - c = \frac{A - c}{2} (2 - \sqrt{4 - 3\theta}).$$

Compare this with the royalty that would support the monopoly outcome, $r_M = p_M - c = (A - c)/2$. The "allowed" royalty as a fraction of the monopoly royalty is given by $\bar{r}/r_M = 2 - \sqrt{4 - 3\theta}$, which is increasing and convex in the strength of the patent, θ .

Given this royalty, we can ask next whether both firms would find the settlement individually rational; put differently, is a fixed payment required to facilitate the optimal agreement, and if so, in which direction does the payment run? This question is easy to answer given the rather stark nature of Bertrand competition. Without a settlement, the entrant would earn no money, even if it wins the patent suit; for the same reason, the entrant earns no money under the license either, since the resulting price just equals the entrant's costs: $\bar{p} = c + \bar{r}$. So the entrant is indifferent to the agreement in the absence of any fixed fees. We know that joint profits are higher under the agreement than under litigation, by construction, so in the absence of any fixed fees, the incumbent captures all of the gains from trade. If we think in terms of Nash bargaining, for example, we would expect the incumbent and entrant to split these gains from trade, which would imply a fixed payment running from the incumbent to the entrant. For the reasons discussed above, such fixed

payments can be abused if antitrust enforcement agencies lack sufficient information to check directly that consumers are not harmed by the settlement in comparison with ongoing litigation, which may be difficult if antitrust enforcers are unable to assess patent strength.

Cournot competition. We turn now to the companion case in which the firms are Cournot competitors. This case differs from the Bertrand case in that consumers benefit less from the possibility of successful entry, and the entrant makes positive profits if it wins the patent suit. I continue to assume that the entrant is equally efficient to the incumbent (and marginal costs are constant), thus abstracting away from issues of how to achieve production efficiency in the context of a settlement.

The Cournot duopoly price is given by $p_C = (A + 2c)/3$. Performing the same type of calculation as was done above in the Bertrand case, the settlement price under Cournot competition is

$$\bar{p}_C = A - \frac{A - c}{6} \sqrt{16 - 7\theta}.$$

This outcome could be supported through a settlement in which the entrant pays per-unit royalties sufficient that the resulting Cournot equilibrium gives this price. If the royalty rate is r , the resulting Cournot equilibrium price is $(A + 2c + r)/3$. The royalty rate is

$$\bar{r} = \frac{A - c}{2} \left(4 - \sqrt{16 - 7\theta} \right).$$

Again we can ask whether the two firms would find this royalty individually rational, or more generally what range of fixed fees would be required to make this agreement mutually acceptable. I focus on the challenger's profits, to see whether a payment from the patentholder to the challenger is needed to make settlement at the optimal royalty rate acceptable to the challenger. With no settlement, the challenger earns the Cournot duopoly profits if it wins the patent suit. Since the profits of each Cournot duopolist are given by $\pi_{1C} = \pi_{2C} = (A - c)^2/9$, the challenger's expected profits from litigating are given by $(1 - \theta)(A - c)^2/9$. The challenger's profits from agreeing to pay royalty rate \bar{r} can be shown to be

$$\pi_{2S} = \frac{(A - c)^2}{9} \left(\sqrt{16 - 7\theta} - 3 \right)^2.$$

The challenger's gains from settlement are given by $\pi_{2S} - (1 - \theta)\pi_{2C}$. Substituting and simplifying, we get the gains from settlement to the challenger as

$$G_2 = 6\pi_{2C} \left(4 - \theta - \sqrt{16 - 7\theta} \right).$$

Of course, there are no gains to the challenger if there is no uncertainty, i.e., if $\theta = 0$ or $\theta = 1$, because the settlement just replicates the certain outcome in these polar cases. For all intermediate values of θ , the challenger is strictly better off under the settlement in the absence of a fixed payment running in either direction. This is important, because it implies that no fixed payment from the patentee to the challenger is required to implement the optimal settlement. Since it is exactly such payments that raise antitrust concerns, this is an encouraging result: the optimal royalty should be acceptable to the challenger without the use of a fixed payment.

Rationalization of production. In both of these cases I assumed that the two firms had constant and equal marginal costs of production, so the mix of production between the two firms was irrelevant from the perspective of production efficiency. What about the case in which the two firms have, in general, different costs as a function of output, and in which marginal costs are not constant?

If there are economies of scale in production, the optimal settlement involves one firm shutting down. But for this settlement to meet the consumer-surplus constraint, the other firm will

typically have to produce more than the monopoly output. Assuming that the antitrust authorities do not want to engage in direct regulation of the remaining firm, some type of supply agreement is needed to ensure that output is sufficiently large that consumers are not disadvantaged. Such supply agreements can work well in theory, enabling competition while taking advantage of economies of scale, but they present a range of problems in practice, especially if costs are uncertain and likely to change over time.

Even in the absence of strong scale economies, efficiency can still be promoted by rationalizing output across the two firms, especially if one firm is considerably more efficient than the other. Consider, for example, the case in which the patentholder is primarily a research outfit that can engage in some manufacturing, whereas the challenger has a large, established manufacturing presence (for related, noninfringing products, say) and is far more efficient at production. The obvious solution here is for the challenger to obtain a license to the patent. But some degree of “inefficient” production by the patentholder may be needed to protect consumers from paying monopoly prices set by the licensee. Although structuring the license with a higher fixed fee and lower running royalties may ensure that the resulting price is low enough to satisfy the consumer-surplus constraint, even with no running royalties, consumers may not be made whole unless the (inefficient) patentholder also engages in some production.

Differentiated products. In many cases, the challenger offers a product that is distinct from the offerings of the patentholder. Clearly, efficient provision of consumer surplus may well involve preserving both products under the terms of the settlement. In general, if efficiency requires both products to be produced, the two parties can use per-unit royalties and fixed-fee payments to move around their reaction curves and thus induce a Bertrand pricing equilibrium that replicates the optimal settlement. However, the resulting contract may require per-unit royalties running from the patentholder to the challenger, which tends to raise its own antitrust concerns. (In this respect, settling litigation in which each party is asserting patents against the other can provide more flexibility to the settling parties.) If such payments are prohibited, the primary tool remaining to influence the Bertrand equilibrium is the royalty rate pay by the challenger to the patentholder. But the fully optimal settlement may not be obtainable as an induced Bertrand equilibrium. And an outright merger might well not satisfy the consumer-surplus constraint, unless the merger would generate its own efficiencies.

4. Mergers and the patent competition index

■ I turn now from a general discussion of the benefits from settlement to an analysis of specific types of settlements, which occupies the remainder of the article. I begin in this section with the most inclusive form of settlement, namely an outright merger between the two parties to the patent dispute. Two real-world examples illustrate the types of mergers studied here. The first example is the acquisition by Boston Scientific of Cardiovascular Imaging Systems (CVIS). Boston Scientific, a large company that makes a range of medical equipment, was producing and selling certain imaging catheters that CVIS, a small company, claimed infringed its patents. To settle the dispute, Boston Scientific acquired CVIS. As a condition for approving the merger, the FTC required Boston Scientific to license the CVIS patents.¹⁶ A second, more recent example is the acquisition by Gemstar of TV Guide. Gemstar asserted that TV Guide’s “interactive program guides,” basically on-screen interactive information about television program listings, infringed

¹⁶ See “Boston Scientific to Help Launch New Maker of Cardiac Catheter, to Settle FTC Charges Over CVIS, Scimed Acquisitions,” February 24, 1995 (available at www.ftc.gov/opa/predawn/F95/boston.scient.htm). See also www.ftc.gov/opa/1995/9505/boscvis.htm, announcing the license with Hewlett-Packard. But note also that the DOJ later sued Boston Scientific for failing to license all the patents and provide all the interface information to Hewlett-Packard as required under its agreement with the FTC; see www.ftc.gov/os/2000/10/bsscnp.htm.

Gemstar's patents. After years of litigating and competing against Gemstar, TV Guide agreed to be acquired. The DOJ did not challenge this merger.¹⁷

□ **The patent competition index.** Suppose that the two firms engaged in the patent dispute agree to merge, completely eliminating competition. Consumer surplus under a merger, assuming no efficiencies, is just the level from monopoly, S_M . Expected consumer surplus from ongoing litigation has already been defined as \bar{S} , so the merger will cause a loss of consumer surplus of $\bar{S} - S_M$.

Compare this to the loss of consumer surplus that would result from a merger between these two parties if the challenger were known *not* to be infringing the patent, i.e., if the challenger were not under the cloud of possible infringement. Call consumer surplus in the state of nature in which the patent is held to be valid but not infringed, S_D , reflecting duopoly competition between the two firms. Under these conditions, the loss of consumer surplus from a merger would be $S_D - S_M$. I call the ratio of these two measures of harm to consumers the "patent competition index" (PCI), which is meant to calibrate the degree of competition between the two firms in comparison with a conventional merger without the patent component. The PCI is defined as

$$PCI = \frac{\bar{S} - S_M}{S_D - S_M}. \quad (6)$$

If the patent is valid but the challenger would surely be found not to have infringed, then $\bar{S} = S_D$ and the PCI is one. If the patent is valid and totally blocks the challenger, then $\bar{S} = S_M$ and the PCI is zero. If the patent would certainly be found *invalid*, enabling more entry, then $\bar{S} > S_D$ and the PCI is *greater* than unity.

Besides giving a sense of how much competition may be lost as a result of a proposed settlement involving a full merger, the PCI also tells us how large the *efficiencies* associated with the merger must be, relative to the efficiencies that would be required to justify a merger *without* the patent overlay. Call the extra consumer benefits flowing from efficiencies associated with the merger E ; typically these efficiencies result from cost savings that lead to lower prices. Call the minimum such benefits that would make a conventional merger attractive to consumers E_D : $S_M + E_D = S_D$. Likewise, call the minimum such benefits that would make the proposed settlement attractive to consumers \bar{E}_D : $S_M + \bar{E}_D = S_D$.

Now we can take the ratio of these two efficiency measures, \bar{E}/E_D , to compare the efficiency hurdle that the settlement must clear, versus a conventional merger, to be in the interests of consumers. Since $\bar{E} = \bar{S} - S_M$ and $E_D = S_D - S_M$, this ratio is precisely the patent competition index: $PCI = \bar{E}/E_D$. In other words, the PCI can also be interpreted as the magnitude of efficiencies required to make the settlement procompetitive, calibrated to the efficiencies that would be required of a conventional merger in which the challenger does not face a patent "cloud" over its ability to compete. I record this simple arithmetic in the form of a proposition.

Proposition 4 (efficiencies necessary for consumers to benefit from a merger settlement). Suppose that a conventional merger between two firms would need to generate efficiencies causing extra consumer surplus of E to benefit consumers. Then a patent settlement between these two firms must generate efficiencies causing extra consumer surplus of $PCI * E$ to benefit consumers, where PCI is the patent competition index, from (6).

So long as there is *some* degree of competition between the merging parties, either prior to the settlement or prospectively in the absence of a settlement, a merger cannot benefit consumers unless it generates some efficiencies. In other words, the fact that there is a patent cloud of some type hanging over the challenger reduces, but does not eliminate, the need for merger efficiencies to offset anticompetitive effects.

¹⁷ For more information about Gemstar and its business plans, see "Henry Yuen Wants to Be Your TV Guide," *Fortune*, August 14, 2000, pp. 200–208 and "TV Guy: Will Gemstar-TV Guide's Henry Yuen Take Control of Your TV Set?" *Business Week*, March 12, 2001, pp. 66–76.

An extension to this article would be to integrate this analysis with traditional structural merger analysis based on measures of market concentration. The safe-harbor provisions in merger enforcement can reasonably be viewed as indicating the magnitude of efficiencies that are credited to merging parties as a matter of course. Presumably, a somewhat greater increase in concentration would be permitted if the acquired firm is operating under a patent cloud. One could, in principle, calculate and apply new safe-harbor concentration measures that permit greater increases in concentration when the PCI is lower.

□ **Calculating the patent competition index.** I now derive the PCI in several specific cases.

Immediate resolution of patent litigation. Suppose that the patent litigation can be resolved instantly and there are only two possible outcomes. With probability θ , the patentholder wins, in which case the challenger is totally blocked from the market. With probability $1 - \theta$, the challenger wins, in which case the challenger is found not to infringe the patent (but the patent is still valid). In this simple case, we have $\bar{S} = \theta S_M + (1 - \theta)S_D$, and direct calculations show that $PCI = 1 - \theta$. In other words, the PCI directly tracks the weakness of the patent.

Consider a variation on this example in which the challenger's winning means that the patent is held invalid, thus leading to more competition than would arise under duopoly. Calling consumer surplus with the invalid patent S_I , we have $\bar{S} = \theta S_M + (1 - \theta)S_I$. Suppose that patent invalidity generates a multiple k incremental surplus to consumers as does duopoly (in comparison with monopoly): $S_I - S_M = k(S_D - S_M)$, where $k > 1$. Then we have $PCI = k(1 - \theta)$. If the patent is not too strong, and if invalidity generates significantly more consumer surplus than mere noninfringement, i.e., if other firms can take advantage of the invalidity ruling, then a merger causing the patent suit to be dropped may face a *higher* hurdle in terms of efficiency than a plain vanilla merger involving the same companies in which the challenger is known not to infringe.¹⁸

Next, consider the case in which losing the patent litigation does not completely exclude the challenger from the market but simply imposes a cost penalty on the challenger as a result of the need to invent around the patent.¹⁹ We can capture this in reduced form through the consumer-surplus function $S(b)$, this being consumer surplus if the challenger's cost of inventing around the patent is b . In terms of our earlier notation, $S(0) = S_D$ and $S(b) = S_M$ for large values of b . Consumer surplus under litigation is thus $\bar{S} = \theta S(b) + (1 - \theta)S_D$.²⁰ As above, calibrate $S(b)$ according to $S(b) - S_M = g(S_D - S_M)$, where now $g < 1$. In this case, we get $PCI = \theta g + (1 - \theta)$. Note that in this case the PCI is no less than g , even if the patent is ironclad, and of course it equals unity if the patent is very weak ($\theta \approx 0$).

We can always convert the PCI to specific efficiencies that the merger must enable, given enough structure in terms of cost and demand functions and oligopoly behavior. To illustrate, consider again the case with linear demand, $D(p) = A - p$, constant marginal cost for each firm of c , and Bertrand competition. The duopoly outcome involves price at marginal cost, in comparison with the monopoly price of $(A + c)/2$, so $E_D = S_D - S_M = 3(A - c)^2/8$. Returning to the basic case where $PCI = (1 - \theta)$, we know that the settlement must generate efficiencies of $\bar{E} = (1 - \theta)E_D = 3(1 - \theta)(A - c)^2/8$. If the settlement permits a reduction of marginal costs of Δ , then the resulting price is $(A + c - \Delta)/2$ and the associated efficiencies are $\bar{E} = \Delta[2(A - c) + \Delta]/8$. To meet the required level of efficiencies, the per-unit cost savings must satisfy $[\Delta/(A - c)][2 + \Delta/(A - c)] = 3(1 - \theta)$. Since consumer benefits are convex in the cost saving, the per-unit cost saving necessary if $PCI = 1/2$ is *more* than half of the cost saving that would be required to justify a conventional merger.

¹⁸ A complete analysis would account for the probability that another challenger would continue ahead with litigation and prove the patent invalid. Of course, the holder of a weak patent might settle with a series of challengers to avoid just this outcome.

¹⁹ Inventing around could also take time. See the next subsection for a discussion of cases in which competition varies over time.

²⁰ I have returned to the assumption that a victory by the challenger means that the challenger is held not to be infringing, but leaves the patent intact. I do not develop here the general case in which there are many possible outcomes of the patent litigation, involving different values of b and perhaps patent invalidity as well.

Delayed resolution of patent litigation with no interim competition. What about the realistic case in which the patent litigation takes some time? More precisely, what about the case in which the settlement occurs well before the patent litigation would likely be resolved in the courts?

In this case, we must explicitly keep track of the passage of time and recognize that competitive conditions can change over time. Call the date of the settlement time $t = 0$, and the expiration of the patent date $t = 1$. Define the interim period to be the period $[0, T]$ until the patent litigation would be resolved in the courts.²¹ To keep things simple, I assume that demand conditions, and the presence or absence of other firms, do not vary over the time interval $[0, 1]$, and I will assume a zero interest rate. Relaxing these assumptions would be straightforward but add considerable complexity to the resulting expressions.

To focus on the timing and the possible differences between the interim period $[0, T]$ and the subsequent time period $[T, 1]$, let me return to the case in which a victory by the challenger would lead to a standard duopoly situation (i.e., would mean a finding of patent validity but noninfringement). If the challenger would choose *not* to compete during the interim period,²² then consumer surplus under litigation would be $\bar{S} = T S_M + (1 - T)(\theta S_M + (1 - \theta)S_D)$. In this case, the PCI is $(1 - T)(1 - \theta)$.²³ We get the same result if the challenger is not allowed to compete because the court has issued a preliminary injunction ordering the challenger to cease its possible infringement, as tends to occur automatically in pharmaceutical cases under the provisions of the Hatch-Waxman Act.

Delayed resolution of patent litigation with interim competition. A more interesting situation arises if the challenger would choose to compete during the pendency of the patent litigation. For example, in the Gemstar/TV Guide merger, TV Guide continued to offer its interactive program guides to cable companies, along with patent indemnification, while under the threat of Gemstar's patent suit. If the actual level of competition and consumer surplus during the pendency of the patent suit can be observed, the PCI is relatively easy to calculate directly.

To study how competition is likely to play out under the shadow of patent liability, we must specify the liability to which the challenger is potentially exposed by infringing the patent. Clearly, such liability will tend to impede the challenger's ability to compete effectively. Call the consumer surplus resulting from competition between the incumbent and the challenger facing potential liability for infringing S_L . Then we have $\bar{S} = T S_L + (1 - T)(\theta S_M + (1 - \theta)S_D)$. As above, calibrate the consumer benefits associated with the "impeded" duopoly, S_L , according to $S_L - S_M = PCI_L(S_D - S_M)$, where $PCI_L < 1$ and where PCI_L is smaller, the stronger is the patent and the more favorable to the patentholder are the damages rules in patent infringement cases. With these definitions, the PCI turns out to be $T * PCI_L + (1 - T)(1 - \theta)$. Now consumers get some benefits even during the interim period, depending upon how strongly the challenger will compete under the shadow of possible liability.

I now explore how vigorously the challenger would compete if it decides not to withdraw and wait, i.e., what determines PCI_L . In one central case, I establish that $PCI_L > 1 - \theta$, so consumer benefits during the interim period are even larger than they will be (on average) after the patent dispute is resolved.

Of course, to analyze the challenger's behavior, we need to specify the legal rules governing the calculation of damages in the event the patent is subsequently found to be valid and infringed. I shall work with a legal rule that awards lost profits to the patentee.²⁴ Denote by π_M the patentee's

²¹ In practice, this period can be broken down further. For example, there may be preliminary rulings that cause the parties to update their beliefs about the ultimate outcome of the resolution. There may also be a ruling in the lower court that is subsequently appealed.

²² For example, makers of generic drugs challenged by patentholders typically do not enter until after receiving a favorable ruling on infringement, because their damage exposure—based on the profits lost by the maker of the branded drug—tend to be far higher than the profits that the generic maker can earn through entry.

²³ If we allowed for a positive interest rate, then the PCI would be somewhat lower, reflecting the fact that the period of competition comes only after the period of monopoly.

²⁴ I am assuming here that damages are equal to lost profits, not a multiple of lost profits. In the case of willful infringement, damages can equal three times the lost profits.

profits in the absence of competition from the challenger. If the patentee's (flow) profits during the interim period are π_1 , then the (flow) damages due are $\pi_M - \pi_1$.

To illustrate these points, I now derive the Bertrand and Cournot equilibria when the challenger faces potential liability for infringing. I believe that this analysis is of independent interest, as it characterizes price and quantity competition in the face of uncertainty about liability for infringement.

□ **Bertrand pricing game between patentholder and challenger.** Consider the case of homogeneous goods and Bertrand competition, perhaps better thought of as bidding competition to serve the market (or a single customer). As is often the case, Bertrand equilibrium with homogeneous products is quite a fragile concept. Denote by $\pi_i(p)$ the profits to firm i from winning at price p . If the two firms are equally efficient, $\pi_1(p) = \pi_2(p)$.

I now show that an equally efficient challenger will simply not enter the market under these conditions. Indeed, the entrant must be much more efficient than the incumbent to make entry profitable. As a consequence, even a weak patent can yield monopoly profits. This is a new twist on the well-known fact that an entrant with even a very small fixed cost of entry will not enter just to compete on price against an equally efficient rival, and strong postentry competition tends to deter entry. Here, there is effectively a fixed cost of entry, namely the expected liability costs of participating in the market.

If the challenger enters the market and wins at price p , then the challenger's payoff is given by $\pi(p) - \theta\pi_M$, where π_M are monopoly profits. If the challenger enters the market, bids p , and loses to the incumbent who sets the lower price q , then the challenger's payoff is given by $-\theta(\pi_M - \pi(q))$.²⁵ Since the expected liability costs, $\theta\pi_M$, are independent of the entrant's bid, they do not influence bidding. Since $\pi(p) > \theta\pi(q)$ for values of p near to q , the challenger is always better off undercutting the incumbent rather than losing, for any value of q that exceeds marginal cost.²⁶ So, given the entrant's presence in the market, the only equilibrium is for both firms to bid down to marginal costs. Anticipating this outcome, an equally efficient challenger would not choose to enter the market in the first place.

If the challenger enjoys a sufficiently large cost advantage, entry may be profitable. In this case, the incumbent will bid down to the lowest cost at which it can break even, so the price satisfies $\pi_1(p) = 0$. Call this price p_1^* . The entrant captures profits of $\pi_2(p_1^*)$ but incurs expected liability costs of $\theta\pi_M$. The entrant must be sufficiently more efficient than the incumbent that the profits it can earn at the incumbent's break-even price cover these expected liability costs.

This analysis gives quite a strong, even striking result: the patentholder can capture the full monopoly profits, even if the patent is very weak. Here, a key question is whether the potential entrant can bid for business without exposing itself to liability in the event that the incumbent meets or beats the entrant's prices and thus wins the bidding. But this seems even harder to do than the corresponding strategy in the Bertrand pricing game with a fixed entry cost, where the entrant may be able to adopt a bidding strategy that allows it to avoid incurring the fixed entry costs if it loses the bidding. Here, if the entrant's bid clearly induces a response from the incumbent, liability for infringement will be hard to avoid.

Proposition 5 (pricing competition in the shadow of liability). Suppose the patent liability rule awards lost profits, and the two equally efficient firms compete as Bertrand rivals. Then the challenger will not enter the market, and the patentholder will capture the monopoly profits, even if the patent is arbitrarily weak (θ very small but still positive).

In this situation, since $PCI_L = 0$, consumers receive no benefits from interim competition and settlements look relatively attractive.

²⁵ Here I am assuming damages equal to actual lost profits. The result would not change if damages were equal to the minimum possible lost profits given the entrant's price p , namely $\pi_M - \pi(p)$.

²⁶ With the damages rules less generous to the patentholder, where damages equal $\pi_M - \pi(p)$, the challenger is still better off undercutting the incumbent than losing the bidding.

□ **Output game between patentholder and challenger.** In the corresponding Cournot game, the firms simultaneously set outputs, resulting in a price (and profits) during the interim period. Then, if the challenger is found to have infringed, the challenger owes damages to the patentholder equal to the difference between monopoly profits and the profits actually earned by the patentholder. Calculations available from the author establish the following proposition.

Proposition 6 (competition in the shadow of liability). Suppose the patent liability rule awards lost profits, and the two equally efficient firms with constant marginal costs compete as Cournot rivals facing linear demand. Then competition prior to the determination of infringement yields *greater* benefits to consumers than will arise on average after the determination of infringement.

For example, when $\theta = 1/2$, we get $PCI = 0.57$. A patent with a 50% chance of winning generates more than 50% of the benefits to consumers from full-fledged duopoly. The extra benefits consumers get on average from interim competition, in comparison with subsequent competition, are greatest when infringement is very much in doubt, i.e., when θ is in the neighborhood of one-half.

Proposition 6 tells us that the overall PCI, calculated above to be $T * PCI_L + (1 - T)(1 - \theta)$, is increasing in the time it takes for the patent litigation to be resolved, T , since the $PCI_L > 1 - \theta$. This is very different from the Bertrand pricing case, in which $PCI_L = 0$, so consumers benefit from an earlier resolution of the patent litigation.

5. Settlements specifying a date of entry

■ One way in which a patentholder and a challenger engaged in a patent dispute can settle their dispute is to negotiate a date at which the challenger can enter the market. Settlements of this type between incumbent patentholders selling branded drugs and potential entrants offering generic versions of these same drugs have been intensively scrutinized by the Federal Trade Commission.²⁷ These settlements have the unique feature that, under the Hatch-Waxman Act, settling with the first generic challenger can protect an incumbent drug company from other generic challengers as well for a specified period of time, because the first generic challenger may enjoy exclusive rights for a period of time. The FTC has expressed great skepticism about settlements in which the incumbent makes payments to the challenger as part of a negotiated entry date. The FTC recently released a broader study of competition between generic and branded drugs; see FTC (2002).

To keep things simple, let us suppose again that there are only two outcomes of the patent litigation: the patent is valid and infringed with probability θ , or the patent is valid but not infringed with probability $1 - \theta$. As above, call the date of the settlement time $t = 0$, the expiration of the patent date $t = 1$, and denote by T the date at which the entrant could enter if successful in the patent litigation. Restricting attention to settlements simply involving an entry date (but no royalties), it is easy to see which settlements leave consumers whole.

Absent any settlement, expected consumer surplus is $\bar{S} = T S_M + (1 - T)(\theta S_M + (1 - \theta) S_D)$, where S_D again is consumer surplus under duopoly.²⁸ Consumers benefit from a negotiated entry date t if and only if $t < T + \theta(1 - T)$. Assuming that duopoly profits are less than monopoly profits, however, there is little reason to expect the firms to find such entry dates mutually attractive. If the firms are risk neutral, a reasonable assumption for large, publicly traded firms if not individual managers at those firms, and ignoring litigation costs, there are simply no gains from settlement under these conditions when the only available instrument is the entry date. Factoring

²⁷ In March 2000, the FTC challenged settlements between Abbott and Geneva (www.ftc.gov/os/2000/03/abbottcmp.htm) and between Hoechst Marion Roussel (now Aventis) and Andrx (www.ftc.gov/os/2000/03/hoechstandrxcplmt.htm). In April 2001, the FTC challenged settlements between Schering-Plough and Upsher-Smith and American Home Products (www.ftc.gov/os/2001/04/scheringpart3cmp.pdf). In July 2002, an administrative law judge dismissed the FTC's complaint against Schering-Plough and Upsher-Smith.

²⁸ I assume the challenger will not enter during the litigation phase; this assumption fits the facts well in the case of pharmaceutical settlements.

in litigation costs, there will typically be a range of mutually acceptable dates in the neighborhood of $T + \theta(1 - T)$. To the extent that the patentholder believes the patent is stronger than does the challenger, settlement is made even more difficult, as the patentholder will insist on a later entry date and the challenger will not agree to wait so long to enter.

In this simple model, a naked cash payment flowing from the patentholder to the challenger (in excess of avoided litigation costs) is a clear signal that the settlement is likely to be anticompetitive. Presumably, the patentholder would not pay more than avoided litigation costs unless it believed that it was buying later entry than it expects to face through the litigation alternative. For this reason, the FTC has a sound basis for its skepticism about “reverse cash payments” from the patentholder to the challenger. This is not to say that such payments are necessarily anticompetitive if other factors are brought into the analysis, such as risk aversion and asymmetric information about market conditions, as “reverse cash payments” may be important in more complex settings for successful settlement.²⁹

Some of the settlements challenged by the FTC also involved the transfer of noncash assets from the challenger to the patentholder. These side deals pose some additional, interesting questions. If the noncash assets have a well-defined market value, then they can be treated much like cash. The proper test then involves comparing the *net* payment from the patentholder to the challenger to avoided litigation costs. A large net payment running from the patentholder to the challenger is inherently suspicious. A net payment running from the challenger to the patentholder should be quite welcome to antitrust officials, although it raises a tricky question: presumably, the challenger is paying for earlier entry than would occur (on average) from litigation, but if this is the case, why is it mutually profitable for the firms to agree to earlier entry if entry dissipates joint profits? One benign answer is that joint profits rise with entry because the challenger brings complementary assets to the market or because the entrant earns more profits by taking business from *other* firms than it reduces the incumbent’s profits.

If the noncash assets received by the patentholder have no well-defined market value, it becomes necessary to estimate their value to the patentholder. If the patentholder is receiving more in value, as seen through its own eyes, than it is giving up, the patentholder is making no net payment to the challenger, and there is no basis for presuming that the settlement delays entry in comparison with litigation.

6. Patent pools

■ Patent pools are another form of settling patent disputes.³⁰ Famous patent pools over the years include the pool of sewing machine patents in the mid-19th century and the World War I pool of airplane patents.

The recent pool involving patents for laser eye surgery, which was forced to dissolve by the FTC, illustrates the general pattern. Each of the two companies forming the pool, Summit Technologies and Visx, claimed that it held patents essential to manufacture machines that perform laser eye surgery. Each sued the other for infringement. To settle their dispute, Summit and Visx placed their relevant patents into a pool, called Pillar Point Partners. The pool then licensed these patents as a package back to Summit and Visx as well as to third parties seeking to sell laser eye surgery machines. The pool charged \$250 for each procedure performed using licensed machines. The FTC charged that “in the absence of the [pool], Visx and Summit could have and would have competed with one another in the sale or lease of . . . equipment by using their respective patents, licensing them, or both.”³¹ In contrast to this case, the DOJ issued business review letters

²⁹ In the FTC’s case against Schering-Plough, Robert Willig testified that such payments may be necessary as part of procompetitive settlements in the presence of risk aversion and asymmetric information.

³⁰ The more common settlement form of cross-licenses could be analyzed along very similar lines.

³¹ See paragraph 8 in the FTC’s March 1998 complaint, available at www.ftc.gov/os/1998/9803/summit.cmp.htm. Summit and Visx later agreed to dissolve their patent pool. See www.ftc.gov/os/1998/9808/d09286ana.htm.

approving the pooling of patents necessary to make products complying with the MPEG standard and two standards for DVDs.³²

Patent pools can easily be studied using the framework developed above. Suppose that firm 1 and firm 2 each holds a patent that it asserts is essential to the manufacture of a given product. If the firms are themselves manufacturers, as in the Summit/Visx case, the two firms will be suing each other. If the firms seek to license their patents to manufacturers, then both firms will be asserting their patents against these manufacturers. Either way, in the absence of some agreement, we have a classic and inefficient situation involving Cournot complements if each firm sets per-unit royalties, as are commonly used.

In the absence of a settlement, suppose that litigation will be resolved, as above, at date T . Prior to that time, let us suppose that the firms will license their patents independently using per-unit royalties. This is the standard Cournot complements problem, yielding consumer surplus of S_C .

After the litigation is resolved, one or both patents may be declared invalid. If each patent has strength θ , then we have effectively three possible outcomes. With probability θ^2 , both patents are valid and we presume that the companies would then be allowed to form a pool, leading to the monopoly outcome, with surplus S_M . With probability $2\theta(1 - \theta)$, precisely one patent is valid, in which case we again get the monopoly outcome. With probability $(1 - \theta)^2$, both patents are invalid, in which case we get the perfectly competitive outcome (either via Bertrand pricing between the two firms or through free entry of manufacturers), with consumer surplus S_I . We know in general that $S_I > S_M > S_C$.

Putting all of this together, consumer surplus from ongoing litigation is given by

$$\bar{S} = T * S_C + (1 - T)(\theta^2 + 2\theta(1 - \theta))S_M + (1 - T)(1 - \theta)^2 S_I.$$

The settlement to be evaluated is the formation of a patent pool. Under the terms of the pool, the two patents are licensed as a package for a specified per-unit royalty rate, r , with the license fees then divided up between the two firms. With the patent pool, consumers receive surplus of $S_P(r)$. The pool is beneficial for consumers if $S_P(r) > \bar{S}$.

One question to ask is whether a pool that replicates the monopoly outcome is beneficial to consumers. This is certainly possible, since $S_M > S_C$. To focus on this question, suppose that monopoly provides a fraction μ_M of the incremental benefits to consumers over Cournot complements, in comparison with perfect competition. In other words, $\mu_M \equiv (S_M - S_C)/(S_I - S_C)$. With this definition, it is not hard to show that the most profitable pool, i.e., the pool replicating the monopoly outcome, is beneficial to consumers if and only if

$$\mu_M > \frac{(1 - \theta)^2(1 - T)}{T + (1 - \theta)^2(1 - T)}. \quad (7)$$

Since competition can only arise after date T , and even then only occurs with probability $(1 - \theta)^2$, the monopoly pool can easily raise consumer surplus, if monopoly offers even a modest increase in surplus relative to Cournot complements.

More generally, we can define $\mu(r) \equiv [S(r) - S_C]/(S_I - S_C)$. Then the maximum acceptable royalty rate is found by plugging $\mu(r)$ into equation (7). Since $\mu(0) = 1$, we know that a pool with a sufficiently low royalty rate will always be beneficial to consumers. But, as just noted, even a pool with the monopoly royalty rate can lead to higher consumer surplus than ongoing litigation, especially if the litigation will be protracted, so that T is large, or if the patents are strong.

Proposition 7 (royalties charged by patent pools with two members). Even a patent pool that

³² The 1997 MPEG case can be found at www.usdoj.gov/atr/public/press_releases/1997/1173.htm. The first DVD pool was approved in December 1998; see www.usdoj.gov/atr/public/press_releases/1998/2120.htm. The second DVD pool was approved in June 1999; see www.usdoj.gov/atr/public/press_releases/1999/2484.htm. The DOJ insisted that only "essential" patents be included in these pools.

replicates the monopoly outcome can benefit consumers by replacing a Cournot complements outcome; equation (7) is the necessary and sufficient condition for this to occur. More generally, putting $\mu(r)$ into equation (7) gives the maximum royalty rate that a pool can charge and make consumers better off than litigation.

These same methods can also be used to evaluate the effects of pools with more than two members. The equations are necessarily more complex, as we need to keep track of consumer surplus if k out of n patents are found valid, for $k = 0, 1, \dots, n$. But the same principle articulated in Proposition 7 applies to these larger pools: if litigation is protracted and/or the patents are reasonably strong, even a pool replicating the monopoly outcome can be favorable for consumers. When this condition is met, antitrust concerns about proposed patent pools should be greatly reduced.

7. Conclusions and extensions

■ In this article I have proposed and explored the following simple antitrust rule governing settlements of intellectual property disputes: a settlement cannot lead to lower expected consumer surplus than would have arisen from ongoing litigation. I argue that this rule respects intellectual property rights while encouraging efficient settlements. Under extremely general conditions, there exists a settlement that leaves consumers better off and raises the joint profits of the two firms engaged in the dispute. I then apply this general test to several types of settlements: licensing; mergers; agreements specifying the timing of entry; and patent pools.

Although this article has covered a lot of ground, there are many more unresolved issues and questions in this and related areas. I close by noting a few of these outstanding questions.

First, I have focused here on situations involving a single patent, or, in the case of patent pools, just two patents. In many patent disputes, at least one party has a whole portfolio of patents.³³ It remains to be seen how the presence of multiple patents affects my analysis. This would have implications for the incentives of firms to assemble, or perhaps disassemble, patent portfolios. I conjecture that there are diminishing returns to the number of patents held by one party in a given field.³⁴

Second, I have explicitly avoided introducing asymmetric information between the two parties to an intellectual property dispute. Asymmetric information, and resulting differences in beliefs, are important factors that can make settlement difficult. Another rich area of exploration involves marrying the analysis in this article with the extensive literature on bargaining, signalling, and self-selection in the presence of asymmetric information. Risk aversion could also be included.

Finally, I have focused my attention on a single challenger to a patent, while recognizing that other potential competitors may benefit if a patent is held invalid or interpreted narrowly. Another valuable extension would be to explicitly model multiple challengers, recognizing that the patentholder may have an incentive to settle with the strongest challenger, hoping that weaker challengers will later settle on more favorable terms.

References

- AYRES, I. AND KLEMPERER, P. "Limiting Patentees' Market Power Without Reducing Innovation Incentives: The Perverse Benefits of Uncertainty and Non-injunctive Remedies." *Michigan Law Review*, Vol. 97 (1999), pp. 985–1033.
- COHEN, W.M., NELSON, R.R., AND WALSH, J. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (Or Not)." National Bureau of Economic Research, Working Paper no. 7552, 2000.
- COOTER, R.D. AND RUBINFELD, D. "Economic Analysis of Legal Disputes and Their Resolution." *Journal of Economic Literature*, Vol. 27 (1989), pp. 1067–1097.
- GALLINI, N. "The Economics of Patents: Lessons from Recent U.S. Patent Reform." *Journal of Economic Perspectives*, Vol. 16 (2002), pp. 131–154.

³³ Lanjouw and Schankerman (2002) report that patentholders with larger portfolios are much less likely to litigate their patents than are those who hold only one or a few patents.

³⁴ In fact, I am aware of a situation in which one firm spun off part of its patent portfolio so that the two entities, each controlling essential patents for certain applications, could separately seek royalties from licensees.

- GILBERT, R.J. AND SHAPIRO, C. "Antitrust Issues in the Licensing of Intellectual Property: The Nine No-No's Meet the Nineties." *Brookings Papers on Economics: Microeconomics*, (1997), pp. 283–336.
- AND TOM, W.K. "Is Innovation King at the Antitrust Agencies? The Intellectual Property Guidelines Five Years Later." *Antitrust Law Journal*, Vol. 69 (2001), pp. 43–86.
- GRINDLEY, P. AND TEECE, D.J. "Managing Intellectual Capital: Licensing and Cross-Licensing in Semiconductors and Electronics." *California Management Review*, Vol. 39 (1997), pp. 1–34.
- HALL, B. AND HAM ZIEDONIS, R. "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979–1995." *RAND Journal of Economics*, Vol. 32 (2001), pp. 101–128.
- KATZ, M. AND SHAPIRO, C. "On the Licensing of Innovations." *RAND Journal of Economics*, Vol. 16 (1985), pp. 505–520.
- KLEIN, J.I. "Cross-Licensing and Antitrust Law." 1997, available at www.usdoj.gov/atr/public/speeches/1123.htm.
- KORTUM, S. AND LERNER, J. "Stronger Protection or Technological Revolution: What Is Behind the Recent Surge in Patenting?" *Carnegie-Rochester Conference Series on Public Policy*, Vol. 48 (1998), pp. 247–304.
- LANJOUW, J.O. AND SCHANKERMAN, M. "Characteristics of Patent Litigation: A Window on Competition." *RAND Journal of Economics*, Vol. 32 (2001), pp. 129–151.
- AND ———. "Enforcing Intellectual Property Rights: Suits, Settlements and the Explosion in Patent Litigation." Mimeo, Department of Economics, Yale University, June 2002.
- MEURER, M. "The Settlement of Patent Litigation." *RAND Journal of Economics*, Vol. 20 (1989), pp. 77–91.
- SHAPIRO, C. "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting." In A. Jaffe, J. Lerner, and S. Stern, eds., *Innovation Policy and the Economy*, Vol. 1. Cambridge, Mass.: MIT Press, 2001.
- UNITED STATES DEPARTMENT OF JUSTICE AND FEDERAL TRADE COMMISSION. *Antitrust Guidelines for the Licensing of Intellectual Property*. April 1995, available at www.usdoj.gov/atr/public/guidelines/ipguide.htm.
- . *Antitrust Guidelines for Collaborations Among Competitors*. April 2000, available at www.ftc.gov/os/2000/04/ftcdojguidelines.pdf.
- UNITED STATES FEDERAL TRADE COMMISSION (FTC). "Generic Drug Entry Prior to Patent Expiration." July 2002, available at www.ftc.gov/os/2002/07/genericdrugstudy.pdf.